Revisiting the capitalization of public transport accessibility into residential land value: An empirical analysis drawing on Open Science

José Rafael Verduzco Torres
BArch, MSc, MArch

Submitted in fulfilment
of the requirements of the Degree of
Doctor of Philosophy (PhD)

School of Social and Political Sciences
College of Social Sciences
University of Glasgow

March 2023

Abstract

Background: The delivery and effective operation of public transport is fundamental for a for a transition to low-carbon emission transport systems'. However, many cities face budgetary challenges in providing and operating this type of infrastructure. Land value capture (LVC) instruments, aimed at recovering all or part of the land value uplifts triggered by actions other than the landowner, can alleviate some of this pressure. A key element of LVC lies in the increment in land value associated with a particular public action. Urban economic theory supports this idea and considers accessibility to be a core element for determining residential land value. Although the empirical literature assessing the relationship between land value increments and public transport infrastructure is vast, it often assumes homogeneous benefits and, therefore, overlooks relevant elements of accessibility. Advancements in the accessibility concept in the context of Open Science can ease the relaxation of such assumptions.

Methods: This thesis draws on the case of Greater Mexico City between 2009 and 2019. It focuses on the effects of the main public transport network (MPTN) which is organised in seven temporal stages according to its expansion phases. The analysis incorporates location-based accessibility measures to employment opportunities in order to assess the benefits of public transport infrastructure. It does so by making extensive use of the open-source software OpenTripPlanner for public transport route modelling ($\approx$ 2.1 billion origin-destination routes). Potential capitalizations are assessed according to the hedonic framework. The property value data includes individual administrative mortgage records collected by the Federal Mortgage Society ($\approx$ 800,000). The hedonic function is estimated using a variety of approaches, i.e. linear models, nonlinear models, multilevel models, and spatial multilevel models. These are estimated by the maximum likelihood and Bayesian methods. The study also examines possible spatial aggregation bias using alternative spatial aggregation schemes according to the modifiable areal unit problem (MAUP) literature.

Results: The accessibility models across the various temporal stages evidence the spatial heterogeneity shaped by the MPTN in combination with land use and the individual perception of residents. This highlights the need to transition from measures that focus on the characteristics of transport infrastructure to comprehensive accessibility measures which reflect such heterogeneity. The estimated hedonic function

suggests a robust, positive, and significant relationship between MPTN accessibility and residential land value in all the modelling frameworks in the presence of a variety of controls. The residential land value increases between 3.6% and 5.7% for one additional standard deviation in MPTN accessibility to employment in the final set of models. The total willingness to pay (TWTP) is considerable, ranging from 0.7 to 1.5 times the equivalent of the capital costs of the bus rapid transit Line-7 of the Metrobús system. A sensitivity analysis shows that the hedonic model estimation is sensitive to the MAUP. In addition, the use of a post code zoning scheme produces the closest results compared to the smallest spatial analytical scheme (0.5 km hexagonal grid).
Conclusion: The present thesis advances the discussion on the capitalization of public transport on residential land value by adopting recent contributions from the Open Science framework. Empirically, it fills a knowledge gap given the lack of literature around this topic in this area of study. In terms of policy, the findings support LVC as a mechanism of considerable potential. Regarding fee-based LVC instruments, there are fairness issues in relation to the distribution of charges or exactions to households that could be addressed using location-based measures. Furthermore, the approach developed for this analysis serves as valuable guidance for identifying sites with large potential for the implementation of development-based instruments, for instance land readjustments or the sale/lease of additional development rights.

Table of Contents

Chapter 1: Introduction **1**
1.1 Rationale and aim 1
1.2 Background 2
1.3 Objectives and Research Questions 4
1.3.1 Objectives 4
1.3.2 Research Questions 5
1.4 Contributions and key findings 5
1.5 Structure of the thesis 7

Chapter 2: The recovery of land value, land rent, and accessibility . . **11**
2.1 Land value capture 12
2.1.1 LVC rationale 12
2.1.2 LVC in practice 13
2.1.3 LVC instruments in the context of transport 14
2.1.4 The missing link between LVC literature and land rent 20
2.2 Land rent theory and the residential market 21
2.2.1 Early fundamentals on the theory of land rent 21
2.2.2 Residential land rent through the monocentric city model 24
2.2.3 Employment outside the CBD 29
2.3 Accessibility 32
2.3.1 The concept of accessibility 33
2.3.2 Operationalizing accessibility measures 35
2.3.3 Location-based measures 40
2.3.4 Accessibility in the context of residential land value 47
2.3.5 Measures of accessibility and public transport systems 48
2.4 Summary and final remarks 49

Chapter 3: The capitalization effects of accessibility: A review of literature **51**
3.1 Nonmarket valuation via revealed preferences 52

3.1.1 The residential market through the hedonic framework 53
3.1.2 Further considerations and challenges of the hedonic approach . 61
3.1.3 Equilibrium sorting model 64
3.2 Empirical analyses on public transport accessibility 65
3.2.1 The classic empirical approach 66
3.2.2 An alternative approach: The impact of intra-urban accessibility on land value 72
3.2.3 Net benefits in empirical studies 80
3.3 Spatial aggregation bias 80
3.3.1 MAUP in the study of accessibility 82
3.3.2 MAUP in hedonic property models 84
3.4 Summary and final remarks 85

Chapter 4: Study area: Greater Mexico City 87
4.1 An overview of Greater Mexico City 88
4.2 Mobility and the public transport system in GMC 91
4.2.1 The transport system and the main public transport network . . 94
4.3 Housing in Mexico 104
4.3.1 The national housing policy 104
4.3.2 The housing stock in Mexico and GMC 106
4.4 The politics of LVC in Mexico 108
4.5 Summary and final remarks 111

Chapter 5: Methodology 113
5.1 Research strategy 114
5.1.1 Transparency and reproducibility 115
5.2 Spatial analytical framework 117
5.2.1 Spatial units of aggregation of the sources 117
5.2.2 Proposed spatial analytical scheme 118
5.3 Property data: Federal Mortgage Society (SHF) 120
5.3.1 Source and structure of data 121
5.3.2 Sample selection 121
5.3.3 Geocoding process 124
5.4 Modelling framework 127
5.4.1 Hedonic model 129
5.4.2 Estimating the hedonic function 129
5.4.3 Multilevel and spatial multilevel models 129
5.4.4 Bayesian inference 133
5.4.5 Model assessment 137

5.4.6 Benefit measurement 137
5.5 Environmental measures 139
5.5.1 Neighbourhood measures 139
5.5.2 Accessibility measures 142
5.6 Summary and final remarks 145

Chapter 6: The role of the main public transport network in shaping accessibility to employment. An exploratory spatial data analysis . 147
6.1 Introduction and Research Question 147
6.2 Research strategy and estimation 149
6.2.1 Overview 149
6.2.2 Accessibility measures 150
6.2.3 Data 151
6.2.4 Public transport travel time modelling 160
6.2.5 Spatial decaying function and calibration of accessibility parameters 162
6.3 Exploratory analysis 165
6.3.1 Overview of accessibility to employment 165
6.3.2 Spatial aggregation of accessibility 168
6.3.3 Temporal variability of accessibility 173
6.4 Discussion and limitations 180
6.5 Conclusions 183

Chapter 7: The value of public transport accessibility 185
7.1 Introduction and Research Question 185
7.2 Research strategy and estimation 187
7.2.1 Data and measures 187
7.2.2 Model specification and estimation 189
7.3 Results 194
7.3.1 Descriptive statistics 194
7.3.2 The base hedonic property value model 199
7.3.3 Incorporating a measure of public transport accessibility to employment to the hedonic property value model 199
7.3.4 Multilevel models over time 203
7.3.5 The potential for LVC 211
7.4 Discussion 214
7.4.1 Location-based accessibility in the hedonic model 214
7.4.2 The implications of the MPTN in the housing market system . . 216
7.4.3 Limitations and future research 217

7.5 Conclusions 219

Chapter 8: Conclusion 221
8.1 Introduction and review of Research Questions 221
8.2 Addressing the Research Questions 222
8.2.1 What is the role of the main public transport network in shaping accessibility to employment in Greater Mexico City? 222
8.2.2 What is the willingness to pay for the accessibility generated by the main public transport network in the residential land market between 2009 and 2019? 224
8.3 Implications and contributions to knowledge 225
8.3.1 From the empirical point of view 225
8.3.2 From the public policy point of view 226
8.3.3 From the methodological view 228
8.4 Limitations 230
8.5 Future work 231
8.6 Final thoughts 234

Appendix A: List of source code repositories 237

Appendix B: Model specifications to adjust private car travel times . 241

Appendix C: Mexibus (MXB) services 243

Appendix D: Estimation procedure of economic activity spatial reallocation 247

Appendix E: Correlation of surface area and accessibility 249

Appendix F: Distribution of accessibility according to Gini index . . . 251

Appendix G: Prior sensitivity analysis 253

Appendix H: Full base hedonic property value model results 255

Appendix I: MAUP sensitivity analysis 259
I.1 Introduction and objective 259
I.1.1 Data and measures 259
I.1.2 Model specifications 260
I.1.3 Implicit prices and willingess to pay 262
I.2 Results 262
I.2.1 Descriptive statistics 262

I.2.2 MAUP and spatial decay in the main public transport network . 264
I.2.3 MAUP and implicit prices in the spatial multilevel hedonic model 265
I.2.4 MAUP and the willingness to pay for accessibility to employment 269
I.3 Final remarks 271

References **273**

List of Tables

2.1 Land value capture instruments in the context of public transport . . . 15
2.2 Accessibility measures 36

3.1 Common hedonic functional forms and marginal implicit prices. 57

4.1 Composition of Greater Mexico City, 2015. 90
4.2 Distribution of employed people by sector. Greater Mexico City, 2014. . 92
4.3 Modal split. Number of journeys that use one of the following modes. Greater Mexico City, 2017. 93
4.4 Classification of urban public transport modes by ROW and major technological features 96
4.5 Bus rapid transit (BRT) corridors in Greater Mexico City, 2019. 99
4.6 Metro system in Greater Mexico City, 2019. 101
4.7 Housing stock by state. Greater Mexico City, 2020. 107
4.8 Housing stock characteristics. Greater Mexico City, 2020. 109

5.1 City blocks summary. Greater Mexico City. 118
5.2 Proposed spatial analitycal schemes. 120
5.3 Sample 1 geocoding summary 126
5.4 Sample 2 geocoding summary 127

6.1 Main public transport network extensions between 2010 and 2019 in Greater Mexico City grouped by temporal stage. 155
6.2 Trip distribution model estimates 165
6.3 Parameters employed in accessibility models 165
6.4 Summary statistics of accessibility models. 165
6.5 Spatial correlation over various SAS according to Moran's I test. 172

7.1 Prior parameters used for the sensitivity analysis. 193
7.2 Descriptive statistics 195
7.3 Main public transport network (MPTN) accessibility to employment in the hedonic property value model. 200

7.4 Regression estimate comparison across various models for Stage 7 (2018 to 2019) for PT accessibility. . . . 204
7.5 Spatial multilevel model results for the Besag-York-Mollié (BYM) specification. . . . 208

B.1 OLS summary. Dependent variable is Google Maps (GM) travel time by car in minutes . . . 242

C.1 Mexibus corridor services, 2019 . . . 243

E.1 Bivariate correlation between accessibility and surface area in various SAS according to the Pearson correlation coeficient. . . . 249

F.1 Gini index for all temporal stages of the MPTN between 2010 and 2019 for measures of accessibility considerin all and matching opportunities. 252

G.1 Prior sensitivity analysis. Stage 1 (Before 2010). . . . 253

H.1 House value determinants. Greater Mexico City, 2009-2019. . . . 256

I.1 Descriptive statistics . . . 263
I.2 MPTN spatial decay parameter across various spatial analytical schemes. 264
I.3 Spatial multilevel Besag-York-Mollié (BYM) model results for various grid scales. . . . 267
I.4 Total willingness to pay summary . . . 269

List of Figures

2.1 Consumer choice. 26
2.2 Bid rent function for land. 29
2.3 Polycentric city. 31
2.4 Relationship between components of accessibility. 34
2.5 Impedance functions. 44

3.1 Hedonic equilibrium. 55
3.2 Land value uplift coefficients for residential in United States studies. . . 70

4.1 Greater Mexico City in the regional context. 89
4.2 Distribution of time spent on individual journeys by purpose in Greater Mexico City. 94
4.3 Main public transport network (MPTN). Greater Mexico City, 2019. . . 103
4.4 Institutional housing system in Mexico. 105

5.1 SHF Sample Selection Process. 122
5.2 Size and value, 2009-2015. 124
5.3 Modelled travel time by car in Google Maps and OpenTripPlanner. . . 144

6.1 Conceptual route-map for the estimation of accessibility models. 149
6.2 Temporal stages of the main public transport network. Greater Mexico City, 2010 to 2019. 154
6.3 Population density aggregated at various spatial analytical schemes. Greater Mexico City, 2010. 156
6.4 Level of education aggregated at various spatial analytical schemes. Greater Mexico City, 2010. 157
6.5 Employment density aggregated at various spatial analytical schemes. Greater Mexico City, 2010. 160
6.6 Distribution of accessibility measure by temporal stage and SAS. 168
6.7 Accessibility generated by the main public transport network aggregated at various spatial analytical schemes. 169
6.8 Distribution density of accessibility measures over various SAS. 173

6.9 Relative change of mean accessibility to employment over various temporal stages (reference is 'Before 2010') according to all and matching employment opportunities measures. 174
6.10 Variability of accessibility to employment over various temporal stages. Greater Mexico City, 2010-2019. 176
6.11 Mean Gini index for each MPTN temporal stage between 2010-2019. . 179

7.1 Absolute accessibility to employment gains after the MB-L7 extension in Stage 7 featuring top benefited post code zones (reference is Stage 6). 212

C.1 Stations serviced in the corridor Mexibús I. 244
C.2 Stations serviced in the corridor Mexibus II. 245
C.3 Stations serviced in the corridor Mexibus III. 246

I.1 Estimated spatial decay parameter for various grids. 265
I.2 TWTP uncertainty. 271

Acknowledgements

This research was funded by the 2018-000037-01EXTF-00438 scholarship awarded by the National Council of Science and Technology of Mexico (Conacyt), and jointly supported by the University of Glasgow.

I am grateful to all the members of my supervision team for their exceptional support. David McArthur was not only approachable for academic discussions, but was also available for amiable conversations. I am deeply indebted to him for his generosity, patience, and care. I also want to thank Jinhyun Hong for encouraging and supporting me to publish my first academic article at an early stage in the programme. Likewise, I am thankful for the genuine support and commitment of David Waite from the first time I contacted him, even when he was not formally a part of the supervision team. His friendship and thought-provoking conversations have been an important part of this process.

I am also grateful for the kindness, support, and feedback of my unofficial mentors and colleagues in the Department of Urban Studies and the Urban Big Data Centre (UBDC). I especially want to thank Mark Wong for his thoughtful words of encouragement, which motivated me to continually improve, Nick Bailey for his friendship and the opportunities offered to PhD members of the UBDC, Luis Serra, Joanna Stewart, Fraser McMillan, and all the other colleagues who have always made me feel welcomed.

This work would not have been possible without the love and unconditional moral support of my family and partner. The weekly calls with my Dad eased the difficulties of leaving behind my life in Mexico to make this dream come true, and the spontaneous text messages from my Mum made the Scottish rainy days happier. My brother Oscar and my sister Liz were sources of inspiration for seeking growth. I cannot be more fortunate to have Pau, the most supportive and understanding partner, accompanying me on this journey. My nephew Oscar and niece Juli are constant reminders of hope. Lastly, my sincere thanks to the friends who are always there to listen and who keep me in their thoughts.

I dedicate this work to my Grandma Lupita, who sadly passed away while I was writing this thesis.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: José Rafael Verduzco Torres
Signature:

List of abbreviations

AIC	Akaine information criterion
AMM	Alonso-Muth-Mills
B-P test	Breusch-Pagan test
BLUE	Best linear unbiased estimator
BRT	Bus rapid transit
BYM	Besag-York-Mollie
CAR	Conditional autoregressive
CBD	Central business district
CETRAM	Multimodal transport hub (Centro de Transferencia Modal)
CONAVI	National Housing Commission (Comisión Nacional de Vivienda)
DENUE	National Directory of Economic Units (Directorio Estadístico Nacional de Unidades Económicas)
DIC	Deviance information criterion
DID	Difference-in-difference
CPU	Central processing unit
ENIGH	National Household Income Survey (Encuesta Nacional de Ingresos y Gastos de los Hogares)
ESM	Equilibrium sorting model
FONHAPO	National Fund for Social Housing (Fondo Nacional de Habitaciones Populares)
FOVISSSTE	Housing Fund of Institute of Social Security of State Workers (Fondo de la Vivienda del Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado)
GBP	Great british pound
GDP	Gross domestic product
GLA	Greater London Authority
GM	GoogleMaps
GMC	Greater Mexico City (Zona metropolitana del Valle de México)
GTFS	General Transit Feed Specification
GWR	Geographically weighted regression
HPD	Highest posterior density

INEGI	Institute for National Statistics (Instituto Nacional de Estadística y Geografía)
Infonavit	Institute of the National Workers' Housing Fund (Instituto del Fondo Nacional de la Vivienda para los Trabajadores)
INLA	Integrated nested Laplace approximation
LRT	Light rail transit
LVC	Land value capture
MAUP	Modifiable areal unit problem
MCMC	Markov chain Monte Carlo
MB	Metrobús
MLM	Multilevel model
MPTN	Main public transport network
MXB	Mexibús
MXN	Mexican peso
NAICS	North American Industry Classification System
NLM	Nonlinear model
NYC	New York City
OD	Origin-destination
OLS	Ordinary least squares
Onavis	National housing organisms (Organismos nacionales de vivienda)
Orevis	State housing organisms (Organismos estatales de vivienda)
OSM	OpenStreetMap
OTP	OpenTripPlanner
PUMA	Public Use Microdata Area
PSM	Propensity score matching
RESET	Regression specification error test
RGR	Regional or commuter rail
RMSE	Root-mean-squared error
ROW	Right of way
RQ	Research question
RUM	Random utility maximization
SAR	Spatial autoregressive model
SAS	Spatial analytical scheme
SEM	Spatial error model
SCINCE	Census Information Consultation System (Sistema de Consulta de Información Censal)
SD	Satandard deviation
Sedatu	Ministry of Agriculture, Land and Urban Development (Secretaría de Desarrollo Agrario, Territorial y Urbano)
Semovi	Transport department for Mexico City (Secretaría de Movlidad de la Ciudad de México)

SHF	Federal Mortgage Society (Sociedad Hipotecaria Federal)
SITRAMyTEM	Mass transportation system for the state of México (Sistema de Transporte Masivo y Teleférico del Estado de México)
SIV	National housing system (Sistema nacional de vivienda)
SNIIV	National system for housing statistics (Sistema Nacional de Información e Indicadores de Vivienda)
Sofoles	Limited Financial Societies (Sociedades Financieras de Objeto Limitado)
Sofomes	Multiple Financial Societies (Sociedad Financiera de Objeto Múltiple)
STE	Metro Operating Agency (Sistema de Transporte Colectivo)
TAZ	Traffic analysis zone
TC	Transaction costs
TfL	Transport for London
TIF	Tax increment financing
TTM	Travel time matrix
TWTP	Total willingness to pay
UK	United Kingdom
US	United States
VIF	Variance of inflation factor
WHO	World Health Organization
WTP	Willingness to pay

Chapter 1

Introduction

1.1 Rationale and aim

In Mexico, the transportation sector was the largest energy consumer constituting almost half (47%) of the total energy use in 2016 (SIE, 2019).[1] The demand for energy in the transport sector grew by more than 10% in only 2 years (from 2014 to 2016), and virtually all the energy sources used were fossil fuels (only 0.20% was powered by electricity). Within the transport sector, road transportation (including passenger and freight) was by far the largest energy consumer at the national level, accounting for 90%.[2] Cities are playing important roles in fuel consumption trends for transport and its consequent carbon footprint (Holden, Banister, Gössling, Gilpin, & Linnerud, 2020; Stokenberga & Schipper, 2012). As shown in different contexts, effective public transport systems can aid the reduction of fossil fuel consumption (Suzuki, Murakami, Hong, & Tamayose, 2015).

The provision of public transport infrastructure often involves large public expenditures (Gómez-Ibáñez & Liu, 2022). Financing these projects has been particularly challenging for developing countries with high urbanization rates (Abiad, Farrin, & Hale, 2019; Blanco, Moreno, Vetter, & Vetter, 2016; Guerra, 2014). In addition, it is not only the initial infrastructure costs that are increasingly difficult to cover, but the operational and maintenance costs as well (Cervero & Murakami, 2008; Falcocchio, Malik, & Kontokosta, 2018). While governments often face difficulties funding transport infrastructure, public investments generate benefits that are often capitalized in the value of land (Aveline-Dubach & Blandeau, 2019). Land value capture

[1]The final consumption figures exclude energy used for activities dedicated to its own generation. In particular, these include transformation, distribution and consumption. The final consumption figures include activities classified as non-energetic and energetic activities. This last classification refers to the following productive sectors: industry, transport, agriculture, public, commercial, and residential use.

[2]The Energy Information System (SIE, 2019) uses the following classification to describe the transport sector: road transport (including passenger and freight), rail, aerial, maritime, and electric.

(LVC) mechanisms aim to assemble all or some of the economic gains generated by actions other than the landowners for the community at large through a broad range of strategies (Gómez-Ibáñez, Hong, & Du, 2022; Smolka, 2013). For example, these can be fee-based or development-based (Suzuki et al., 2015); the arrangement can be negotiated, compulsory or voluntary (Muñoz Gielen & van der Krabben, 2019); and the implementation can take place before or after the construction of the infrastructure; or before or after the urban development (Zhao, Iacono, Lari, & Levinson, 2012).

The present work aims to advance the current literature on the economic benefits generated by public transport infrastructure brought about by the capitalization of the value of residential land. It will do so by taking advantage of innovative resources developed under the Open Science concept—a "transparent and accessible knowledge that is shared and developed through collaborative networks" (Vicente-Saez & Martinez-Fuentes, 2018, p. 434). This approach encompasses methods, tools, and sources that, in combination, allow the following:

- Thorough examination of accessibility by enabling the detailed modelling of origin-destination routes by public transport (Higgins et al., 2022);
- Curation of large-scale and complex datasets, for example those derived from raw administrative property data records and their respective geolocation;
- Implementation of appropriate inferential statistical methods considering the scale characteristics of the data, and;
- Improvement of transparency and reproducibility of research.

To summarize, this research project seeks to revisit the willingness to pay for public transport accessibility to employment in the residential land market by integrating to the analysis key and novel elements chiefly addressed in the fields of urban and transport geography in relation to the concept of accessibility . For this purpose, the analyses draw on the case of *Zona Metropolitana del Valle de México* (Greater Mexico City) over a timespan of ten years.

1.2 Background

Following the considerations set out above, the question of how LVC mechanisms are supported by theory then comes to the fore. Some of the key concepts necessary for exploring this question are highlighted in the the Alonso-Muth-Mills (AMM) model (Alonso, 1964; Brueckner, 1987; Mills, 1972; Muth, 1969). In short, the AMM model reinterprets Von Thünen (1966) central town as the *central business district* (CBD), where employment and urban services are concentrated and where residents commute to from their respective residential locations. In this model, the willingness to pay

(WTP) per unit of land decreases as a function of increased transportation costs. These are assumed to rise gradually as a result of reduced accessibility with respect to the CBD. One critique of this theory is its oversimplification (Malamis et al., 2016). In particular, access to employment assumes a monocentric structure in contrast to the spatial dispersion of opportunities found in reality (Ahlfeldt, 2011; Anas, Arnott, & Small, 1998; Duranton & Puga, 2015).

The impact of transportation infrastructure on property prices has attracted a lot of attention in empirical studies (see Debrezion, Pels, & Rietveld, 2007; Higgins & Kanaroglou, 2016; Mohammad, Graham, Melo, & Anderson, 2013). Nonetheless, the consideration of the individual characteristics of public transport stations and the architecture of the network system has often been overlooked (Higgins & Kanaroglou, 2016; S. Ryan, 1999). In a growing approach, some studies have been employing comprehensive accessibility measures to assess the level of the public transport system services and their relation to land value with the intention of accounting for the spatial dispersion of opportunities and the topology of the transport network (e.g. Ahlfeldt & Wendland, 2016; Cordera, Coppola, Dell'Olio, & Ibeas, 2018; Iacono & Levinson, 2017; Osland & Thorsen, 2013).

A pioneering work employing this approach was conducted in Belfast, Northern Ireland (Adair, McGreal, Smyth, Cooper, & Ryley, 2000). This empirical study showed a small contribution of the accessibility indexes on property price variance at the citywide level. However, results were particularly relevant for predicting housing prices in sub-markets. Similarly, in a regional analysis in Norway (Osland & Thorsen, 2008), gravity-based parameters for labour-market accessibility were tested and compared to the performance of travelling time to the CBD as a benchmark. The results show that both values are roughly equal in their ability to explain variations in housing prices. The conclusions emphasize that gravity-based measures are useful for capturing the irregular distribution of the labour market. Ahlfeldt (2011, 2012) offers important contributions to this approach. In these findings, the neoclassical bid-rent theory (embedded in the AMM model) has been recognized as prevalent when appropriate measures are tested for polycentric structures. The conclusions suggest that gravity-type measures acknowledge the heterogeneity of stations and offer a superior performance when compared to 'conventional' estimates.

Although recent empirical studies adopting comprehensive accessibility metrics to examine the impacts of transport infrastructure on the land value have made relevant contributions, there are important gaps that remain open. First, the availability of detailed and standardized information about the operation of public transport and the tools to analyse this have grown considerably over the past few years. However, these resources have rarely been used or explicitly discussed in the context of land value capitalization. Second, theoretical and empirical literature suggest that accessibility is not perceived equally between population subgroups (e.g. on the demand side of the labour market: age, gender, income, education) and the type of opportunities available (e.g. on the supply side of the labour-market: position, wages, or educational requirements) (Geurs & van Wee, 2004; Levinson & Wu, 2020; Páez, Scott, & Morency, 2012; M. Ryan, Lin, Xia, & Robinson, 2016; Thériault, Voisin, & Des Rosiers, 2013). Nonetheless, when examining the property market, many empirical studies assume both jobs and workers as being homogeneous. For instance, accessibility has been operationalized as the total number of jobs accessed within a time-threshold (Iacono & Levinson, 2017), travel time to the nearest employment centre (Mulley, 2014), or the sum of all types of jobs from a specific location discounted by a function of distance (as in gravity-based accessibility measures) (Diao, 2015; Osland & Thorsen, 2008). Gjestland, Osland, & Thorsen (2020) specifically highlights the potential relevance of accounting for the different categories of jobs and workers when studying the housing market. A few researchers have studied these implications, which include a focus on accessibility generated by private car, e.g. Osland (2010a) in relation to gender or Thériault et al. (2013) who considers the type of household and purpose.

Accordingly, the present research calls for a comprehensive approach which addresses the heterogeneity of public transport infrastructure as well as the spatial dispersion of opportunities and their characteristics in the study of residential land value.

1.3 Objectives and Research Questions

1.3.1 Objectives

The general objective established for the present research is to:

- Examine the extent of the potential for land value capture (LVC) in the residential land context as a financing tool for public transport, drawing on the empirical case of Greater Mexico City.

The specific objectives are disaggregated according to their type of contribution as follows:

- To explore the distribution of accessibility to employment enabled by the main public transit network in Greater Mexico City between 2009 and 2019.
- To evaluate the willingness to pay for the accessibility benefits derived from the main public transport network (MPTN) in the residential land market considering the adequacy of location-based measures while acknowledging the characteristics of both the offer (salaries paid at potential destination) and the demand (education level at the origin) in the labour market between 2009 and 2019.

1.3.2 Research Questions

Based on the objectives defined, the main Research Question is:

- *How are the economic benefits generated by the main public transport infrastructure distributed in the residential land market of Greater Mexico City?*

Derived from the above, the specific Research Questions (RQ) are:

- RQ1. *What is the role of the main public transport network in shaping accessibility to employment in Greater Mexico City?*

- RQ2. *What is the willingness to pay for the accessibility generated by the main public transport network in the residential land market between 2009 and 2019?*

1.4 Contributions and key findings

The primary contribution of this thesis is the explicit incorporation of concepts, tools, and methods developed in the fields of urban and transport geography (under the Open Science framework) into an examination of the role of public transport accessibility in the residential land market. This approach both draws on and sheds new light on land rent theory. By doing so, it supports the bridging of a broader gap, identified in literature, between the narratives of land value capture and land rent (Vejchodská et al., 2022). The concept of accessibility serves as a critical link in this process providing a well-developed analytical framework for jointly discussing these parallel strands of the literature. Despite previous attempts to connect these concepts at a theoretical level, this is the first time the gap in knowledge has been addressed from an empirical perspective using public transport accessibility as a common framework.

The empirical findings highlight the spatially and perceptually heterogeneous character of public transport accessibility in relation to employment. For example, Chapter 6 illustrates the differentiated role of distinct public transport modes in combination with the overall and local land use. That is, the physical and operational characteristics, such as service frequency, average speed, or distance between stations, in combination with local employment density or overall proximity to the main employment hubs, are reflected in a variety of accessibility patterns. These findings are consistent with accessibility theory. This finding contributes to the discussion of the extent and magnitude of accessibility changes generated by the expansion of the MPTN in the context of environmental valuation in an empirical case. Another contribution is the examination of the assumption that all employment opportunities are equally attractive to all segments of the population. A direct comparison shows that, while a generic accessibility measure attributes a more critical role to the public transport system, a matching employment-resident measure emphasises the local land use. In addition, the network effects resulting from introducing a new public transport line extend to broader areas adjacent to existing public transport stations when it is assumed that local residents would value all types of employment opportunities, as opposed to the limited effects resulting from matching the characteristics of residents with employment. These findings are highly relevant for reconsidering the study of land value in relation to public transport as accessibility has been oversimplified in both theoretical and empirical literature.

The considerations taken in estimating public transport accessibility in Chapter 6 are directly incorporated into the empirical study of the WTP in Chapter 7 through a series of hedonic analyses. While the methods for estimating differentiated location-based measures are adapted from the transport geography literature (Pereira, Banister, Schwanen, & Wessel, 2019), this study represents the first attempt to incorporate them into the environmental valuation literature. Overall, the findings suggest a robust association between public transport accessibility, as measured by location-based measures, and residential land value. This is supported by a significant and positive correlation between both measures that consider the matching characteristics of residents and employment, as well as those that do not, across all temporal periods of the public transport network studied in GMC. This association is observed across different regression techniques and after considering the spatial structure of the data. The use of an accessibility measure that accounts for matching perception improves all the hedonic models that do not account for local heterogeneity (including unstructured or spatially structured effects at the post code level). In addition, the accessibility parameters estimated using property value data are very similar to those estimated entirely exogenously, i.e. using commuting flows in a gravity model framework. Nevertheless, considering the local heterogeneity mitigates the differences in model performance.

An estimate of the net benefits associated with the introduction of a bus rapid transit corridor (*Metrobús Línea 7*, MB-L7) from a partial-equilibrium approach, as presented in Chapter 7, suggests a capitalization of between MXN$4,321 million (GBP£157 million) at current prices in the partial scenario and MXN$1,900 million (GBP£69 million) in a full scenario.[3] This is equivalent to 0.7 to 1.5 times the capital cost of implementing the infrastructure. These results readdress some of the previous neutral findings reported in the region (e.g. D'Elia, Grand, & León, 2020; Flores Dewey, 2011). It is argued that location-based measures, as opposed to buffers or lines

[3]Considering an exchange rate of 27.60 MXN per 1 GBP, according to the currency exchange published by the Central Bank of Mexico (*Banxico*) on the 15/09/2021 on the website https://www.banxico.org.mx.

around stations, can control for the 'effective' accessibility benefits of public transport, accounting for the various components involved, namely transport (operational and physical), land use, and individual perception. In addition, the findings support LVC by drawing on theoretically robust principles of land rent theory.

This thesis demonstrates how innovative resources, such as public transport operational data in standard formats (i.e. GTFS), multi-modal open-source tools (i.e. OpenTripPlanner), administrative digital data, and increased computing power can support the analysis of land capitalization. These allow us to comprehensively address the concept of accessibility, a central element in land rent theory that has often been trivialized in empirical studies evaluating public transport infrastructure. By incorporating these resources, the analysis gains robustness and strengthens the theoretical framework adopted for the main objective of the study.

1.5 Structure of the thesis

This thesis is structured into eight chapters, which includes the present introduction.

Chapter 2 establishes the theoretical framework used to address the Research Questions. It opens the discussion by expanding on the LVC approach, including aspects related to its rationale, role in practice, and the variety of instruments available. Later, it reviews its supporting theoretical framework through the lens of land rent theory, where the concept of accessibility emerges as a central component. However, this perspective does not explicitly address accessibility (e.g. Alonso, 1964; Brueckner, 2011; Duranton & Puga, 2015). To broaden the understanding of accessibility, this thesis draws on literature from urban and transport geography, specifically location-based measures. This practical approach accommodates many of the aspects suggested in the theory of accessibility. An extensive survey of the literature confirms the disjointed nature of these complementary pieces of knowledge, which are useful for addressing the RQs. Thus, accessibility is adopted as a linking element across these strands of the literature.

Chapter 3 focuses on the empirical study of accessibility capitalization into land value found in the literature. For this purpose, the fundamentals of nonmarket valuation are established. Two main approaches are identified: the hedonic method and the equilibrium sorting model. The former has been employed in most empirical studies over the last fifty years. The latter, however, is relatively new and represents important potential for these types of studies. Further on, the results from the empirical studies are reviewed according to a main distinction made in the definition of accessibility in Chapter 2. Namely, those focusing on measures of mobility or transport infrastructure and those adopting a broader view including both mobility and land use elements.

The implications of the choice of the spatial analytical framework are discussed as a cross-cutting issue.

Chapter 4 introduces the study area, namely Greater Mexico City. This provides a brief overview of the political and administrate arrangement, the definition of the metropolitan area, the geographic location and regional context, demographics, extension, and the economy. In addition, it provides an overview of the mobility dynamics and describes the public transport system in detail. After a categorization of the modes operating in GMC, it provides a delimitation of the main public transport network (MPTN), the major object of study. Later, a description of the housing sector is offered. This is useful for introducing the main characteristics of the housing stock in the study area from a national perspective. This chapter draws the main delimitations of the study and provides an outline of the regional and local dynamics which will be useful for interpreting the results of the empirical analysis conducted.

Chapter 5 sets out the methodology of this thesis. Specifically, an outline of the research strategy is presented. This includes transparency and reproducibility considerations for this work in the context of Open Science. In addition, this chapter defines the spatial analytical schemes employed for the analyses based chiefly on the considerations pointed out in the modifiable areal unit problem (MAUP) literature addressed in Chapter 3. The source of information of the property value data used is also presented, including a description of the curation process carried out to harmonize the raw administrative data. This chapter details the main modelling framework adopted in order to address the main Research Question, drawing on the hedonic model presented in Chapter 3. While the equilibrium sorting model is considered as an alternative, there are limitations to its implementation in this research. In addition, there is a detailed description of the estimation procedures employed and the technical considerations made. Finally, the sources of information and measures used for the independent variables in the analyses are detailed.

Chapter 6 presents the results for RQ1 which aims to explore the distribution of accessibility to employment enabled by the MPTN between 2009 and 2019. The analysis is based on location-based measures which are developed from both a spatial and a temporal perspective. The spatial dimension is approached by using five alternative spatial analytical schemes (SAS), defined in Chapter 5. The temporal aspect captures the changes introduced by the extensions to the MPTN by identifying seven temporal breaks in the period studied (denominated MPTN temporal stages). Accessibility measures are not only estimated for all types of employment opportunities, but also by considering the characteristics of the population at the origins (education level) and of employment opportunities at the destinations (salaries paid), following the literature (Levinson & Wu, 2020; Pereira et al., 2019). The strength of the spatial decay is adjusted according to the observed commuting flows in GMC. The results evidence the spatial heterogeneity from both the temporal and spatial perspective. The accessibility outcomes clearly reflect the differentiated role of the MPTN according to the characteristics of the elements considered in accessibility measures, i.e. the transport (characteristics of the modes or connectivity of the MPTN), land use (spatial distribution of employment), and the individual (valuation of opportunities according to their characteristics and the travel time needed to reach them) (Geurs & van Wee, 2004).

Chapter 7 addresses RQ2 by introducing location-based measures to the hedonic model framework for the assessment of the willingness to pay (WTP) for MPTN accessibility between 2009 and 2019. It does so by acknowledging the characteristics of both the offer (salaries paid at the potential destination) and the demand (education level at the origin) of the labour market. For its main inputs, this analysis uses the travel time matrices estimated for the accessibility models developed to answer RQ1, as well as administrative mortgage records at the individual level (N=~800,000). The results consistently reflect a positive and significant relationship between MPTN accessibility and residential land value, even after considering a series of control variables and spatial local heterogeneity. The parameter estimates for accessibility are considerably larger in inferential techniques that do not model the spatial structure of the data than in those that explicitly model it. This chapter also provides an empirical illustration of the potential for LVC according to the aggregated willingness to pay (TWTP) for the benefits introduced by the bus rapid transit (BRT) Line 7 of the Metrobús system (MB-L7). The aggregated benefits are considerable, ranging from the equivalent of 0.7 to 1.5 times the capital cost of the MB-L7.

Chapter 8 concludes this thesis by outlining the Research Questions and addressing each of them. This chapter discusses the empirical, policy, and methodological implications and contributions to knowledge of this work, as well as acknowledging its overall limitations. Recommendations for future work are discussed before closing with some final thoughts.

Chapter 2

The recovery of land value, land rent, and accessibility

This chapter sets the theoretical framework guiding the assessment of potential land value increments introduced by public transport infrastructure, a core element supporting the land value capture (LVC) approach.

Recovering all or a portion of potential land value increase induced by public actions can provide financial support for the maintenance and expansion of public transport systems (Mathur, 2019; Mathur & Gatdula, 2020; Medda, 2012). The LVC idea is expanded in Section 2.1, incorporating the main elements of its rationale, empirical examples of its implementation for public transport projects, and the main instruments for its application found in literature. The review of LVC literature from the perspective of urban planning or planning law shows that these contributions rely on the assumption of positive gains caused by public actions. Yet, this strand of literature seldom interacts with the contributions from the perspective of urban economics, as suggested in earlier work (Vejchodská et al., 2022).

Section 2.2 introduces the land rent theory which offers the basis to quantitatively explain the full value of residential land and potential changes to it, supporting the main assumption of the LVC approach. Specifically, the section introduces the early fundamentals of land rent. It proceeds by outlining the monocentric city model, one of the most influential contributions of the field. Later, the discussion addresses the case of location of employment outside the central business district (CBD), one of the limitations of this model.

Section 2.3 primarily elaborates on the case of employment outside the CBD. For this purpose, it draws on the notion of accessibility which has been amply developed in the field of transport geography. This concept formalizes the hypothesized improvements introduced by public transport highlighted by LVC literature in line with the main elements provided by the theory of land rent. The section first reviews and harmonizes the concept of accessibility. Later, it provides a thorough overview of the taxonomy of accessibility measures using a framework based on its theoretical definition. Section 2.4 presents a summary of this chapter and provides some final remarks.

In sum, the main contribution of this chapter is that it integrates different perspectives around LVC in the context of public transport that had remained disjointed in existing literature, a gap identified in earlier work (Vejchodská et al., 2022). The concept of accessibility enables a unified discussion since it articulates the ideas developed from different scientific communities, namely urban planning and planning law on one side, and urban economics on the other. This contribution sets a robust underlying theoretical framework for revisiting the capitalization of public transport accessibility into residential land value, as proposed by the main objective of this thesis.

2.1 Land value capture

The present section offers a review of LVC focusing on public transport. First, it presents the overall rationale of the LVC approach. Next, some practical examples at the international level are discussed. This is aimed at illustrating the potential of the LVC approach based on previous empirical applications. Later, seven selected LVC instruments are discussed and grouped according to various dimensions identified by literature. The details of these instruments set the basis for the interpretation of the assessment of land value changes in the context of public transport. The section closes by linking the conceptual idea of LVC to the theory of land rent and the principles of the residential market.

2.1.1 LVC rationale

The LVC idea can be traced back to the labour theory of value developed by political economists of the 19th century. This theory implies that the source of commodities' value laid on the extent of individuals' effort (Fainstein, 2012; King & McLure, 2015). From that view, pure land value was excluded from the conceptualization of commodities since this is not a product of labour. In line with this, land value is conceived as a social or public asset which is taken by landowners through land rent. Since the source of revenue received by landowners is not a product of their own effort, it was therefore termed as *unearned income* (Alterman, 2012). In consequence, some influential ideas included: the purchasing of private land (at present value) to keep future uplifts for the state by John Stuart Mill (Mill, 1909); an imposition of a tax on land at the full rate of rent by Henry George (George, 1934); and the nationalisation of land by Karl Marx (Haila, 2016).

In contrast to the distributive justice (Fainstein, 2012) and moral (Andelson, 2000) concerns among political economists of the 19th century, recent advocates have shifted the main discussion to a *pragmatic* rationale (Alterman, 2012; Crook, Henneberry, & Whitehead, 2016; Muñoz Gielen & van der Krabben, 2019). From this view, the motivation is in terms of efficiency and occasionally in equity (Chapman, 2016; Fainstein, 2010; Mathur, 2014; Medda, 2012; Vejchodská et al., 2022). There is good consensus on the idea that a fiscal burden on land does not lead to distorting effects on the economy (Oates & Schwab, 2009). Thus, this is considered as an efficient source of public revenue. Furthermore, Ingram & Hong (2012) consider that recapturing the land value is efficient because the distribution of public costs can be addressed to the beneficiaries. Another aspect associated with efficiency is that a burden on land may act as an incentive for the landowner to use it productively or transfer it to one who will, instead of holding a site without use for speculative purposes (Andelson, 2000; Borrero, 2013; Fischel, 2015). Chapman (2016) adds that this is also efficient because it prevents undervaluation of public goods. In a similar vein, Vadali et al. (2018) acknowledge that the distribution of costs is equitable as well since the internalization of public benefits and services in land value can be returned to the public through government action.

Under the pragmatic view, the return of an increase in the land value rests on the difference between private and public goods (Vadali et al., 2018; Webster, 2010). It is argued that whilst private goods can be allocated through markets, public goods and services provided by government present characteristics that make difficult their allocation under the same mechanisms. Accordingly, two principles are identified as the basis for the payment of those goods:

1. The beneficiary pays principle
2. The cost principle.

The former concept seeks to identify those who receive a special advantage from a public action, so that they should contribute to the source of the benefit. Meanwhile, the latter focuses on those whose actions or activities impose a cost on government (or the public), and who should therefore compensate for those burdens.

2.1.2 LVC in practice

Some practical experiences in major cities at the international level have been able to, or are in the process of, mobilizing considerable sums to fund large transit projects (Salon, Sclar, & Barone, 2019). For example, the Greater London Authority (GLA) set a business rate supplement (BRS) and a mayoral community infrastructure levy (MCIL). These two mechanisms are aimed at supporting the funds of the *Crossrail* project (Greater London Authority, 2018; Roukoni & Medda, 2012). The GLA is expecting to collect a total of £4.1 billion and £0.6 billion, respectively, which together represent t 32% of the project cost. The Grand Paris Express project, a 200-kilometre regional rail in France, is planning to raise €21.8 billion based on a development tax on office space (Salon et al., 2019). This represents about 80% of the capital costs. Washington, D.C., and New York City offer significant examples. A relevant case is the Subway 7 Line Extension, in New York City. This project is estimated to be paid almost entirely (98%) by the Hudson Yards redevelopment for an amount of US$2.37 billion (McSpiritt, 2012; Salon et al., 2019). At a relatively smaller scale, the Washington Metropolitan Area Transport Authority (WMATA), in Washington D.C., has received US$730 million for the Dulles Metrorail Silver Line Expansion, representing around 14% of the budget (Salon et al., 2019). Similarly, the WMATA is receiving 28% of the funding for the construction of the New York Avenue Metro Station, equivalent to US$25 million (ibid).

In addition to the previous examples, it is worth mentioning the Asian experience. In particular, two of the main cases are Hong Kong, Special Administrative Region (SAR), China, and Tokyo, Japan. Hong Kong has been amply recognized for its 'rail plus property' (R+P) programme, which was crucial for expanding its existing regional rail network without undermining public financing (Cervero & Murakami, 2009). Research suggests that the Mass Transit Railway Corporation (MTRC) has effectively shifted their LVC strategy to a management-based model, assembling HK$7.4 billion in 2014, which represented about 60% of their income for that year (Aveline-Dubach & Blandeau, 2019). Likewise, the railway network in Tokyo has been substantially supported by LVC strategies (Yoshino & Stillman, 2018). From 2003 to 2012, one of the largest operating companies obtained 34% of its net income from real estate operations (Suzuki et al., 2015).

The subsection below elaborates on the instruments for the implementation of LVC.

2.1.3 LVC instruments in the context of transport

The specific strategies developed for the implementation of LVC encompass a wide variety of instruments. These are grouped according to several dimensions and characteristics in literature. First, it is possible to distinguish them into two main types (Suzuki et al., 2015; Walters, 2012): 1) fee- or tax-based, and 2) development-based, as shown in Table 2.1. This dimension can be distinguished because the first group relies on direct monetary contributions which are regularly required to be paid compulsorily. Meanwhile, the second can recover land value uplifts 'implicitly,' that is through the sale, lease, or trade of land or development rights. Under this scheme benefits can be

	Rationale			Arrangement			Cost type		Contributor		Infrastructure improvement		Land development	
	Direct	Indirect	Macro	Compulsory	Negotiated	Voluntary	Capital	O&M	Landowner	Developer	Before	After	Existing	New
Tax- or fee-based														
Land value tax or split rate tax	•	•		•			•	•	•		•	•	•	•
Betterment charges/levies	•	•		•			•		•		•	•	•	
Tax increment financing		•		•	•	•	•		•		•		•	•
Development-based														
Land sales or leases			•			•	•	•	•	•	•	•		•
Air rights sales or leases	•	•				•	•	•		•		•		•
Joint development		•			•	•	•	•		•	•	•		•
Land readjustment or redevelopment schemes			•	•	•	•	•		•	•	•	•		•

Note:
Source: the author based on Suzuki et al. (2015); Alterman (2012); Zhao et al. (2012).

Table 2.1: Land value capture instruments in the context of public transport

recovered not only in cash but also in-kind, e.g., public infrastructure, social housing, etc.

Additionally, there are further dimensions grouping LVC instruments identified in literature, namely: *rationale*, *arrangement*, *cost* type, *contributor*, *timing of infrastructure improvement* and *timing of land development* (Table 2.1). Identifying these dimensions is relevant since: (1) they can influence the feasibility of their application in relation to their socio-political support (Muñoz Gielen & van der Krabben, 2019); and (2) they reflect flexibility within implementation under various empirical cases, e.g. whether they are required when infrastructure already exists, or the condition of the land development.

The *rationale* under which a LVC emerges is important since this can influence the extent of their feasibility for their implementation. Alterman (2012) suggests that instruments can recover some of the value 'directly' based on the 'unearned' income argument standing alone, or 'indirectly' using alternative rationales such as cost-recovery, or mitigation of environmental or social impacts. A 'macro' rationale (also termed as public or public-private land assembly) is referred to as a broad land governance ideology, in which authorities play an 'active' role in developing land or partnering with landowners and/or developers (Muñoz Gielen & van der Krabben, 2019).[1]

[1]It should also be noted that one or similar versions of LVC instruments can emerge under a direct or indirect rationale. A mix of these is also possible and it is often observed in practice (Alterman,

The *arrangement* denotes the sort of relation between the contributor and the collector (state or transit agency) in which the LVC instrument is carried out (Muñoz Gielen & Lenferink, 2018; Muñoz Gielen & van der Krabben, 2019). This relation can be compulsory, that is a non-negotiable contribution imposed under legal or regulatory frameworks in place. A different type is the negotiated, in which there is room to define the extent of the participation. A voluntary arrangement can also occur when a developer, for instance, freely agrees to deliver in-cash or in-kind contributions in exchange for (additional) rights on land use.

The *cost type* is used in literature to identify the main kind of expenditure that the proceedings of the instrument is expected to cover (Zhao et al., 2012). This can be oriented to defray capital costs, that is directly on the initial investment, or it can be dedicated to operation and maintenance (O&M) of the transit system. The periodicity of the instrument, whether a one-time or recurrent payment, can help to elucidate the type of cost to be covered. For instance, a recurrent base can be more adequate for covering operating costs than instruments that are implemented only on a one-time basis.

The *contributor* denotes the agent that is subject to be charged (Zhao et al., 2012). As discussed above, this includes not only owners of land who capture the special benefits derived from public actions, but also other agents play strategic roles in the development process. This dimension distinguishes who is to be asked to contribute, whether the landowner or the developer.

The other two dimensions, *infrastructure development* and *land development*, refer to the timing in which it would be most appropriate to introduce these tools (Levinson & Istrate, 2011; Zhao et al., 2012). First, there are some instruments that can be implemented before or after the transit infrastructure is deployed. This is related to a retroactive or proactive character of the strategy. The former requires calculating the change in land value that can be attributed to transit-related investments, whereas the latter implies an estimate of possible future uplifts (Vadali et al., 2018).

Seven selected LVC instruments are presented and characterised by the dimensions described above, namely: (1) Land value tax or split rate tax; (2) Betterment charges and special assessment districts (SAD); (3) Tax increment financing; (4) Land sales or leases; (5) Air rights sales or leases; (6) Joint development; and;(7) Land readjustment or redevelopment schemes.

Tax- or fee-based instruments

Property tax is a very common and widespread practice that can recover a portion of land value increments. As a general practice, an equal rate is applied to buildings

2012).

and land to estimate its amount, i.e. 1:1 ratio. A *land value tax* or a *split rate* tax proposes to distinguish between the rate that is to be applied to land and that to buildings (Junge & Levinson, 2012). Under this scheme, the split rate suggests more emphasis on land rather than on buildings, for example by establishing a 2:1 ratio. The proportion can vary to any degree or ultimately use only the value of land as the basis (i.e. 1:0), whether the site is developed or not, and exempt any burden on buildings. When the assessment of the basis is up to date and aligned with market values, this instrument can incorporate all the benefits that are internalized in land value (Ibid). This implies that a portion of the land value induced by transit-related investments can be recovered regardless of the time when the transport improvement was built. Likewise, this method can be used in existing and underdeveloped areas, as long as the transit investment is reflected in the land value. The arguments for this are often related to the efficient use of land. The reason is because a continuous burden on land, whether developed or not, is expected to inhibit the under-use of land which may eventually lead to more compact and denser use of land (Plassmann & Tideman, 2000). This has also been associated with more affordable land values since a passive landowner would prefer to transfer the property rather than speculate with a future value (Andelson, 2000).

Betterment charges or *betterment levies* are compulsory fees aimed at recovering induced value in a defined area of influence that receives unique and special benefits, such as increased accessibility generated by public transport infrastructure. From this view, this is a direct instrument and may be accompanied by redistributive and social justice motivations (Alterman, 2012; Muñoz Gielen & van der Krabben, 2019). Due to the straight association with a specific action by the public (e.g. transit development) this levy is charged one time, even though the collection can occur over a middle-term span, and it is more adequately matched with the financing of capital infrastructure. This instrument can perform well in existing or consolidated urban areas that are to implement new transport infrastructure (Levinson & Istrate, 2011), although it can take either a retroactive or proactive character. In some legislations it is allowed to establish the fee up to 4 years before the infrastructure is implemented (Vadali et al., 2018). Accordingly, in this case the landowner is the targeted contributor. Special assessment district (SAD) is a similar instrument used in the U.S. Several authors regard it as equivalent to that of betterment charges (Medda, 2012; Suzuki et al., 2015; Vadali et al., 2018). Still, they display certain differences. Whereas the base for the betterment is only the difference in value with and without the infrastructure, the base for the SAD is usually the full value of the property including land and buildings (Vadali et al., 2018). Another important dissimilarity is the rationale, as SAD is typically implemented as a cost recovery instrument, which makes it an indirect type

(Muñoz Gielen & van der Krabben, 2019).

Tax increment financing (TIF), also termed as *accessibility increment contribution* (AIC), has been mostly used in the U.S. (Haider & Donaldson, 2016). Conceptually, future revenues attributed to increased accessibility are used to finance present expenditures in defined areas (can also apply jurisdiction-wide). The instrument is well-suited for redeveloping areas (that is, in existing urban areas), and charges occur after infrastructure has been built. After the declaration of a TIF a base-line value assessment is fixed, the proportion of the property revenues associated with the increments above the baseline are directed to a special fund to cover infrastructure expenditures of the project. Once the collection target (or period) is met, the further proceedings go to general local funding (Ibid). The rationale is noticeably indirect since it is used primarily to cover capital costs. The instrument is closely related to the property tax, and it is therefore collected from the landowner as fees.

Development-based instruments

Public *land leases and sales* occur as permanent transfers of public land or by long- or middle-term leases. These forms are well-suited in regimes where the state is the landowner or in a transition phase to a freehold system, but it is not limited to those cases (Bourassa & Hong, 2003; Peterson, 2009). Broadly, the rationale of these practices has been identified as macro since other objectives and ideologies beyond LVC are involved (Alterman, 2012). A closer examination shows that two strategies can be distinguished. One is when public land management occurs at a large scale, that is within a broad governance regime, and the other when governments under freehold systems take a more active role through different forms (Muñoz Gielen & van der Krabben, 2019). For example, governments and transit agencies own land that, under an adequate plan, could support transit systems' finances and simultaneously accomplish desirable land development goals (e.g. Knowles, 2012; Knowles & Ferbrache, 2016). Likewise, governments can acquire land around transit infrastructure through different mechanisms and sell it or lease it after infrastructure has been provided (via the strategy known as land banking), effectively capturing the increments (Alexander, 2008; Chava & Newman, 2016). In some cases, auctioning can even take place before the infrastructure is developed, helping to finance future infrastructure (Peterson, 2009). In both cases (leases and sales), the immediate contributor is the developer that acquires the rights on the land. This is also linked with the type of costs that these instruments are more apt to support. Whereas land sales are one-time revenue, leases are more flexible since they can be one-time (e.g. auctioning of long-term leases), or recurrent in middle-term leases (with periodic income e.g. yearly, quarterly, etc.), or a mix. Therefore, leases may support initial capital investments and/or O&M costs, whilst sales are apt to support capital investments. Both leases and sales are more appropriate in new urban development than in consolidated areas. However, they are not limited to the previous since they can also be implemented for redeveloping areas.

The implementation of *air rights sales* or *leases* is closely related to land use regulations (Gómez-Ibáñez et al., 2022; Mathur, 2019; Mathur & Gatdula, 2020). Like land sales or leases these are development-based and are usually paid in cash by a developer. Air rights are defined as the space above land or extensions to the existing development rights which are mainly compounded by two aspects of land use: intensity and type. Specifically, intensity is understood in terms of the net area of a site surface that it is allowed to be developed (commonly expressed as a building coverage factor), and the gross developable area, which is the effective total area that can be developed in a site (generally expressed as a factor known as floor area ratio, FAR). The type of land use determined by the legal framework or regulations limits or allows aspects such as whether a site is used for rural/agricultural, urban, residential, commercial, or industrial purposes, among others. Both aspects, intensity, and type of land use, are key under this approach. This is because they can unlock the potential use of a site induced by accessibility and thus increase value, according to the land rent theory. This instrument mostly responds to a direct rationale. An indirect approach, for instance to compensate extended development rights for the higher demand of services or infrastructure, would be understood as exactions or development fees. As conceived today, the arrangement of this instrument occurs voluntarily, e.g. developers bid for, trade, or directly buy additional development rights only if it is intended to develop a site beyond the 'basic' limit, as it is in CePACs in Brazil (H. Kim & Song, 2018). The appropriate timing of this instrument is when infrastructure already exists and a new development will take place. Sales of air rights are better suited to financing capital costs, whereas leases can aid both capital and O&M costs.

Joint development is a coordinated intervention that usually involves a mix of tools for its implementation based on a mutual recognition of transit infrastructure benefits between government (or transit agencies) and developers (Zhao, Das, & Larson, 2012). This instrument is closely related to, and often involves, lease of land and air rights grants in addition to other complements such as connection fees, that altogether promote real estate development around transit (Cervero et al., 2004). A key distinction of joint development, in contrast to leases of land or air rights, is that the public sector takes the lead by constraining and directing the main aspects of the project (Vadali et al., 2018). These types of arrangements are defined case-by-case and can include a variety of land uses and terms. The contributions by the developer are usually in-kind and cover some capital costs, for instance constructing a station, but also give room to negotiate O&M costs. These investments are recovered via the sale or lease of real estate which is usually under the administration of a private owner. When the public agency is not partnering (sharing risk and benefits with a private entity), the gains recovered by the public agency are limited to those of the cost of its implementation, which make it an indirect LVC instrument.

Land readjustment is both a LVC instrument and a land management tool which can be useful to restructure and/or consolidate urban areas, assemble land for public purposes (e.g. rights of way, parks, services, etc.), and supply public infrastructure (Muñoz Gielen & Mualam, 2019; Sorensen, 2000). Its application requires cooperative agreement among various landowners in adjacent areas, investors and/or developers, usually coordinated by a government agency. The process consists of various steps (Y.-H. Hong & Needham, 2007). Generally, it starts with recognition of potential to redevelop a site. This is followed by the development of a plan and design of the project. Later, the original land is pooled to re-develop it under the integrated project. During this stage, original landowners give up a portion of their land (previously agreed), which is sold after land is restructured and serviced to recover the expenses. When the development is complete, the landowner receives the redeveloped land which although smaller in size, it should be similar in value to the original property. The instrument is appropriate for new land development, but also can be employed to regenerate or upgrade existing urban areas. As shown, the main contributor is the landowner who provides the land. Even though this does not exclude the usual participation in cash by governments and the private investments. The timing in relation to the infrastructure provision is before improvements are in place. The rationale of the instrument does not necessarily respond to recover the 'unearned' income nor to the cost-recovery or other similar 'practical' objectives, but responds to broader objectives. Therefore, it is considered as a macro instrument (Alterman, 2012).

2.1.4 The missing link between LVC literature and land rent

LVC capture literature is well-developed from the perspective of urban planning and planning law (Vejchodská et al., 2022), as shown in the previous sections. This often assumes positive changes in value for a variety of reasons (i.e. public actions), overlooking the theoretical foundation. Land rent theory has been developed in the field of economics almost in parallel (ibid). Incorporating the view of the latter can be useful to confirm, support, and provide a theoretical framework to quantitatively measure the assumed changes. This can advance the discussion of the effective implementation of LVC instruments.

The next section addresses land rent theory and the principles of the residential land market, chiefly from the perspective of urban economics. This represents the formal framework which sets the foundation for assessing the potential of LVC in the context of public transport, in line with the main objective of this thesis.

2.2 Land rent theory and the residential market

A commonly acknowledged process determining the market value of land is that resulting from the anticipated amount of future income generated by the services provided by a piece of land (Freeman, Herriges, & Kling, 2014; McDonald & McMillen, 2010). In other words, the market value of land (V) is the present value of the *stream of income* earned by the asset, and it can be expressed as follows:

$$V = \frac{R_1}{1+r} + \frac{R_2}{(1+r)^2} + \frac{R_3}{(1+r)^3} + ... \tag{2.1}$$

The periodic income received by the landowner is known as *land rent* (R, *rent*, hereafter). R_1 is the rent for the first year, R_2 and R_3 are the subsequent income for the respective following years. r is the interest rate (in decimals) corresponding to a return equivalent to that earned in alternative investments of similar risk (McDonald & McMillen, 2010). This process is often referred to as the capitalized rent or capitalization of rents.

As observed, the market value of land reflects the magnitude of rent. The latter arises from two key notions: (1) that the supply of land (with particular attributes) is fixed; and (2) that the *productivity* of land is different across sites (McDonald & McMillen, 2010). Both internal and external characteristics of land may affect or define its productivity. Accordingly, the concept of rent is key to understanding how urban (dis)amenities are capitalized in the value of land. The present section expands these fundamental ideas from a theoretical point of view.

2.2.1 Early fundamentals on the theory of land rent

The differential land rent

Some key concepts outlined by David Ricardo remain useful foundations for the study of land rent according to the theoretical framework known as *differential land rent* (e.g. DiPasquale & Wheaton, 1996; Erba, 2013; Severen, Costello, & Deschênes, 2018). The underlying notion of this view is that the quality of land and its productivity varies across sites, resulting in differential rents. Although the original theory was framed around agricultural land given the time when these ideas were proposed (i.e. 1822), this framework is useful for highlighting important factors intervening in the land market. Here, the original model is introduced first and the subsequent subsections re-shift the focus of the discussion to the context of the present work.

The original differential land rent model notes the following assumptions (McDonald, 2018, p. 3):

- The farmland market is perfectly competitive, i.e. 'no landlord or renter has market power';
- Rent is the amount paid to the landowner for the use of the 'original and indestructible powers' of land;
- Matters of uncertainty and transition from one state to another are set aside, focusing on the long run;
- Capital and labour inputs are perfectly elastic (i.e. any amount is supplied at the same price);
- Demand (for the product of land) is perfectly inelastic (i.e. demand does not change in response to changes in price).

According to the model, in the initial stage land is free of rent (Ricardo, 2001). This would be possible if the demand for products of land is low relative to the capacity of production since there would be no need to compete for land. In this scenario, only the most productive land (N. 1) is cultivated. As the demand for a product increases, the activity needs to extend beyond N. 1 to land of lower quality, namely land N. 2 and land N. 3. In this case, the same amount of capital and labour gradually yields a smaller amount of product, i.e. 100, 90, 80, respectively. The difference between the produce of land n and the highest quality land (N. 1) will reflect the magnitude of rent. Namely, the rent in land N. 1 is 20, in N. 2 is 10, and 0 in N. 3. Considering this, cultivators are indifferent between paying rent or harvesting a smaller volume of crop in land of lower quality. A central aspect that should be noted from the previous is that rent arises because of the inelastic supply of land and the differentials in productivity among these types of land (McDonald, 2018).

Ricardo's model describes the differentials in terms of the fertility of agricultural land, which represent systematic *advantages*. This fundamental can and has been extended to other contexts including the urban (McDonald & McMillen, 2010). Namely, any special characteristic that makes a site more productive can drive the emergence of differential rent. This is the case of location (Fujita & Thisse, 2013), as will be discussed in the next subsection.

The spatial dimension of land rent

In explaining land rent, the main contrast to *standard* economic approaches has been the emphasis on the role of spatial features (i.e. distance and area) (Blaug, 1997; Proost & Thisse, 2019). This perspective is crucial for the present work since transport is required to overcome space (i.e. the geographic separation between two different locations). Thus, transport infrastructure is a fundamental component of this analytical approach (Proost & Thisse, 2019).

A paradigmatic contribution in the construction of the theory of land rent is provided by Heinrich von Thünen (1783-1850) (Frambach, 2012; Fujita, 1989; Fujita & Thisse, 2013). Following the same notion of Ricardo's rent theory based on fertility differences, von Thünen developed a model based on location differences, assuming homogeneous fertility of land (Todaro, 1978). Thus, the key assumptions in Ricardo's theory are essentially reversed in von Thünen's model (McDonald & McMillen, 2010). Specifically, the model devises a single town located in a featureless plain, surrounded by land of identical fertility, where the production costs are the same everywhere, there is no transportation infrastructure or navigable canal, and the only way of moving goods is using pulled carts which can move in all directions. Also, the system considers that exchange of products occurs only at the market located in the town and that this operates in a closed economic system (i.e. there are no imports or exports).

In this idealized setting, farmers close to the market incur lower transport costs. Since producers are responsible for shipping their products and the land is equally fertile everywhere, it is evident they will be willing to pay more for sites close to the market (McDonald & McMillen, 2010). Considering that the price of a product (Product 1) in the market is fixed and exogenous, the rent per unit of land (R) in the town will thus equal the transport costs compared to the furthest site of production. In this fashion, rent gradually decreases to zero at the edge of the cultivated land (the furthest location from the town) as distance from the market increases. The previous can be formally expressed as follows (Blaug, 1997):

$$R = e(p - a) - efk. \tag{2.2}$$

Here, R represents land rent, e the output (yield) per unit of land, p the fixed price per unit of product, a the cost of production per unit of product, f the transportation rate per unit of distance of the product, and k the distance from the market. The value of R in Eq. (2.2) represents the amount that farmers will be willing to pay for renting a site. This is referred to as the *bid rent*. As in Ricardo's model where farmers were indifferent to the various types of land quality after paying rent, producers are equally well-off in this model at different locations (in economic terms they receive *zero profits*). The result from the bid rent continuum as distance changes constitutes a *bid rent function*.

From Eq. (2.2), it follows that a further product (e.g. Product 2 and Product 3) with a different output per unit of land e incurring different transport costs will give rise to a distinct bid rent function (given the changed values of f and e, and potentially p and a too). This process introduces competition for land among crops since one product can outbid the other given the increased transportation cost and a resulting higher bid rent. Specifically, land will be allocated to the product with the highest bid rent, assuming that the new crop has a higher transportation cost. This results in a structure where the town is located at a central point surrounded by concentric rings, since farmers seek to minimize transport costs (efk) (Balchin, Isaac, & Chen, 2000; Von Thünen, 1966).

The novelty and relevance of von Thünen's model is given by the set of simplifying assumptions which highlight the role of proximity (or analogously, transport costs). in determining the use and rent of land. Specifically, this model gives rise to a trade-off between land rent and transport costs. This constitutes a fundamental concept which is still valid in contemporary literature (Fujita & Thisse, 2013; McDonald & McMillen, 2010; Proost & Thisse, 2019).

2.2.2 Residential land rent through the monocentric city model

Having defined the fundamentals of land rent, attention is now turned to the residential context. During the first half of the twentieth century, the study of land-use in the urban context gained attention in response to the intensive and rapid growth of urban areas (Fujita, 1989). A 'new version' of land rent theory was consolidated building on von Thünen's work (Anas et al., 1998). This perspective repositions the spatial factor at the centre of land rent. Also, it translates the core notions from the agricultural sector to the industrial, residential, and commercial sectors (Brueckner, 1987; McDonald, 2007). Another innovation worth noting is that this model shifts the gaze from the regional scale to the intra-city level. In the remainder of this section, the pivotal ideas on urban land rent are presented, keeping the focus on residential land use.

The prevalence of a single point of trade in a featureless plain in the revisited theory explaining urban structure led to what it is now known as the *monocentric city model*. This constitutes a core piece of theory in contemporary urban economics (Duranton & Puga, 2015; Proost & Thisse, 2019). This quantitative model is constructed from various contributions. One of the most influential is the work of Alonso (1964). Later, it was extended in separate efforts by Muth (1969) and Mills (1967, 1972). An important difference between these is that in Alonso's model, residents get benefits directly from land (McDonald & McMillen, 2010). In further extensions it is proposed that residents benefit from housing which is provided by the residential development sector. As a result, the unified monocentric city model (also referred to as the Alonso-Muth-Mills model or AMM model) consists of two broad complementary sides, namely: the consumer or household, and the housing developer or producer (Brueckner, 1987, 2011; Glaeser, 2008; McDonald & McMillen, 2010). The model outlined in this chapter draws on that presented by Brueckner (1987, 2011) and McDonald & McMillen (2010).

Model assumptions

The monocentric city model simplifies several of the details in the urban context. This is aimed at capturing the essential features of the urban structure. In the simplified city the basic assumptions are:

- Production and consumption takes place at a single point, referred to as the central business district (CBD). This is where all jobs are located. In the simplest version of the model, this is a dimensionless point that does not consume space. The location of it is given exogenously.
- The featureless plain includes a dense radial road network which minimizes the distance from every location to the CBD.
- In this city, all households are identical. These consist of a single resident of equal income that commutes to work in the CBD. This also implies that residents have the same preferences over consumption goods.
- Furthermore, the residents consume only two types of goods: a composite good and housing. The former is an abstraction that amalgamates the variety of commodities other than housing (e.g. food, clothes, etc).

As farmers are responsible for shipping costs in von Thünen's model (determining important land use patterns and bid rent), here, households are in charge of their commuting costs. These result from the distance from the resident's home to the CBD (x) and a cost per unit of distance t. Thus, the larger x, the higher the commuting cost is. Each household gets an equal income, y. Therefore, commuting costs at location x limit the money available to spend on goods. This relationship is given by the expression $y - tx$ and it is known as the household *disposable income*. This is a crucial aspect in the model since increased x implies a smaller disposable income. The implications of this are discussed below.

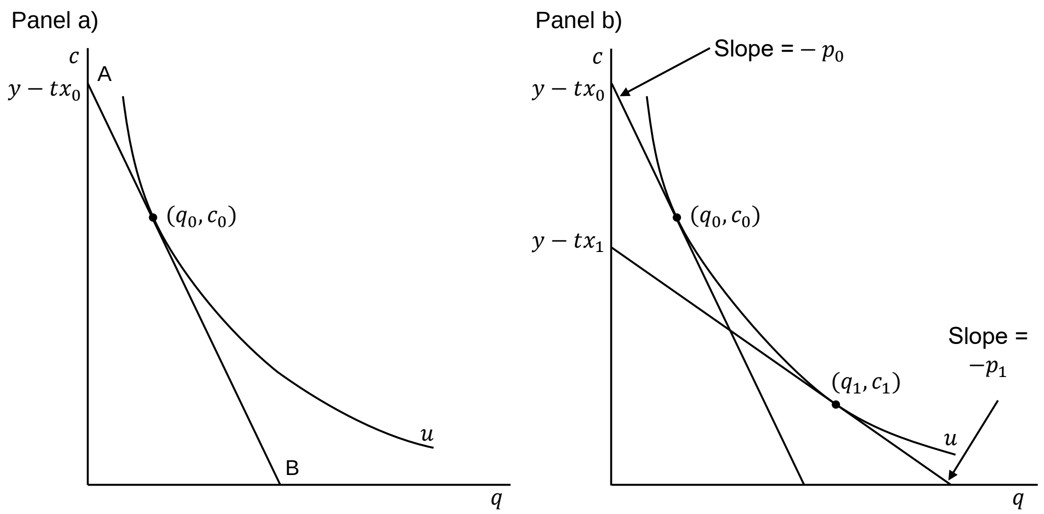
Household analysis

To develop the household side of the analysis, it is useful to elaborate on the only two goods assumed to be consumed by the household. First, the composite good is simplified to a common unit denoted by c. Also, this is normalized to a price of 1 per unit. Thus, it is not necessary to distinguish between the amount and price. Second, the consumption of housing (q) is complex because it embodies a bundle of characteristics (e.g. construction quality, age, amenities). According to the analytical framework, this is measured by a single unit. Often, the preferred measure is floor space. Having selected a unit for housing consumption, a rental price per square unit p can be attached to q.

With these elements in place, a *budget constraint* for the household is defined as

$$c + pq = y - tx. \tag{2.3}$$

The above implies that the sum of consumption of composite goods c and expenditure on housing pq equals disposable income $(y - tx)$. By trading-off consumption in c for q, a constant level of satisfaction can be achieved. The balanced continuum of this arrangement can be depicted in a so-called *utility function* (denoted by $u(c, q)$). Along this curve, the household is indifferent between consuming more c and having less housing space q, as illustrated by the convex curve (u) in Figure 2.1.



Source: the author based on Brueckner (2011, p. 30).

Figure 2.1: Consumer choice.

According to the theory, the household maximizes utility (the highest level of satisfaction) by consuming a particular combination of (c, q) subject to a budget constraint. This is illustrated in Figure 2.1 where the vertical axis represents the amount of c and the horizontal q. Consider a household that is located at x_0. For this resident, the budget constraint is given by $y - tx_0$ which is depicted by the line between A and B in Panel a). The optimal consumption bundle is the highest tangential point between the indifference curve and the budget constraint, namely $c(q_0, c_0)$. This represents a household that is located near to the CBD which consumes a high amount of the composite good but a small amount of housing.

The relationship described above among the composite good c, housing space q, and location from the CBD x, gives rise to a bid rent function in which the price per unit of housing decreases with a larger distance from the CBD. This is possible only under the following fundamental location equilibrium condition:

> *"Consumers must be equally well off at all locations, achieving the same utility regardless of where they live in the city."* (Brueckner, 2011, p. 28).

The relevance of this requirement is illustrated in Figure 2.1 Panel b). Here, a further household is located at $x = x_1$, which is at a greater distance from the CBD than the former resident, $x_1 > x_0$. Since commuting costs are higher, this household is left with a smaller disposable income, namely $y - tx_1$. Now, utility is maximized at the tangent point (q_1, c_1). For the second household to maintain the same level of utility, the price per unit of housing should decrease as x increases. This compensates for the increased commuting costs (and reduced disposable income) following the condition of spatially uniform utility. A further aspect arising is that as p decreases (with distance), the household consumes a larger amount of housing space q.

As observed in Panel b) of Figure 2.1, the slope at the suburban location $(-p_1)$ is flatter than the central location $(-p_0)$. Expanding this implication to the multiple locations while maintaining the same level of utility generates the so called *bid rent function* for housing. Accordingly, the price peaks near the CBD and gradually diminishes at distant locations, decreasing at a fast rate first and slowing down towards the edge of the city. The exact slope can be found by differentiating the maximum utility subject to the budget constraint with respect to x, and it is expressed as[2]

$$-\frac{\partial p}{\partial x} = -\frac{t}{q}. \tag{2.4}$$

This relation confirms the intuitive result presented above. Specifically, the ratio considering a constant transport cost rate t denotes a steep slope at a short distance from the CBD where the space of housing consumed q is relatively small compared to suburban locations. Also, it implies that the decline in housing price as one moves further away from the CBD should be offset by the increase in the cost of transport.

Housing production analysis

Having reviewed the household side, attention is now turned to the housing production component. Specifically, this subsection focuses on the activities of housing developers, who are assumed to rent housing space to households.

[2] In Brueckner (1987)'s analysis, totally differentiating $\upsilon(y - tx - pq, q) = u$ yields

$$-\upsilon_1\left(t + \frac{\partial p}{\partial x}q + p\frac{\partial q}{\partial x}\right) + \upsilon_2\frac{\partial q}{\partial x} = 0.$$

Since $\upsilon_2 = p\upsilon_1$, the partial derivative is $\frac{\partial p}{\partial x} = \frac{-t}{q} < 0$.

The objective of developers is to produce housing floor space, Q. The inputs can be reduced to two essential components: capital K and land l. The former is an abstraction of the multiple elements that producers require in reality, i.e. building materials (the basic model ignores the role of labour and machinery). The combination of these outputs is denoted by the production function $Q = H(K, l)$.

In the production function, two further economic assumptions should be noted: (1) diminishing returns of capital; and (2) constant returns to scale. The former refers to the fact that although land can be substituted for capital to a certain extent to produce the same amount of Q (by expanding vertically), the rate of return gradually decreases. The second property means that scale economies are not relevant in this process. In other words, doubling the inputs does not lead to more than doubling of floor space. Under these circumstances, the developer chooses the combination of capital and land to maximize profits.

Since the developers rent the housing to households, their revenue is obtained by the produced floor space multiplied by the household's bid price for housing, denoted by $pH(K, l)$. Following the rental orientation of the model, the producers rent capital and land. The rental price per unit of capital is denoted by i and for land by R. Thus, the production cost is given by $iK + Rl$. Accordingly, the developer's profit is given by:

$$\text{Profit} = pH(K, l) - iK - Rl. \tag{2.5}$$

Here, the profit is simply determined by the revenue minus the production cost. While p is higher the closer to the CBD, the rental price per unit of capital i is assumed not to vary in space. Thus, the profit would be higher in central locations than in the suburbs in the absence of competition among producers.

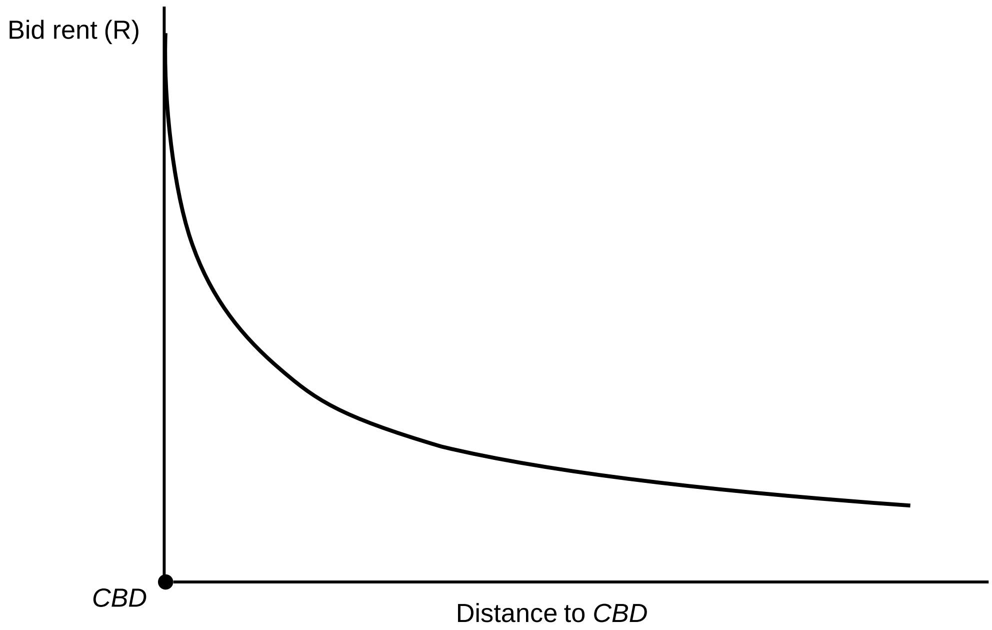
Yet, in an open market developers compete by offering additional land rent for locations which lead to higher revenue, until they make zero economic profits in the long run (still, the inputs in production make normal rates of return). This leads to a second spatial equilibrium condition in which

> *developers are equally well-off and thus indifferent between sites for producing housing.*

The bid rent per unit of land can be found by solving for R in Eq. (2.5) and setting profits to zero, as shown below:

$$R = \frac{pQ - iK}{l}. \tag{2.6}$$

From the above, it is evident that the bid rent per unit of land falls with increased distance from the CBD. Figure 2.2 illustrates the bid rent function for land. Qualitatively, this curve is analogous to that of the price per unit of housing. Yet, the former falls at a higher rate than the latter. This is because of the possibility of substituting l for K in the production function while sustaining Q at the same level. In line with production theory, as the price of one input increases developers tend to substitute this for other inputs. High land rents close to the CBD incentivise the shift away from land in favour of capital. This has two consequences. First, the savings in land would lead to increased profits in the short run. However, competition via bid rent for land ensures that any extra profits are exhausted in the long run. Thus, the higher the capital-to-land ratio (K/l), the more convex the land rent curve will be. Second, the increased capital and reduced amount of land results in taller buildings which translates into higher densities near the CBD. This is a further regularity predicted by the model which qualitatively resembles that of p or R with respect to location.



Source: the author based on McDonald and McMillen (2010, p. 102).

Figure 2.2: Bid rent function for land.

The two main observations from the developer's analysis are the following: land per unit falls as distance from the CBD increases; similarly, density decreases with distance from the CBD.

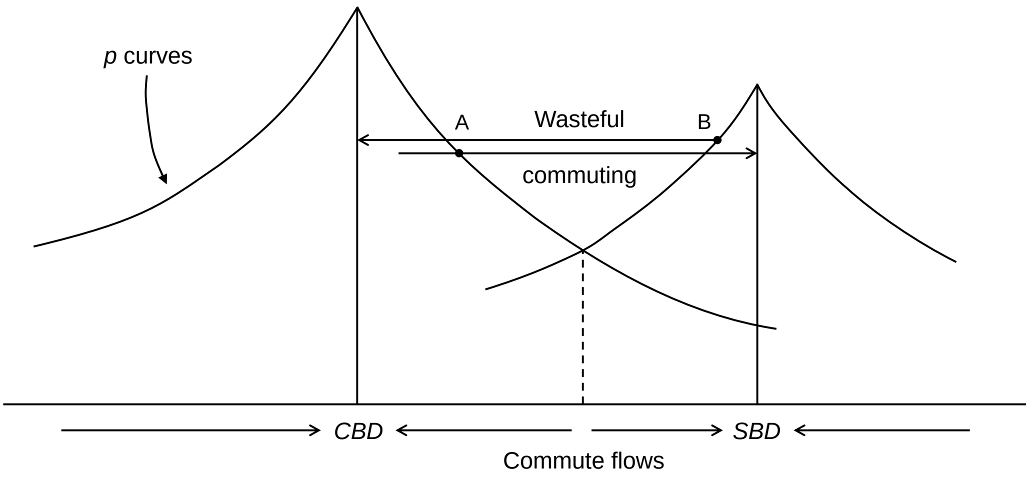
2.2.3 Employment outside the CBD

An obvious limitation and consequent major critique of the AMM model is its monocentric structure (Duranton & Puga, 2015). Although employment and economic activities tend to concentrate in urban areas, there is employment located outside the CBD to various extents. This observation has given rise to several studies examining the implications for the basic AMM model. Brueckner (2011) identifies the following cases: (1) dispersed employment; (2) employment subcentres; (3) employment decentralization and spatial mismatch; (4) commuting in the information age.

In the first and fourth case the implications of the basic model remain practically unchanged. The argument for the first case relies on a further regularity of urban economics known as the *wage gradient* (Chapelle, Wasmer, & Bono, 2021; Eberts, 1981; McMillen & Singell, 1992). This states that wages decline as proximity to the CBD falls. Thus, the savings in commuting resulting from a job located outside the CBD are compensated by a lower wage and the monocentric structure prevails. Teleworking is becoming common in the information age, as discussed in the fourth case. Essentially, this implies a fall in transport costs given the reduction in the number of days that commuting is required (Rhee, 2009). Accordingly, the implications can be analysed using the standard AMM model, given a reduction in the parameter t, e.g. lower density, and an increased urban area covered (the latter effect referred to as *sprawl*).

Employment decentralization occurs when firms choose to locate in the suburbs (Gobillon & Selod, 2021; Gobillon, Selod, & Zenou, 2007). One of the arguments for this decision is based on lower transport costs derived from transport improvements and the avoidance of high land rents near the CBD. This results in some workers relocating near to suburban jobs, further attracted by the capacity for higher housing consumption. However, disadvantaged residents are left behind since they are outbid by higher income ones. One of the main results is the *employment mismatch* in which poor households face longer commutes and are more likely to be unemployed.

The employment subcentres case is of special relevance for the present work. The existence of secondary business districts (SBD, also referred to as subcentres in literature) is well-documented in literature (Bartosiewicz & Marcińczak, 2020; Giuliano & Small, 1991). This can be approached by extending the logic of the basic monocentric city model (Brueckner, 2011). Effectively, subcentres are joined to the main CBD as a second city resulting in a *polycentric* city, as illustrated in Figure 2.3. The scheme shows that residents will commute to the nearest centre, either the CBD or SBD. The approach also allows for different incomes paid in each of the centres. This can generate different p curves. Furthermore, the area of the CBD would be larger than that of the SBD, and thus so would the distance to the respective edge of each city. The main observation is that the price of housing p along with land rent r declines at greater distance from each of the centres.



Source: the author based on Brueckner (2011, p. 59).

Figure 2.3: Polycentric city.

The polycentric city model gives rise to an observation known in the literature as wasteful commuting (or excess commuting) (Rouwendal, 2021), also depicted in Figure 2.3. This occurs if a resident located at A commutes to the SBD and a second in location B commutes to the CBD instead of to the nearest centre, as assumed by the polycentric model. This is considered to be wasteful because residents could swap locations to optimize their travel patterns. Empirical data suggest that a higher proportion of residents engage in longer commutes than they *should* do, questioning the rationale of the monocentric city model (Hamilton & Röell, 1982).

Research has responded to this observation with several explanations, including the potential subjectivity of measuring 'wasteful' commuting (Kanaroglou, Higgins, & Chowdhury, 2015). Two relevant approaches for the present work are the *multiple household worker* and *employment uncertainty*. The former explains the phenomenon by elaborating on information about a second working member of the household, who may influence the location choice (Surprenant-Legault, Patterson, & El-Geneidy, 2013). Meanwhile, the second approach suggests that households do not only consider their current job to make a residential location decision, but also the uncertainties of the employment market. Crane (1996) puts it as follows:

> "The ultimate choice of a place to live is therefore based not only on the current job, but also on the expectation of (i) where the next job will be, and (ii) how often the household will choose to move for other reasons." (p. 343)

This view effectively introduces a spatial probabilistic aspect of employment which is not explicitly discussed in the monocentric city model.

The monocentric city model continues to represent an important tool in urban analyses. Even the most basic model establishes the ground for foreseeing the effects of changes in its parameters. This is important for the present work because it sets the basis to hypothesize about the implications of changes in transport costs or analogous changes in accessibility. Furthermore, although the basic model includes strong simplifying assumptions, it constitutes the foundation for extensions incorporating additional elements of reality, e.g. travel time in commuting costs to the CBD, two-income groups, household structure, or employment uncertainty (e.g. Brueckner, 2011; Duranton & Puga, 2015; Fujita, 1989).

The following section focuses on the measurement of accessibility beyond the classic simplification often adopted in urban economics (i.e. distance to the CBD) as a way to address some of the limitations noted in literature.

2.3 Accessibility

As discussed in the previous section, a long-standing framework explaining the value of land has been the differential rent established for the intra-urban context in the monocentric city model. Differential rent can arise in the residential context for various reasons (e.g. land use restrictions, quality of air, or quality of amenities), one of the most important being location with respect to employment (DiPasquale & Wheaton, 1996; McDonald & McMillen, 2010). In the exposition of the model commuting costs (measured as distance to the CBD) equates to the inverse of accessibility (Duranton & Puga, 2015; Proost & Thisse, 2019). The concept of *accessibility* provides a theoretical framework to characterize location in the urban context from a richer perspective than that referred to in the basic monocentric city model Gjestland et al. (2020). This broader view can address employment dispersion, as discussed in Section 2.2 and in literature (Duranton & Puga, 2015).

The present work adopts a comprehensive conceptualization of accessibility for the assessment of the potential changes in the value of residential land, relevant to support the LVC approach in the context of public transport. Additionally, it extends this view by employing Open Science resources, such as open-source software and open access data. This may be particularly relevant for the study of public transport systems given their multiple physical and operational characteristics which affect the level of accessibility.

This section first reviews the concept of *accessibility* in literature. Then, it elaborates on the measurement of it. It proceeds by focusing on location-based measures and some important elements that characterize them, e.g. impedance functions and the calibration of its parameters. Furthermore, this section discusses the contributions of introducing location-based accessibility measures to account for employment located outside the CBD as an extension of the monocentric city model in the residential land market. Lastly, the relevance of these measures is discussed in the context of public transport systems, connecting the concept of accessibility with the main objective of this thesis.

2.3.1 The concept of accessibility

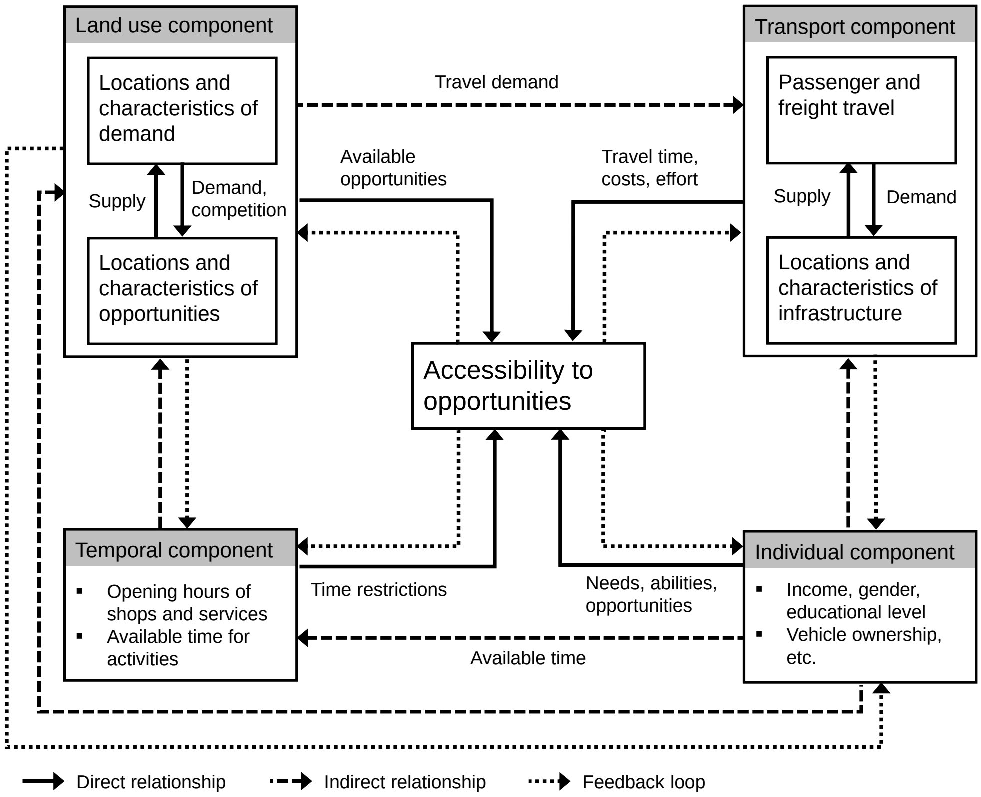
Conceptualizations of accessibility include a variety of elements of the regional/urban layout. A paradigmatic contribution was proposed by Hansen (1959), defining it as "the potential of opportunities for interaction" (p. 73). This idea shaped an important tradition in the literature drawing on two basic elements: (1) the 'spatial distribution of activities' and (2) the 'ability and desire' to overcome spatial separation between these. Later, Geurs & van Wee (2004) referred to accessibility as "the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)" (p. 128). Similarly, for Páez et al. (2012), the notion is described as "the potential for reaching spatially distributed opportunities while considering the difficulty involved in travelling to them" (p. 141).

More recent literature emphasizes further aspects involved in spatiotemporal interaction. For example, Thériault et al. (2013) conceptualize accessibility as "the ability that persons living in that area can, within an acceptable travel time, reach a satisfactory diversity of activity locations that are important to them" (p. 234). Levinson & Wu (2020) simply put it as "the ease of reaching valued destinations" (p. 130). A good degree of overlap can be noticed between both mobility aspects and the location and characteristics of opportunities for individuals, which incorporate time considerations. As such, this can be defined as "the extent to which land-use and transport systems enable (groups of) individuals to reach activities or destinations by means of a (combination of) transport mode(s)" (Geurs & van Wee, 2004, p. 128).

Geurs & van Wee (2004) delineate a conceptual framework that accommodates some of the key components referred to in the definition of accessibility. This is schematically illustrated in Figure 2.4.[3] According to this analysis, accessibility to opportunities is directly shaped by four components, namely: (1) land use, (2) transport; (3) temporal; and (4) individual. The first component on the top-left captures the spatial dispersion of both demand (e.g. users or residents) and opportunities (e.g. employment or shopping). In addition to location, the characteristics of these are sug-

[3] For the purpose of the present review, the attention is on the fundamental elements distinguished within the concept. This differs from the emphasis on potential relationships between components, as in the original work.

gested to influence their availability. The supply and demand (or competition) between users and opportunities are shown to play a role too.



Source: the author based on Geurs and van Wee (2004, p. 129).

Figure 2.4: Relationship between components of accessibility.

The top-right quadrant in Figure 2.4 shows that the transport component depends on passengers and freight travel as well as the location and characteristics of infrastructure. Demand and supply may also play a role, e.g. road congestion may affect travel time, cost, or effort. Land use indirectly influences this component. The individual component, on the bottom-right, influences accessibility according to the need, abilities, and opportunities of people. This includes aspects such as income, gender, educational level, or vehicle ownership. This is indirectly related to the transport and land use component, e.g. by affecting the supply-demand balance. The temporal component, on the bottom-left, represents time restrictions according to opening hours of shops and services, for example. Those restrictions are indirectly related to the land use component. Lastly, there are some feedback loop effects suggested. For instance, each of the main components can respond to the level of accessibility to opportunities.

2.3.2 Operationalizing accessibility measures

There are several relevant pieces of literature contributing to the establishment of a classification of accessibility measures (e.g. Dumolard, 2013; Geurs & van Wee, 2004; Handy & Niemeier, 1997; Janić, 2019; Lei & Church, 2010; Páez et al., 2012; M. Ryan et al., 2016). Table 2.2 summarises the main forms identified in earlier reviews including a general description of each and some examples applied in empirical research. A broad distinction between the type of measures included in the present analysis is according to two of the main components identified in theory. Namely, (1) measures that only consider aspects of the transport component (referred to in the table as 'Infrastructure- or spatial configuration based') and (2) measures that include the both the land use and transport component (referred to in the summary as 'Land use and transport infrastructure-based'). A second aspect emerging from this review is the effort to reconcile the different labels used interchangeably in literature. The alternative names used in literature are entered using the forward slash character ('/').

Table 2.2: Accessibility measures

Type	Subtype	Description	Examples and applications
Infrastructure- or spatial configuration-based			
Infraestructure-based[2]/ Spatial separation[7]		Chiefly concerned with the spatial separation of activities based on the state of the infrastructure.	One or more travel time or cost measures, e.g. level of congestion or average travel speed on the road network. See further examples below.
	System accessibility[5]	Focused on the physical proximity to some components of a transport system.	Time to closest motorway (Mulley, 2014); Walking distance to closest transit station (M. Ryan et al., 2016); Presence of a transit line within a quarter-mile (Azar, Ferreira, & Wiggins, 1994).
	System facilitated[5]	Considers an origin and a destination and the cost associated to travel between these two points.	Transit time based on average speed (S. Liu & Zhu, 2004); Shortest path distance (O'Sullivan, 2012).
	Relative[5]	Considers competing modes or type of users, usually computed as cost relative to other modes.	Ratio between travel time by transit and driving (Lei & Church, 2010).
	Graph theory[1]	Focus on the topological properties of the infrastructure represented as a network.	Space syntax, topological distance of road network (Hillier, 2007); In rapid transit network: ratio between links and vertices (Kim & Song, 2018).
	Surfaces of friction[1]	Based on 'spatial impedance' or 'spatial friction', characterized because it considers the whole surface (in contrast to network analyses), but it also considers 'barriers' (topography, costs, or general environment) and infrastructure (roads, speed, etc.)	Knight's case movement (Delamater, Messina, Shortridge, & Grady, 2012). It is an index of travel cost in raster GIS surface using cell's dimension and weights assigned.
Land use and transport infrastructure-based			
Location-based[2]/ Integral[5]/ Interaction between users and infrastructure[1]		A broad family of measures which includes at least a form of activities and spatial separation in the context (or origin and destination).	See examples below.

(Continued on next page...)

Table 2.2: Accessibility measures *(continued)*

Type	Subtype	Description	Examples and applications
	Cumulative[3,6]/ contour[2]/ isochrones[4]	The sum of opportunities that can be reached within a fix travel threshold.	Number of jobs by economic strata and opportunity type reachable within different time thresholds (Pereira, 2019).
	Gravity-type[3,4,5,6,7]/ Potential accessibility[2]	Estimates the size of opportunities from one area to all others, weighing them by a function of travel cost.	Accessibility to employment weighed by an exponential function of distance (Hansen, 1959; Osland & Thorsen, 2008).
Person-based[1]/ Constraints-based[7]/ Space-time geography[5]		Unlike aggregated approaches in location-based measures, this is focused at the individual (household) level, and does not assume a single origin.	Feasible opportunity set given by the space-time constraints (Kwan, 1998); Total accessible opportunities mediated by available service time (Kim & Kwan, 2003; Miller, 1991).
Utility-based[2,3,4,5,6,7]		Assessed (economic) benefits that people derive from the access to activities. Utility may be affected by the characteristics of individuals and transport options.	The denominator of the multinomial logit model, known as the logsum (Niemeier, 1997); Expected benefits per trip generated in doubly constrained entropy models (Martínez & Araya, 2000).
Accessibility by distributed artificial intelligence[1]		Based on the relationship between agent's behaviour and rules of territorial dissemination in automated processes.	Agent's variable perception over time in meta-multiagent mode (Banos & Thévenin, 2013; Zöllig & Axhausen, 2012).

Source:
1 = Dumolard (2013); 2 = Geurs and van Wee (2004); 3 = Handy and Niemeier (1997); 4 = Janić (2019); 5 = Lei and Church (2010); 6 = Paez et al. (2012); 7 = Ryan et al. (2016).

The first panel in Table 2.2 groups the measures dealing with infrastructure or spatial configuration only (Geurs & van Wee, 2004). These have also been referred to as measures of mobility, which only denote the costs of travelling. Thus, it is argued that this conceptualization provides a partial perspective by overlooking the urban opportunities and services (El-Geneidy & Levinson, 2006; Levinson & Istrate, 2011). This type includes metrics such as system accessibility or system facilitated accessibility (Lei & Church, 2010) which refer to the presence of transport infrastructure or to the degree it enables mobility, respectively. Related to the previous, the relative forms consider different transport modes as competitors, which can be operationalized through travel time ratios of different modes, for example. Other types of measures in this cluster draw on the concepts of graph theory. This is chiefly concerned with optimization of the network as in shortest path algorithms (Dijkstra, 1959) or the topological distance given by the configuration of the road network or the built environment in Space Syntax modelling (Hillier, 2007). A different approach is associated with the analysis of surfaces of friction. This considers the territory as a continuous layer recognizing natural and built environment constraints as well as facilitating infrastructure. This last approach has been achieved by considering travel costs in uniform grids (raster) in geographic information systems (GIS) (Dumolard, 2013). As noted, the previous computations do not explicitly consider, or sometimes assume, the location and characteristics of opportunities, and focus on the ease of movement.

The second broad set of accessibility measures in Table 2.2 considers at least both the land use component and the transport infrastructure in some form. Some of this also considers the individual and temporal components. A grouping of these measures is not clear-cut in literature. Whilst some authors have made a sharp distinction between location-based and person-based (Dijst, de Jong, & van Eck, 2002; Geurs & van Wee, 2004; Kwan, 1998), others have suggested an implicit character and simple variation of the approaches (Handy & Niemeier, 1997; Páez et al., 2012). For the purpose of the present review, person-based measures are considered in a separate group to location-based ones.

The first type of measure within the land use and transport infrastructure-based grouping is location-based (Geurs & van Wee, 2004; Kwan, 1998), alternatively referred to as integral (Lei & Church, 2010), or measures of interaction between users and infrastructure (Dumolard, 2013). For consistency with most of the literature, the remainder of this work refers to these as *location-based* measures. The most common form of this kind is the so-called cumulative opportunities measure (*cumulative*, henceforth), also found under the contour or isochrone labels. This essentially aggregates the total number of opportunities reachable within a travel threshold. Similarly, gravity-type, or potential accessibility, assesses the opportunities available from one location, discounting their importance by the travel cost required to reach them. Unlike the previous, space-time geography, person or constraint-based metrics are characterized by the establishment of a viewpoint from the individual considering spatial and temporal constraints. Furthermore, these recognize individuals as mobile agents. Thus, it conceptualizes accessibility as a dynamic value and not a constant from a fixed location of departure. The last has been substantially less operationalized in empirical research, possibly due to the fine-grained data required.

A further measure in the same grouping is the utility-based approach. This is developed from the perspective of economics, viewing individuals as consumers and potential destinations as a set of choices (Lei & Church, 2010). In general, it aims to estimate the expected maximum utility that a person would obtain from a transport system (Nassir, Hickman, Malekzadeh, & Irannezhad, 2016). Two methods for implementing this approach are found in literature (Geurs & van Wee, 2004), namely: 1) according to the random utility theory (via the logsum) (e.g. Handy & Niemeier, 1997; Nassir et al., 2016); and 2) based on a doubly constrained entropy model (Martínez & Araya, 2000). Lastly, some methods based on decentralized artificial intelligence (DAI) have been recognized as a flexible and dynamic alternative (Torrens, 2003, 2012). Essentially, the analytical unit can be at the individual (or household) level, represented by independent agents which simulate human behaviour based on a set of characteristics, for instance, rules of behaviour, demographics, interaction with other agents, and the environment (Dumolard, 2013). Common techniques within this approach are cellular automata and multiagent system modelling.

The remainder of the present work focuses on location-based measures. This is because of their (1) consistency in relation to theory, (2) their adaptability to incorporate further dimensions (e.g. not only land use and transport components but individual and time, as will be shown later), (3) ease of communication and computation, and (4) relatively accessible data requirements (Geurs & van Wee, 2004; Páez et al., 2012; Pereira et al., 2019).

2.3.3 Location-based measures

This subsection expands the review on location-based measures. Although the specific focus is on the subtype known as gravity-type, cumulative measures are briefly discussed for convenience as a way to introduce the former. The cumulative type is defined by the sum of opportunities available within a travel threshold from one origin (Levinson & Wu, 2020). The conceptual formulation is sensible in relation to the analysis of travel behaviour since it incorporates individuals' limitations within their ability to move in space. Additionally, it incorporates the notion of matching probabilities for services or a job from a particular point of departure. A common mathematical formulation is given by the following expression (Kwan, 1998):

$$A_i = \sum W_j f(\cdot), \text{ subject to } f(\cdot) \begin{cases} d_{ij} \leq \bar{t} \\ 0 \text{ otherwise} \end{cases} . \tag{2.7}$$

Here, i and j represent an origin and destination, respectively. A_i is the cumulative accessibility for location i. W_j represents the attractiveness of opportunities at potential destinations, e.g. number of jobs, shops, etc. Here, a travel threshold is represented by $\bar{t}$. This is the maximum distance or time that a person is willing to travel between i and j to enjoy opportunities W. This type of constraint is referred to as a *rectangular* function (Kwan, 1998). Here, opportunities within the threshold are evaluated equally regardless of the travel cost between i and j.

Cumulative measures can offer a range of advantages when compared to the alternatives available to assess accessibility. One major advantage is that it can be easily communicated (Geurs & van Wee, 2004; Pereira et al., 2019). Inputs and outputs used are often in terms of the number of jobs and/or services available within distance or time spent, which are easy to interpret due to their familiarity to policy makers and stakeholders. Another important characteristic is related to its operationalisation, which has been recognized as an easy to calculate index (Lei & Church, 2010).

Despite the advantages mentioned above, this approach has limitations. The main is that it assumes homogeneous desirability of opportunities in relation to travel costs within the travel threshold. It can be argued that people may be willing to travel further (or less) depending on the type and size of the opportunity. Some constructions have attempted to address this limitation by discounting opportunities by a negative linear function according to the travel cost required to reach them (Kwan, 1998). A further drawback is that the analyst must select a threshold which may be considered as 'arbitrary' (Geurs & van Wee, 2004; Niemeier, 1997; Pereira et al., 2019).

Gravity-type measures are part of the broader location-based measures family together with the cumulative measure (Páez et al., 2012). Thus, their specification and inputs look similar. The original concept of modelling spatial interaction based in a Newtonian analogy can be traced back to Stewart (1947). Initially, it was argued that the idea of attraction force between two bodies mediated by their distance could be useful to explain some of the dynamics in social sciences. By the end of the 1950s the construction was formalized in empirical applications for land-use and transport interaction analyses as an accessibility measure by Hansen (1959). The original accessibility specification by Hansen is expressed as

$$A_1 = \frac{S_2}{T^x_{1\sim 2}}. \tag{2.8}$$

Here, accessibility in *Zone 1* is denoted by A_1. This is given by the activity in *Zone 2*, represented by S_2. S represents the size of the activity, usually measured by number of jobs, inhabitants, or retail supply. In general, this is known as a measure of *attractiveness*. $T_{1\sim2}$ expresses the travel cost between Zone 1 and Zone 2 (time in minutes, in Hansen). Finally, in this formulation x is a deterrent of travel cost, to be estimated. When more than two zones are considered, the formulation above can simply be extended as following:

$$A_1 = \frac{S_2}{T^x_{1\sim2}} + \frac{S_3}{T^x_{1\sim3}} ... + \frac{S_n}{T^x_{1\sim n}}. \tag{2.9}$$

Nowadays, there are several extensions drawing on Hansen's original idea (Levinson & Wu, 2020). These can accommodate multiple components and parameters. The general formulation to construct location-based measures consists of the sum of the product of two functions, namely $g(\cdot)$ and $f(\cdot)$, as shown in the following equation (Páez et al., 2012):

$$A^p_{ik} = \sum_j g(W_{jk}) f(c^p_{ij}). \tag{2.10}$$

In this expression, accessibility A is viewed from a fixed location i (considered to be the origin) to opportunities of type k, from the perspective of person type p. As outlined previously, the measure reflects the attractiveness of opportunities W at location j of type k. Finally, the term c_{ij} refers to a function of travel cost between i and j, as experienced by person p. As shown, both opportunities and travel cost can take several ad hoc forms.

Páez et al. (2012) suggest that the generic formulation in Eq. (2.10) can also be acknowledged as a person-based measure since it incorporates: (1) individual's profiles; 2) how different profiles perceive or experience the space; and 3) constraints that may prevent segments of the population to access some type of activities. Some of these individual aspects involved in the construction of accessibility measures have been found to be relevant in empirical studies examining the residential market. For instance, gender in Osland (2010a) or age and gender in Thériault et al. (2013). The empirical studies are reviewed in depth in Chapter 3.

Gravity-type measures can overcome some limitations presented by cumulative ones. First, these do not require a time threshold or cut-off. In addition, their formulation is still relatively easy to understand and calculate (Geurs & van Wee, 2004; M. Ryan et al., 2016). A further characteristic is that this type of measure can evaluate the combined effect of opportunities and the transportation system (Geurs & van Wee, 2004), since the magnitude of opportunities is regulated by the travel costs. Finally, this type of measure is able to incorporate individual perception by using an ad hoc function of distance decay for different cases of study (a point that will be discussed in more detail below), or by recognizing the different demographic profiles, among other techniques.

Notwithstanding the adaptability and the wide range of applications of potential or gravity-type measures, these come with limitations. To start with, these measures are difficult to communicate when compared to cumulative ones (Geurs & van Wee, 2004). This may be particularly challenging when the targeted audience includes policymakers, or nonspecialized stakeholders.

Another aspect to be considered is the level of spatial aggregation of the data inputs (e.g. census tracts or traffic analysis zones). Spatial aggregation is worth discussion for two reasons (Torrens, 2003, 2012): the selected spatial areas can fall in the modifiable areal unit problem (MAUP); and the ecological fallacy. One aspect related to the MAUP is the scale effect, which is associated with the number of areal units used to analyse a given zone. This can result in divergent effects in relation to statistical descriptions (e.g. same mean values and declining variance) which can translate into a loss of information (Dark & Bram, 2007). The second aspect of MAUP is related to the effects of zoning, resulting fundamentally from the way in which the spatial analytical units are geometrically defined (e.g. orientation, size, proportion, etc.). Here, the outcomes can differ unexpectedly for every modification of the aggregated unit, which can lead to misleading inferences about the population contained in the areal unit (Dark & Bram, 2007). Since the MAUP is a ubiquitous concern in spatial analysis, this discussion is extended in Chapter 3 in light of empirical literature.

Meanwhile, the ecological fallacy is particularly concerned with inferences made from aggregated information which may require the assumption of homogeneity of individual observations (Openshaw, 1984). Certainly, those aspects should be addressed in this type of research. Empirical examples have shown the extent of the effects of some of these issues, encouraging the development of good practices such as sensitivity analyses to reach informed conclusions or decisions in policy making (Pereira et al., 2019; Tan & Samsudin, 2017).

Impedance functions

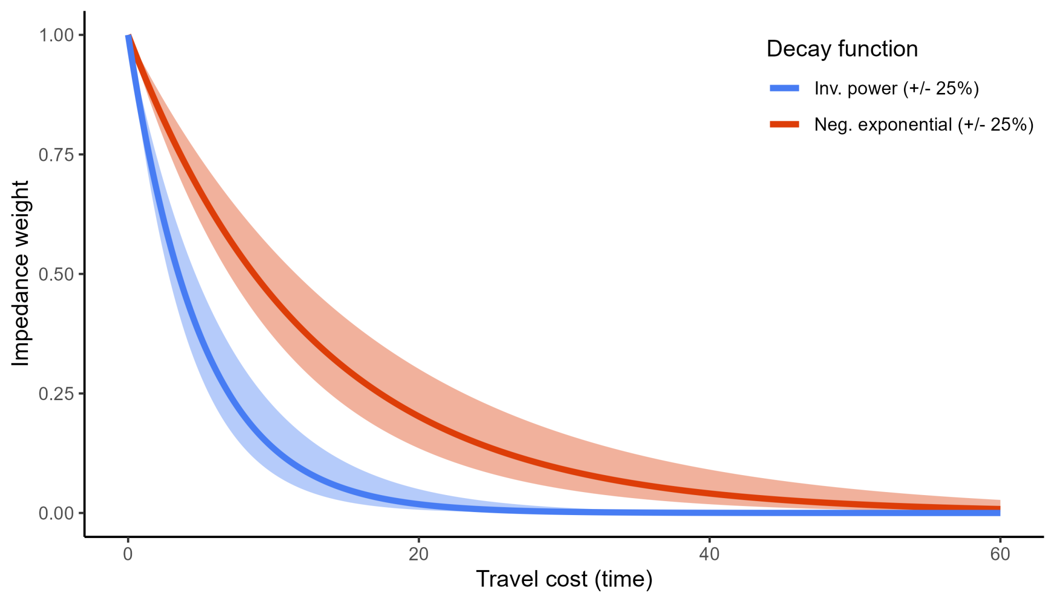
This section discusses the functional form that travel costs can adopt in Eq. (2.10) (i.e. $f(c_{ij})$). This term is referred to as *impedance*. Broadly, this represents the way in which people discount the relevance of opportunities as travel costs increase. Accordingly, opportunities in close proximity are considered to be attractive, while those further away become negligible. It should be noted that travel costs consist of multiple aspects including physical characteristics of the environment, individuals' perceptions, pecuniary costs, abilities, and profiles. In practice, these have been represented by a variaty of measures, such as Euclidean distance, shortest path distance in a network, travel time, or a combination of these. Among these, travel time is preferred over distance, given the ability to reflect characteristics of the mode, infrastructure, as well as changes in demand and supply (Miller, 2018).

In practice, an *impedance* function $f(c_{ij})$ consists of two basic elements: (1) a decaying functional form and (2) one or more parameters regulating the rate of discount associated to distance (commonly referred to as a distance deterrence or spatial decay parameter). These elements are grounded on travel behaviour theory (Handy & Niemeier, 1997) and are aimed at *correcting* the estimates based on observed action.

Regarding the functional form representing spatial decay, literature often refers to two main types (Higgins, 2019; Ortúzar & Willumsen, 2011; Rosik, Stępniak, & Komornicki, 2015):

- $f(c_{ij}) = c_{ij}^{-\beta}$ the inverse power function, and;
- $f(c_{ij}) = e^{-\beta c_{ij}}$ the negative exponential.

Here, β is a parameter to be estimated and is discussed in more detail below. Generally, the inverse power function curve is characterized by a rapid rate of decline when compared to a negative exponential, as qualitatively illustrated in Figure 2.5. In practice, it has been observed that the exponential function is often used for regional analyses (long distances), whereas at metropolitan or local level both functions have been employed (Rosik et al., 2015).



Source: the author based on Higgins (2019).

Figure 2.5: Impedance functions.

In essence, β should reflect people's perception of space between an origin and a potential destination (Handy & Niemeier, 1997). These are useful to adapt to each context of study. Here, the smaller the value, the less sensitive a commuter is assumed to be to travel costs. In Figure 2.5, the lower bound of the shaded area of each of the curves represents more sensitivity to travel costs than the upper bound.

β has taken different values in the negative exponential function when estimated empirically. It is argued that this parameter can vary according to the scale of the area of study. In a review of the β parameters employed in research at different spatial scales, Rosik et al. (2015) found the following values for a negative exponential function: at the European level this is between 0.0058 and 0.010 (by car); at the national between 0.009 and 0.049; and at the local, from 0.068 to 0.289. In the context of the residential land market, Ahlfeldt (2013) observed similar empirical values according to a negative exponential function within a range between 0.05 and 0.10 for Rogaland (Norway), Berlin (Germany), and London (U.K.). For the inverse power function, the optimal β values tend to be greater than 2 (e.g. Ahlfeldt & Wendland, 2016; Osland & Thorsen, 2008).

The decaying functions presented above are monotonic in shape. This implies that the attractiveness of opportunities is assumed to constantly diminish as distance increases. The adequacy of a nonmonotonic function is discussed in Chapter 3.

Calibration of parameters

Literature suggests that the value of β should ideally be estimated based on empirical data reflecting ‘typical’ travel patterns (Levinson & Wu, 2020; Stępniak & Rosik, 2018). Different modes and travel purposes can affect the strength of distance decaying parameters (Tahmasbi & Haghshenas, 2019). An optimal parameter can be estimated according to a trip-distribution problem using aggregate commuting flows obtained from local travel surveys (Handy & Niemeier, 1997). The gravity model, more adequately referred to as the spatial interaction model (SIM), is a commonly used framework to approach the trip-distribution problem. While this is useful for forecasting demand in classic transport modelling (Ortúzar & Willumsen, 2011), this is also valuable for transport behavioural studies (Reggiani, Bucci, & Russo, 2011; Thorsen & Gitlesen, 1998). The simplest version of the SIM is the *unconstrained* version, which is denoted as follows (Oshan, 2021; Wilson, 2010):

$$T_{ij} = kV_i^{\mu}W_j^{\alpha}f(c_{ij}), \tag{2.11}$$

The intuition behind it is as follows: the size of the aggregate flow T between origin i and destination j is determined by the travel cost c_{ij} between these locations, in addition to the magnitude or characteristics of the origin V_i and the destination W_j, while k is a scaling factor and μ and α are parameters to be estimated. The travel costs can be entered in a variety of functions. The main are the negative exponential and the inverse power, as described in the previous subsection. Both of these include a distance deterrence parameter β, which is the parameter of primary interest in gravity-type accessibility measures.

Originally, the SIM proponents suggested deriving it via nonlinear optimization techniques (i.e. according to the entropy maximization principle) (Wilson, 1971, 2010). In this way, Ortúzar & Willumsen (2011) refer to two calibration techniques: one is manual, and the other is through an iterative process. The first suggests 'arbitrary' manual adjustments until the modelled trip length distribution (MTLD) and the observed trip length distribution (OTLD) are satisfactorily close. The second is based on an iterative process where the initial value β_0 is a resultant of the mean OTLD. Then, each recalculation adjusts the value of β and the values for OTLD and MTLD are compared until a satisfactory prediction is reached.

Alternatively, the gravity model can be estimated in a regression framework (Oshan, 2021). Accordingly, Eq. (2.11) can be linearised employing a logarithmic transformation on both sides of the equation, as in the following:

$$\ln T_{ij} = k + \mu \ln V_i + \alpha \ln W_j - \beta \ln c_{ij} + \epsilon. \tag{2.12}$$

Here, the notation follows the same as in Eq. (2.11) and ϵ is a random normal error term. Common fitting techniques for this specification include ordinary least square (OLS) and maximum likelihood (Silva & Tenreyro, 2006, 2010).

Based on the theoretical notion that land value should reflect accessibility (as reviewed in Section 2.2), a small strand of literature which is decidedly relevant for the present work has calibrated the distance decaying parameter endogenously using property price data (e.g. Ahlfeldt, 2013; McArthur, Osland, & Thorsen, 2012; Osland & Pryce, 2012). This analysis is referred to as the employment potential capitalisation model. A general specification of this approach is given by the following expression:

$$\ln(P_i) = \alpha + \beta_1 \left(\sum_j g(E_j) f(c_{ij}, \beta_2) \right) + X_i B + \epsilon_i. \tag{2.13}$$

In the equation above: P is the observed price of property at location i; X_i is a matrix of control variables of property i; E is the number of jobs at location j, and c_{ij} is the travel cost between i and j which enters to the equation as one of the functions described in the previous section; β_1 is a regression coefficient that represents the capitalization effect of accessibility to employment on land, which is jointly estimated by the distance decaying parameter β_2; B is a vector of regression coefficients corresponding to implicit hedonic prices (Rosen, 1974); α is the intercept; and ϵ_i the random error term. This model is estimated in a regression framework using nonlinear methods (e.g. nonlinear least square or maximum likelihood).

In an empirical study, Ahlfeldt & Wendland (2016) compared the distance decaying estimates derived from the SIM and those by the employment potential capitalisation model, i.e. β and β_2, respectively. In the study, it is suggested that these are directly comparable parameters. The conclusions suggest a good degree of equivalence in the results produced by these two methods.

2.3.4 Accessibility in the context of residential land value

There are important contributions introducing land use and transport infrastructure-based measures to account for employment located outside the CBD as an extension of the monocentric city model. The first explicit attempts appeared long after the original development of the basic monocentric model (Adair et al., 2000; Srour, Kockelman, & Dunn, 2002). In general, these have found positive relationships between access to employment and residential land value, as expected from theory.

In line with this view, the performance of gravity-type measures to employment by car in relation to housing prices was tested for the case of Rogaland, Norway (Osland & Thorsen, 2008). One of the main innovations on top of the findings of previous contributions was the calibration of the spatial decay parameter endogenously using property value data as shown in the previous subsection. The results suggest that location-based measures are not able to replace the usual measure of proximity to the CBD. This can be due to the mono-centric-like structure of this region referred to by the authors. Yet, it is suggested that both measures are complementary and produce differentiated effects in the residential market.

A similar study aimed at accounting for the dispersion of employment using gravity-type measures examined residential land values in Berlin, Germany (Ahlfeldt, 2011). Two sets of models were built for the analysis: one using conventional proxies for accessibility, i.e. Euclidean distance to CBD and shortest distance to transit stations; and a second incorporating access to employment in gravity-type measures by different modes, i.e. driving, rapid-transit, and walking. The results show that the second modelling set outperformed the first in terms of goodness-of-fit. Furthermore, distance to CBD and distance to nearest public transport station are rendered insignificant in the presence of gravity-type measures. Other usual forms of accounting for accessibility reversed the direction of the effect when accounting for potential to employment measures, i.e. access to highways and ratio of car ownership. The argument is that gravity-type measures successfully capture the positive effects of accessibility while the other measures reflect negative externalities such as traffic congestion, noise, or crime (in the case of rail stations). This study supports the validity of neoclassic land rent model frameworks by generalizing the monocentric structure to a polycentric one via the use of location-based measures to employment.

Collectively, these contributions represent important advancements for the theory of land rent by connecting the advancements of accessibility developed somewhat in parallel. Still, contributions dealing with the particularities of accessibility generated by public transport remain limited. Some of these are discussed in the next subsection.

2.3.5 Measures of accessibility and public transport systems

Location-based accessibility measures are increasingly used to evaluate the performance of public transport systems in recent research (Conway, Byrd, & Van Eggermond, 2018; Ermagun, 2021; L. Liu, Porr, & Miller, 2022; Malekzadeh & Chung, 2020; Palmateer, Owen, Levinson, & Ermagun, 2021). A measure based on realistic travel times (representing travel costs c) should capture key characteristics of a public transport system (according to the transport component referred in Geurs & van Wee (2004)). This is because travel time by public transport is affected by several aspects of the system, such as:

- The topology of the network (e.g. connectivity, centrality);
- Capacity of the infrastructure (e.g. degree of right of way segregation, capacity and design of stations);
- Spatial coverage (e.g. availability of infrastructure), and;
- Operational characteristics (e.g. frequency of services, overlapping routes, express services, transfer wait times).

Furthermore, this type of travel time estimates allows addressing the time component by considering variability resulting from travelling by public transport at different times of departure in a day, as well as the related uncertainty in services during different time-windows (Owen & Levinson, 2015).

Computing and methodological advances developed from the perspective of transport geography (e.g. Conway et al., 2018; Pereira, Grégoire, & Karner, 2019) make feasible the incorporation of these detailed elements. These are important contributions towards an ideal accessibility measure, according to theory (Levinson & Wu, 2020; Miller, 2018). The implementation of such resources can contribute to a theoretically consistent study of the accessibility-land rent trade-off, particularly for the case of public transport innovations (Higgins & Kanaroglou, 2016).

2.4 Summary and final remarks

The present chapter constitutes the theoretical foundation for addressing the proposed research questions. Section 2.1 reviewed the LVC approach in the context of public transport. The range of strategies for the recovery of land value increments is wide. A key aspect for their successful application entails the adequate assessment of the benefits to differt extents. It is noted that LVC literature from the perspective of urban planning and planning law is limited in its interaction with the theory of land rent, resulting in disjointed contributions (Vejchodská et al., 2022). Section 2.1 expanded the fundamental assumption of the LVC approach by addressing the essential elements determining residential land value from the perspective of urban economics. Land rent literature positions accessibility at the core of the analysis (Duranton & Puga, 2015; Proost & Thisse, 2019). Yet, the accessibility concept had remained abstract since the formalization of these ideas until a relatively recent time. Section 2.3 introduced a comprehensive view of accessibility, incorporating advancements from the perspective of transport geography. The integration of location-based accessibility measures in the study of residential land rent has led to important contributions aimed at establishing a generalization of a monocentric structure to a model which successfully addresses the spatial dispersion of employment opportunities (outside the CBD) (e.g. Adair et al., 2000; Ahlfeldt, 2011; Osland & Thorsen, 2008). Such an approach can be aided by the digital advancements made available under the Open Science framework. These allow a rich assessment of public transport innovations.

Accessibility thus represents a key element bridging two quasi-parallel perspectives in literature for the case of LVC for public transport, namely urban planning or planning law and urban economics. Revisiting the capitalization of accessibility into residential land value by integrating the richness of the concept and cutting-edge resources empirically can thus represent relevant contributions for LVC in the context of public transport. The next chapter presents the main analytical approaches for evaluating such capitalizations empirically. Also, it provides a review of existing studies examining this relationship, putting special attention on those approaches acknowledging the comprehensive effects of accessibility included in this review.

Chapter 3

The capitalization effects of accessibility: A review of literature

Chapter 2 established the theoretical framework explaining the formal relationship between accessibility and residential land value. This framework is useful in guiding the assessment of land value increments associated with transport improvements. The evaluation of the economic benefits of these improvements is relevant for the land value capture (LVC) approach because this evaluation lies at the core of its philosophical argument. The present chapter focuses on the measurement of such a relationship.

The main contribution of this chapter is the articulation of accessibility in the context of land-rent theory and its empirical evaluation of the residential land market. Specifically, an updated discussion of the hedonic model and the equilibrium sorting model (ESM) is provided. This contribution is supported by a review of the results from the empirical literature. In this context, it is found that many studies overlook important elements of the accessibility concept. An emerging approach of adopting location-based measures illustrates how aspects highlighted in theoretical conceptualizations can be incorporated into the study of land value capitalization (e.g. Ahlfeldt & Wendland, 2016; Osland & Pryce, 2012; Osland & Thorsen, 2008). In addition, this review of the empirical literature finds relevant gaps. These include the explicit consideration of the role of the person component in location-based measures, the implications of the spatial unit of aggregation chosen, and the potential advantages of modelling accessibility based on detailed public transport timetables.

This chapter is organised as follows: Section 3.1 provides an outline of the nonmarket valuation approach (this section discusses the hedonic property value model and the ESM); Section 3.2 reviews the relevant results of empirical studies focusing on the relationship between accessibility and public transport; Section 3.3 discusses the spatial aggregation bias, known in the literature as the modifiable areal unit problem (MAUP); the last section (3.4) summarises the contents of the present chapter and provides some final remarks.

3.1 Nonmarket valuation via revealed preferences

As discussed in Chapter 2, LVC focuses on the recovery of increments in the value of land that are associated with a specific action by the government or local authority, e.g. the delivery of public transport infrastructure for the purpose of this research. As shown, LVC entails the challenge of distinguishing and measuring specific components from the full value of a property.

Environmental, or nonmarket valuation, can provide relevant practical resources with which to approach this task. Broadly speaking, these aim at "valuing environmental goods and services that are not traded in a market" (Champ, Boyle, & Brown, 2017, p. 1). In the urban context, environmental goods include not only natural resources but also those found in the local built environment. Collectively, these are usually studied under the *public good* concept. These goods are characterized either by their nonexcludability (i.e. it is difficult to prevent people from consuming it) or their jointness consumption nature (i.e. additional people can consume it at no additional cost) (Holcombe, 1997). Therefore, accessibility can be considered as a nonmarketable commodity (Webster, 2010) which is consumed through housing, as shown in the AMM model in Chapter 2.

An implicit trade in public goods or services arises when they are indirectly accessed through marketable goods (Champ et al., 2017). As Freeman et al. (2014) put it, "... people can choose the level of consumption of local public goods through their choice of a jurisdiction to reside in; thus, the housing market functions also as a market for the purchase of local public goods" (p. 104). This link makes it possible to infer the economic value of public goods through observed consumers' behaviour. This concept can be shown in *revealed preference* valuation methods. In contrast, *stated preference* valuation methods directly ask consumers about the value they place on such goods or services. However, some of the most salient issues identified with the latter method are the potential incentives for overestimation or underestimation, and the low cost-effectiveness ratio (Loomis, 2011).

In this thesis, the focus will be on revealed preference methods. Within this category, the literature has identified two main approaches. The first is the hedonic property value method, formalized in Rosen (1974). The second is the residential, or the ESM (Kuminoff, Smith, & Timmins, 2013). Although both methods describe the same equilibrium between consumers and producers in the residential market (Klaiber & Kuminoff, 2014), the hedonic method focuses on the outcomes and the sorting model focuses on the process (Freeman et al., 2014).

This section will first introduce the principles of consumer behaviour in the residential market according to the hedonic method. This discussion includes aspects relating to this method's estimation, environmental variable measurement, and its use for the evaluation of net benefits resulting from changes in environmental quality, e.g. public transport accessibility for this purpose. Secondly, some key considerations and challenges associated with the hedonic approach will be discussed. Lastly, the ESM is introduced as an alternative paradigm.

3.1.1 The residential market through the hedonic framework

Before proceeding onto the details, some of the underlying assumptions within this type of analysis must first be established. Firstly, it is important to consider housing as either a unitary differentiated commodity or a heterogeneous commodity (L. O. Taylor, 2017). These can be distinguished from each other through their special and specific characteristics, however, they are still traded in the same market.

The single market refers to both spatial and temporal dimensions. In terms of the spatial dimension, the requirement is that households can potentially choose any location within the definition of the market . In practice, however, this definition ranges from an entire nation (e.g. Nuñez, Paredes, & Garduño-Rivera, 2017) to an urban district (e.g. Guzman, Enríquez, & Hessel, 2021), although Bishop et al. (2020) suggest that it is likely that the 'one price function' principle can be maintained within a metropolitan area. One simple strategy for countering this issue may be of entering dummy variables into the hedonic function which represent a priori identified submarkets (e.g. Gibb, Osland, & Pryce, 2014; Williams, 1991). For in-depth applications, Pryce (2009, 2013) proposed a measure for the identification of submarkets based on the substitutability of dwellings. Meanwhile, the temporal definition assumes that the preferences of households remain stable over time.

Rosen (1974) formalized the hedonic method by considering an equilibrium solution arising from the interaction between housing developers (producers) and house buyers (consumers). With this view, the market is in equilibrium—neither producers nor consumers have incentives to change their respective choices (Klaiber & Smith, 2011). However, this comes under an implicit assumption of perfect competition (Kumbhakar & Parmeter, 2010) where prices are *determined* by demand, but households' independent actions are unable to modify the price schedule (Klaiber & Smith, 2011).

Liang et al. (2021, p. 251) identified a further three premises as being prevalent within the literature concerning house buyers. Firstly, that house buyers have full information about the housing market, meaning they are aware of the alternative options and respective characteristics of the market. It is suggested that home buyers devote considerable resources to assembling as much information as possible since the acquisition of a dwelling involves a large investment and sets the conditions for other long-term aspects of dwellers' lives (Phaneuf & Requate, 2017). Meanwhile, search costs have been hypothesized as preventing participants from acquiring complete information, potentially undermining the market efficiency (Kumbhakar & Parmeter, 2010). The second premise is that moving costs are not significant, allowing households to be fully mobile. These types of costs can result not only from the logistics involved, but also from restrictive reallocation policies (e.g. Liang et al., 2021). The latter has been shown to produce biased estimates when these are high (especially at the regional level). The third premise identified is that the market is not exclusionary and house buyers make location choices within an environment free from discrimination. Relevant constraints affecting this premise in reality can take a variety of direct or indirect forms, e.g. circumstances making it difficult to obtain a loan, perception of the ability to pay, prejudice, etc. (ibid.). Further implications will be discussed further on.

Hedonic function

It is within the conventional framework of the hedonic function that the majority of the literature has adopted its standpoint. In this literature review , Rosen's model will be introduced first, according to L. O. Taylor (2017).

Formally, a dwelling Z of a specific class can be described by a vector of its characteristics $\underline{z} = z_1, z_2, z_3, ..., z_n$. These include not only the internal attributes and the plot of land (*structural*), but also the characteristics of the immediate surroundings (*neighbourhood*) and the overall environment (*locational* characteristics usually described by accessibility measures).[1] In a perfectly competitive market, producers and consumers determine together the price schedule of the differentiated dwellings $P(\underline{z})$. Therefore, there are two complementary sides, namely consumers who *bid* in the housing market and producers who *offer* dwellings.

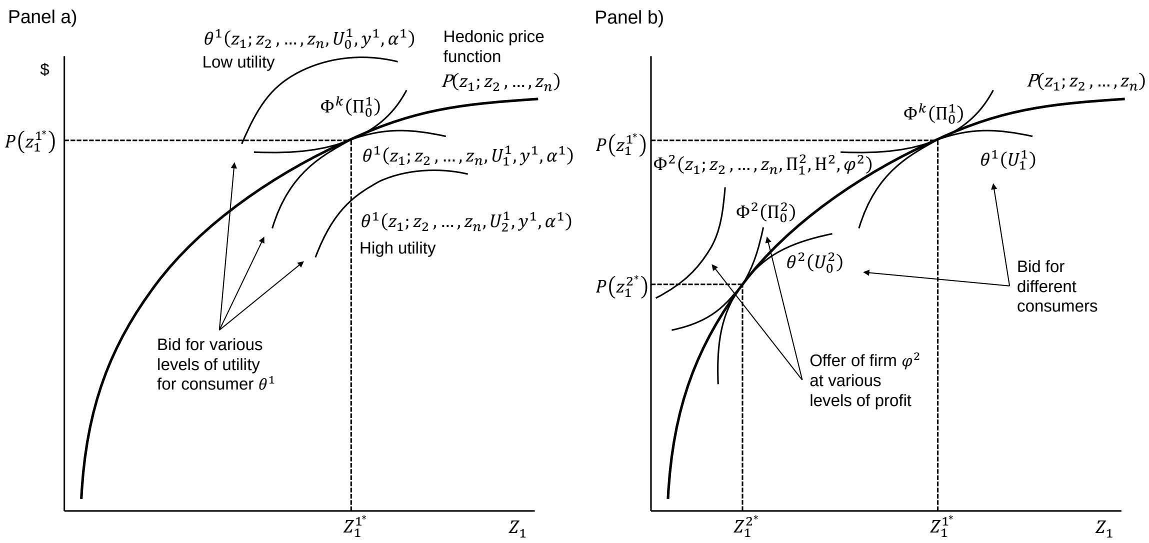
The process of interaction between these two sides is illustrated in Figure 3.1 for characteristic z_1, while holding all other attributes of Z constant. The approach assumes that a household's utility is given by consuming only two types of goods, namely a composite good x and only one dwelling, Z. Thus, a consumer's budget Y is constrained as the following $y = x + P(\underline{z})$. In Panel a) there are three bid functions, θ, for consumer 1 according to three different levels of utility, i.e. $U_2^1 > U_1^1 > U_0^1$ (a lower bid results in higher utility since there is more money to spend in x). The consumer

[1] According to L. O. Taylor (2017), the difference between neighbourhood features and locational ones is purely 'semantic.' Therefore, these characteristics are grouped within a single category in empirical studies (e.g. McArthur et al., 2012; Osland & Thorsen, 2008).

seeks to maximize utility, therefore,

> "the optimal choice of z_i is where the consumer reaches the lowest possible bid function while still being able to participate in the market" (L. O. Taylor, 2017, p. 205).

On the supply side, producer 1 generates convex offer functions denoted by Φ. In Panel a), the offer function is according to the level of profit Π_0. Since the producer seeks to maximize the profit, the equilibrium between these two agents occurs when the bid and offer functions touch each other at only one point.



Source: the Author based on Taylor (2008).

Figure 3.1: Hedonic equilibrium.

The previous idea is expanded in Panel b) of Figure 3.1. A further set of offer functions from a second developer (2) is shown, i.e. φ^1 and φ^2. For a developer with characteristics δ^k the profit is given by $\Pi = H * P(\underline{z}) - C(H, \underline{z}, \delta^k)$, where H is the number of housing units produced, and $C(\cdot)$ is the cost function. A curve closer to the horizontal axis denotes a lower level of profit. Since producers seek to maximize profit,

> "the firm's optimal choice of what level of z_i to produce in each of its H units of the differentiated product is where the firm reaches the highest possible offer function while still being able to participate in the market" (L. O. Taylor, 2017, p. 240).

Also in this panel, there is a second consumer (2) who generates the bid θ^2. The additional set of developers and house buyers illustrate the equilibrium process among the multiple agents on both sides of the market. This is graphically depicted in unique points of contact between alternative offer and bid functions.

The final and most important piece in the process, presented in Figure 3.1, is the combined outcome of the interaction between multiple producers and consumers. Panel b) illustrates that the hedonic price function is an envelope of the equilibrium interactions between all producers and house buyers, as depicted by the thick-line curve. Thus, in practice, this function can take any form; usually a nonlinear form (Ekeland, Heckman, & Nesheim, 2004). The interaction described also implies that at the tangential point the slope describing the buyer's marginal willingness to pay for a small change in z_i (through the slope of the bid curve) equates to the price that each consumer must pay for a small change in the dwelling's characteristic z_i (the price function slope) (Bishop et al., 2020). Formally, this equivalence is given by the partial derivative of the functions as follows:

$$\frac{\partial P}{\partial z_i} = \frac{\partial U/\partial z_i}{\partial U/\partial x}. \tag{3.1}$$

Estimation

Often, empirical researchers focus on the estimation of the parameters of Eq. (3.1) (referred to as the *first stage* of the hedonic model) (Parmeter & Pope, 2013). This is aimed at measuring the WTP for a change in an urban environmental characteristic associated with a property z_i. The hedonic function is commonly estimated within a regression framework. For instance, a linear relationship with the specification can be written as the following (L. O. Taylor, 2017):

$$P = \alpha + \sum_{i=1}^{h} \beta_i H_i + \sum_{j=1}^{n} \beta_j N_j + \sum_{k=1}^{L} \beta_k L_k + \epsilon, \tag{3.2}$$

Where P is the observed price of a dwelling, H represents structural characteristics of the house and lot, e.g. floor space, number of bedrooms, or lot size. N includes characteristics of the neighbourhood, e.g. median income, quality of schools and L captures locational characteristics, e.g. distance to the CBD, or accessibility to a range of amenities. α is a constant or intercept, and ϵ is a random error term. β is the regression coefficient of the respective attributes of the house. This is commonly estimated by ordinary least squares (OLS) or maximum likelihood techniques (Phaneuf & Requate, 2017).

Table 3.1 presents common functional forms of the hedonic function, namely the linear, log-linear, and log-log forms, and their respective interpretation as implicit prices. Although there are other flexible functional forms in the literature (i.e. Box-Cox transformations), these are relatively uncommon and add complexity to the estimation

Name	Relationship	Implicit price
Linear	$P_i = ... + \beta z_i + ... + \epsilon_i$	$\partial\hat{P}/\partial z_i = \beta$
Log-linear	$\ln P_i = ... + \beta z_i + ... + \epsilon_i$	$\partial\hat{P}/\partial z_i = \beta \times P_i$
Log-log	$\ln P_i = ... + \beta \ln z_i + ... + \epsilon_i$	$\partial\hat{P}/\partial z_i = (P_i/z_i) \times \beta$

Note:
Source: the Author based on Taylor (2017) and Phaneuf & Requate (2017).

Table 3.1: Common hedonic functional forms and marginal implicit prices.

and interpretation processes. Therefore, such approaches are not discussed in this review. In a linear specification, the implicit price for any of the attributes z_i is simply the estimated regression coefficient, i.e. $\beta_i = \partial P(\underline{z})/\partial z_i$. The interpretation of the log-linear relationship as marginal prices should be evaluated at some point of house prices , P (Phaneuf & Requate, 2017; L. O. Taylor, 2017). Here, it is implied that β is a constant proportion of P that people are willing to pay for an increase in z_i. Finally, the log-log relationship with β is interpreted as an elasticity of price with respect to z_i.

There are several aspects to be considered in the practical estimation of the hedonic function in addition to the functional form, e.g. the definition of the market, characteristics of data sources (such as quality and measurement), or the econometric specification including potential endogenous regressors. These considerations seek to achieve the main goal of the researcher, namely to estimate unbiased implicit prices of a particular characteristic in the housing market (L. O. Taylor, 2017). Addressing the recommendations and implications of each of these aspects in empirical research is out of the scope of the present review. Yet, it is worth discussing the potential for omitted variables since this is argued to be a first-order concern for introducing bias (Parmeter & Pope, 2013; L. O. Taylor, 2017). This issue often arises due to unobserved characteristics or measurement errors . Two common approaches to mitigate this issue are the implementation of spatial modelling techniques and the adoption of quasi-experimental research designs (Freeman et al., 2014; L. O. Taylor, 2017).

Spatial modelling techniques can mitigate bias generated by the omission of spatial variables by incorporating information from neighbouring dwellings. These techniques explicitly acknowledge the *spatial dependence* processes of data (LeSage, 2015; LeSage & Pace, 2009). In the context of hedonic models, this is useful when omitted spatial variables are thought to be independent from regressors (spatial error correlation) or when omitted variables are correlated with regressors (spatial autoregression) (L. O. Taylor, 2017). In both approaches, the information from neighbouring observations is abstracted into a matrix representing the spatial relationship between each observation and all others (referred to as spatial weight matrix and commonly represented by $\boldsymbol{W}$). In the case of the former, the spatial error model (SEM) can be used to 'adjust' the error term ϵ in Eq. (3.1), which was previously assumed to be independent. In the latter, a spatial autoregressive model (SAR) can include the information of neighbours in the form of an additional explanatory variable referred to as a spatial lag. SEM and SAR are two common techniques used in hedonic modelling. Nevertheless, spatial econometric literature offers a broad range of alternatives (see Anselin & Lozano-Gracia, 2009; Brady & Irwin, 2011; Osland, 2010b).

At the same time, quasi-experimental designs have been gaining relevance in hedonic modelling (Bishop et al., 2020; Parmeter & Pope, 2013). These approaches stem from the classic experimental design where randomized observations are exogenously placed in 'treated' or 'control' groups in a regulated environment. However, these conditions are not feasible in the study of the housing market for ethical and economic reasons. Still, the implementation of a policy or an environmental change may recreate treated and untreated groups if only some dwellings are affected (e.g. a local regulation, amenity, or disamenity). The endogenous selection imposed by the urban environment generates appropriate conditions for the adoption of a quasi- or natural-experimental design. This approach requires the counterfactuals information to estimate the effect on the treated dwellings. This represents a hypothetical situation for the treatment group as if it were not treated . A common design that allows this is the difference-in-difference (DID) approach (L. O. Taylor, 2017). This aims to measure the differences between both groups—*after* and *before* the observations are affected—producing an *average treatment effect* which is interpreted as the impact of the policy change. Another common approach is the regression-discontinuity design. Its specification is similar to the DID with the exception that the instrumental variable denoting timing (i.e. after/before) is replaced with a spatial boundary. Both approaches exploit discontinuities in time or space to recover the treatment effect.

Environmental variable measurement

The hedonic literature acknowledges some of the complexities around assigning amenity levels to houses (Bishop et al., 2020). Upon an initial examination, this involves the 'objective' measurement of spatial variation in the amenity of interest that can be assigned to an individual dwelling. A first challenge stems from the availability of data on space in terms of both coverage and the spatial granularity of the information. Representing the characteristics of the physical environment is usually approached by employing spatial interpolation or satellite image-based techniques. Spatial granularity, on the other hand, implies the generalization of a characteristic in space which is assumed to be applicable to a group of houses. This aspect is further discussed in

Section 3.3.

A second aspect in the measurement of the amenity level is the perception of house buyers. Bishop et al. (2020) say that the concern is “whether homebuyers’ beliefs about the amenity coincide with objective measures and, if not, to consider alternative ways of modeling buyer beliefs” (p. 267). In this regard, Klaiber & Smith (2011) assert that “these measures are proxy variables. Usually we do not know how households conceptualize the services each receives from these spatially delineated environmental resources” (p. 229). This is where the developments in accessibility literature, reviewed in Section 2.3, can offer relevant insights into the method. In particular, additional knowledge on accessibility can help to elucidate the relationship between the benefits resulting from public transport infrastructure and land value. As shown in Chapter 2, there is literature relating to the implications and principles of the adequate measurement of accessibility including not only *objective* aspects of the built environment, but also the person and temporal components.

Environmental variable measurements are especially relevant for the evaluation of changes in quality, for example, improvements in public transport accessibility. This is because the analysis conveys the implicit assumption that “there is a consistent link to changes in the underlying perceived services” (Klaiber & Smith, 2011, p. 229).

Net benefit measurement

There is considerable interest in transport policy analysis regarding the value that households put on changes in public transport accessibility (e.g. the implementation of a new corridor to the network). For such analyses, the strategy of establishing a correspondence between the implicit prices and the WTP depends on the situation. The observations discussed below are consistent in Phaneuf & Requate (2017), L. O. Taylor (2017), and Palmquist (2005). In examining these changes, the literature considers property owners and renters separately (under this view an owner can ‘rent from oneself’).

If the change in accessibility affects a relatively small portion of the households (it is localized) and the moving or transaction costs are irrelevant, the renter would choose an alternative bundle with the original characteristics (without being affected or benefited). If in addition the size of the change is marginal, then the implicit price will reflect the amount that the owner would be willing to pay to retain the improvement. In such a case, the owner realizes the gain, and the total WTP is simply the sum of the implicit prices across all the affected owners. This is the simplest situation and the one which is often adopted in the empirical literature evaluating public transport benefits (e.g. Ahlfeldt, 2013; Diao, 2015; McIntosh, Newman, Trubka, & Kenworthy, 2015; McIntosh, Trubka, & Newman, 2014). If the change is nonmarginal, the owner’s

WTP is equal to the sale price plus the new price of the amenity, minus the price at the initial level (namely $P_1 - P_0$). In prospective policy analyses, the new price (P_1) is an estimated forecast using the hedonic function (e.g. Gjestland, McArthur, Osland, & Thorsen, 2014).

The estimate above would represent an *upper bound* if the assumption of costless transactions is relaxed. This implies that the gains would be overstated and the losses would be understated. Conceptually, the net benefits would discount the transaction costs (TC) required for the renter to move to a new dwelling with exactly the same characteristics, maintaining an unchanged level of utility (i.e. $P_1 - P_0 - \text{TC}$). Note that these situations assume that the market equilibrium, and by implication the demand for the amenity, remain unchanged. A further assumption is that there is a continuum of alternatives in the housing market and all the combinations of characteristics are feasible. If the renter does not find an optimal location, the utility loss should be considered a correct net benefit estimate. Otherwise, the latter would also represent the upper bound (i.e. having discounted the TC). Finally, in the case that the renter chooses not to move (e.g. the moving costs are too high), the net benefits cannot be estimated using this approach. However, the difference above would reflect the upper bound. This is a partial adjustment, even if consumer increases the rent paid to the owner, since the gain in utility does not align with the optimal choice of the renter.[2]

Estimating the net benefits in the case of nonlocalized changes in an environmental characteristic (i.e. it affects a large proportion of the market) is more complex than the cases discussed above. The challenge essentially stems from the fact that demand is observed only at a single point in the hedonic function (where the willingness to pay reflects the inverse demand). Since it is likely that large changes imply adjustments to the levels of consumption, utility, and rent, these would result in a new hedonic equilibrium schedule (which is unknown in prospective analysis). Most of the strategies for evaluating a nonlocalized change require either relying on a series of assumptions or modelling heterogeneity in individual demand (Banzhaf, 2020). For instance, the latter approach draws on Rosen's second stage. This essentially regresses the WTP estimates on the quantities consumed and demographic characteristics (Bishop et al., 2020). However, this has been shown to be problematic since there are important endogeneity concerns (Bishop & Timmins, 2019; Epple, 1987). If moving is not allowed, the literature suggests that an upper bound can be estimated by aggregating an individual WTP following a similar procedure, as described above (Bartik, 2008; Bockstael & McConnell, 2007). An alternative approach, in which residents are assumed to have identical preferences, suggests the possibility of obtaining additional information about the demand curve by comparing two or more similar housing markets (Bishop et al.,

[2] Phaneuf & Requate (2017) discusses this situation at detail.

2020; Phaneuf & Requate, 2017; L. O. Taylor, 2017). This procedure identifies more than one point of the demand curve, allowing the analyst to 'connect the dots' of a function (Banzhaf, 2020; Bishop & Timmins, 2018).

The most difficult situation is when residents are allowed to move (Palmquist, 2005). One alternative is to use the previous measures as the lower bound (ibid.). Another one is extending the hedonic method by including additional information about consumers to model demand and address endogeneity (Phaneuf & Requate, 2017). However, these options, which draw on the hedonic framework, come with several limitations that are discussed in the next section. Alternative approaches to the hedonic method, such as the residential sorting model, can address the issue of uncompensated demand and other limitations, as will be reviewed in the next section.

3.1.2 Further considerations and challenges of the hedonic approach

Critical reflections on the hedonic method are numerous. Bishop et al. (2020) suggests "the literature has [. . .] generated more insights about the 'dos' and 'don'ts' of hedonic modeling than what has been observed for other economic frameworks that are used to analyse policy" (p. 275). This subsection discusses the main aspects and further considerations that constitute the key challenges in this method. First, it briefly enumerates the challenges associated with the first stage of the hedonic model. Then, it reviews additional conceptual issues chiefly related to heterogeneous preferences in the residential market. Altogether, these aspects serve as the main motivation for introducing the residential sorting model in the next subsection.

At this point, most of the key issues relating to the estimation of the first stage have been discussed in the previous sections. Specifically, Phaneuf & Requate (2017) enumerated them as follows: the extent of the market, functional specification, unbiased estimation, variable measurement, spatial econometrics, and timing of impacts. Excluding the timing of impacts and variable measurement, these key issues are treated as overall issues in the hedonic method. The aforementioned exceptions are further discussed in the specific context of public transport in Section 3.2 in light of results from the empirical literature.

The models discussed so far maintain the assumption of a representative household with equal income and preferences. Heterogeneous household income accounts for an important source of discussion in the literature. From the trade-off point of view established in the standard monocentric city model (discussed in Chapter 2), high-income residents would consume more land than lower income residents given that housing is assumed to be a normal good. If commuting costs are the same for all groups, low-income households will be sorted near to the CBD in small houses and high-income households placed on the periphery (Duranton & Puga, 2015). This placement is coupled with the premise that commuting costs are relatively less important for the latter group. Still, it is plausible that high-income households will outbid low-income ones in central areas if other commuting costs are considered, such as the opportunity-cost of time for commuting. Thus, the resulting bid-rent curve will reflect the upper envelope of the respective curves given by the different groups. In addition, the equilibrium of two (or more) income groups thus implies perfect segmentation (ibid.). In the context of the hedonic property model, Ahlfeldt (2013) reports that an increase in income implies a small but significant reduction in the WTP for accessibility in Berlin (1.2% less than the average). These results reflect a flatter land-rent gradient for affluent households, as hypothesized by land-rent theory. This is consistent with other empirical findings in North America (Kestens, Thériault, & Des Rosiers, 2006).

Alternatively, the spatial sorting according to different population groups can be approached from Tiebout's perspective (1956). In the Tiebout model, the incentives for high-income residents to suburbanise stem from fiscal and social problems. Specifically, Mieszkowski & Mills (1993) referred to high taxes, low quality public schools and other government services, racial tensions, crime, congestion, and low environmental quality. It is hypothesized that this pattern is reinforced cyclically given that central areas then deteriorate further, re-emphasizing the original incentives. Furthermore, the model establishes that high-income suburban residents seek to create (internally) homogeneous communities. This is motivated by their preferences (e.g. being among individuals of similar income, education, ethnicity, etc.), taste, or specific demands for local public goods. Additionally, it states that homogeneity itself can feed back into the local environment, for example the quality of education through peer-group effects. The latter represents potential endogeneity issues for the hedonic model (Banzhaf, 2021; Bishop & Timmins, 2019).

Regardless of the perspective, the literature generally recognizes the segmentation of the residential market according to households' characteristics. For instance, specialized literature has drawn on the 'homophily' concept—a basic organizing principle which suggests that the degree of interaction between similar people is higher than among dissimilar people (McPherson, Smith-Lovin, & Cook, 2001). Similarity can manifest in a variety of social characteristics, e.g. age, sex, religion. However, race and ethnicity have been identified as the largest division in social networks in the United States. Homophily has been found to have implications in the residential choice process. Bakens & Pryce (2019) reported that not only does the internal ethnic composition of a neighbourhood attract house buyers—meaning people of the same background as themselves—but so does the composition of surrounding neighbourhoods. This effect, referred to as the homophily horizon, is highly relevant as it can potentially emphasize spatial segregation patterns. Dean & Pryce (2017) offered an innovative method for measuring homophily as the perceived substitutability of a dwelling and social integration based on network analysis and revealed preferences. This method found a strong level of religious homophily in the case of the Glasgow housing market. Similar research in the context of North America supports these results (Bayer, Fang, & McMillan, 2014).

Some aspects of this quasi-experimental hedonic design are worth discussing considering the dynamics of the residential market discussed above. While this approach has emerged in order to address some of the limitations of the basic model (e.g. omitted variable bias), it also implies new challenges. Indeed, the assumption of interpreting capitalization effects obtained from a quasi-experimental analysis as benefit measures has been referred to as 'heroic' (Klaiber & Kuminoff, 2014). L. O. Taylor (2017) suggested that DID estimates are appropriate for estimating the net benefits if the change is nonmarginal and localized. Most of these issues essentially arise because the hedonic function may shift between the periods of time observed (Banzhaf, 2021; Bishop et al., 2020). Furthermore, this approach relies on the assumption that the amenity of interest and the initial conditions are uncorrelated (Parmeter & Pope, 2013). This assumption may be challenged if, for example, public transport accessibility improvements incentivise homeowners to modify dwellings, such as improving the quality or extending them (i.e. creating additional floor space), as this may be more economically advantageous under the new conditions. As a result, the treatment effect may also capture unobserved variations (Banzhaf, 2021). In our example, the treatment effect may conflate unobserved structural changes with accessibility improvements.

To summarize, the implications for the hedonic approach are numerous especially those related to the welfare measurement for changes in environmental quality. For instance, a considerable change in the spatial distribution of public transport may incentivize adjustments in demand elasticity according to income or preferences motivated by demographic characteristics such as ethnicity or age. These imply the potential re-sorting of residents and a consequent new equilibrium. Although the literature has proposed a range of strategies for estimating the compensated demand, many aspects remain difficult to implement in the hedonic model given the complex dynamics of the residential market discussed above. These challenges make relevant the possibility of a two-way modelling approach between people and their urban environment (Klaiber & Kuminoff, 2014). The next subsection introduces the ESM which provides a suitable framework for such a task.

3.1.3 Equilibrium sorting model

As suggested earlier, the hedonic model and the ESM essentially describe the same equilibrium of the residential market (Klaiber & Kuminoff, 2014). Specifically, these types of models build on the conceptual Tiebout model where household preferences and taste are integrated into the location decision process. This explicitly addresses some of the challenges implied in the hedonic model, e.g. recognizing the homophily effects in the residential sorting process. An important formal difference with the standard hedonic model is that, while this assumes that both structural and environmental characteristics exist as a continuum, ESM treat them as discrete choices (ibid.). Residential selection as a discrete choice problem has been chiefly developed in random utility maximization (RUM) models for a long time (Kuminoff et al., 2013)). From this view, households maximize utility, subject to a budget constraint, by choosing a house bundle comprised of the dwelling itself and its respective neighbourhood (a bundle comprised of private and public goods in Kuminoff's et al. (2013) words). This refers to some level of amenities, e.g. tax revenue, environmental services, proximity to urban attractions, and demographic composition. Thus, the focus shifts from a single dwelling to the neighbourhood.

A special feature of the ESM, compared with the standard RUM framework, is that the former considers and additional step for modelling equilibrium (Conway, 2021; Phaneuf & Requate, 2017). This is based on the market-clearing condition where demand (represented by the modelled probability to choose a dwelling in a certain neighbourhood) equates to the supply for each neighbourhood (L. O. Taylor, 2017). This condition allows the simulation of residential location choice for a new equilibrium (and a corresponding new price schedule) given a nonmarginal change in dwellings, amenities, or preferences. In other words, households can re-sort in response to an environmental or other type of shock (Conway, 2021). In these models, supply is often assumed to be fixed.[3] This flexibility overcomes some of the main challenges presented by the standard hedonic model, as discussed in the previous section (3.1.2).

Although the ESMs were formalized nearly twenty years ago (Bayer, McMillan, & Rueben, 2004), their application remains limited. The ESMs mostly focus on the valuation of nonmarket goods, such as open space (Klaiber & Phaneuf, 2010), air quality (Liang et al., 2021; Tra, 2010, 2013), school quality (Bayer, Ferreira, & McMillan, 2007), or cultural amenities (van Duijn & Rouwendal, 2013), although they are not limited to such a task. For instance, in the context of transport and land use, Conway (2021) examined the effect of housing supply on vehicle ownership. They have also been used to estimate the impact of public transport quality on car ownership (Mulalic & Rouwendal, 2020). The studies focused on nonmarket valuation have found that the

[3] The assumption of fixed supply is relaxed in Conway (2021), for example.

average WTP tends to be similar between ESMs and hedonic methods (Klaiber & Kuminoff, 2014). However, relaxing the assumption of a perfectly elastic demand implied in some hedonic strategies for net benefit measurements can add credibility (ibid.). In this regard, the measurement of net benefits can substantially differ between one method and another (Bayer et al., 2007; Sieg, Smith, Banzhaf, & Walsh, 2004). Based on our current understanding, there are not yet studies employing this methodology to evaluate the benefits of public transport accessibility.

Flexibility comes at a cost. As expected, given what has just been discussed, the data requirements to estimate ESMs are considerable (Pryce, 2021). They usually require longitudinal data collection on households and individuals. This type of information is expensive to collect and, in some cities, it is simply unavailable. In addition, due to data privacy regulations, official datasets tend to reference this type of information in large spatial zones. For example, in a recent application in Southern California, the spatial analytical unit is limited to the Public Use Microdata Area (PUMA) level (including not less than 100,000 inhabitants each). In practice, the method is adapted on almost a case-by-case basis and the data employed differs substantially. This improves adaptability; however, it also poses challenges. For example, effectively communicating the methods and data utilized may demand elaborate explanations (e.g. Bayer et al., 2007; Klaiber & Phaneuf, 2010). Furthermore, the diversity of applications adds complexity for researchers outside of the field of environmental economics.

Additionally, there are also technical aspects worth considering. For example, the models need to be coded (essentially) from scratch (Klaiber & Kuminoff, 2014). Thus, ESMs still require both advanced programming skills in addition to solid foundations in econometrics and statistics. Finally, fitting these models demands time and computational resources (see Appendix B in Conway (2021) for technical details). Thus, this can become an onerous task, even for relatively simple situations.

ESMs represent significant advancements in the field of urban economics (Pryce, Wang, Chen, Shan, & Wei, 2021). They address many of the limitations associated with the standard and expanded versions of the hedonic model. One of their key advantages is that they provide a theoretically correct estimate of net benefits for policy analysis. However, there are conditions that make it challenging to generalize their application.

3.2 Empirical analyses on public transport accessibility

This section offers a review of the empirical work focusing on the association between public transport infrastructure and land value. The hedonic property value model presented in the previous section is the most common analytical tool employed in these types of studies. To the extent of our knowledge, there are no studies in the literature that have implemented the ESM for this purpose. The review of the literature is chiefly structured around the strategy chosen to approximate the environmental measure representing public transport improvements or changes. This is due to the interest in addressing a theoretically robust measure of accessibility, as established in Chapter 2.

Therefore, this section is organised based on the two broad distinctions made in Section 2.2 of Chapter 2 in accessibility measures. These are: (1) *Infrastructure- or spatial configuration-based* measures, and (2) *Land use and transport infrastructure-based* measures. In the former, accessibility is chiefly represented by the transport component, and it is focused on mobility. Since most of the literature falls under this category, these studies are discussed under the section labelled the *classic* empirical approach. The second block approaches the accessibility concept from a comprehensive viewpoint, namely it includes studies which consider at least the transport and land use component. Since this approach is less frequently used than the former, it is discussed in the section labelled the *alternative* approach. The studies that consider the transport network as a system or acknowledge the spatial heterogeneity of the transport infrastructure either explicitly or implicitly are discussed in the alternative approach.

Due to methodological relevance, some studies assessing other modes of transport (i.e. cars) are also considered. Similarly, some studies examining property and land markets other than residential are included for the same reason.

3.2.1 The classic empirical approach

Mohammad et al. (2013) conducted a meta-review using empirical studies that focused on the impact of rail projects on land and property values. The findings indicated that the use of developed property data values generated lower estimates than vacant land values. This result, however, should be considered with care since it is not clear whether the studies considered appropriate variables to control for the structural characteristics of the property, as suggested by the hedonic modelling literature (L. O. Taylor, 2017). In addition, it was reported that there was not a significant difference in the studies using purchase or rent values as the dependant variable, which is in line with the theory of land-rent (Chapter 2). Regarding the type of use, it was shown that offices do not differ from residential properties. However, commercial properties were associated with larger effects than residential ones. This finding is partially in line with previous analyses (Debrezion et al., 2007), which suggest that this is true only at close proximity to public transport stations (i.e. within a quarter of a mile).

In the meta-review, external factors focused on the effects of the transport scheme, i.e. transport mode (Mohammad et al., 2013). Commuter rail was consistently associated with larger effects on land value than in other modes while the impact of heavy rail was associated with smaller effects . The impact of the commuter rail is supported by earlier meta-analysis while the effects of heavy rail remain unclear (Debrezion et al., 2007).

The timing of capitalization is disputed in the literature. The findings of Mohammad et al. (2013) suggest that there is not a significant difference between the stages of a project's implementation (i.e. from the announcement to the opening), except for the stabilization stage. This could be because the expectation of house buyers was high in the initial stages but then adjusts to the conditions of the market. In contrast to this finding, more recent empirical studies in Australia show that it is in the consolidation stage (equivalent to stabilization) when further capitalizations are likely to occur (Mulley & Tsai, 2016; Yen, Mulley, Shearer, & Burke, 2018). Other studies suggest that the timing of capitalization depends on the rate of ownership (Ahlfeldt, 2013; Gibbons & Machin, 2005). It is concluded that if the proportion of ownership is high, capitalizations occur during the opening stage.

Public transport accessibility is often represented in areas of influence around public transport stations or corridors referred to as *catchment areas* (Heyman, Law, & Berghauser Pont, 2019). Mohammad et al. (2013)'s findings suggest that there are no significant effects in catchment areas of 0 to 200 metres nor areas of 201 to 500 metres. Yet, a significant and positive relationship arises in studies which consider catchment areas larger than 805 metres (half a mile). The effects of accessibility as represented by catchment areas are subject to debate (Guerra, Cervero, & Tischler, 2012; Páez et al., 2012; Petheram, Nelson, Miller, & Ewing, 2013). While some studies focusing on the size of the area of influence report negative effects at short distances from the stations (i.e. 100 m) (Mulley, 2014; Mulley & Tsai, 2016), others argue that the definition of catchment areas should follow an augmented criteria considering further factors other than proximity to facilities (e.g. score matching techniques) (Yen, Mulley, & Shearer, 2019). Accessibility generated by cars has been examined based on similar approaches, e.g. the proximity to highway intersections. Contrary to the expectation based on land-rent theory, this proxy has been negatively related to the value of land (Debrezion et al., 2007; Mohammad et al., 2013). Accessibility in empirical studies will be re-addressed in a special section below given their relevance to the main topic of the present work.

The choice of analytical method employed in empirical studies shows a variety of outcomes. A study in the Tyne and Wear area in the UK found that geographically weighted regression (GWR) improved the fit of the model compared to the performance of ordinary least squares (OLS) (Du & Mulley, 2012). What is more, the authors argue that the incorporation of GWR methods help to clarify the weak association of accessibility and property market previously found in a study conducted in the same area (Du & Mulley, 2007).

A different study, assessing the impact of the access generated by a BRT line in Sydney, Australia, compared the performance of 'global' OLS models with 'local' models through GWR (Mulley, 2014). In coincidence with the above , the study concluded that local models showed improvements over the global models in terms of goodness-of-fit and disaggregation of the results. A further analysis focused on how temporal variations impact the transport infrastructure in the same context (Mulley & Tsai, 2016). Here, multilevel modelling was identified as an appropriate method to account for spatial dependence. Furthermore, multilevel modelling outperformed the results generated by alternative methods such as spatial lag or spatial error models. Acknowledging the mixed outcomes and potential issues that different methods can introduce, Yen et al. (2019) evaluated the performance and adequacy of different techniques, drawing on the case of the light-rail line on the Gold Coast in Queensland, Australia. The study contrasted multilevel and DID approaches. It concluded that the multilevel regression model produced better goodness-of-fit than the DID models. However, the ability of the methods used to mitigate potential bias was not discussed in these studies.

Shifting the focus to the Latin American context, it is observed that most of the studies focus on *bus rapid transit* (BRT) systems given the limited extension of rail-based infrastructure in the region (Stokenberga, 2014). Munoz-Raskin (2010) reported that in Bogotá, Colombia there was an 8.7% premium in properties located within five minutes walking distance from a feeder BRT line compared to those within ten minutes walking distance. However, it was found that only the middle-income housing positively valued the immediate proximity to the BRT system and it is argued that the ticket fare is a possible barrier for low-income dwellers. A further study in the same context employing spatial modelling techniques in a DID design, estimated an average treatment effect of US$1,950 in Bogotá and US$1,800 in Barranquilla (i.e. for affected dwellings located less than 500 m away) (Perdomo Calvo, 2017).

A follow-up study in Bogotá implementing specific methods for the identification of control observations in a DID approach (i.e. *Coarsened exact matching*) reported positive effects on land value in low-income neighbourhoods in both close and medium proximity catchment areas, i.e. under 200 metres and between 200 and 500 metres, respectively (Guzman et al., 2021). However, some areas of influence which qualitatively identified as middle- and high-income areas did not show significant effects nor negative effects in some cases. The authors suggest that these also enjoyed good overall accessibility levels attributable to conditions other than the BRT infrastructure, e.g. high concentration of employment. Furthermore, the results of this study contrast those in Munoz-Raskin (2010). Guzman et al. suggest that the lack of capitalization in low-income dwelling could be attributed to the bus fare as a barrier. However, the findings in Guzman et al. (2021) question this argument.

The literature studying these types of impacts in the context of Mexico is still very limited. Furthermore, existing studies focus on BRT systems only, although the extension of the Metro network in Mexico City is considerable (approx. 220 km long). One contribution, however, includes a study of the municipality of *Ecatepec*, in *Estado de México* (a non-core municipality of Greater Mexico City) (Flores Dewey, 2011). The analysis aims at estimating possible capitalizations in housing prices after the announcement of the construction of the first BRT line in this state, i.e. *Mexibus I*. The study assesses the housing sale prices reported to the local record valuations according to a DID design. The definition of the treatment area is given as a 1 kilometre buffer along the proposed BRT line, while the control area includes observations of an adjacent avenue in the same municipality. The analysis did not find significant impacts on 'low-quality' properties. By contrast, 'high-quality' construction properties showed a negative average treatment effect.

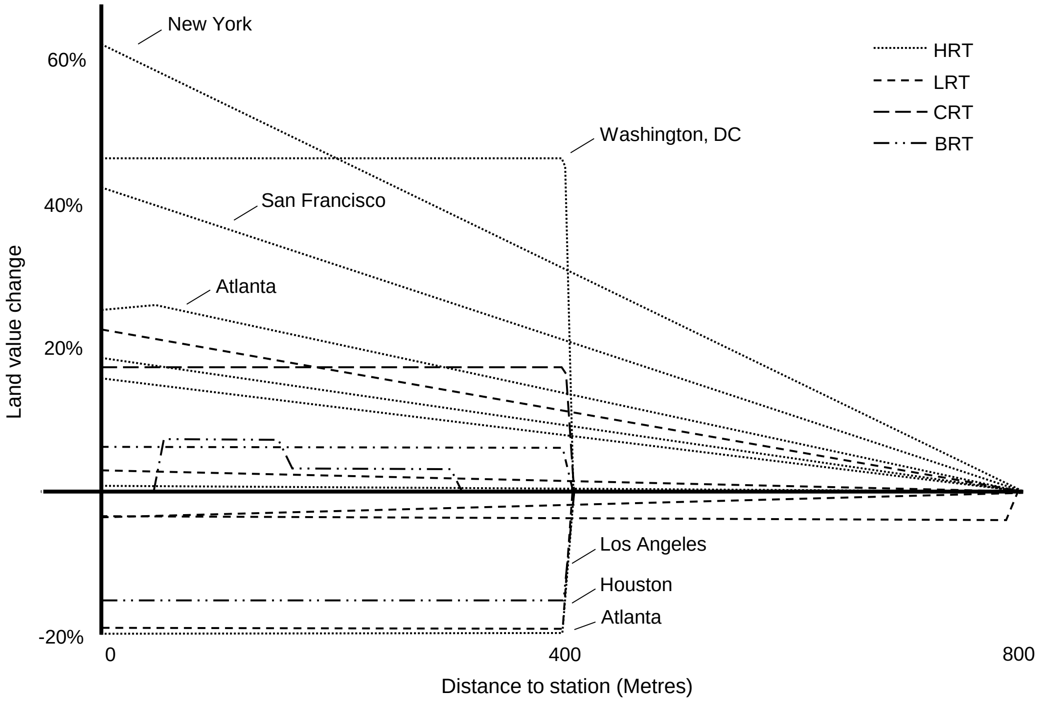
These findings may be responding to several conditions. Flores Dewey (2011) suggests factors such as land-taking rumours and transitional effects during the construction as potential elements leading to these results. In addition, since it was the first system of this type, its benefits could have been unclear to residents. It is worth pointing out that the timing of capitalizations have been contested in the literature (Mohammad et al., 2013), that is, whether the residential market responds to the implementation of the infrastructure at the date of the announcement, start of construction works, opening, or stabilization/consolidation stage. Research suggests that considerable capitalization effects occur close to the opening date of the transport infrastructure in contexts of high owner-occupancy (Ahlfeldt, 2013; Gibbons & Machin, 2005). Thus, the findings in Flores Dewey (2011) may also respond to the timing of capitalizations.

Another example which draws on the case of *Ciudad de México* (the core state of Greater Mexico City) evaluates the effects of the construction of the *Metrobús* BRT Line 1 and Line 2 on residential land value following a DID approach (Velandia Naranjo, 2013). The main results show positive increases of between 14.6% and 15.6% in plots within 500 metres from the BRT Line 1 in the northern section (closer to the financial business district) compared with the control areas. As an aside, it was reported that the announcement and the beginning of the construction works of the nearby Metro line (Line-12) did not show any significant effect.

Heterogeneity of the outcomes

As observed above, and in line with previous reviews (Debrezion et al., 2007; Higgins & Kanaroglou, 2016; Mohammad et al., 2013), a prevalent concern is the variability and, at times, the conflicting results of the impacts of public transport projects on residential land value. As shown, empirical studies have reported mostly positive outcomes, but also neutral and sometimes negative effects.

This heterogeneity of the outcomes in empirical studies is illustrated in Figure 3.2. The plot shows the estimated coefficients for residential property values reported in relation to the distance to public transport stations for several cities in the US according to various public transport modes, i.e. heavy rail transit (HRT), light rail transit (LRT), commuter rail transit (CRT), and BRT (Higgins & Kanaroglou, 2016). It is clear that the range of variance is wide. For instance, the estimated effects go from -20% up to approximately 60%. In particular, HRT shows the largest positive impacts (above 25% and up to 60%). Still, some neutral or negative impacts can be observed for this mode as well as in the other modes presented in the plot.



Source: the Author based on Higgins and Kanaroglou (2016, p. 614). *Note:* HRT = Heavy rail transit; LRT = Light rail transit; CRT = Commuter rail transit; BRT = Bus rapid transit.

Figure 3.2: Land value uplift coefficients for residential in United States studies.

An explicit or implicit assumption in these types of studies is that transport infrastructure generates accessibility improvement. Therefore, it is expected that this will be reflected in the value of the land. Despite the advancements of accessibility literature, several empirical studies still rely on simplifications. A systematic review of empirical literature quantifying the frequency of the use of different forms of operationalizing accessibility in hedonic studies supports this view (Heyman et al., 2019). For instance, about two out of three (65%) of the measures employed in the sample reviewed used an *infrastructure- or spatial configuration-based* type (e.g. shortest distance to the opportunity of interest).[4] What is more, only 16% of these consider the network distance and 17% use travel time for its computation. The rest is built upon Euclidean or unspecified metrics. It is also worth noting that in these types of measures, public transport facilities (e.g. stations) are often accounted for as the actual destination (representing 31% of opportunities) and not as a component enabling accessibility.

A critical examination of the methods employed by many empirical studies adopting a *classic* approach may suggest possible misspecification of the accessibility concept from both the perspective of the land-rent theory and accessibility literature. As noted by Higgins & Kanaroglou (2016), one concern is that the effects are often reported as global or average figures, assuming homogeneity of the benefits produced by public transport stations. This is also in line with the review of the empirical literature presented in the previous section. It is observed that many studies adopting the approach of homogeneous catchment areas overlook the land-use component as well as important aspects of the transport component often referred to within accessibility literature (Section 2.2).

In this vein, Ryan (Higgins & Kanaroglou, 2016; 1999) asks: *"Do the evaluated stations or lines in fact serve the needs of the local users?"* In other words, does the assessed infrastructure effectively increase accessibility and consequently reduce the transportation costs perceived by households as suggested in land-rent theory? Higgins & Kanaroglou (2016) argue that the approach employed in most of the empirical studies have captured only the 'tip' of a complex relationship where accessibility has been playing a minor role. Thus, these critical views call for the re-examination of aspects of empirical specificity in future studies. Specifically, the authors point out the following aspects: (1) revisiting the accessibility concept using 'more behaviourally relevant specifications' which move beyond aggregated measure (or mobility-focused measures, as termed in this work); (2) controlling for land use planning at the local specific contexts, and; (3) assessing and controlling for spatial dependencies and heterogeneity in the data. The present work builds on this view, putting special attention on the first recommendation.

[4] Infrastructure- or spatial configuration-based are referred to as 'spatial separation' in the original source.

3.2.2 An alternative approach: The impact of intra-urban accessibility on land value

A distinctive characteristic of the research presented in this subsection is the focus on, and acknowledgement of, the transport network as a system and the consequent spatial heterogeneity of transit stations. This is in addition to the characteristics of the land use. From this perspective, the assessment of accessibility in relation to land value is studied at two broad scales: the city or regional level and the neighbourhood or district level. Therefore, the review of this type of literature in the coming subsection is organised according to this distinction.

The city and regional perspective

The seminal work adopting this approach dates back to the work of Adair et al. (2000). This study explicitly employs a gravity-type measure in cross-sectional analyses of the housing market in Belfast, UK. The authors constructed an accessibility index based on a spatial interaction model using observed commuting flows for different journey purposes, i.e. work, non-home based, and other. The estimated coefficient for the spatial deterrence parameter for work journeys was 0.055. The results were mixed. At the city level, the relationship between housing prices and accessibly was not significant. Yet, a closer examination of spatial sub-market and housing type produced significant coefficients for some segments. Specifically, among the different types of dwellings, only terraced properties responded to accessibility. Regarding the spatial sub-market, dwellings in west Belfast and north Belfast (inner) sectors showed significant coefficients. The authors argued that the mixed results were due to the almost 'universal' access generated by private cars. It was, therefore, suggested that this makes some dwellers less sensitive to variations in accessibility.

One regional study examined housing prices with a focus on the adequacy of location-based accessibility measures in Rogaland, Norway (Osland & Thorsen, 2008). The analyses tested several gravity-type measures to account for the spatial dispersion of the labour market according to the following alternative specifications:[5]

$$A_i = \sum_{k=1}^{w} D_k \exp(-\beta_e c_{ij}), \tag{3.3}$$

$$A_i^e = \sum_{k=1}^{w} D_k^{\gamma_e} \exp(-\beta_e c_{ij}), \tag{3.4}$$

$$A_i^p = \sum_{k=1}^{w} D_k^{\gamma_p} c_{ij}^{-\beta_p}. \tag{3.5}$$

$$\bar{c_j} = \sum_{k} \frac{D_k}{D} c_{jk}, \tag{3.6}$$

In Eq. (3.3), A_j denotes the accessibility at the zone of origin j; D_k is the total number of jobs in the zone of destination k; c_{jk} represents the travel cost as the estimated travel time by car between k and j; and β_e is a spatial deterrence parameter within a negative exponential function. Eq. (3.4) is modified version of the *usual* gravity-type measure adds a further parameter γ_e directly affecting employment opportunities. Eq. (3.5) differs from the previous parameter, entering the spatial deterrence parameter β_p in an inverse power impedance function. A further specification is assessed according to Eq. (3.6), which essentially represents the average distance to job opportunities. Meanwhile, the availability of urban amenities is assumed to be at the central area of the region, and it is thus measured as the travel time to the CBD.

The analyses in Osland & Thorsen (2008) showed that the poorest results are produced by the average distance to job opportunities (as in Eq. (3.6)). The rest of the gravity-type measures (from Eq. (3.3) to Eq. (3.5)) improved the results. Specifically, it was shown that the consideration of a γ parameter helps to improve the goodness-of-fit and mitigates econometric issues (e.g. misspecification bias and spatial autocorrelation). The difference between the two tested functional forms of impedance was argued to be negligible. In general, the authors advocate for a differentiation of the effects of the measures capturing accessibility to employment and proximity to the CBD in the housing market. Furthermore, it was concluded that access to employment could not be considered as a 'replacement' for proximity to a CBD.

Osland & Thorsen (2008) provides several relevant contributions. One is the implementation of nonlinear techniques to simultaneously calibrate the spatial decay accessibility parameters together with the coefficients suggested in hedonic studies using property price data (as detailed in Section 2.2). This is in contrast to the stepwise procedure where these values are estimated before being entered into the hedonic model. This approach is relevant because it allows flexibility to capture the house-buyer's perception according to the context-specific valuation of accessibility, as

[5] The notation presented here is adapted from the original source to maintain consistency with the previous notation used in the present work.

discussed previously (Subsection 3.1.2). Another extension to the literature is the examination of the performance of alternative functional forms. Although this had been extensively examined in the context of spatial interaction models, the implications of it has remained understudied in the context of hedonic models. The integration of concepts borrowed from spatial interaction models and accessibility theory represents an important advancement in this strand of the literature.

One empirical analysis employing a similar approach as the study referred to above used a set of multi-modal gravity-type measures of employment aimed at generalizing the basic monocentric city model into a polycentric structure (Ahlfeldt, 2011). It focused on the residential land market drawing on the case of Berlin, Germany. Several aspects can be highlighted from the results. For instance, the effect of distance to the nearest rail station deserves special attention. It was shown that this measure is negative and significant in the absence of controls for the *effective* level of public transport accessibility. However, this became less significant after an explicit public transport accessibility measure was entered. It is argued that when a comprehensive measure of accessibility is not accounted for, the distance to the nearest station reflects the net effect including not only location advantages but also disamenities e.g. noise or crime. This result illustrates how possible biases may arise due to an error in measurement of accessibility in some empirical studies. This is possibly related to the heterogeneity of the effects reported in empirical studies relying on infrastructure-based accessibility measures.

Other measures usually found in the hedonic literature are worth discussing in light of their empirical findings (Ahlfeldt, 2011). For instance, distance to the employment centre also became insignificant in the presence of location-based accessibility controls. Furthermore, car ownership measured as the distance to the nearest main road becomes a negative and significant coefficient. The argument for this effect is that these variables are said to reflect both positive and negative externalities such as traffic congestion. The overall conclusion suggests that "gravity models facilitate the application of rent theory to real world settings where the simplifying assumption of a perfectly monocentric city is relaxed" (Ibid., p. 335).

There are also some relevant contributions which expand on the findings in relation to negative externalities in the housing market correlated with accessibility. Osland & Pryce (2012) suggests that a flexible *nonmonotonic* function integrated into a location-based measure can capture negative externalities associated with proximity to employment nodes, e.g. congestion, noise, air pollution. The nonmonotonic effect can be included by adding an extra parameter directly affecting the size of opportunities

(W) before being exponentiated as shown in Eq. (3.7) by the term c_{ij}^{θ}.

$$A_i = \sum W_i c_{ij}^{\theta} \exp(-\beta c_{ij}) \tag{3.7}$$

This specification has been shown to capture negative externalities in housing prices while accounting for the dispersion of employment using data from Glasgow, UK (Gibb et al., 2014; Osland & Pryce, 2012). In a later contribution, Osland, Thorsen, & Thorsen (2016) argued that local spatial heterogeneity can reflect such negative externalities and can be modelled using a spatially structured random effect at the neighbourhood level.

Thériault et al. (2013) used hedonic models to validate a set of accessibility indexes for the case of Quebec City, Canada. Although the focus of this study was not the perspective of nonmarket valuation, it is valuable in terms of the way it shows how accessibility is perceived by house buyers. This approach examines the relevance of two types of accessibility measures, namely a 'measure of centrality' (operationalized as the sum of the modelled generated trips in a spatial interaction model) and a set based on cumulative opportunities measures for various purposes, i.e. employment, school, shopping, grocery shops, and healthcare. The latter defines the impedance threshold according to information obtained from the travel survey, namely the reported travel time by car from different types of household, i.e. women, men, children, families and/or adults. Later, these are normalized and aggregated to represent the 'sum of satisfactory opportunities' to avoid collinearity. The study suggests that both measures are appropriate to model accessibility in a polycentric urban context in line with previous findings (Ahlfeldt, 2011). Furthermore, the set of accessibility indicators supplement the effects of centrality in coincidence with previous results (Osland & Thorsen, 2008). Broadly, the study shows how the perception of accessibility by different types of households is reflected in the value of residential land.

As shown, most of the studies employing location-based accessibility measures have adopted a cross-sectional design. An empirical study compares the performance of cross-sectional and first-difference approaches to illustrate a possible contrast between these two techniques (Iacono & Levinson, 2017). The first-difference models are estimated according to changes in both accessibility and property price between the years 2000 and 2005, while the cross-section is limited to one point in time, i.e. 2000. The study used data from Minneapolis-St Paul, US. Here, accessibility is measured using cumulative opportunities to employment by private car, public transport, and on foot according to a twenty-minute time threshold. Additionally, the study includes car accessibility to workers using the same threshold . The cross-sectional results shows that accessibility to employment by car and by foot only are positive and significant. In contrast, access to workers and access to employment by public transport show negative signs. It is worth noting that the aim of the authors is to show the potential differences of the methodological approach. In the cross-sectional case there are important issues of collinearity between accessibility variables as well as the potential for omitted variables, which may challenge the validity of the estimates.

The second part of the analysis in Iacono & Levinson (2017) compares the results of the cross-sectional and first-difference models. A first aspect to note is that the goodness-of-fit substantially decreases in the latter. It also shows that the first-difference approach mitigates collinearity. While the variables accounting for the changes in accessibility to employment show a positive sign in the final model, none are significant. The authors argue that the lack of relationship may arise because housing is a durable asset and the value may reflect a previous urban structure, such as the monocentric structure.

In addition to the explanation offered by the authors in relation to the interpretation of the results of the previous study, some methodological aspects can be considered. Firstly, it is noted that the research used the traffic analysis zone (TAZ) as the analytical framework and the median values as 'representative' figures. However, the level of within variance of these zones is not clear, nor is it clear whether the median value is an appropriate assumption for generalizing observations within zones. In addition, the functional form of the econometric specification of all models is linear-linear, which has been shown in previous research to have impacts on the coefficient (Mohammad et al., 2013). The study also reported high collinearity (a variance of inflation factors over 20 for accessibility variables) in the cross-sectional models. Moreover, potential spatial autocorrelation, which has been shown to be important in this kind of study, is not discussed or treated (Anselin & Lozano-Gracia, 2009; Yin, Zhang, Patterson, Silverman, & Wu, 2020). A final observation relates to the limited control of neighbourhood variables, as emphasized by the hedonic modelling literature, to prevent omitted variable bias .

Previous research has also reported the impact of public transport accessibility in different intra-urban structures, i.e. Rome, Italy, and Santander, Spain (Cordera et al., 2018). This study considers various forms of accessibility by public transport, namely travel time to the CBD by public transport, public transport coverage, and gravity-type measures. These are estimated within a spatial interaction model framework using commuting flows. The methodology follows a cross-sectional design using property asking prices as the dependant variable. The findings support the use of accessibility measures to explore the incidence of public transport in the capitalization of properties. They also show that the magnitude of the accessibility measures is different for the different cases studied. It is argued that a structure with a single dominant central area, as in Santander, reduces the role of the accessibility compared with travel time to the CBD. This, however, is not the case in Rome, where results confirmed a significant and positive role of public transport accessibility on property asking prices. The study supports the use of measures that capture access to 'end opportunities' instead of infrastructure-based measures, e.g. public transport coverage within the zone.

Most of the previous studies confirm a positive relationship between accessibility to employment and residential land value at the city and regional level in a variety of contexts. These contribute to the extension of the monocentric city model by considering the spatial dispersion of opportunities. In addition, it can be observed that many of the empirical studies of this type focus on accessibility generated by private car, while the role of public transport remains limited. The next section focuses on empirical work adopting a broader view of accessibility.

Local impacts at the local or district level

Some empirical studies examining the residential land value in relation to specific interventions of public transport infrastructure are moving towards the consideration of accessibility effects from a broader perspective. This includes a focus on lower spatial scales, i.e. at local or district level. Some of these examples adopt simple but relevant approaches (e.g. Mulley, Sampaio, & Ma, 2017; Rodríguez & Mojica, 2009). Specifically, some assume accessibility improvements in areas already served by the network due to the expansion of the system in non-adjacent areas (referred to as 'network effects'). Then, the potential effects are assessed in a DID research design. There are also other more complex approaches which draw on explicit measurements of accessibility to assess the potential changes (e.g. Ahlfeldt, 2013). A common characteristic of these studies is the acknowledgement of the accessibility effects generated by a public transport innovation according to a broader understanding of accessibility.

One piece of empirical research of this type examined the effects of two new lines of the BRT network on the residential value of areas already served by the network (referred to as 'non-expansion areas') in Bogotá, Colombia (Rodríguez & Mojica, 2009). The analytical approach was a DID design based on offered property prices collected from a real estate web platform. In this study, the treated observations were assumed to have received accessibility benefits after the construction of two non-adjacent lines of the BRT network. The results suggest that asking prices are between 13% to 14% higher than the control areas after the extension of the network. These findings call for the consideration of wider accessibility effects that are not only defined by the usual 'catchment' areas drawn as buffering zones; they illustrate further insight into the positive and indirect benefits of the improvement of the public transport network.

One study adopting a similar approach to analyse the bus network in Brisbane, Australia, reported coincident results (Mulley et al., 2017). Its aim was to assess the distribution of the capitalization in property prices in different expansion stages of the BRT system in a set of DID analyses. In this research, the treatment group was defined as properties within an 800 metre radius from a station of an existing line (i.e. the South Eastern Busway), while the control groups were defined according to two different approaches: the first was a distance based approach, i.e. properties between 800 metres and 1600 metres away from the station, while the second was based on a propensity score matching (PSM) technique aimed at identifying equivalent properties within a buffering zone from the stations. According to the distance-based model, the average treatment effect was 7% in the stage right after the opening of the extensions (in 2002) and a similar figure four years later (in 2006), relative to the baseline stage (1996). The impact increased to up to 10% when the network was expanded further (in 2011), connecting it to the CBD. Even though both models identified positive results, the PSM model showed differences in the timing of the effects. However, the PSM model did not reflect any increments on the value of land after the completion of the last non-adjacent BRT segment in 2011 (i.e. the Northern Busway).

The so-called ‘network effect’ examined in these studies can be understood from the perspective of accessibility. Specifically, if a modification to the network reduces transport costs or enables mobility to additional areas, it is then expected that more opportunities can be accessed by residents in the affected areas. This is true to the extent that these additional opportunities are located within a ‘reasonable’ travel cost and are ‘valuable’ for people. Thus, the effects on land value as measured by the previous empirical studies can be expected to depend entirely on the conditions of the local context. Specifically, a significant increment in land value is expected to arise only if the assumption of increased accessibility, given the non-adjacent extension of the transport network, is correct. This, of course, depends on the multiple components involved in the concept of accessibility, e.g. the characteristics of the transport network, land use, and perception of residents. Still, the contributions of the previous studies are valuable since they support the idea of increased accessibility effects and its respective heterogeneous distribution across the network. This view contrasts with that of the ‘classic’ approach, where transport interventions are often interpreted in isolation from the rest of the transport system.

Ahlfeldt (2013) fully adopts the view discussed above in order to estimate possible capitalization effects associated to two extensions of the rail network in London, UK. The first step in this study was to estimate the hedonic function from the perspective of a gravity-type model of employment using cross-sectional property data at the city level, similar to some of the studies referred to in the previous subsection (Ahlfeldt, 2011; Osland & Thorsen, 2008). It aims at identifying an appropriate impedance functional form and the respective spatial deterrence parameter (referred to as the spatial decay parameter in the original source and β in Section 2.2) in the area of study. These approximations were validated within a DID design. The estimated hedonic function and the gravity-type parameters were employed to predict property prices for a set of out-of-sample observations at the local level. This was done according to the accessibility changes introduced by the opening of the 1999 extension of the London Underground Jubilee Line and Docklands Light Railway network while considering competing modes (i.e. private car). The predictions were then evaluated against observed property changes in a quasi-panel/repeated sales framework.

Firstly, the results in Ahlfeldt (2013)'s study support the use of gravity-type measures of employment as an adequate substitute for the often-used distance to the CBD measure. This differs from previous findings which suggest that one measure complements the other (Osland & Thorsen, 2008; Thériault et al., 2013). Secondly, the thorough examination of the functional form in the accessibility measure (e.g. using parametric and semiparametric approaches which allow flexibility) shows that the negative exponential is an adequate form, in line with previous studies (Osland & Thorsen, 2008). Furthermore, the calibrated distance deterrence parameter is within a close range from those reported in previous literature (Ahlfeldt, 2011; Osland & Thorsen, 2008), and the estimates produced by cross-sectional data are robust when compared to those calibrated in the DID design.

Thirdly, the study illustrates that the spatial distribution of accessibility benefits and the consequent predicted changes in property prices are heterogeneous across the network. Specifically, it shows that some existing stations receive considerable benefits arising from the network effects discussed previously. This supplements and supports the rationale and findings of some of the empirical studies discussed above (Mulley et al., 2017; Rodríguez & Mojica, 2009). At the same time, some newly opened stations produce small or negligible improvements given the local characteristics of the context, in line with the perspective of accessibility. This finding may support the argument that new public transport infrastructure lacks effect in areas identified with good accessibility levels and high employment density (Guzman et al., 2021). Lastly, the gravity-based approach produces good predictions when only using information extracted prior to the extension of the network. This contrasts with the 'severely' underestimated predictions produced based on the popular measure of distance to the nearest station, which was estimated as a benchmark.

Altogether, these studies highlight the adequacy and relevance of incorporating the notions of accessibility theory into the study of the effect of public transport infrastructure on residential land value. This is in line with relatively recent reviews which support this idea as a way to advance these types of studies (Higgins & Kanaroglou, 2016).

3.2.3 Net benefits in empirical studies

While several empirical studies focus on the estimation of the hedonic function, only a few provide estimates for the net benefits (or aggregate welfare change) associated with the public transport infrastructure improvements. Below, some findings are presented with the purpose of setting an overall benchmark for their scope.

The empirical findings available suggest that gains are considerable in relation to capital costs. In a governmental report issued by Transport for London (TfL), the transport public agency for the Greater London Authority (GLA) in the UK, it was reported that eight large transit projects could produce land value uplifts equivalent to 2.4 times their implementation costs (TfL, 2017). Likewise, a study in Perth, Australia, estimated 'project-induced' revenues of a rail line which range from 0.60 times the capital cost in a "no intensive" scenario (without increasing development use) to 1.3 times the capital cost in an intensification scenario (McIntosh et al., 2015). In the case of Shanghai, China, it was suggested that by 2020 property owners could accrue more than 50% of the economic benefits produced by the metro network (W. Liu, Wang, & Wang, 2018). Therefore, property owners are the largest beneficiaries compared with the rest of the recipients, i.e. users, government, and commercial companies. In line with the overall findings, a recent study suggested that the gain is equivalent to 1.2 times the cost of the Second Avenue Subway extension project in New York City (NYC) (i.e. $5.53 billion USD in a base-line scenario versus $4.5 billion USD for the cost of the project) (Gupta, Van Nieuwerburgh, & Kontokosta, 2022).

None of the examples referred to above have explicitly modelled demand (as in Rosen's Second Step or the ESM). Thus, they follow the assumptions discussed in Section 3.1.

3.3 Spatial aggregation bias

According to the location-based approach discussed in Chapter 2, the level of accessibility in a region S is most commonly evaluated by subdividing it in J discrete nonoverlapping areal units {i.e. $S = (S_1, ..., S_J)$} (Kwan & Weber, 2008; Levinson & Wu, 2020). Under this concept, individual observations $i = (1, ..., n)$ (e.g. households) represented as points in space that fall within a zone j are assumed to have equal levels of accessibility. Following this process, empirical studies adopting location-based measures for the study of capitalizations of transport on land value use a variety of areal definitions. These include traffic analysis zones (TAZ) (Adair et al., 2000; Bourassa, Hoesli, Merlin, & Renne, 2021; Diao, 2015; Iacono & Levinson, 2017; Srour et al., 2002), voting precincts (Ahlfeldt, 2011), post code zones (Ahlfeldt, 2013; Gjestland et al., 2020; Osland & Thorsen, 2013), or census blocks (Bourassa et al., 2021). Often, these boundaries are defined beforehand for administrative or political purposes and then adopted by these analyses rather than a spatial analytical scheme which follows the underlying process under investigation, e.g. by designing optimal zoning systems (Openshaw & Rao, 1995).

Despite the fundamental role of the spatial definition in accessibility measures (Levinson & Wu, 2020), these types of studies have rarely discussed the potential implications of their choice explicitly. Yet, the literature in many fields, including in transport studies, show that this decision can influence results (Horner & Murray, 2004; Kwan & Weber, 2008; Ortega, López, & Monzón, 2012). This issue is referred to as the modifiable areal unit problem (MAUP) and it is well-known in the field of geography (Openshaw, 1984; Wong, 2009b, 2009a).

The MAUP literature distinguishes two main types of issues, namely (1) *scale* effects and (2) *zoning* effects (Wong, 2009a).

- The scale effects are related to a variability of the results in spatial analyses derived from the level of spatial resolution used. For example, an increased number of zones in S constitutes higher resolution schemes.
- Meanwhile, the zoning effects result from the spatial configuration of the areas chosen while maintaining the number of zones in S fixed, i.e. how the boundaries delineating the areal units are drawn. In addition, the zoning problem has been described as the variability in results generated by the different regrouping of smaller zones of a given scale (Dark & Bram, 2007; Kwan & Weber, 2008). The former view is pertinent when the source of the information is disaggregated at the point level, whereas the second is suitable when the source is already spatially aggregated in smaller areal units.

There has been considerable effort to understand the effects and structural patterns of MAUP. The clearest effects are according to simple univariate statistics. Broadly speaking, it has been suggested that scale effects are easier to foresee than zoning effects. General observations suggest that the mean values can remain more or less constant in various scales while the variance declines (Dark & Bram, 2007; Wong, 2009b). Also, the lower the spatial resolution, the narrower the size of the range. However, the structure of the effects has been shown to dissipate in multivariate regression (Fotheringham & Wong, 1991).

It has been suggested that zoning effects are difficult to predict (Dark & Bram, 2007). Some argue that there are further aspects involved, i.e. spatial autocorrelation and the aggregation mechanism (e.g. mean, sum, median) (Wong, 2009b). Accordingly, it is argued that the zoning effect is minimal if there is a random spatial pattern, i.e. low spatial autocorrelation. In contrast, the results are highly sensitive to zoning if there is a strong positive spatial structure in the data. However, the effect is minimized if the definition of areal boundaries follows the pattern of the *underlying spatial process*. Recent studies have confirmed this idea (S.-I. Lee, Lee, Chun, & Griffith, 2019). Specifically, it is suggested that higher spatial autocorrelation increases the overall MAUP effects. In particular, those resulting from zoning are found to be 'severe.'

The regularities of the effects in bivariate analyses seem more consistent than those in multivariate analyses, e.g. regression analyses. When both variables are spatially aggregated at higher levels (i.e. lower spatial resolution) in bivariate analyses, the relationship will tend to be stronger (Wong, 2009b). Since most of the outcomes in multivariate regression draw on co-variances, Wong (2009b) argues that these effects will be reflected in more complex statistical analyses. This is illustrated in a classic piece of research which uses incremental scales to examine the effects on the results of multiple regression, concluding that "it is possible to find any desired level of accuracy [measured according to adjusted R-squared] simply by aggregating the data sufficiently" (Fotheringham & Wong, 1991, p. 1041). The effects on regression analyses have been reported within a considerable range, i.e. from a dire view, where it is suggested that the problem is severe and 'essentially unpredictable' (Fotheringham & Wong, 1991), to one recommending an emphasis on the modelling specification (Briant, Combes, & Lafourcade, 2010; Ye, 2022; Ye & Rogerson, 2022). For instance, Ye & Rogerson (2022) argues that if the model is correctly specified, the regression coefficients remain unbiased but less efficient in the presence of the MAUP.

Applied research in related fields provide relevant contributions. In the context of economic geography, Briant et al. (2010) examined the scale and zoning effects on frequently used modelling specifications employed in the field, i.e. spatial concentration (e.g. Gini indices, Ellison and Glaser indices), agglomeration economies, and gravity equations (also known as spatial interaction models). According to the empirical results, it is argued that scale is a second order concern, although still important at larger scales. In this context, the model specification is a first-order issue. Meanwhile, zoning remains a third-order concern. Based on these findings, it is suggested that "when zoning systems are specifically designed to address local questions, as is the case for French employment areas, we definitely argue that they should be used. Those who are left with other administrative units should not worry too much however, as long as the aggregation scale is not too large" (p. 300).

3.3.1 MAUP in the study of accessibility

It has been argued by Kwan (1998; 2008) that space-time accessibility measures are scale independent. This is because they use a 'nonzonal' or frame-independent approach. As such, they are not affected by the scale effects identified in MAUP literature. However, these types of analyses require considerable additional data which are rarely available (Stepniak & Jacobs-Crisioni, 2017). Thus, location-based measures still represent a relevant option. The MAUP is particularly relevant for location-based measures. Despite the potentially relevant role of MAUP in this approach, there is a surprisingly limited amount of literature examining the specific implications and potential mitigation alternatives for the estimation of location-based measures, as noted earlier (Stepniak & Rosik, 2015). This subsection presents some of the relevant findings in this regard.

An empirical study drawing on a central region of Poland (Mazovia) examined three alternative gravity-type accessibility measures using information estimated at various spatial scales (Stepniak & Rosik, 2015). Specifically, the main analytical spatial unit was the municipal level. The base-line model simply computed accessibility at the municipality level using population as a measure of attractiveness and negative exponential impedance function. The other two models disaggregated further the internal spatial composition of the municipality in a uniform grid, i.e. 1 km^2. In the second model, accessibility was estimated using grid-level information and then re-aggregated by the corresponding municipality to facilitate the comparative analysis. The last model is similar to the previous except that inter-zone travel times were weighted by the population of the destination. Here, all models used the same spatial decaying parameter value which was exogenously estimated. The self-potential accessibility was estimated using the area of the zones (this refers to the level of accessibility considering only the within-zone opportunities).

The findings in Stepniak & Rosik (2015) show that large areal units (municipality-based) produce smoother estimates than the alternative approaches, which draw on disaggregated inputs (grid-based) in line with the MAUP literature (Wong, 2009b). Additionally, the analyses examined the differences between the self-potential estimates and concluded that the differences observed can be attributed to the elements involved in the gravity-type measures (e.g. transport and land use component) and not to the process of estimating self-potential accessibility. Since the measures employed add a considerable level of complexity to this study, it is challenging to interpret the MAUP effects in isolation to the other processes (e.g. population weights on both the size of opportunities and travel time, or different procedures of self-potential).

Further work emphasizes that a relevant source of error in the context of location-based accessibility and spatial interaction analyses is the aggregation of figures to account for spatial separation between observations (commonly represented by travel time) (Stepniak & Jacobs-Crisioni, 2017). This view identifies two types of source errors (drawing on Hillsman & Rhoda (1978)), namely (1) the aggregation to account for travel time between locations within the same zone (intra-zone) and (2) between locations in different zones (inter-zone). An empirical comparison of common forms of aggregation found in the literature with a benchmark based on disaggregated travel times (i.e. 1 km grid obtained from Stepniak & Rosik (2015)) provides guidance for minimizing the effects. The findings suggest that a population-weighted average distance to a zone's population-weighted centroid is more accurate than the areal-based method according to a half radius for intra-zone measures. In the case of inter-zone travel time, it is suggested that distance between respective population-weighted centroids is preferable over geometric centroids.

Although the previous research represents valuable contributions in the study of accessibility, there are further aspects suggested in the MAUP literature that remain understudied, e.g. the interaction of MAUP with autocorrelation, or zoning effects.

3.3.2 MAUP in hedonic property models

Even in a best-case scenario, where property value data is disaggregated at the individual level, the MAUP continues to be relevant in the context of hedonic modelling (D. Lee, 2016). This is not only related to the measurement of accessibility but also to the set of relevant environmental measures that are assigned to individual dwellings. As discussed above, these measures should represent households' perceptions (Bishop et al., 2020; Klaiber & Smith, 2011). Thus, both scale and zoning can play important roles.

A recent study illustrates how the choice of observational units that spatially aggregate data can considerably impact the estimated WTP (Bivand, Sha, Osland, & Thorsen, 2017). It argued that aggregating the property value data according to the data generation process of the variable of interest (air quality) not only mitigates misspecification concerns but also produces markedly higher WTP estimates. These are over three times larger than the benchmark when including spatially lagged independent variables.[6] These findings are in line with Wong (2009b)'s argument suggesting that estimates are particularly sensitive to zoning and support the idea of using areal configurations that follow the spatial pattern under study.

D. Lee (2016) showed that the MAUP effects are considerable even if the property price and structural characteristics are entered at the individual level and only the locational and neighbourhood characteristics are spatially aggregated. Firstly, it was found that scale and zoning effects are present in both standard hedonic models and spatial hedonic models (Spatial Lag Model or SAR). However, in terms of model fit and scale, the results were ambivalent. While the accuracy of models fitted by OLS tends to increase with variables aggregated at higher scales, up to a certain level of aggregation,

[6] The study in Bivand et al. (2017) re-examines the original estimates of a classic study by Harrison & Rubinfeld (1978).

the precision of the results produced by spatial models tend to worsen. Zoning also affects the predictive power of models. The results suggest that administrative zones perform worse than optimally generated zones (minimizing the aggregated dwelling price deviation) or randomly created zoning schemes (which maintain a similar scale). Regarding the estimation of coefficients, structural variables are slightly affected while locational and neighbourhood vary indistinctly. It is thus concluded that no general pattern can be found for coefficients according to the MAUP effects, in line with the literature (Fotheringham & Wong, 1991). D. Lee (2016) further observes that neighbourhood variables present a higher degree of correlation at lower resolutions, which may increase collinearity and, consequently, affect the coefficient estimated.

One recent research project examined the potential special case of the Simpson's paradox in the context of scale effects (Fotheringham & Sachdeva, 2022). This paradox supposes that spatial aggregation may not only affect the precision of the coefficients but also reverse their sign. It is argued that scale plays a distinct role in global and local models. Specifically, it is suggested that spatial processes are not spatially constant. Thus, a renovated approach should shift the focus from the data to attention to the process. Local models are argued to be more adequate for the latter.

With the purpose of illustrating this potential effect, Fotheringham & Sachdeva (2022) estimated a hedonic function using a global model (OLS) and a local model at the individual level in addition to a set of models aggregating data of eleven spatial scales (square uniform grids from 400×400 m to $4,000 \times 4,000$ m). The focus was on how the age effects the dwelling's price. The results suggest that age is positive in the global model while the individual local model predicts negative effects in most locations. Furthermore, it was suggested that the sign of the coefficient is shifted to a positive when data is aggregated to units equal to or larger than one squared kilometre. The suggestion of approaching scale effects heterogeneously in space appears to be relevant. However, in the context of hedonic modelling, the illustration may be affected by the well-documented omitted variable bias referred to in the literature (Bishop et al., 2020; L. O. Taylor, 2008, 2017). This is because the model specification does not consider general location controls and minimum neighbourhood variables. Thus, age may be reflecting a location's advantages.

3.4 Summary and final remarks

A key argument of LVC rests on the land value increments produced by actions other than the landowner (Gómez-Ibáñez et al., 2022; Smolka & Maleronka, 2018). Therefore, it is pertinent to advance discussion on the distribution of the economic benefits triggered by transport infrastructure. Environmental economics provide useful analytical tools for the valuation of nonmarket goods. The hedonic model has a long tradition in this field. However, it presents several econometric challenges, such as the selection of an adequate functional form and potentially omitted variable bias (Klaiber & Smith, 2011; Malpezzi, 2003). Furthermore, the evaluation of net benefits for nonmarginal or nonlocalized environmental changes depends on important assumptions, critically those related to the elasticity of demand (Klaiber & Kuminoff, 2014). More recently, equilibrium sorting models have emerged as an alternative with enormous potential to overcome many of these issues. Still, they are limited in their application chiefly due to important data requirements and technical barriers.

Empirical studies have mostly found positive effects associated with public transport accessibility using the hedonic method. Yet, there is an important degree of variability in the magnitude of the effects. The approaches to operationalize the level of accessibility induced by transport infrastructure vary widely. Location-based measures have emerged as a more theoretically robust alternative than those based on oversimplified assumptions attributing homogeneous benefits to public transport stations. The hedonic model and location-based approach are not free of challenges in their application. For instance, a known issue is the bias generated by the arbitrary spatial aggregation of neighbourhood and locational characteristics of a dwelling studied under the MAUP umbrella.

Although the empirical literature studying transport infrastructure and land value is vast, there are several aspects that remain limited. First, the accessibility measures employed are often inconsistent with respect to accessibility theory, overlooking the land use component or relevant characteristics of the transport system. Second, some studies are moving towards the incorporation of comprehensive accessibility measures. Still, there are aspects to be clarified. For example, the potential role of the person component, the implications of the spatial unit of aggregation chosen, and the potential advantages of modelling accessibility based on detailed public transport timetables. These aspects are explored empirically in subsequent chapters drawing on the case of Greater Mexico City. Before proceeding to the analytical part of the thesis, the next chapter introduces the area of study. The details provided draw interpretation in line with the theoretical framework established in Chapter 2 as well as the considerations presented in this chapter.

Chapter 4

Study area: Greater Mexico City

The implementation of land value capture (LVC) instruments is linked to several important contextual conditions at both the city and the local level (as per the review in Chapter 2). These generally include political and administrative arrangements, the dynamics of the housing market, or the characteristics of the transport system. Additionally, the application and interpretation of environmental valuation methods, as presented in Chapter 3, are subject to assumptions determined by these contexts. This chapter introduces the area of study, Greater Mexico City (GMC), and provides the relevant contextual details.

The main contribution refers to the delineation of an essential city profile of GMC. This is comprised of its overall political-administrative and socio-economic contextualization, main mobility trends, housing sector, and local politics of LVC. A key component of this contribution includes the classification of the public transport system, drawing on an academic framework and the characteristics of the local services. This concludes with the definition of the *main public transport network* (MPTN), the object of main interest of this research. This is the first of its kind in academic literature. This contribution is useful to inform the discussion about the potential of LVC in line with the main objective of this thesis in subsequent chapters. Additionally, some of the methodological choices will be guided by the local conditions of the area of study.

The findings suggest various specific features of the area of study. First, although the economic activity is attracted to central areas, empirical studies suggest the existence of relevant sub-centres. The travel survey also highlights the role of public transport modes in the overall mobility system. These characteristics provide further support for the adequacy of location-based accessibility measures for addressing the main objective of this research. Secondly, the current housing system operates mostly as financial enabler. The policies in place actively seek easing frictions within the housing market. The housing sector appears to be settling in GMC, considering that the stock grew at a lower pace than it did at the national level during the last decade. The housing stock is characterized by new, detached dwellings, occupied by the owners. Thirdly, although the legal framework only implicitly serves as the foundation for LVC, there are explicit administrative instruments relative to public transport infrastructure with the potential to recover at least a portion of land value increments. Taken altogether, these confirm the appropriateness of this study and settle the ground for discussing empirical findings in subsequent chapters.

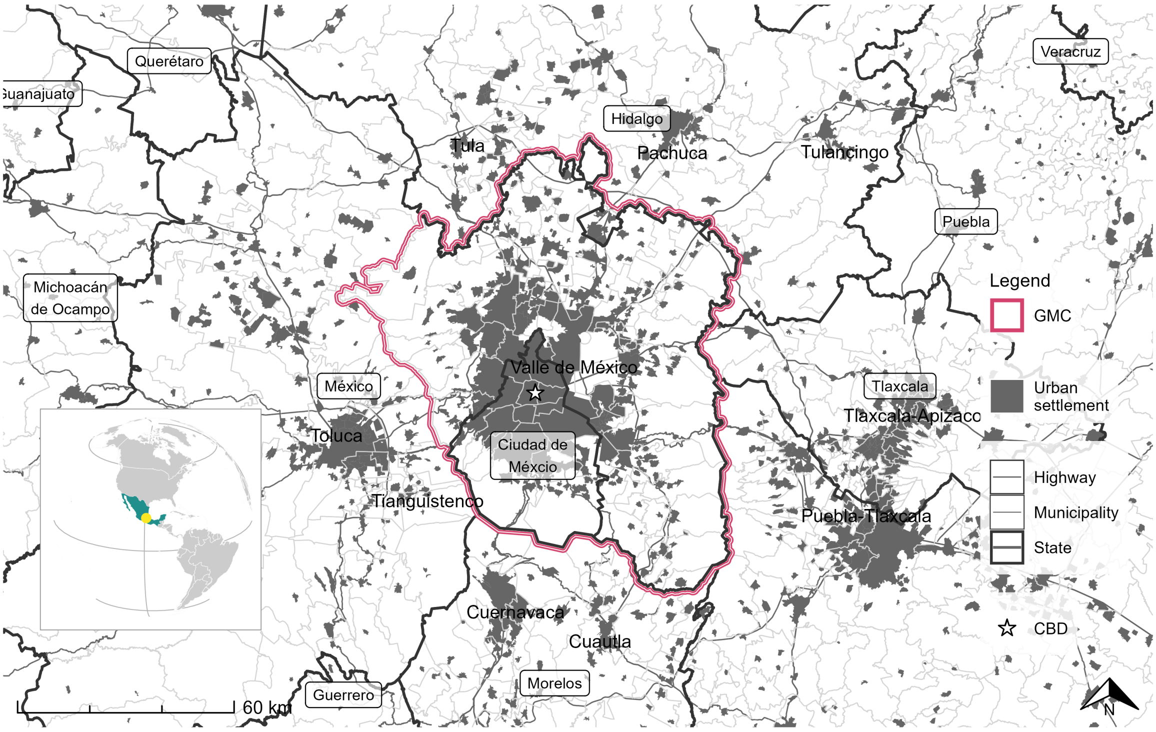
The chapter is split into five sections. Section 4.1 provides an overview of the metropolitan area, including the main political, administrative, demographic, and economic characteristics of GMC. Section 4.2 describes some of the most important mobility patterns and the central characteristics of the local modes of public transport. Specifically, it provides a definition of the main public transport network (MPTN). Section 4.3 provides an overview of the housing system in GMC in the broader context at the national level. It also provides relevant aspects of the national housing policy and outlines the nature of the housing stock. Section 4.4 discusses the politics relating to LVC in the local context. The final section (Section 4.5) summarises the chapter.

4.1 An overview of Greater Mexico City

The United States of Mexico (Mexico) is a federal republic comprised of 32 states (formerly 31 states and one federal district). The political and administrative structure consist of three hierarchical levels of government: the national (or federal) at the top; federal states (*Estados*), which are sovereign territories, at the subnational level; and municipalities at the lowest level (these are the basic units in Mexico and are designated as *alcalidias* within the state of *Ciudad de México*, and as *municipios* across the rest of the country). Mexico had 120 million inhabitants in 2015 (INEGI, 2015b), of whom 99 million (83%) resided in urban areas (Sedatu, Conapo, & INEGI, 2018).

Greater Mexico City (GMC) (*Zona Metropolitana del Valle de México*) is located in an extensive valley in the south-centre of Mexico, as shown in the location map (bottom-left) and as to its regional context in Figure 4.1. The traditional central business district (*Zócalo*) lies at coordinates 19° 25’ N and 99° 08’ W. GMC is the second largest urban agglomeration in Latin America (after São Paulo, in Brazil), and it is by far the largest in Mexico.[1] In 2015, there were almost 21 million inhabitants (INEGI, 2015a; United Nations, 2015). The population grew rapidly during the 1990s, at a median annual rate of 1.7%. This pace of growth decreased, to an annual figure of 0.9% between 2000 and 2010, and to 0.8% between 2010 and 2015.

[1]GMC is followed in population size by the Metropolitan Area of Guadalajara, which had 4.9 million inhabitants in 2015 Sedatu et al. (2018).



Source: the Author based on Marco Geoestadístico Nacional 2014 Versión 6.2 (INEGI, 2014); Delimitación de las zonas metropolitanas México 2015 (INEGI, 2018); Red Nacional de Caminos (SCT, 2015); Infraestructura de Datos Espaciales Abiertos (IDEA), Universidad Nacional Autónoma de México (IDEA, 2022).

Figure 4.1: Greater Mexico City in the regional context.

The official delimitation of GMC consists of 76 adjacent municipalities shown in Figure 4.1 (Sedatu et al., 2018). Each municipality is part of one of the three following states: (1) *Ciudad de México* (16 municipalities); (2) *Hidalgo* (1 municipality), and (3) *México* (59 municipalities). The state of Ciudad de México (formerly the federal district) is the national capital and it is also where the *central city* of the metropolis is located. The administrative arrangement of GMC is shown in Figure 4.1. Ciudad de México shares its borders, on its west, north and east, with the State of Mexico, and its southern border with the state of *Morelos*. The only municipality of Hidalgo that forms part of GMC (*Tizayuca*) is located to the north of the central city. The total area encompassed by GMC is 7,889 km^2, and it is home to 17% of the national population.

Table 4.1 provides a summary as to the composition of GMC by state. This illustrates that the vast majority of the territory (80%) lies within the state of Mexico. However, only slightly more than half of the population resides in this jurisdiction (56%). The number of *employed people* (which approximates the number of employees)

	Municipalities[1]	Population[2]	Employed people[3]	Area[1]
Region/State	*N*	*Million (%)*	*Million (%)*	*Sq. Km (%)*
Greater Mexico City	76	20.89 (100%)	5.08 (100%)	7 866 (100%)
Ciudad de México	16	8.92 (42.7%)	3.60 (70.9%)	1 495 (19.0%)
Hidalgo	1	0.12 (0.6%)	0.03 (0.5%)	77 (1.0%)
México	59	11.85 (56.7%)	1.45 (28.6%)	6 295 (80.0%)

Source:
[1] Delimitación de las zonas metropolitanas de México (SEDATU, CONAPO, & INEGI, 2018);
[2] Encuesta Intercensal 2015 (INEGI, 2015);
[3] Censos Economicos 2014 (INEGI, 2015).

Table 4.1: Composition of Greater Mexico City, 2015.

in Ciudad de Mexico is 2.5 times larger than those in the entire state of México.[2] Hidalgo accounts for 1% or less of the total numbers, with respect to the categories of population, employed people, and land area, of the metropolis.

GMC is the most productive metropolis in the country in economic terms, generating a quarter of the national GDP (OECD, 2015). Table 4.2 presents the distribution of employed people by sector, utilising the North American Industry Classification System (NAICS). The largest sector, with respect to the metropolitan level, is *retail trade* (21%). This is followed by *administrative posts and support and waste management and remediation services* (e.g. those working in offices and in certain public services) and *those working in manufacturing* with each encompassing an equal proportion (14%) of the total number. Other sectors with an important share are *accommodation and food services* (7%) and *finance and insurance* (6%). There are important distinctions to be drawn, when these figures are disaggregated by state. For example, while *retail trade* still plays an important role in Ciudad de México (16%), the *administrative and support and waste management and remediation services* sector dominates the economy of this state (19%). This is the biggest percentage of any of the measured categories in this jurisdiction and is proportionally larger than the figures for this category with respect to the other states, by some considerable distance: e.g. 19% compared with 2% in Hidalgo and with 4% in México. Similarly, the proportion of population employed in the *finance and insurance* sector is 9% in Ciudad de México, whilst this category represents less than 0.5% of those employed in both Hidalgo and México. The local economy of Hidalgo is primarily driven by *manufacturing*, in which almost half of those employed in this jurisdiction (46%) are employed. In the state of México manufacturing also plays a significant role. Here, the economic activity is primarily that of two sectors, *retail trade* and *manufacturing*, representing about a third (33%) and a quarter

[2]The Office for National Statistics in Mexico (INEGI) uses the term *personal ocupado*, which refers to the regular staff associated to an establishment. Not all of these are necessarily paid employees.

(25%) of the total labour force, respectively.

As suggested by the figures provided above, a considerable proportion of employment opportunities are located in central areas of the metropolis. Four of the central municipalities in GMC together concentrate 40% of these (INEGI, 2015a).[3] However, several empirical studies have shown that sub-centres also play an important role in the development of urban dynamics. As such, this suggests an emerging polycentric structure (Fernández-Maldonado, Romein, Verkoren, & Parente Paula Pessoa, 2014; Graizbord, 2008; Montejano-Escamilla, 2015; Muñiz, Sánchez, & García-López, 2015). For instance, based on employment density and commuting flows, Muñiz et al. (2015) identify the following seven sub-centres as doing so: *Cuautitlán*, *Ecatepec*, *Tlanepantla*, *Aeropuerto*, *Pantitlán*, *Central de Abastos*, *Santa Fe*, and *Tlalpán*. These findings provide support for the adoption of location-based measures as a means of capturing the spatial dispersion of these opportunities in this context.

4.2 Mobility and the public transport system in GMC

The urban setting described in the previous section generates mobility patterns that represent the travel of significant numbers. According to the 2017 Travel Survey (INEGI, 2018), it is estimated that there are 34.6 million journeys starting or ending in GMC on any regular weekday (Monday to Friday). The purpose of about one in every five of these journeys is going to *work* (22%). This is the second most significant motive for travel, with journeys that involve returning home representing around half of all travel flows (47%). *Education* is also an important motivator, generating one tenth of the total volume (12%). Other relevant reasons for travelling are *to accompany someone somewhere to pick them up* (7%), and *shopping* (6%). The motives that generate the least displacement are *social, recreation, and sport* (3%), *health* (1%), and *administrative services* (>1%).

The journeys in GMC, just outlined, were completed via a range of, and sometimes a combination of, transport modes. Two thirds of the journeys (66%) involve utilising only one mode of transport. The other third utilised a combination of two or more modes. The average number of modes utilised, for a journey undertaken, is 1.5. As summarised in Table 4.3, of the total travel journeys, the following modes were used for at least one stage (either exclusively for 66% of journeys, or in combination with one or more other modes):[4] 64% walking or bicycle (22.4 million);

[3] *Cuauhtémoc*, *Miguel Hidalgo*, *Azcapotzalco*, and *Benito Juárez* host about 40% of the employed population of GMC.

[4] On any given weekday i.e. Monday to Friday.

Sector (NAICS code)	Ciudad de México	Hidalgo	México	Total
	Thousands (%)	*Thousands (%)*	*Thousands (%)*	*Thousands (%)*
Agriculture, forestry, fishing and hunting (11)	0 (0.0%)	- (-)	0 (0.0%)	0 (0.0%)
Mining, quarrying, and oil and gas extraction (21)	3 (0.1%)	- (-)	0 (0.0%)	4 (0.1%)
Utilities (22)	108 (3.0%)	- (-)	1 (0.1%)	109 (2.1%)
Construction (23)	92 (2.6%)	2 (6.1%)	13 (0.9%)	107 (2.1%)
Manufacturing (31-33)	361 (10.0%)	12 (45.8%)	359 (24.6%)	731 (14.4%)
Wholesale trade (43)	211 (5.9%)	1 (3.5%)	89 (6.1%)	301 (5.9%)
Retail trade (46)	577 (16.0%)	4 (17.4%)	476 (32.8%)	1 058 (20.8%)
Transportation and warehousing (48-49)	173 (4.8%)	1 (2.9%)	38 (2.6%)	211 (4.2%)
Information (51)	132 (3.7%)	0 (0.2%)	7 (0.5%)	139 (2.7%)
Finance and insurance (52)	315 (8.7%)	0 (0.3%)	6 (0.4%)	321 (6.3%)
Real estate and rental and leasing (53)	39 (1.1%)	1 (2.2%)	18 (1.2%)	58 (1.1%)
Professional, scientific, and technical services (54)	209 (5.8%)	1 (2.2%)	24 (1.7%)	233 (4.6%)
Management of companies and enterprises (55)	24 (0.7%)	- (-)	1 (0.1%)	25 (0.5%)
Administrative and support and waste management and remediation services (56)	674 (18.7%)	0 (1.7%)	58 (4.0%)	732 (14.4%)
Educational services (61)	133 (3.7%)	1 (2.8%)	68 (4.7%)	202 (4.0%)
Health care and social assistance (62)	96 (2.7%)	1 (2.3%)	41 (2.8%)	137 (2.7%)
Arts, entertainment, and recreation (71)	40 (1.1%)	0 (0.5%)	20 (1.4%)	61 (1.2%)
Accommodation and food services (72)	263 (7.3%)	1 (5.0%)	114 (7.8%)	378 (7.4%)
Other services (except public administration) (81)	152 (4.2%)	2 (6.2%)	107 (7.4%)	260 (5.1%)
Aggregated for privacy protection (NA)	1 (0.0%)	0 (0.8%)	15 (1.0%)	17 (0.3%)

Source: Censos Economicos 2014 (INEGI, 2015); NAICS = North American Industry Classification System.

Table 4.2: Distribution of employed people by sector. Greater Mexico City, 2014.

Mode	Journeys *Million*	%
By foot	21.64	62.61
Bus or trolley (Regular bus and trolleybus)	12.43	35.95
Private car or motorcycle	6.98	20.19
BRT or rail (Metro, LRT, regional rail, or aerial tramway)	5.41	15.66
Taxi (public, APP, bicitaxi, mototaxi)	2.00	5.78
Bicycle	0.72	2.08
Special school/work transport	0.32	0.92
Other	0.04	0.12

Source: Encuesta Origen Destino en Hogares de la Zona Metropolitana del Valle de México (EOD) 2017 (INEGI, 2018).
Note: The sum of journeys adds up to more than the total (34.6 million) since each trip can use more than one mode; BRT = Bus rapid transit; LRT = light rail transit. Regular bus includes: 'Colectivo/Micro', 'Autobús RTP o M1', or 'Autobús'. Rail includes: 'Metro', 'Metrobús', 'Mexibús', 'Tren ligero' (LRT), 'Tren suburbano' (regional rail), and 'Mexicable' (aerial tramay).

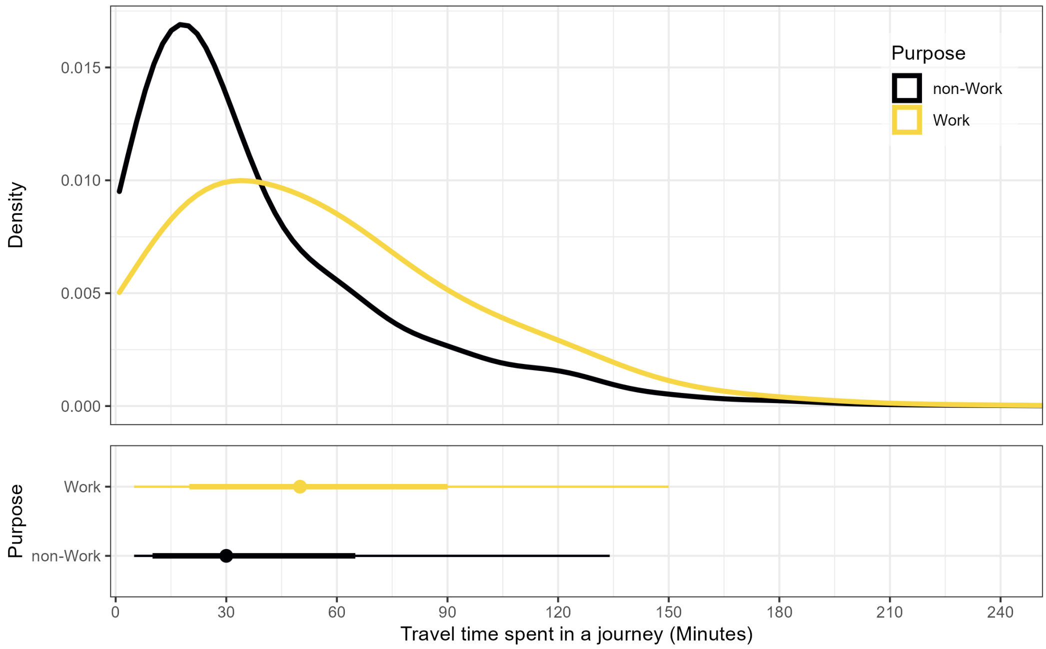
Table 4.3: Modal split. Number of journeys that use one of the following modes. Greater Mexico City, 2017.

16% bus rapid transit (BRT) or rail (5.4 million);[5] 36% bus or trolley (12.4 million); 20% private car or motorcycle (7 million); 6% taxi (2 million); 1% special school or work transport (0.3 million); and less than one percent (0.1%) utilised other forms of transport. The sum total of these figures adds-up to more than the total number of journeys (34.6 million), since 33% of trips used more than one mode of transport.[6] A further disaggregation indicates that the purpose of about a quarter (25%) of the journeys that use a rapid transit mode is travel to work (or half (50%) if journeys back to home are not considered).

In GMC, the median time spent making a journey, with respect to all purposes of travel, is 30 minutes. This figure increases to almost an hour (50 minutes) with respect to travel flows relating to work. Figure 4.2 illustrates the difference between the time spent on travel to work, when compared with that spent for non-work purposes. The upper panel illustrates that while the relative number of flows to work slowly declines with respect to journeys longer than 30 minutes (the overall median), the relative frequency of non-work trips drops at a faster pace. In the lower panel it is shown that the preponderance of the non-work trips take between 10 and 60 minutes, whilst a similar proportion of trips to work take between 20 and 90 minutes. It would seem to

[5] Rail modes include metro, light rail transit (LRT), regional or commuter rail; and aerial tramway.
[6] The source of this information does not identify a 'main mode.' In line with this structure, the journeys are treated as multi-modal.

be the case that residents are more likely to consider time taken for non-work travel to be a barrier to their undertaking such trips, than is the case with regard to travel for the purposes of work.



Source: the Author based on 2017 Travel Survey (INEGI, 2018).

Figure 4.2: Distribution of time spent on individual journeys by purpose in Greater Mexico City.

4.2.1 The transport system and the main public transport network

Before proceeding to characterise the transport system of GMC, the following subsection will provide operational definitions for the different modes of transport and outline their respective characteristics that are utilised with respect to this research.

Within the context of developing countries, Verma & Ramanayya (2015) distinguishes the following three categories of urban passenger transportation, defined in terms of the type of operation and the nature of its usage:

1. Personal or private transportation;
2. For hire or intermediate public transportation, and;
3. Public or mass transportation.

In terms of usage, the *personal* or *private* is characterized by the restricted availability of service for the owner only, the route and time-schedule determination are flexible, and the operation and costs are absorbed by the user. Walking, bicycle, motorcycle, and automobile (car) are the most common modes here. The *intermediate public* *transport* service, also designated as *paratransit*, is supplied by a carrier. This is available to either all individuals or small groups of users who meet contractual conditions (e.g. pay a prescribed rate). Here, the route and time-schedule are most often flexible and the cost is given as a fixed rate. Some modes in this category are taxicab (a type of rental car), auto or cycle rickshaws (a light three-wheeled vehicle) and pulled (human or animal traction) rickshaws or carriages. In developing countries, these are also provided by minbuses or shared taxis, which can adjust their route upon request of users (on-demand). *Public transport* services, also referred to as *transit*, are provided by one or more carrier or directly by the transport authority. The usage is available for the general public and the cost price is a fixed rate for users, although these are normally funded by government subsidies. Some of the main distinctions from paratransit services are that in the latter case the routes and schedules are fixed. Furthermore, the service is shared by 'strangers without private arrangement' (ibid., p. 41). Common modes in this category include bus, light rail transit, and rapid transit or metro, among a wide variety of modes. The remainder of this section elaborate on public transport.

There are several characteristics that influence the definition of a public transport mode, i.e. *right-of-way*, *system technology*, and *type of service* (Vuchic, 2007). For the present work, the focus is on a grouping scheme fundamentally based on the right-of-way (ROW). This is because it is suggested as a major factor substantially influencing both performance and costs (Verma & Ramanayya, 2015; Vuchic, 2007). Additionally, the type of ROW implies some degree of permanence of a transport service, which can translate into certainty in the future availability of a service for the housing market. This characteristic of the infrastructure is referred to as *immobility* in literature (Gómez-Ibáñez & Liu, 2022).

Specifically, the ROW in the context of public transport refers to the strip of land on which the vehicles operate. There are three ROW levels distinguished by their degree of separation from other traffic, namely (ibid.):

- Category C implies a mixed on-street operation with traffic. This can include on-pavement or other road signs controlling the mixed operation between traffic and public transport;
- Category B includes physical elements along the ROW separating the operation of public transport services from traffic. Still, there is interaction mainly at road intersections with pedestrian crossings or other motorized private transportation modes;
- Category A is fully controlled ROW excluding other vehicles or persons from it by physical and often legal means.

Table 4.4 presents a classification matrix of public transport modes by ROW category and technology, i.e. type of *support* and *guidance* (Vuchic, 2007). Here, the support

ROW	Technology (Support and guidance) Highway Driver-Steered	Rubber-tired guided Partially guided	Rail	Specialized
C	Paratransit Shuttle bus Regular bus (on street)	Trolleybus	Streetcar/tramway Cable car	Ferryboat Hydrofoil
B	Bus rapid transit (BRT)	Guided bus	Light rail transit (LRT)	(Cog railway)
A	Bus on busway only	Rubber-tired metro Rubber-tired monorail Automated guided transit PRT	Light rail rapid transit Rail rapid transit/Metro Regional/commuter rail Monorail Schwebebahn	Cog railway Fonicular Aerial tramway

Source: Urban Transit Systems and Technology (Vuchic, 2007, p. 51).
Note: PRT = Personal rapid transit (Not in use).

Table 4.4: Classification of urban public transport modes by ROW and major technological features

technology refers to the type of vertical contact between the vehicle and the surface of traction, e.g. rubber tyre on concrete, asphalt, or other, and steel wheel on steel rail. The guidance technology relates to the lateral guidance of the vehicle. The most common are highway vehicles *steered* by the driver, and rail vehicles guided by flanges and the form of the wheel. Other specialized technologies are included in column four, e.g. water, or suspended. The literature proposes three generic classes of public transport based mostly, but not entirely, on the ROW category utilized, namely: *Street transit*, *Semi-rapid transit*, and *Rapid transit*, for modes operating in ROW category C, B, and A, respectively (Vuchic, 2007). Many of the public transport modes presented in the matrix operate in GMC. A description of these is presented for the local context below.

The broad classification of most of the modes (excluding the specialized) in the first row in Table 4.4 is known as *street transit* (Vuchic, 2007). The most common public transport modes in GMC operate on mixed traffic (i.e. ROW C) and use highway driver-steered technology. According to the 2017 Travel Survey (INEGI, 2018), one third of the journeys (33%, 11.5 journeys) used *paratransit*- or *shuttle*-like modes ('Colectivo/Micro'). The vehicles used for these modes are vans (referred to as *Combi*) or minibuses (locally known as *Microbús*). These are small vehicles of between 6 to 7 metres long propelled by gasoline motors or exceptionally diesel (Vuchic, 2007). For the year 2015, Hernández-Moreno & Mugica-Álvarez (2013) estimate that the vans' fleet size was 56,208, and 32,412 for minibuses. The type of service of these modes is usually as paratransit (or dial-a-ride), feeder to major transfer stations, and regular public transport services (Islas Rivera, 2000). In the former case, there is some degree of flexibility on routes. The schedule tends to be regular although not fixed. Another *street transit* mode operating in the area of study is the *regular bus* (or standard bus).

This is used less frequently than the previous modes in GMC. About 1.3 million journeys used this alternative (3.6%) (INEGI, 2018). This mode is usually operated by single-decker buses of between 10 to 12 metres long. The number of seats is around a maximum of 49 and the passenger capacity per unit is between 50 to 80 spaces, although, literature suggests a crush capacity of 100 for developing countries (Verma & Ramanayya, 2015), in line with government reports (Semovi, 2020). The total number of buses in GMC for the year 2015 was estimated to be 50,809 (Hernández-Moreno & Mugica-Álvarez, 2013). The vast majority of these vehicles are run by diesel motors. Regular bus services are operated by private companies through concessions, which, most of the time, are under regulated conditions including fixed routes and fares (C. Rojas & Yojan, 2016). In addition to this type of arrangement, Ciudad de México directly operates a system of buses named *Red de Transporte de Pasajeros* (RTP). The latter comprises 94 routes and 1,139 buses (Semovi, 2020). RTP serves about a third of the journeys in this modality (408,507) (INEGI, 2018). Apart from regular buses, the government of Ciudad de México also runs eight *trolley* routes through a state-owned agency (*Servicio de Transportes Eléctricos de la Ciudad de México*, STE). This mode operates with vehicles of similar characteristics to those of regular buses (Semovi, 2020), with the exception that trolleys are propelled by electric motors. The operation depends on additional electric infrastructure (overhead wires) and is therefore classed as partially guided (Vuchic, 2007). The 353-fleet serves 146,479 journeys a day (Semovi, 2020). The rest of the modes utilizing the ROW category C namely *tramway*, *cable car* or none of the specialized do not operate in GMC.

Most modes utilizing ROW category B (second row in Table 4.4) fall under the *semi-rapid transit* categorization (Vuchic, 2007). *Bus rapid transit* (BRT) and *light rail transit* (LRT) are the only modes within this public transport categorization found in GMC. A description of these modes and a discussion of their role in the local context is provided below. Since *guided bus* is not present in the area of study, a specific discussion of it is not provided due to the low relevance for the purpose of the present work.

BRT is a highway driver-steered technology mode (first column of Table 4.4). This consists of an integrated system incorporating special physical and operational elements which allow higher capacity, superior performance, and better image than regular buses (Ibid). According to a report by Transportation Research Board*Case studies in bus rapid transit* (2003), the ROW (equivalently referred to as *running ways* in the source) is a key element of this mode. BRT systems use mostly ROW category B, and in limited sections, C (Vuchic, 2007). This permits substantial independence from other traffic. Another important feature of this mode is the stops/stations. These are clearly designated and the spacing between them is much longer than regular buses (400 to 600 m apart and in no case should be less than 300 m) (Vuchic, 2007; Wirasinghe, Kattan, Rahman, Hubbell, & Thilakaratne, 2013). A further distinctive component is the type of vehicles used. These can be articulated (16 to 18 m length), double-articulated (22 to 24 m length), or double-decker buses (10 to 12 m length). Some influential characteristics of the vehicles in terms of performance are size and body structure, doors, interior design, floor elevation, and vehicle propulsion (Wirasinghe et al., 2013). The offered line capacity per direction of BRT systems is usually between 4,000 and 8,000 passenger spaces per hour (sps/h). This can reach up to 20,000 sps/h when multiple parallel lines and overtaking at stations are available (Vuchic, 2007). Regarding the investment cost per kilometre this can range from 5 to 10 million US dollars per pair of lanes. Given the cost-capacity ratio (among other aspects), this mode has become a popular alternative in fast-growing cities (Wirasinghe et al., 2013).

There are two BRT systems operating in GMC, namely *Metrobús* (MB) in Ciudad de México and *Mexibús* (MXB) in the state of México. These are administered by decentralized public agencies and operated directly by state-owned companies or by private companies through concessions. Both systems together serve about 3.1% of the total number of journeys in GMC (INEGI, 2018). As shown in Table 4.5, MB operates seven corridors which add up to a total of 240 kilometres in length (Semovi, 2020). The total MB fleet consists of 657 buses of different types, of which 58% are articulated, 18% are double-articulated (operating only in MB Line 1), 14% are double-decker (operating only in MB Line 7), and 10% are low-platform (operating only in MB Line 4). In the year 2019, the MB system transported 1.4 million passengers on average per working day (Monday to Friday). The first BRT corridor to be implemented was Line 1 in June 2005 along *Insurgentes* avenue (Metrobús, 2021). Also, this is currently the longest and most used, with 54.5 kilometres and 600,000 passengers a day. The shortest corridor in the MB system is Line 5 (MB-L5) with 19.5 kilometres long. Line 7 (MB-L7) is the most recently implemented BRT corridor. This is a 28-kilometres line which started operations in March 2018. This runs along one of the main business corridors in Ciudad de México (*Paseo Reforma*), transporting more than 137,000 passengers a day.

Meanwhile in the state of México, the MXB system operates three BRT corridors as shown in Table 4.5, namely MXB line 1 (MXB-L1), MXB line 2 (MXB-II), and MXB line 3 (MXB-III). The length covered by this system sums 57.3 kilometres, enabling almost 300,000 journeys a day (SITRAMyTEM, n.d.). The fleet consists of 180 articulated buses, which is substantially smaller than the MB system. The first MXB corridor to be implemented (MXB-L1 or locally denominated as *Mexibús I*) consists of a 17-kilometre line which started operations in October 2010. MXB-L2 (or *Mexibús II*) is the most recently implemented line. This opened to the public

Agency	Corridor/Line	Length (Km)	Passengers	Opening	Fleet
MB	1	54.50	522 749	Jun-05	185
MB	2	36.74	200 480	Dec-09	120
MB	3	32.26	185 051	Feb-11	72
MB	4	31.34	70 748	Apr-12	70
MB	5	19.52	96 336	Nov-13	28
MB	6	37.68	222 305	Jan-16	92
MB	7	27.82	137 403	Mar-18	90
MXB	1	16.80	111 573	Oct-10	51
MXB	2	22.30	97 932	Jan-15	74
MXB	3	18.20	78 147	May-13	55

Source: Metrobús (2021); Semovi (2020); SYTRAMyTEM (2019); SYTRAMyTEM (2021);
Note: MB = Metrobús; MXB = Mexibús; Passengers is the average number of passengers in a working day (Mo-Fi) in the year 2019 for the MB system and 2018 for the MXB system; Opening refers to the date operations first started (does not consider extension of a line); Fleet is the total number of buses including vehicles in operation and in mantainance.

Table 4.5: Bus rapid transit (BRT) corridors in Greater Mexico City, 2019.

in January 2015 and runs along a 22-kilometre corridor. Empirical research shows that the planning and delivery process has been less transparent for the MXB BRT system than it has been for MB (Lámbarry Vilchis, Rivas Tovar, & Peña Cruz, 2011). Additionally, in the MXB system the engagement during the planning and operation processes of pre-existing public transport operators in the area has been limited. Consequently, these observations may be reflected in some deficiencies in the quality of the service, e.g. prevalence of competing public transport modes (e.g. *combis* and *minibuses*), lack of fare integration within the same system, poor connectivity to access stations from sidewalks, or deficient materials.[7]

Another form of *semi-rapid transit* present in GMC is LRT. This mode utilizes mostly ROW category B (sometimes A and exceptionally C), and its support and guidance technology is fully guided on rail tracks (second row and second column in Table 4.5) (Vuchic, 2007). Even though the operation of LRT is substantially segregated from other traffic, this still interacts with other modes, mostly at main road intersections. The electric vehicles can couple up to four cars with multiple doors for easy/rapid access. The individual capacity is between 100 to 720 spaces and the line capacity ranges from 10,000 to 24,000 sps/h (Ibid.). Concerning costs, the investment

[7]Literature assessing the quality of service of these systems is scarce. Yet, academic bodies and transport experts have reported glitches of the MXB system in the local media, e.g. https://elpoderdelconsumidor.org/2019/10/a-nueve-anos-de-la-entrada-del-servicio-de-mexibus-continuan-las-deficiencias-en-su-operacion/; https://www.eluniversal.com.mx/metropoli/obras-de-mexibus-elevan-polucion-ong.

per kilometre of a pair of lanes is between 10 to 50 million US dollars. In GMC, there is a 13-kilometre LRT line owned and operated by a state agency called *Servicio de Transportes Eléctricos* (Electric Transport Service, STE for its acronym in Spanish) (Semovi, 2020). This operates in the south of Ciudad de México from *Tasqueña* to *Xochimilco*. This line evolved from a former tramway to a LRT by the end of the 80s (Islas Rivera, 2000). Currently, the fleet consists of 24 double-articulated trains (including in operation and out of service vehicles) with individual capacity of between 292 to 372 passengers (Semovi, 2020). The average number of users in a working day is 93,174. The stations with higher demand in this system are those at the extreme of the line (namely *Xochimilco* at the south-end and *Tasqueña* at the north-end).

Public transport modes utilizing only ROW category A are categorized as *rapid transit* (third row in Table 4.5) (Vuchic, 2007). There are four types of *rapid transit* modes operating in GMC, namely *rubber-tyred metro*, *rail rapid transit*, *regional/commuter rail*, and *aerial tramways*. A description and characterization of these in the local context is provided below. The rest of the modes within this category in Table 4.5, namely *bus on busway only*, *rubber-tired monorail*, *automated rapid transit*, *personal rapid transit* (PRT), *monorail*, *cog railway*, and *funicular*, are not discussed here due to its low relevance for the present work.

Rubber-tired metro and *rail rapid transit* are similar *rapid transit* transport mode systems utilizing different support and guidance technology. The former uses rubber tires on a concrete or steel surface whilst the latter employs steel wheels running on steel rails. This difference is reflected mainly in technical aspects related to comfort, initial or maintenance costs, adhesion between tires/wheels and guideway, acceleration, etc (Vuchic, 2007). Since the functional characteristics are the same between these two modes, these are discussed generically as one type and referred to as *metro* for the purpose of the present work. Metro systems use electric trains of between three to ten cars. The individual vehicle capacity ranges from 720 to 2,400 spaces while the line capacity is between 40,000 and 70,000 sps/h. This is the highest performance urban mode and the most expensive. The capital cost ranges from 40 to 100 million US dollars per kilometre of pair of lanes (ibid). This type of investment enables the use of fully controlled ROW which permits a reliable and safe service. This is one of the main advantages of the metro. Furthermore, stations are often accessed using pre-paid systems which allow rapid and controlled access of passengers.

Mexico's City public agency *Sistema de Transporte Colectivo* (STE) owns and directly operates 12 metro lines in GMC shown in Table 4.6. The system started operations in 1969 with Line 1 (Metro-L1, from *Zaragoza* to *Chapultepec*) (STC, 2022). The fleet consists of 393 trains from which the majority (84%) uses rubber-tired technology and only a small proportion (16%) uses steel rail support and guidance

Agency	Corridor/Line	Length (Km)	Passengers	Opening	Fleet
STC	1	17.68	726 262	Sep-69	49
STC	2	22.01	810 228	Aug-70	41
STC	3	22.59	696 720	Nov-70	54
STC	4	10.74	87 106	Aug-81	14
STC	5	15.68	264 760	Dec-81	25
STC	6	13.00	158 075	Dec-83	15
STC	7	18.40	349 365	Dec-84	32
STC	8	19.37	404 499	Jul-94	30
STC	9	14.44	351 401	Aug-87	34
STC	A	17.19	341 879	Aug-91	33
STC	B	22.25	454 834	Dec-99	36
STC	12	24.63	427 255	Oct-12	30

Source: Semovi (2020); STC (2021);
Note: STC = Sistema de Transporte Colectivo; Length includes operational segments and excludes garage and service; Passengers is the average number of passengers in a working day (Mo-Fi) in the year 2019; Opening refers to the date operations first started (does not consider extension of a line); Fleet is the total number of trains in operation.

Table 4.6: Metro system in Greater Mexico City, 2019.

technology. The capacity of trains ranges from 1,020 to 1,530 spaces. The total length of the metro network is 218 kilometres (excluding garage tracks), where the longest line is Line 12 (Metro-L12, 25 Km). Metro-L12 is also the most recent service to open, in October 2012. Line 4 (Metro-L4) is the shortest, running along an 11-kilometre corridor. The total number of users a day of the metro system averaged five million passengers on a working day in 2019 (Semovi, 2020). The most in demand line is Line 2, which transported 810,228 passengers on an average working day. Meanwhile, Metro-L4 served about a tenth of these journeys only (80,000). The metro is an important transport system in GMC, enabling 13% of all journeys (INEGI, 2018).

Regional rail (RGR) or *commuter rail* is another form of *rapid transit*. This operates with electric trains in a fully controlled ROW (A) (Vuchic, 2007). Exceptionally, trains can run on diesel (e.g. in North American cities). RGR serves long journeys typically connecting core metropolitan areas to suburban town centres. Thus, the rail road tracks often form a radial-like network. The space between stations is long, ranging from 1.2 kilometres up to 4 kilometres. The frequency of the RGR service is lower and the operating speed is higher than metro systems. The trains in this mode can couple up to 10 cars, offering an individual capacity of between 150 to 1,800 spaces. The line capacity is 25,000 to 40,000 sps/h. The specialization of this infrastructure is costly, generating investment costs of 50 to 120 million US$ per kilometre of pair of lanes. At the present time, the *Tren Suburbano* (SUB, henceforth) is the only RGR line operating in GMC. The 27-kilometre line connects Ciudad de México from *Buenavista* station to *Cuahutitlán* in the state of México. This service opened in 2008 and it is operated through a concession scheme by the private company *Construcciones y Auxiliar de Ferrocarriles SA* (CAF) (CAF, n.d.). The three- to four-car electric trains offer an individual total capacity of 842 to 1,138 spaces. In the year 2017, the average number of users a day was 189,000 (INEGI, 2018). The SUB is frequently combined with other *rapid transit* modes. For instance, 43% and 21% of the journeys that use the SUB are combined with metro or BRT, respectively.

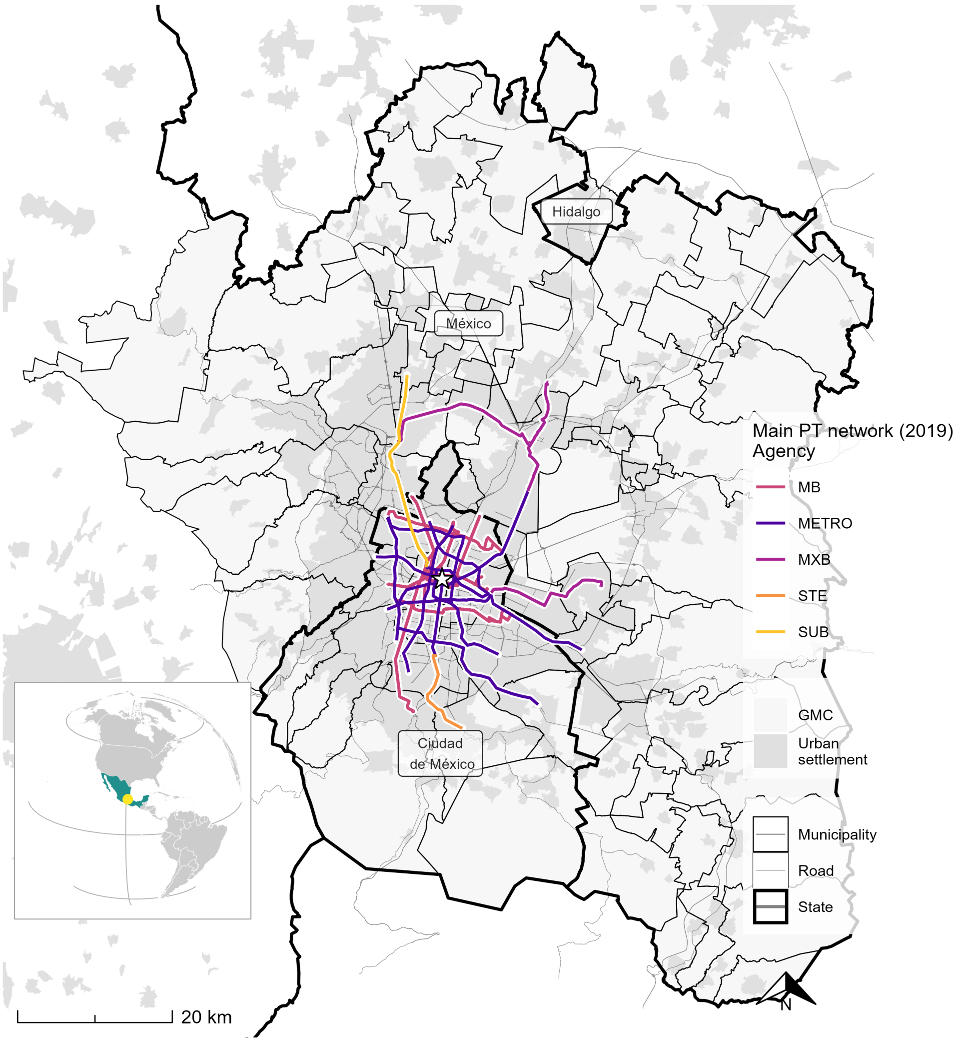
The *aerial tramway* is a specialized *rapid transit* mode consisting of cable-suspended cars which are pulled by a closed-loop cable (Vuchic, 2007). The operation is powered by electric motors. This form is useful for overcoming geographical and topological barriers such as rugged terrain or bodies of water. The offered capacity reported in the literature ranges from 1,500 sps/h in the Roosevelt Island-Manhattan Line (Ibid.) to 3,000 sps/h in the *Metrocable* in Caracas, Venezuela (Tischler & Mailer, 2019). Until the year 2019, *Mexicable* was the only *aerial tramway* operating in GMC.[8] This system is a 4.7-kilometre line located in the state of México (SITRAMyTEM, n.d.). This connects the low-income residential area (*Sierra de Guadalupe* in the municipality of *Ecatepec*) to the BRT MXB-L1 and other *street transit* services in GMC. The *Mexicable* started operations in 2016. The offered capacity and demand is lower than other *rapid transit* systems operating in GMC. For instance, *Mexicable* transported 7,400 passengers on average in a weekday in the year 2017. Therefore, its role at the metropolitan level is low.

For the purpose of the present research, the focus is limited to *rapid transit* and *semi-rapid transit* modes excluding the specialized (i.e. *aerial tramway*). This choice is based on the immovable nature of these modes, a key characteristic of infrastructure that allows the capitalization of their benefits on land (Gómez-Ibáñez & Liu, 2022). Hence, the modes which do not operate in a segregated ROW (i.e. *street transit* such as regular bus, trolleybus) are not included explicitly in the accessibility models. This decision is further motivated by the lack of detailed information on bus itineraries. Furthermore, the *street transit* modes (e.g. *combi*, *micro*, or regular buses) cover practically all the area of analysis and offer equivalent uniform level of service, as found in the 2017 Travel Survey (INEGI, 2018). Finally, including multiple accessibility variables generated by these modes (e.g. car, rapid and semi-rapid transit, and street transit) can generate possible collinearity issues, as suggested in previous studies (Dubé, Thériault, & Des, 2013)

The combination of public transport modes considered for this work (i.e. *rapid*

[8] Two additional lines opened in Ciudad de México in the year 2021. Yet, these are out of the temporal analytical scope of this work and are not discussed in the present research.

transit and *semi-rapid transit*) is referred to as the *main public transport network* (MPTN). Figure 4.3 illustrates the MPTN in GMC for the year 2019. This consists of a variety of modes operated by five main different agencies as following: The metro system, directly operated by the state-owned company STC and designated as METRO; two BRT systems, namely MB in Ciudad de México and MXB in the state of México; One LRT line operated by STE; and One RGR designated as *SUB*.



Source: The author based on Marco Geoestadístico Nacional 2014 Versión 6.2 (INEGI, 2014); Delimitación de las zonas metropolitanas México 2015 (INEGI, 2018); Secretaría de Movlidad de la Ciudad de México (SEMOVI, n.d.); Sistema de Transporte Masivo y Teleférico del Estado de México (SITRAMyTEM, n.d.); OpenStreetMap (OSM, n.d.).

Figure 4.3: Main public transport network (MPTN). Greater Mexico City, 2019.

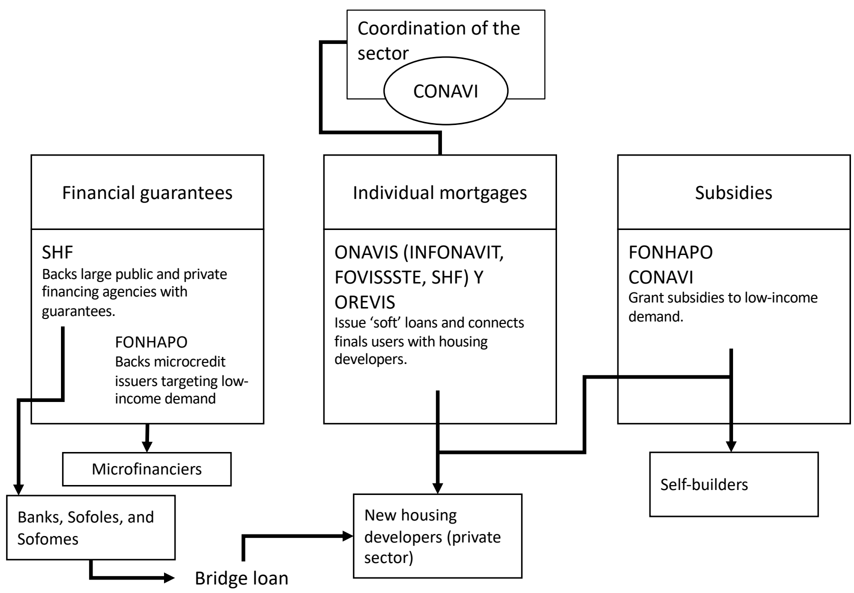
4.3 Housing in Mexico

4.3.1 The national housing policy

The foundations of the current institutional housing system (*sistema nacional de vivienda*, SIV) in Mexico were established during the 1970s (PUEC-UNAM, 2012). Some of the most relevant actions at that time were the creation of lending funds for both salaried and non-salaried employees. In this first stage (1972-1989) the state actively promoted the production of housing. In a later phase (1990-2000), this approach shifted to a liberal model where the promotion of housing was delegated to the private sector (Eibenschutz & Rodríguez, 2013; Medina Ciriaco & Kunz Bolaños, 2013; Monkkonen, 2011). Under this approach, the function of the public organisations are mostly as financial enablers. Currently, the SIV operates according to the liberal model, and it is chiefly delivered by the following organisations (Ziccardi & González Reynoso, 2015):

- The National Housing Commission (*Comisión Nacional de Vivienda*, CONAVI), which is a central organisation aimed at defining the housing policy, formulating and managing a sexennial National Housing Program, and coordinating the organisations in implementing the housing policy. It also grants housing subsidies and supports the low-income housing sector;
- National housing organisations (ONAVIS) which operate as financial entities for individual mortgage loans. The most important (according to the number of credits issued) are:
  - The Institute of the National Workers' Housing Fund (INFONAVIT, for its acronym in Spanish) addressing salaried employees in the private sector, and
  - the Housing Fund of Institute of Social Security of State Workers (FOVISSSTE) for the government's employees;
- Banking, Limited Financial Societies (Sofoles), and (Sofomes), which act as mediators to issue bridge loans for housing developers and also offer mortgage loans, and;
- Development banking which issues financial guarantees through the Federal Mortgage Society (SHF) to private banking, Sofoles and Sofomes, using federal resources.

Figure 4.3 illustrates the operational structure of the previous and other complementary organisations according to their main function in the SIV (PUEC-UNAM, 2012). Under this scheme, there are four main functions, namely: coordinating the sector, granting mortgage loans, subsidizing, and issuing guarantees. At the top, the CONAVI coordinates the national policy, having direct influence on ONAVIS and other state housing organisations (OREVIS) that finance individual mortgage loans. ONAVIS and OREVIS chiefly address the needs of salaried employees, while the National Fund for Popular Housing (FONHAPO) and CONAVI provide subsided loans targeting low-income sectors directly, for self-building housing solutions or through developers of new housing in the private sector. SHF issues financial guarantees to Sofoles and Sofomes which function as intermediates financing new-housing developers through bridge loans. Both Sofoles and Sofomes cover middle-income as well as low-income households. Meanwhile, FONHAPO grants guarantees backing microcredits targeted to low-income households who are not necessarily formally employed.



Source: The author based on PUEC-UNAM (2012).

Figure 4.4: Institutional housing system in Mexico.

Since the housing reforms in the 90s, the focus of the national housing policy has been on making available completed new dwellings.[9] Between 2001 and 2006, 60% of all the loans were issued for the acquisition of complete dwellings (followed by 30% for physical improvement, 5% for core housing units,[10] and 1% for financial improvement)

[9]This contrasts with other housing solutions offered in a previous stage of the housing policy. Until the housing reforms of the 90s, much of the efforts to provide housing consisted of incremental development schemes (without permits) (Monkkonen, 2011).

[10]Core housing units (*vivienda inicial* or *píe de casa* in Spanish) refers to and initial structure of a housing built of permanent materials usually consisting of one or two rooms which is intended to be the core unit of a progressively-built dwelling.

(PUEC-UNAM, 2015). From the loans issued for complete dwellings, 90% were for new houses. Yet, these trends have been shifting in more recent years towards favouring used housing. For instance, while complete new dwellings represented 94% of the number of loans in 2003, this figure decreased to 58% in 2010 (Ibid).

These growth patterns have been criticized for their predominant location on the urban periphery and their limited access to urban amenities (Monkkonen, 2008, 2011). This implies poor availability of the services provided by the MPTN (as the MPTN chiefly serves central areas in GMC, as shown in Figure 4.3), and the consequent reliance on *paratransit*- or *shuttle*-like public transport modes (*Combi* or *Microbus*) for many of the new housing developments. This is relevant for the present work since the accessibility generated by the MPTN becomes scarce in the housing market and could possibly have important impacts on the value of residential land.

4.3.2 The housing stock in Mexico and GMC

The housing sector was considerably active in Mexico during the time period studied. In the year 2010 the national housing stock consisted of 35.6 million dwellings increasing to 43.9 million in 2020 (SNIIV, n.d.). This change implies a net growth of 8.3 million units in a ten-year period. Accordingly, the average annual growth rate of the housing stock was 2.3%, whilst the population growth rate averaged 1.2% in the same period.[11] The rapid growth of the offering relative to the population is also reflected in a decreasing number of inhabitants per dwelling. For instance, this indicator was 3.9 in 2010, decreasing to 3.7 in 2015, and 3.5 in 2020 (INEGI, n.d.-b). Considering the overall housing stock, the proportion of occupancy was stable at both the start and end time point of the period studied. About 80% of the dwellings were inhabited, while 15% were not, and 5% were for temporary use in both 2010 and 2020 (SNIIV, n.d.).[12]

Table 4.7 presents the composition of the housing stock by state in GMC for the year 2020 (the end of the time-period studied). The inventory consists of 7 million dwellings which represents about 16% of the national volume. The stock in GMC comprises four million dwellings located in the state of México (57%), three million in Ciudad de México (43%) and only 87,000 in Hidalgo (>1%). At the metropolitan level, 88% of the units are inhabited. This proportion is above the national level, where only 80% of the houses are occupied. Yet, this figure varies by state. For instance, while 91% of the houses are inhabited in Ciudad de México and 86% in the state of México, only slightly more than the half (55%) are in Hidalgo. The low level of occupancy in this part of the metropolitan area can be partially explained by the

[11] The population annual growth rate figure is based on the 2010 and 2020 Population Census results from INEGI (n.d.-b).

[12] Based on the 2010 Population Census and the 2020 Population Census.

Region/State	Inhabited *Thousands (%)*	Non-Inhabited *Thousands (%)*	Temporary use *Thousands (%)*	Total dwellings *Thousands (%)*
GMC	6 231 (87.9%)	649 (9.2%)	209 (2.9%)	7 089 (100.0%)
Ciudad de México	2 756 (90.8%)	207 (6.8%)	72 (2.4%)	3 035 (100.0%)
Hidalgo	48 (54.5%)	30 (34.2%)	10 (11.3%)	87 (100.0%)
México	3 427 (86.4%)	412 (10.4%)	127 (3.2%)	3 966 (100.0%)

Source:
2020 Population Census (INEGI, n.d.)

Table 4.7: Housing stock by state. Greater Mexico City, 2020.

number of houses that are used as temporary. For instance, in Hidalgo about one in ten houses (11%) are for temporary use. This is well above the 3% at the metropolitan level or the 5% in the national stock. In GMC, around 9% of the dwellings are not inhabited, which is lower than the national rate (15%). In the metropolitan area, the lowest sub-utilization occurs in Ciudad de México. In this state only 7% of the houses are not inhabited. This figure is followed by the state of México, where one in ten houses is not inhabited. The figures in Hidalgo contrast with the metropolitan and national figures since about one third of the houses are not inhabited in this part of GMC.

The housing sector was also active in GMC during the period of study. According to the 2010 and 2020 population census, the housing inventory passed from 6.2 million units in 2010 to 7.1 in 2020 (INEGI, n.d.-b). This change implies the introduction of almost one million dwellings (923,570) in ten years. That is an average of about 90,000 units a year. Almost two thirds of this increase occurred in the state of México (64%), slightly less than a third in Ciudad de México (31%), and 4% in Hidalgo. The growth speed of both the population and housing stock was slower in GMC than it was at the national level. The observed growth in the housing stock in GMC represents an average annual rate of 1.5%, while this reached 2.3% nationally. Similarly, whilst the population in Mexico grew on average by 1.2% annually, it grew by 0.8% in GMC. These figures suggest a moderate growth pace in the study area. Yet, the zones defined by state jurisdiction within the metropolitan area show some dissimilarities when compared to each other. While the stock in Ciudad de México grew by 11% and 18% in the state of México, it almost doubled in Hidalgo (85%). In the latter, the inventory passed from 47,000 units in 2010 to 87,000 in 2020. These figures reflect the general spatial trends of the urban expansion in GMC, which has been most concentrated in the north of the metropolis (Flores, 2019).

Regarding the characteristics of the stock, it can be observed that the size of the dwellings in GMC increased during the period of reference. The proportion of dwellings with only one room decreased from 7% to 5% between 2010 and 2020, according to the population census in the respective years (INEGI, n.d.-b). In the meantime, the share of houses with two rooms remained roughly equal (15% in 2010 and 14% in 2020). By contrast, the ones with three or more rooms increased from 78% to 81% during the same period. This is also reflected by the fact that while the number of houses with only one room actually decreased by 6% between 2010 and 2020, dwellings with three or more increased by 24%. These changes are not only the result of the introduction of new houses to the stock but also due to the incremental building practices in the region (Monkkonen, 2011).

The 2020 National Household Income Survey (*Encuesta Nacional de Ingresos y Gastos de los Hogares 2020*, ENIGH) (INEGI, n.d.-b) includes information describing the households' dwelling, which is useful to characterise the housing stock in the area of study. As shown in Table 4.8, 75% of the inventory is made up of independent houses, 17% by flats; 7% by multi-family houses, and about 1% by roof-top dwellings or non-residential buildings. From these, most (64%) had not been used at the time they were acquired. Concerning the type of solution, one in four houses (40%) were already built and about one in three (30%) were instructed for their construction. Another solution in GMC is self-build dwellings, which represent about a quarter of the inventory. Regarding the means of financing, most of the purchased dwellings were accessed using the dwellers' own resources (66%) while almost a quarter (22%) were financed by one of the major national housing organisations (ONAVIS, i.e. INFONAVIT, FOVISSSTE, FONHAPO). Banks, Sofoles and mutual saving banks (microfinanciers) play a moderate role, financing 6% of the stock. Other institutions supported about 4% of the purchases. In the meantime, informal credits supported only 2% of the purchases. In terms of occupancy or ownership, most of the dwellers own the house (57%) while two in ten rent (20%). One other relevant form of occupancy is borrowing (16%). Another 6% of the financed units are owned under an active mortgage.

4.4 The politics of LVC in Mexico

The Constitution of Mexico (*Consitución Política de los Estados Unidos Mexicanos*, Constitution, henceforth) is the supreme law in Mexico.[13] The Constitution grants urban planning competencies to municipalities or exceptionally to the state government of Ciudad de México (Isunza-Vizuet, Castro, & Munévar, 2021). Also, it acknowledges the fiscal independence of municipalities (or Ciudad de México) including the charge

[13] Consulted on 17/09/2022, last reform on 28/05/2021. URL: https://www.diputados.gob.mx/LeyesBiblio/pdf/CPEUM.pdf.

Characteristic	Dwellings *Thousands*	Dwellings *%*
Type of dwelling		
Independent house	4 764.2	75.30
Flat	1 089.3	17.20
Multi-family house	410.9	6.50
Rooftop dwelling	23.4	0.40
Non-residential building	43	0.70
Bought used		
Yes	560.6	36.10
No	991.4	63.90
Type of solution		
Other	182.8	4.70
Bough as complete (built) dwelling	1 552.0	40.10
Contracted for the construction of the dwelling	1 142.7	29.60
Autoconstruted dwelling	988.2	25.60
Type of finiancing		
INFONAVIT, FOVISSSTE, or FONHAPO	817.4	22.40
Bank, Sofol, or mutual savings bank (microcreditor)	216.1	5.90
Other institution	131.3	3.60
Informal credit (family, friend, or lender)	76.4	2.10
Own resources	2 401.4	65.90
Type of occupancy/ownership		
Rents	1 267.0	20.70
Borrows	989.3	16.20
Owns (mortgaged)	349	5.70
Owns	3 516.6	57.40

Source: Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2020. (INEGI, n.d.)
Note: Estimates exclude missing values.

Table 4.8: Housing stock characteristics. Greater Mexico City, 2020.

of contributions over real property, i.e. fees or taxes (art. 115, art. 122). Furthermore, the Federal Tax Code (*Código Fiscal de la Federación*) explicitly recognizes the figure of *betterment contributions*.[14] The rationale of this instrument is *direct*, since it is intended to recover the cost of public works or services.[15]

The criteria for the implementation of the national planning and fiscal framework is set in the local regulation at the state and municipal level. For example, in Ciudad de México, betterment contributions are regulated by the Mexico City Tax Code (*Codigo Fiscal de la Ciudad de México*).[16] Specifically, the local act assigns a weight to distribute the cost of a public investment according to the type of public work and three catchment bands, namely: (A) front, (B) up to 250 m away (excluding fronts), and (C) between 251 m and 500 m. This local regulation explicitly acknowledges transport infrastructure as eligible for the implementation of betterment contributions, such as Metro, BRT, and trolley. Meanwhile, the state of México defines the criteria in its own Tax Code (*Código Financiero del Estado de México y Municipios*).[17] This legislation is broader than the former since it also considers public actions. However, it is not specific in terms of the spatial distribution of the contributions. This is to be determined case-by-case by the government branch implementing the instrument.

The legal framework of Mexico does not explicitly addresses (indirect) land value capture instruments (Isunza-Vizuet et al., 2021). Still, there are at least seven identified mechanisms that can recover land value uplifts (PUEC-UNAM, 2012), namely: (1) property (or council) tax applied when owning a property; (2) Stamp duty, levied when acquiring real property (frequently known as ISAI in Mexico for its acronym in Spanish, *Impuesto Sobre Adquisición de Inmuebles*); (3) *Impuesto sobre la renta* (ISR) applied over the earnings derived from selling or leasing property; (4) betterment contributions (*contribuciones de las mejoras* or *aportaciones de las mejoras*), derived from the benefits of public works or actions; (5) exchange of development rights for land or development of infrastructure; (6) transfer development rights (*transferencia de potencial*); and (7) special development zones (*polígonos de actuación*).

Even though the previous mechanisms are present at different levels of the regulatory framework, they have been underutilized (De Cesare, 2016; Isunza-Vizuet et al., 2021). For example, in 2020 the average municipal income represented 0.2% for betterment contributions, 0.3% for construction permits, and 0.2% for the grant of

[14]Consulted on 17/09/2022, last reform on 12/11/2021. URL: `https://www.diputados.gob.mx/LeyesBiblio/pdf/CFF.pdf`.

[15]The base is not the land value increment but the cost of the public infrastructure, as discussed in Chapter 2. Hence, betterment contributions can be considered a special case of betterment charges.

[16]Consulted on 17/09/2022, last reform on 30/12/2021. URL: `https://data.consejeria.cdmx.gob.mx/images/leyes/codigos/CODIGO_FISCAL_DE_LA_CDMX_4.pdf`.

[17]Consulted on 14/09/2022, last reform 25/05/2022 `https://legislacion.edomex.gob.mx/sites/legislacion.edomex.gob.mx/files/files/pdf/cod/vig/codvig007.pdf`.

development rights (INEGI, n.d.-a). The figures are similar for Ciudad de México, except for betterment contributions which amount to zero. These have not been implemented since the mid- 1980s (Pérez Torres, 2021). The limited experience has been attributed to the fragmentation of the legal framework, discretional application, and high dependence of central transfers (Isunza-Vizuet et al., 2021).

In addition to the instruments discussed above (fee-based on its majority), there are additional land-based instruments both in use and with potential to expand the use of LVC in GMC. These include land sales or leases (based on land banking) and joint developments. An example is given by the multimodal transport hubs (*Centros de Transferencia Modal*, CETRAM, for its acronym in Spanish), which consist of an area where a variety of public transport modes converge typically around a metro station, e.g. street transit, semi-rapid transit, or taxi (Seduvi, n.d.). The land of CETRAMs is usually owned by the state government. Since at least 2013, the state government of Ciudad de México has outlined a programme aimed at improving CETRAMs facilities in exchange for a long-term leasing in the same location (implemented through 30-year concessions) (ibid). Under this scheme, a private investor redevelops the area by improving the transport infrastructure and building commercial space to recover the investment. Unfortunately, this programme is insufficiently documented both in academic literature and official sources. Thus, it is difficult to know its performance and achievements. Still, the conceptual mechanism represents a potential instrument to recover land value increments from public land to improve public transport infrastructure.

4.5 Summary and final remarks

GMC plays an important role in the region and at the national level both demographically and economically. The metropolis consists of 75 municipalities of three states, including the national capital (Ciudad de México). This generates important mobility dynamics, in which 35 million journeys are completed on a typical business day. The public transport system plays a key role, enabling half of these journeys. Based on the classification of the public transport system in GMC included in the city profile, this chapter defines what is referred to in this research as the *main public transport network* (MPTN). For the purpose of the present study, the MPTN includes semi-rapid transit and rapid-transit modes only. This is motivated by their characteristics of immobility (Gómez-Ibáñez & Liu, 2022). Specifically, the MPTN is comprised by the metro system (METRO), two bus rapid transit (BRT) systems (Metrobús, MB, and Mexibús, MB), one light rail transit line (LRT-L1), and one suburban/regional rail (RGR, denominated SUB-L1).

Meanwhile, the institutional housing system in Mexico operates mostly under a top-down approach through a number of consolidated public and private entities. This system smooths potential frictions of the housing market, for example, by enabling access to loans to a different types of residents or reducing transaction costs. The housing stock passed from 6.2 million units in 2010 to 7.1 in 2020. This constitutes a market with a broad range of alternatives for households.

The city profile shows relevant characteristics of the housing stock. For example, from the total units, only 36% are bought used, most of them are semi-detached, and are occupied by owners. Regarding the politics of LVC, it is found that there is not an explicit legal framework in Mexico. However, there is an administrative and legal foundation enabling a number of instruments with potential for recovering land value increments. The combination of the active housing market and important role of the MPTN in the transport system of GMC represent fertile conditions for the implementation of LVC, according to the discussion in Chapter 2. Still, the local policy can benefit from the study of the distribution of the economic benefits generated by such dynamics. The next chapter outlines the methodology adopted for the empirical study of the capitalization of the MPTN onto residential land value in GMC.

Chapter 5

Methodology

The previous chapters have outlined the theoretical framework, the various aspects related to the empirical evaluation of public transport benefits and the actual area of study. A central component of land value capture (LVC) is concerned with the distribution of land value increments triggered by public actions, such as the improvement of accessibility through public transport infrastructure. This chapter sets out the methodology and the main methods used in this research to examine the possible capitalizations of public transport infrastructure onto residential land value.

The major contribution of this chapter is the establishment of the methodology adopted and the selection of the formal methods used to address the specific Research Questions drawing on the Open Science framework. This is guided by the characteristics of the area of study detailed in Chapter 4, the data, and the resources available to conduct this research. A fundamental finding suggests that the research strategy can be substantially enhanced by the adoption of transparency and reproducibility principles included in the definition of the Open Science concept. Many of these principles can be implemented by considering the characteristics of this research. Another important finding indicates that although the equilibrium sorting model represents enormous potential for accomplishing the objectives proposed, there are limitations for its use in this dissertation. The hedonic model has been chosen as the main approach for this research given the ease of the data required, consolidated techniques to estimate it, and its flexibility. The principles of these approaches were discussed in Chapter 3.

The following criteria have been adopted to organise the contents of this chapter with the aim of simplifying the discussion. If a specific method is related to more than one of the Research Questions, it is detailed in the present chapter. Otherwise, it is examined to detail in one of the subsequent chapters, i.e. next to the respective empirical results.

This chapter is structured as follows. Section 5.1 presents the research strategy which guides this study and sets out the transparency and reproducibility considerations adopted. Section 5.2 details the spatial analytical scheme, the sources of information and those adopted for the empirical analyses. Section 5.3 provides a detailed description of the source of data used to account for residential property values and the sample selection process. Section 5.4 details the main statistical modelling framework employed. Section 5.5 outlines the variables and the respective measures used to characterize the neighbourhood and locational characteristics employed in the present study. Section 5.6 summarises this chapter.

5.1 Research strategy

The specific research strategy followed to answer the Research Questions adheres to these broad phases:

1. Assembling representative samples accounting for the housing market in Greater Mexico City between 2010 and 2019. For this purpose, the information accessed consists of property valuation records compiled by the Federal Mortgage Society (*Sociedad Hipotecaria Federal*, SHF, for its acronym in Spanish). These records are also used to compute the national housing price index in Mexico.
2. Assessing the level of service provided by the main public transport network (MPTN) considering land use. This is operationalized by drawing on comprehensive accessibility indicators (Geurs & van Wee, 2004), i.e. location-based measures. This step is further disaggregated in the following broad components:
    - Definition of a spatial analytical scheme (SAS) representing *origin* of accessibility measures. The first SAS is constrained by the geographic details of the housing market data accessed from SHF, as the lowest spatial reference is the post code level. Still, a subset of the SHF property data is referenced at the parcel or property level. This level of spatial disaggregation at the point level allows flexibility in the definition of the SAS. Taking advantage of this characteristic, various uniform grids consisting of hexagonal cells of various sizes are defined following previous research (Pereira, 2018; Pereira et al., 2019; Wong, Lasus, & Falk, 1999).
    - Land use and individual components, which represent *destinations* as perceived by residents at origins, namely: (1) location and characteristics of employment opportunities, which are obtained from the 2014 Economic Census (*Censos Economicos 2014*) (INEGI, 2015a); and (2) location and characteristics of demand, including demographic aspects at specific origins, obtained from the 2010 Population Census (*Censo de Población y vivienda 2010*) (INEGI, 2012).

- The transport component, which consists of the *spatial links* between origin j and destinations k and represent impedance in accessibility-based measures. This is considered as the modelled travel time and should reflect households' commuting costs better than distance, following Ma & Banister (2006). Travel time is modelled for all possible combinations between jk using OpenTripPlanner (OTP),[1], a multimodal open-source routing software platform. This produces a series of travel time matrices (TTM) using information about the road and pedestrian network obtained from OpenStreetMap (OSM) and public transport timetables organized in general transit feed specification (GTFS).

3. Analysing the property data and the key independent variables (i.e. accessibility to employment) to answer Research Question 2 (RQ2) from a quantitative descriptive perspective and in multi-variate regression framework drawing on the hedonic property value model. This approach is observational with a cross-sectional design. The analyses are implemented through a variety of statistical tools which range from descriptive statistics, or ordinary least square methods, to those accounting for spatial effects in a hierarchical or multilevel structure.

5.1.1 Transparency and reproducibility

Transparency and reproducibility are key principles in scientific research (Christensen & Miguel, 2018; Merton, 1979). These values have been recognized as essential for the following aspects: enhancing efficiency, easing self-correction and strengthening credibility (Hardwicke et al., 2020). Although these principles have long been acknowledged, there is growing awareness of the prevalent deviance from these ideals in, for example, social sciences (Hardwicke et al., 2020), economics (Christensen & Miguel, 2018) and geographic information science (Nüst et al., 2018).

Advances in technology mean a greater range of tools and platforms which facilitate this process. These include: the advent of robust open-source and collaborative software (e.g. `R` programming language (R Core Team, 2021), R Studio integrated development environment (IDE) for R,[2] `Python`), and the wider software/coding user community platforms (e.g. Stack Overflow,[3] R Studio Community,[4] GitHub[5]), and open access internet hosting platforms which offer version-control systems (e.g. GitHub or Bitbucket[6]), to name a few of the relevant resources available.

[1] https://www.opentripplanner.org/
[2] https://www.rstudio.com/
[3] https://stackoverflow.com/
[4] https://community.rstudio.com/
[5] https://github.com/
[6] https://bitbucket.org/

Open Science is a phenomenon that addresses and, in some cases, expands the key classic principles for transparency and reproducibility in research. This is a type of knowledge which is distinguished because it is transparent, accessible, shared, and collaborative-developed (Vicente-Saez & Martinez-Fuentes, 2018). The formal definition of this concept based on a systematic review is as follows;

> *"Open Science is a transparent and accessible knowledge that is shared and developed through collaborative networks"* (Vicente-Saez & Martinez-Fuentes, 2018, p. p. 234).

Drawing on the fundamentals of Open Science, the considerations adopted for this work are as follows;

- *Open-source software*. All the data processing and analytical software involved in the present work is open-source. Although there is a variety of software and programming languages involved in these stages, all analyses and processes are implemented from a single platform (i.e. R Studio IDE) and executed through the same programming language (i.e. `R`). This permits the arrangement of the materials in an uninterrupted and sequential order which facilitates the examination or reproduction processes followed. In addition, the communication of results uses open-source software, i.e. R Studio IDE, Knitr (Xie, 2021b), Bookdown (Xie, 2021a), TinyTex (Xie, 2021c) (a light-weigh version of LaTex used for document preparation), and Zotero (for managing biographical references).
- Input or *raw data* will be made available where the licencing permits it. Alternatively, the full reference of external sources is provided.
- The processing and analytical *scripts* will be made open access in a series of GitHub repositories once the main document is published. The list of the permanent links for the repositories can be found in Appendix A. These provide step-by-step details of all the processes and analyses performed. This supplements the verbal descriptions of the processes and analytical processes followed enhancing reproducibility.
- Output files produced are *open format* (also known as free file formats) which do not require commercial software to be consulted, e.g. plain-text files in comma separated values format (`.CSV`), or GeoPackage Encodign Standard (`.GPKG`).
- The final document will be *publicly available* in the Glasgow University Enlighten repository.[7]

Although the restrictions of distributing some of the raw data employed for the analyses conducted in this work are challenging in their ability to verify the results by

[7] `https://theses.gla.ac.uk/`.

pure replication (i.e. using the same specification, identical sample, and the same population), the considerations taken facilitate the reproduction of results (i.e. using the same specification and consistent population, but a different sample) (Clemens, 2015). All-in-all, the actions implemented here towards Open Science add to the important aspects related to efficiency, self-correction, and credibility.

5.2 Spatial analytical framework

This section details the hierarchical spatial framework used by the Office for National Statistics in Mexico (INEGI). This is the main frame of reference by which official data are accessed. In addition, it details the definition of the spatial units of analysis used for the purpose of estimating location-based measures and conducting sensitivity analyses of willingness to pay.

5.2.1 Spatial units of aggregation of the sources

Urban settlements cover approximately half of Greater Mexico City's (GMC) territory (50%, 3,938.3 km^2) (INEGI, n.d.-b).[8] The National Cartographic Framework (MGN, for its acronym in Spanish *Marco Geoestadístico Nacional*) produced by the Office for National Statistics (*Instituto Nacional de Estadística y Geografía*, INEGI) consists of various spatial schemes that are used to aggregate information at various levels, e.g. states, municipalities and census tracts. The smallest geographic definition in the MGN is the city block (block) and the rural settlement. While small rural settlements with populations less than 2,500 inhabitants are represented by spatial points, the blocks are represented by polygons. The boundaries of blocks are delimited by building fronts or roads, and occasionally by physical barriers in the urban context (e.g. rivers or the topography). The MGN is continuously updated by the INEGI. For this work, the information used is aggregated at the block level from two sources which draw on different MGN versions, namely: (1) the 2010 Population Census, and (2) the 2014 Economic Census. The details of these spatial units are provided below.

Table 5.1 presents a summary of the characteristics of two collections of city blocks used in the present work. The 2010 Population Census (INEGI, n.d.-b) identified 158,909 blocks in GMC. These contain 97% of the population (19.6 million inhabitants). The rest (543,616, 3%) is included in rural settlements/villages which are not considered in the analyses conducted here. This is because of their low relevance to the objectives proposed given their non-urban character and their small size. The average surface area of city blocks is 11,762 m^2. These units contain approximately

[8] Estimate based on the *Marco Geoestadístico Nacional 2015* Versión 6.2 (INEGI, n.d.-b)

	City block framework (Year)	
	2010	2015
City blocks (N)	158 909	162 519
Pupulation covered (%)	97.30	-
Area mean (sqm.)	11 761.95	11 609.22
Area SD (sqm.)	45 264.29	44 611.31
Population mean (N)	123.17	-
Population SD (N)	156.86	-
Dwellings mean (N)	37.70	-
Dwellings SD (N)	49.34	-

Source:
Censo de Población y Vivienda 2010 (INEGI, n.d.)
Sistema para la Consulta de Información Censal 2010 (INEGI, n.d.)
Marco Geoestadístico Nacional 2015 Versión 6.2 (INEGI, n.d.)

Table 5.1: City blocks summary. Greater Mexico City.

123 inhabitants and 38 dwellings on average. The city block framework used to aggregate the data for the 2014 Economic Census (published in 2015) includes 162,519 units (3,610 more than in 2010). The size of these is slightly smaller than in the 2010 MGN, averaging 11,609 m^2. The population and housing data are not available for this version as this information is collected at the block level only in the decennial census.

5.2.2 Proposed spatial analytical scheme

Various spatial analytical schemes have been considered for the purpose of estimating accessibility and conducting sensitivity analyses of willingness to pay as follows; According to the MAUP literature discussed in Chapter 3, an experimental design similar to the approach employed in recent empirical research focusing on accessibility is proposed (Pereira, 2018; Pereira et al., 2019). This choice responds to the following criteria:

1. to maintain consistency in this research area;
2. to allow comparability, and;
3. to augment previous findings from the perspective of the assessment of economic impacts of public transport infrastructure.

The details of the procedure adopted here are described below.

The approach adopted in this work to examine the effects of spatial scale referred to in the MAUP literature draws on the method proposed in Wong et al. (1999). First, Greater Mexico City, S, is delimited by its official 2015 metropolitan definition

(Sedatu et al., 2018), as detailed in Chapter 4. Then, S is systematically partitioned in uniform cells, j, of various gradual sizes. This process results in multiple equal-area grid schemes of various scales. Therefore, this study employs hexagonal cells instead of rectangular ones (as originally proposed by Wong et al. (1999)) because posterior literature shows advantages related to modelling applications and visual aspects (Birch, Oom, & Beecham, 2007; Carr, Olsen, & White, 1992). For instance, these include the simplified and 'less ambiguous' definition of the nearest neighbour, and better representations describing paths of movement through the grid. With regard to the size of the grids, the analyses adopted the criteria employed in recent accessibility studies using hexagonal cells of 0.5 km, 1 km, 2 km, and 4 km (Pereira, 2018; Pereira et al., 2019).

The spatial grids were created using the `SP` software (Pebesma & Bivand, 2021) for `R` (R Core Team, 2021) based on the source code used in previous empirical research (Pereira, 2018; Pereira et al., 2019).[9] The cells that exceeded the boundaries of GMC were trimmed according to the boundary of the metropolitan area. Only the polygons larger than 0.05 km^2 were included, following previously used criteria.

With regard to the zoning effects, a further system defined by the official post code zones was considered. This is important because property value records tend to be geo-referenced at the post code level only (or other similar administrative spatial definition). Consequently, empirical analyses are limited in their choice to a single spatial analytical scheme. The polygons used to represent post code zones are those published by the national Post Office (*Correos de México*). The spatial data were manually downloaded from the open-data platform of the Government of Mexico (https://datos.gob.mx/) in October 2019 (corresponds to the *version 5*).

Based on these considerations, five spatial analytical schemes (SAS) have been established, namely: four uniform hexagonal grids of 0.5 km, 1 km, 2 km, and 4 km and one given by post code zones. Table 5.2 shows the main characteristics of each of the proposed SASs. The total number of cells ranges from 36,600 in the highest resolution grid (0.5 km) to 637 in the lowest (4 km). The post codes total 2,567 and cover approximately 82% of the territory. The latter is a condition that is imposed by the source of information. The hexagonal grids contain approximately 97.3% of the total population in GMC and 100% of the population included in the 2010 Population Census blocks. Although post codes are intended to cover all settlements, in practice there are delays in the assignation process. This results in the lack of full coverage of all the population. However, it can be argued that this lag is irrelevant to the area of study. For instance, post code areas include 97% of the total population in GMC and

[9] Source code available in the following GitHub repository: https://github.com/rafapereirabr/thesis.

	Grid 0.5 km	Grid 1 km	Grid 2 km	Grid 4 km	Post code
All zones					
Count	36 600	9 340	2 418	637	2 567
Area covered (%)	99.89	99.97	99.98	99.96	81.95
Poulation covered (%)	97.30	97.30	97.30	97.30	96.84
Population in blocks (%)	100.00	100.00	100.00	100.00	99.52
Selected zones					
Count	13 297	3 972	1 264	413	2 377
Area mean (sq. km)	0.22	0.86	3.44	13.55	1.94
Population mean (N)	1 498.13	5 093.21	16 324.62	50 187.76	8 317.75
Dwellings mean (N)	458.51	1 558.79	4 996.18	15 360.05	2 544.93

Source:
Censo de Población y Vivienda 2010 (INEGI, n.d.)
Sistema para la Consulta de Información Censal 2010 (INEGI, n.d.)
Mapa de Ubicación de Códigos Postales (V5). Correos de México, 2019

Table 5.2: Proposed spatial analitycal schemes.

99.5% of the population included in the blocks.

For the purposes of this thesis, only SAS units that contain at least one inhabitant (year of reference 2010) or one economic unit (year of reference 2014) have been included. The second section of Table 5.2 (*Selected*) provides a summary of the subset of the SAS that satisfies this criteria. The number of units in the highest spatial resolution grid (0.5 km) is reduced to 13,297 (approximately by one third) and the number in the lowest is 413. The number of post code zones including some population or economic activity total 2,377. The mean surface area of the smallest grid (0.5 km) is 0.2 km^2 and this increases to 0.8 km^2 in the 1 km SAS, 3.4 km^2 in the 2 km grid, and 13.5 km^2 in the 4 km grid. The area of the post code zones is between the 1 km and the 2 km grid, averaging 1.9 km^2. The population included in a 0.5 km grid cell totals approximately 1,500 inhabitants and 460 dwellings on average, whilst the 4 km grid contains as many as 50,000 individuals and more than 15,000 dwellings and the mean figures for post code zones are 8,300 inhabitants and approximately 2,500 dwellings.

These SASs serve both as a framework to spatially aggregate information and as representations of locations of origin of residents in the accessibility models to be developed.

5.3 Property data: Federal Mortgage Society (SHF)

This section discusses the data used to represent property values in the area covered by this study, including the source of information, sample selection and geocoding process.

5.3.1 Source and structure of data

The SHF is part of the National Housing Organisations (ONAVIS) detailed earlier in the Study area section. This is a public organisation which has operated at the national level in Mexico since 2001. It promotes the development and provision of housing through guarantees or other financial instruments (SHF, 2021). Although the finance products are issued through third parties (i.e. public or private banks), SHF requires the assessment of the property value by an accredited surveyor as part of its process. The property valuation data collected are also used by the organisation to construct the national housing price index.

In 2008 the SHF established a digital platform to track the valuation records at the national level. This centralizes the information in a standardized database server called *Sistema Maestro de Avaluos* (SMA) (SHF, 2015) and this is fed through a web interface by accredited surveyors. The information was accessed from the SMA through a series of Freedom of Information requests sent between October 2019 and July 2020.[10] The information was received at different time-points between December 2019 and September 2020 .

The SHF data are stored in two databases in the SMA: the first covers the period between 2008 and early 2015 (**SHIF-DB1**) and it is currently inactive for administrative purposes; the second started operating in March 2015 and remains in use up to date (**SHIF-DB2**). The contents and criterion employed for both databases are consistent in many of the key variables and this allowed the combination of the information to cover a longer and uninterrupted period.

One of the core differences between the SMA databases is the depth of information accessed. The SHIF-DB1 is more detailed than the SHIF-DB2 in terms of internal and administrative information. For instance, the former includes location references at the property level, while the latter is limited in terms of administrative information (e.g. type of tenancy of the property or administrative unique identifier), some internal characteristics, and the location references which are limited to the post code level and name of the settlement.

5.3.2 Sample selection

The general sample selection process is illustrated in Figure 5.1. This includes the following broad considerations: geographic and temporal coverage; minimum surface areas; minimum value per square metre; age and other secondary variables; as well as a geocoding process. From the general selection procedure, two samples have been

[10] A copy of the original requests is available in the corresponding GitHub repository listed in Appendix A.

derived, namely: *Sample 1*, which includes information from January 2009 to February 2014 and comes from the SHIF-DB1 only; and Sample 2, which combines information from SHIF-DB1 and SHIF-DB2, and covers a longer period of time, i.e. from 2009 to 2019. Sample 1 covers a shorter period of time but the geographic information is rich, while Sample 2 covers a longer period of time but the geographic information is limited to the post code. The steps and details of this process are detailed below.

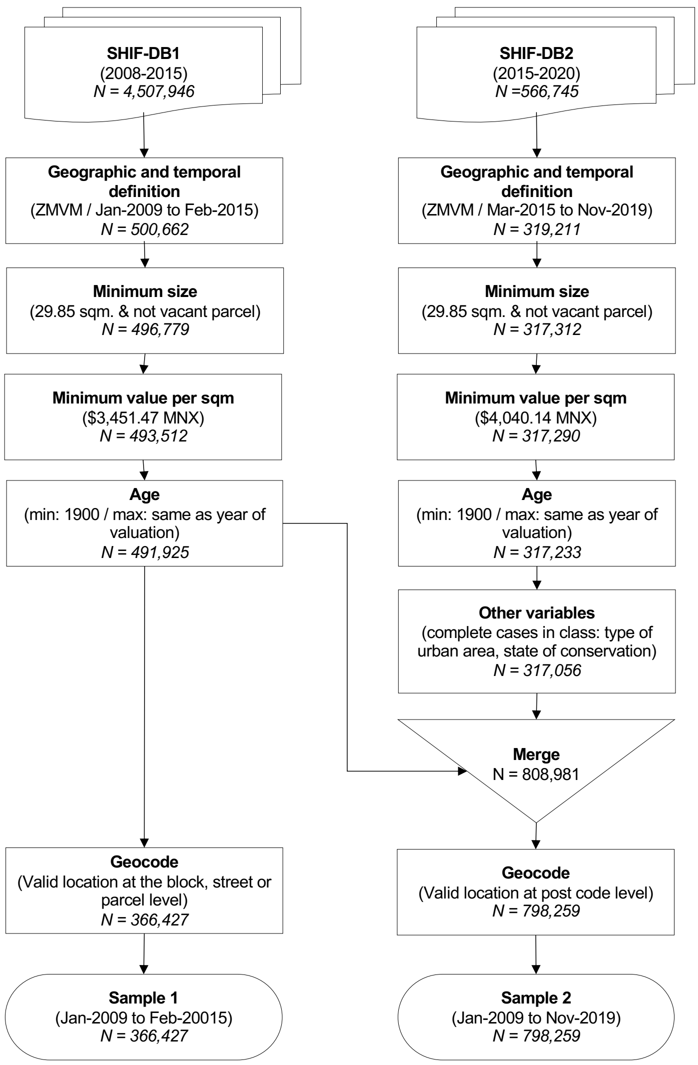


Figure 5.1: SHF Sample Selection Process.

Geographic and temporal coverage

The geographic definition includes all records of the seventy-six municipalities that form the metropolitan area. Any property reported to be located out of this region has been excluded.

The temporal definition of the SHIF-DB1 considers cases that were valued as early as 1st January 2009. This was because the public transport data are available only from the end of 2008. The latest date is February 2015, which is when the SHIF-DB1 stopped receiving information.

The temporal delimitation of the records included in the SHIF-DB2 is based on the date of registration and not the date of valuation, as in the SHIF-DB2. This is because such information was not available for most of the records in this database. The period selected covers from March 2015 until November 2019. The latest date was selected for two reasons: 1) the records from this date and onwards did not include key information (e.g. type of property), and; 2) the atypical circumstances of the year 2020 as a result of the COVID-19 global crisis.

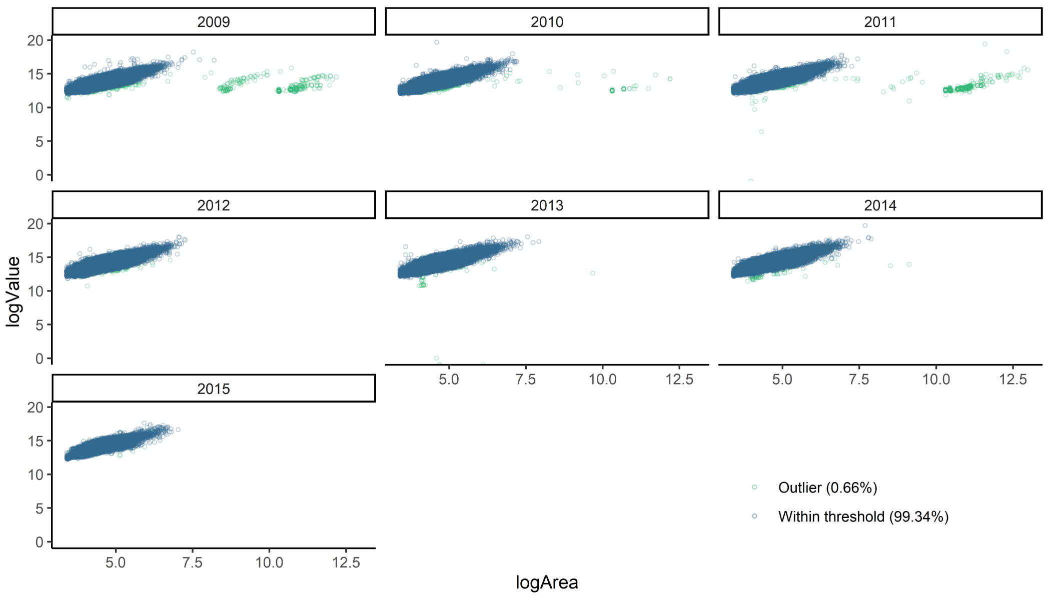
The date of valuation for the records in SHIF-DB2 was assumed to be thirty natural days prior to the date of registration if this information was missing. This assumption is based on the median lag between the date of valuation and the date of registration in the SHIF-DB1.

Surface areas

The databases include four types of surface areas associated with the property expressed in square metres. The first refers to accessory areas (e.g. balconies), and the lowest cut of this was 0 (negative values were assumed to be typing errors). The second is the proportional parcel area which considers values larger than 0, and the third and fourth refer to the built-up area and the saleable area, respectively. These were limited to values equal or larger than 29.85. This is assumed as the minimum dimension of housing unit based on the recommendations of the National Housing Commission (*Comisión Nacional de Vivienda*, CONAVI) (CONAVI, 2017, p. 103). This figure includes the minimum size for the following areas: 1) a superimposed room for living-dining-kitchen; 2) service; 3) bathroom; 4) one bedroom with closet, and 5) an additional 5% for the surface corresponding to walls. Vacant parcel records were not excluded.

Value per square meter

The data included some extremely low values per square metre. This is believed to be as a result of word processing errors when introducing decimal digits. This can be visualized in Figure 5.2, where it can be seen that in some cases (especially from 2009 to 2011) the total value of the property appears to correspond to the main cloud of points (Y-axis), while size is substantially distant from the vast majority of observations (X-axis). This is particularly noticeable for the years 2009, 2010 and 2011.



Source: the Author based on SHF-DB1.

Figure 5.2: Size and value, 2009-2015.

These outliers were excluded by defining a minimum value per square metre. This threshold was established based on the construction cost of a ‘low-cost’ dwelling and the figure is taken from Rocha Chiu & Gama (2007). This value was updated according to the National Consumer Index for Housing (INEGI, 2018) from February 2001 to December 2007 and January 2015, for SHIF-DB1 and SHIF-DB2, respectively.

Age and other variables

A valid age was determined by the year of construction completion. This can be between 1900 and the year of valuation. In addition, the records that did not contain the information of its *class*, *type of location*, or *state of conservation* were removed.

5.3.3 Geocoding process

The key information used to geocode a record was the cadastral code or the post code when the former was not available. The cadastrial code and the address including street and number were available for observations in the SHIF-DB1 only. It is worth noting that the administration of cadaster offices in Mexico corresponds to states and, therefore, the availability and level of detail of this information are heterogeneous. The location of this information was possible by searching through the records in the states of Mexico City and Mexico, but not for Hidalgo (which represents a small fraction of the data). Therefore, this location is excluded from Sample 1.

Geolocation process - Sample 1

The state of Ciudad de México publishes information related to the land use for each parcel. This includes the cadaster code at eight numeric digits, geographic location (a centroid of the plot), name of the street, number of the building or parcel, name of the settlement, post code and municipality. The land use data are published by the Department of Urban and Housing Development of Mexico City (*Secretaría de Desarrollo Urbano y Vivienda of Mexico City*, Seduvi). The data consulted were last updated on the 14th February 2019 and downloaded on the 19th October 2019 and accessed through the open access data platform (https://datos.cdmx.gob.mx/).

The land use information was matched to the SHF records using the cadaster code at the parcel level. There was a second matching round at the city block level using the first six digits of the code. The matches were validated if there was a coincidental common link between the source information (obtained from SHF) and the linked cadaster data in at least one of the following fields: the street number of the building, the full name of the street, one pattern of the street name, the name of the settlement, or the post code.

The state of Mexico publishes the cadastrial data only at the block level using an eight-digit code. This information was published in 2018 by the Institute of Statistics of the State of Mexico (*Instituto de Información e Investigación Geográfica, Estadística y Catastral del Estado de México IGECEM, Secretaría de Finanzas del Estado de Mexico*) and it was manually accessed on the 10th February 2021 through the following web address: http://igecem.edomex.gob.mx/productos-servicios/servicios-catastrales/cartografia-por-manzana-2018. These data are limited to the code and the geographically referenced city block geometry. A geometric centroid was computed for each of the blocks (or set of blocks if it was a multipolygon geometry) to represent the location of a record.

The state of Mexico's cadaster data were linked to the SHF records using the cadastrial code. In the first matching round, a substantial proportion of records in one municipality (*Zumpango*) could not be matched. In this case, records were matched using the street name available in the property records and the name of the street in the road network. In order to validate the matches for the state of Mexico, it was necessary to spatially join (1) the name of the streets in the road network issued by the same source (using a 100 m buffer); (2) the post code geometries, and; (3) the settlement name obtained from the locations of the 2015 National Directory of Economic Units (DENUE) (INEGI, 2015c). A match was considered valid if there was a coincidental common link between the information of the source (i.e. SHF records) and the spatially joined data in at least one of the following fields: the name of the street, a pattern in the name of the street, name of the settlement or post code. Locations using the name

	Valuation records	Geocoded	Percent
Region/State	*N*	*N*	%
GMC	491 925	366 427	74.49
Ciudad de México	169 638	160 483	94.60
Hidalgo	21 223	0	0.00
México	301 064	205 944	68.41

Table 5.3: Sample 1 geocoding summary

of the street were considered as valid.

Table 5.3 shows a summary of the geocoding process by state. The validated location of the records in Mexico City is very high (about 95%), whereas the validated locations in the state of Mexico at the block level is moderate (close to 70%). As seen, Sample 1 does not include observations in Hidalgo. The final size of Sample 1 is 366 427.

Post code matching

As shown in Figure 5.1, Sample 2 was constructed by combining records from the SHIF-DB1 and the SHIF-DB2 and the latter includes location references at the post code level or settlement name only. The process to assign a post code zone to the records in Sample 2 included the following steps:

1. The observations geocoded at a lower spatial reference (in Sample 1) were spatially joined to the post code geometries;
2. The rest of the records were matched to a location using the post code key, and;
3. If there were no coincidental common links found in the previous step , the name of the settlement was used to assign a post code using the matches in the previous steps by filtering the names in the corresponding municipality and assigning the the closest string computing the *Jaro-Winkler* distance (M. P. J. van der Loo, 2014) using the `stringdist` software (M. van der Loo, 2021) for `R`.

The post code matches are considered valid if: (1) the record was geocoded at a lower level (i.e. street, block or parcel), (2) there was at least one coincidental common link in the settlement name, one pattern of the settlement name, or the municipality was the same. Table 5.4 shows that approximately 50% of the observations in Sample 2 are referenced at the post code level; the rest is at a fine-grain level, namely street, block or parcel, and only a small proportion (1%) could not be assigned to a valid location. The final size of Sample 2 is 798 259.

Match level	Records (N)	Percent
Block	203 351	25.13
Parcel	157 770	19.50
Post code	432 915	53.51
Street	4 223	0.52
NA	10 818	1.34

Table 5.4: Sample 2 geocoding summary

5.4 Modelling framework

Chapter 3 provided a discussion about the main approaches used to assess the value residents put on public or environmental goods, such as the service provided by public transport infrastructure. The review identified that the hedonic approach is a widespread method. In addition, it was shown that the equilibrium sorting model (ESM) represents an alternative approach with tremendous potential. The ESM explicitly addresses many of the shortcomings of the hedonic models, such as heterogeneous preferences and non-marginal changes.

The number of applications of the ESM has remained limited since its formalization (Bayer et al., 2004; Sieg et al., 2004). Specifically, the review of the empirical studies revealed that this method has not yet been implemented for the assessment of the net benefits of public transport infrastructure. While this represents an enormous avenue for advancing knowledge in this area, it could also be reflecting some of the challenges associated with this approach. First, the data requirements are substantial. These frequently require socio-economic information at the person or individual household level for different time periods (Pryce et al., 2021). Ideally, this should be coupled with specific information about the characteristics of the dwelling. For this study area, the socio-demographic information from the 2010 Census is not openly available at this level. This is restricted to aggregated figures at the block or census tract. Even when individual-level information can be accessed for the study area, this source does not include information regarding the value of dwellings. Therefore, this should be supplemented by a third source by means of, for example, a representative value or a matching probabilistic approach.

A further aspect concerning data is that individual-level information is usually not disclosed for small spatial references because of personal data protection. Thus, ESM applications tend to be limited to large areal units. For example, a recent study in Southern California is constrained to *Public Use Microdata Areas* (PUMA) which include at least 100,000 inhabitants by design (Conway, 2021). This means that some low-density areas are very large. Therefore, there is a loss of information about the neighbourhood-level attributes in ways that were difficult to foresee, according to MAUP literature discussed in Chapter 3. In addition, this limitation precludes conducting sensitivity analyses concerning the potential spatial aggregation bias.

The implementation of ESM also involves technical challenges. Although there are at least two valuable efforts offering open-source software packages for its estimation, the code is not regularly maintained and the documentation is still limited.[11] This implies that the model still requires to be coded essentially from scratch. Unfortunately, this is out of the time-budget for the present research. In addition, this makes the procedure and estimates prone to error, especially when compared to more established techniques.

Furthermore, estimating ESMs is computationally costly. For instance, a study in Southern California with a similar population size as this study area (19 million versus 21 million in GMC) required the use of a remote private service of considerable capacity (Amazon Web Services, with 128GB of memory and a 16-core AMD EPYC processor) and took fourteen hours using an efficient multi-threading process. In addition, each scenario (allowing residents to re-sort) required between thirteen to twenty hours to simulate using the same setting. For this research, the computational resources were limited to 16GB RAM memory and a 4-core processor at the time of selecting a methodological approach. Therefore, these are far from adequate for the size of the area of study even using optimized computational routines.

On the other hand, the hedonic model presents conceptual and econometric shortcomings reviewed in Chapter 3. However, the method does offer certain advantages for the purpose of this work. Some of these include the readily available property value data at a fine-grained level, allowing the estimation of sensitivity analyses. In addition, the software required to estimate the hedonic model are sound and well documented and usually supported by an active community of users, e.g. a variety of regression frameworks using OLS, maximum likelihood or Bayesian methods. This also permits addressing econometric concerns, such as spatial autocorrelation or heterogeneity issues, using existing software packages with relative ease. Furthermore, this diversity of techniques offers the flexibility required to explore the perception of accessibility revealed by house buyers in simultaneous estimates as proposed in Osland & Thorsen (2008), Osland & Pryce (2012), or Ahlfeldt (2013). This aspect is elaborated in the next section.

Given the advantages and challenges presented by the two main environmental approaches reviewed, it is proposed to use the hedonic method for the purpose of

[11] One option is the `sortingmod` for `R` language (Levkovich & de Graaff, 2017). This fits horizontal models based on Bayer et al. (2004). However, it is not actively maintained. A more recent resource is the `eqsormo` package for `Python` language (Conway, September 6, 2019/2021). This also fits horizontal models based on Tra (2010) and Tra (2013). Currently the documentation is still under development, limiting its use.

the present research. The interpretation of results in the light of the limitations of this approach are discussed in the subsequent chapters. The details of the modelling framework are provided in this section.

5.4.1 Hedonic model

The initial formulation is the general hedonic price function in Eq. (5.1) (McArthur et al., 2012):

$$P_{i,year} = f(z_{s,i,year}, z_{l,i,year}), \tag{5.1}$$

where $P_{i,year}$ is the observed value of house i at time $year$; $z_{s,i,year}$ is structural attributes; and $z_{l,i,year}$ represents locational attributes. These are disaggregated in (1) neighbourhood attributes, and (2) accessibility. The latter is the focus of the present research and the structural and neighbourhood characteristics are considered as control variables as suggested in the literature.

5.4.2 Estimating the hedonic function

Eq. (5.1) is often estimated by the ordinary least squares (OLS) in a regression framework (Phaneuf & Requate, 2017; L. O. Taylor, 2017). This multivariate tool allows the accommodation of the specific structural attributes associated at the individual level of a property, as well as the neighbourhood and accessibility characteristics of it as shown in Eq. (5.2):

$$\ln(P_i) \sim N(\alpha_{+}\beta_S^{\intercal}\mathbf{X}_{iS} + \beta_N^{\intercal}\mathbf{X}_{iN} + \beta_A^{\intercal}\mathbf{X}_{iA}, \sigma_y^2) \text{ for } i = 1, ..., n \tag{5.2}$$

Here, P_i is the value of a dwelling (in MXN) and it is usually entered in the logarithmic form, assuming a non-linear price function according to the best practices for selecting an econometric specification (Bishop et al., 2020). $\mathbf{X}_S$, $\mathbf{X}_N$, $\mathbf{X}_A$ are matrices representing structural, neighbourhood and accessibility characteristics of the dwellings, respectively. β_S, β_N, and β_A are vectors including the respective regression coefficients of the covariates; β_t is the respective regression coefficient; α is a constant term; and σ_y^2 is the error term variance associated to each observation i. This method assumes that the error term is independent and follows a normal distribution.

5.4.3 Multilevel and spatial multilevel models

Property value data usually present some form of spatial structure which is manifested in spatial autocorrelation (i.e. the propensity of close-by areas to share similar attributes (Morris et al., 2019)) or inconsistent variance of the residuals across space (i.e. spatial heterogeneity) (Osland, 2010b). Failing to appropriately model these patterns may result in biased and inconsistent estimates (Anselin & Lozano-Gracia, 2009; LeSage & Pace, 2009; Osland, 2010b). In addition, spatial models are also suitable for capturing preferences in the housing market. For example, the ‘homophily’ principle, reviewed in Chapter 3, states that “[we] are more likely to have contact with those who are close to us in geographic location than those who are distant” (McPherson et al., 2001. p. 430). Therefore, it is necessary to introduce the appropriate considerations to estimate the hedonic model function accounting for possible effects of spatial autocorrelation as detailed below.

In the consecutive definitions, the generic formulation of the methods is presented with the aim of simplifying the technical elements involved. This omits the explicit details of the hedonic model. However, the components of the model are rigorously specified in the corresponding empirical chapters.

One technique that relaxes the assumption of observations’ independence within groups is the multilevel or hierarchical model (Goldstein, 2003). This has been recommended to deal with spatial heterogeneity in urban analysis (Páez & Scott, 2004). In practice, this type of model has been implemented considering the zones in which individual observations are located as the grouping characteristic following the assumption that observations within zones present similarities. Some empirical examples are found in transport geography studies (e.g. J. Hong, Shen, & Zhang, 2014) and hedonic property value analyses (e.g. Habib & Miller, 2008; Mulley & Tsai, 2016; Zolnik, 2019). For these data, a two-level model (multilevel model, MLM) can be considered. Here, the structural characteristics at the individual property level i are defined as the lower-level and neighbourhood and accessibility characteristics are aggregated in one of the SAS zones denoted by j as the upper-level. In the simplest case, the multilevel estimate for a specific zone j with no covariates can be approximated as in Eq. (5.3) (Gelman & Hill, 2007, p. 253):

$$\hat{\nu}_j \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\nu^2}\bar{y}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\nu^2}}, \tag{5.3}$$

where $\boldsymbol{\nu} = (\nu_1, ..., \nu_J)$ is the multilevel estimate, y is a vector expressing the individual value of properties $\ln(P_i)$, n is the number of observations in zone j, σ_y^2 expresses the within-zone variance of the log value of houses, and σ_ν^2 is the variance among the dependent variable of the zones $\bar{y}$. As shown, the multilevel estimate considers individual and zone information, where the information carried by each zone depends on the number of observations contained in j. For instance, zones with small sample size tend to be pulled towards the overall average, whereas large sample size zones are close to the average of the jth zone.

The multilevel estimate ν_j can be treated a zone-level error (usually referred to as a "mixed-effect" model). This estimate can be incorporated in a regression framework as follows (Gelman & Hill, 2007, pp. 264–265):

$$\begin{aligned} y_i &\sim \mathrm{N}(\boldsymbol{X}_i\boldsymbol{\beta} + \nu_{j[i]}, \sigma_y^2), \text{ for } i = 1, ..., n \\ \nu_j &\sim \mathrm{N}(0, \sigma_\nu^2). \end{aligned} \tag{5.4}$$

In this generic formulation, the response variable y_i is assumed to follow a normal distribution; $\boldsymbol{X}$ is the covariate matrix of length $c_1, ..., C$ including a constant term and representing the characteristics of the dwelling detailed in Section 5.4.2; $\boldsymbol{\beta} = (\beta_1, ..., \beta_c)$ is a vector of regression parameters; $\nu_{j[i]}$ represents the zone that contains property i; ν_j is an independent random effect that follows a normal distribution with zero mean; σ_ν^2 denotes the between-zone constant variance, and; σ_y^2 is the individual-level constant variance.

Although the model in Eq. (5.4) acknowledges the grouping characteristics of the data by indexing individual observations to their respective zone, it still assumes independence between the elements defined at the upper-level. Empirical work shows that the capitalization of accessibility and transport-related attributes in land value is affected by nearby areas (Osland et al., 2016). This makes sense since, for example, the level of service generated by transport infrastructure is not exclusive to specific areas, unless there are major physical barriers which prevent potential users from crossing from one zone to another. The exogenous spatial connectivity between zones can be represented by $\mathbf{W}$, a $J \times J$ neighbourhood or adjacency matrix. This is a key element which controls the spatial autocorrelation between zones. $\mathbf{W}$ abstracts the spatial structure of zones by assigning a positive value denoting the spatial closeness between j and a potential neighbouring zone k if these are neighbours or 0 otherwise. Here, neighbouring elements in $\mathbf{W}$ are expressed as w_{jk}.

There are several criteria used to determine the spatial relationship in w_{jk} (Bivand, Pebesma, & Gómez-Rubio, 2013). The most common are the contiguity- or adjacency-based (i.e. if the elements share boundaries), graph-based (e.g. k-nearest), or distance-based (centroids are within x distance). A graph-based approach has been used in this study, where the spatial structure is determined according to the six nearest neighbours. This choice responds to the following reasons:

1. the official source of information defining the post code geometries presents defects at the micro level where the boundaries defining zones show small overlaps or gaps. This makes the use of an adjacency-based approach very challenging;
2. there are few multi-polygon definitions (i.e. post code IDs represented by more

than one closed polygon). Thus, the adjacency-based criteria would not be reliable;

3. high resolution SASs (e.g. 0.5 km) are less likely to contain observations in contiguous cells. Therefore, some may not be considered as neighbours even if they are close to other polygons or other may be left as *islands* (i.e. without neighbours);
4. to choose an appropriate K number, a preliminary sensitivity analysis was conducted considering three- to seven-nearest neighbours where the six-nearest criteria produce the most adequate results for the post code-based spatial scheme;
5. the hexagonal shape of the uniform grids makes this decision natural, according to the potential nearest neighbour interaction as described in Birch et al. (2007).

The values in $\boldsymbol{W}$ can express the spatial cost between one element and another in the *spatial weight* matrix. For this work, a binary weight matrix for spatial modelling purposes was used, which is the most commonly used type of weight (Bivand et al., 2017; e.g. Osland et al., 2016). Formally, this is expressed in Eq. (5.5):

$$w_{jk} \begin{cases} 1 \text{ if areas } j \text{ and } k \text{ are neighbours} \\ 0 \text{ otherwise,} \end{cases} \tag{5.5}$$

where a neighbouring area is given by the K-nearest zone considering their population weighted centroids. In addition, one spatial element cannot be neighbour to itself. This is why $w_{jj} = 0$. The neighbourhood structure and the weight matrix are constructed using the `spdep` package (Bivand, 2021) for the `R` programming language (R Core Team, 2021).

There are several model specifications that can account for the spatial structure of data (Lawson, 2009). One of the most popular is the intrinsic conditional autoregressive (ICAR) model (Besag, 1974; Besag, York, & Mollié, 1991) (also referred as the *Besag* model in literature). This is given by:

$$\upsilon_j \mid \boldsymbol{\upsilon}_{-j}, \mathbf{W}, \sigma_\upsilon^2 \sim N\left(\frac{\sum_{j=1}^{J} w_{jk}\upsilon_k}{\sum_{j=1}^{J} w_{jk}}, \frac{\sigma_\upsilon^2}{\sum_{j=1}^{J} w_{jk}}\right). \tag{5.6}$$

Here, $\boldsymbol{\upsilon} = (\upsilon_1, ..., \upsilon_J)$ denotes a spatially structured random effect for zone j. As shown, the conditional distribution of υ_j is given by the following two elements: (1) the conditional expectation, which is the mean of the random effects of the neighbouring zones of j, and (2) the conditional variance, which is inversely proportional to the number of neighbours of j according to the variance parameter σ_υ^2.

Using the above model alone has been criticized as overdispersion is modelled as a spatial autocorrelation producing biased estimates (Breslow, Leroux, & Platt, 1998; Riebler, Sørbye, Simpson, & Rue, 2016). To mitigate this, a commonly used specification for the random effects is the convolution of the unstructured element ν_j which accounts for non-spatial heterogeneity and the spatially structured element υ_j in a composite random effect $\xi_j = \nu_j + \upsilon_j$. This model was initially outlined by Besag et al. (1991) and it is referred to in literature as the Besag-York-Mollie CAR model (BYM). This can be incorporated in the multilevel model in Eq. (5.4) as follows:

$$\begin{aligned} y_i &\sim \mathrm{N}(\boldsymbol{X}_i\boldsymbol{\beta} + \nu_{j[i]} + \upsilon_{j[i]}, \sigma_y^2), \text{ for } i = 1, ..., n \\ \nu_j &\sim \mathrm{N}(0, \sigma_\nu^2), \\ \upsilon_j \mid \boldsymbol{\upsilon}_{-j} &\sim N\left(\frac{\sum_{j=1}^J w_{jk}\upsilon_k}{\sum_{j=1}^J w_{jk}}, \frac{\sigma_\upsilon^2}{\sum_{j=1}^J w_{jk}}\right). \end{aligned} \tag{5.7}$$

Because of the complexity of the spatially structured random effects, this type of models is most commonly implemented in a Bayesian setting (Lawson, 2009).

5.4.4 Bayesian inference

Bayesian inference is achieved by updating *prior* beliefs of unknown *parameters* based on observed data in order to derive a *posterior* distribution of these parameters (Havard Rue et al., 2017). As detailed by Blangiardo & Cameletti (2015), the uncertainty of a random variable Y can be modelled using a density function represented by θ. This idea is expressed by the *likelikhood* function which is denoted as $L(\theta) = p(Y = y|\theta)$. This specifies the distribution of the data y in light of the model indexed by θ. For simplicity, the likelihood is expressed as $p(y|\theta)$. On one hand, the variability of y is subject to the sample selection which generates uncertainty because of the assumption that this is only one random sample from all other possibilities. On the other, θ is an unknown parameter which is modelled via a corresponding prior probability distribution denoted by $p(\theta)$ and this reflects previous knowledge before Y is observed. Considering these two elements (the likelihood and prior) the inferential problem can be solved using Bayes Theorem—

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}. \tag{5.8}$$

Here, the aim is to obtain the *posterior* distribution $p(\theta|y)$ which reflects uncertainty about θ after observing the data. Therefore, the parameter of interest $\{\theta\}$ is conditioned on y. $p(y)$ is the *marginal distribution* of the data and it can be seen as a normalization constant as it is independent from θ (Blangiardo & Cameletti, 2015). The marginal distribution integrates the uncertainty on θ. This can be obtained from the sum of the weighted probabilities of mutually exclusive and exhaustive events for discrete parameters (based on the law of total probabilities) or by integral calculation for continuous variables as follows:

$$p(y) = \int_p p(y|\theta)p(\theta)d\theta. \tag{5.9}$$

All in all, the posterior distribution will be somewhere between the prior and the likelihood distribution. Therefore, the prior distribution can play an important role in Bayesian inference especially when using small samples (Lawson, 2009). In these models, all parameters are assigned a prior probability distribution (including both fixed and random terms). Here, there are two elements that should be considered (Blangiardo & Cameletti, 2015): the type of the distribution (i.e. functional form), which reflects the nature of the parameter of interest (e.g., if the variable describes a probability, a distribution ranging from 0 to 1 would be appropriate), and the parameters associated to this, which provide the level of information at hand describing the distribution. The usual type for the covariate's coefficients is the Gaussian as this can go from $-\infty$ to ∞ while the most common prior distributions for variance precision parameters are the gamma, inverse gamma, or uniform families (Lawson, 2009).

Having identified an appropriate functional form of the prior distribution, the parameters can define it as informative or *noninformative*. The noninformative do not imply heavy preferences over values of the variables (Lawson, 2009). Therefore, these take a flat shape, assigning the same probability on all the values within a feasible range, as presented in the *Bayes–Laplace postulate* (Blangiardo & Cameletti, 2015). In practice *vague* distributions are often used to approximate a non-informative prior. These assume ignorance on a portion of the parameters where the likelihood is at a great distance from 0. By contrast, an informative prior distribution is used when there is consolidated, previous experience in the topic or alternative based on experts' opinion (Blangiardo & Cameletti, 2015). This type reflects a considerable narrow distribution which is centred on previous knowledge (LeSage & Pace, 2009). It is worth noting that the influence that the prior exerts on the posterior distribution depends on the size of the empirical data (it should be recalled that the prior belief is updated based on the observed data) (Lawson, 2009; LeSage & Pace, 2009). Therefore, the choice becomes more relevant for small sample sizes.

The integrated nested Laplace approximation (INLA)

Commonly, Bayesian inference is carried out based on simulation approaches, e.g. Markov Chain Monte Carlo (MCMC) simulations via the Metropolis and Metropolis-Hastings algorithm or the Gibbs Sampler algorithm (Lawson, 2009). One of the major drawbacks is that these methods are highly time- and resource-consuming despite the increased computing power available today (Havard Rue et al., 2017, p.

396). This is especially true for complex models and large datasets. An alternative method for computing fully Bayesian inference is the integrated nested Laplace approximation (INLA) (Håvard Rue, Martino, & Chopin, 2009). The main advantage of this approach is the substantial reduction in computing time and resources while maintaining the flexibility required to fit complex models and to provide accurate approximations (or even more accurate than the classic simulation methods, as its proponents claim) (Havard Rue et al., 2017; Håvard Rue et al., 2009). Comparative studies show the equivalence of results between simulation-based approaches and INLA in applied studies in the spatial statistics field including hedonic models (e.g. Bivand, Gómez-Rubio, & Rue, 2015; Bivand et al., 2017; Gerber & Furrer, 2015).

The INLA approach is restricted to *latent Gaussian models* (LGM). This is a subset of structured regression additive models which in its most generic form can be expressed by (Blangiardo, Cameletti, Baio, & Rue, 2013; Håvard Rue et al., 2009)

$$\eta_i = \alpha + \sum_{c=1}^{C} \beta_c x_{ci} + \sum_{l=1}^{L} f_l(z_{li}). \tag{5.10}$$

In these type of models the response variable, $\boldsymbol{y} = (y_1, ..., y_n)$, is assumed to be of the exponential family, which is linked to the model by a parameter φ_i (usually the mean) via a link function $g(\cdot)$, so that $g(\varphi_i) = \eta_i$. α represents the intercept; the coefficients $\boldsymbol{\beta} = \beta_1, ..., \beta_C$ denote the linear effect of the covariates $\boldsymbol{x} = (x_1, ..., x_C)$ on the response, and; $\boldsymbol{f} = f_1(\cdot), ..., f(\cdot)_L$ is one or more functions of the covariates $z = (z_i, ..., z_L)$. $f_l(\cdot)$ can take different forms, including non-linear effects, a varying random intercept, or spatial random effects. Therefore, LGMs cover a wide variety of the most commonly used models, from the (generalized) linear to spatio-temporal. The vector parameters of interest for the inference (also referred in literature as latent or nonobservable components) in Eq. (5.10) are represented in a set of parameters defined as $\theta = \{\alpha, \boldsymbol{\beta}, \boldsymbol{f}\}$. A second vector represents the K hyperparameters as $\boldsymbol{\psi} = (\psi_1, ..., \psi_K)$, typically this is shorter in length than θ.

Bayesian inference focuses on the estimation of the posterior marginal distribution for each parameter vector

$$p(\theta_i \mid \boldsymbol{y}) = \int p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y}) p(\boldsymbol{\psi} \mid \boldsymbol{y}) d(\boldsymbol{\psi}), \tag{5.11}$$

and for each element of the hyperparameter vector

$$p(\psi_k \mid \boldsymbol{y}) = \int p(\boldsymbol{\psi} \mid \boldsymbol{y}) d\boldsymbol{\psi}_{-k}. \tag{5.12}$$

Therefore, it is necessary to compute (Blangiardo & Cameletti, 2015; Blangiardo et al., 2013):

1. $p(\boldsymbol{\psi} \mid \boldsymbol{y})$, from where the appropriate marginals $p(\psi_k \mid \boldsymbol{y})$ can be obtained, and;
2. $p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})$, which is required to estimate the posteriors of the marginal parameter $p(\theta_i \mid \boldsymbol{y})$.

INLA's efficiency is because of its deterministic approach to compute 1 and 2 using nested approximations based on the Laplace integration method, rather than a simulation approach. The method exploits the assumption of the joint distribution as a *Gaussian Markov Random Fields* (GMRF) with a precision matrix being the sum of the precision matrix of the fixed effects and the rest of the components of the model specified in Eq. (5.10) (Havard Rue et al., 2017).

While INLA is limited to a specific class of models, the advantages for applied researchers are numerous. For instance, the following are directly applicable to this data and research process:

1. INLA's agility permits a richer and more in-depth exploration of data than the simulation-based approach by facilitating the iterative modelling process referred in Havard Rue et al. (2017). This includes testing the stability of alternative variables in terms of consistency and significance of coefficients in preliminary stages;
2. in addition, it is possible to generate more robust results as sensitivity analyses become feasible, and;
3. this offers the capability of fitting large datasets similar to these that otherwise could simply not be possible to fit in an ordinary CPU using other alternative simulation-based software because of the time needed and technical resource limitations (e.g. preliminary tests were conducted using the `HSAR` V0.5.1 package (Dong, Harris, & Mimis, 2020) for `R`. However, the time required to fit a model considering the size of the sample used for the present work meant the use of such simulation-based approach was not possible).

Therefore, INLA is the preferred method to estimate the spatial multilevel models in the present work given by Eq. (5.7), which can be re-written following INLA's notation as:

$$\eta_{ij} = \alpha + \sum_{c=1}^{C} \beta_c x_{cij} + \nu_j + \upsilon_j + \epsilon_{ij}. \tag{5.13}$$

The implementation of this approach is through the `INLA` software (V. '21.7.10.1') (Havard Rue, Lindgren, & Teixeira Krainski, 2021) for the `R` programming language (R Core Team, 2021).

5.4.5 Model assessment

With regard to the goodness-of-fit, the most usual criterion in OLS models is the R^2. This measure is expressed as $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, where SS_{res} is the sum of squares of residuals of a fitted model and SS_{tot} is the total sum of squares. This can be understood as the proportion of the variance explained. For OLS models, the adjusted R squared, denoted as $\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$ was reported. Here, n is the sample size and p is the number of explanatory variables.

For models fitted using the method of maximum likelihood, the log of the maximized likelihood function $\{\ln(L)\}$ and the *Akaine information criterion* (AIC) were reported. The latter is calculated as $\text{AIC} = -2\ln(L) + 2k$, where k is the number of predictors in the model. According to this measure, models with lower AIC values are preferred. The analogue version of the AIC in Bayesian models is the *deviance information criterion* (DIC) (Gelman & Hill, 2007). This is expressed as $\text{DIC} = D(\bar{\theta}) + 2p_D$, where $D(\bar{\theta})$ is the mean deviance and p_D is the number of effective parameters. As shown, both AIC and DIC are a form of a penalized fit according to the complexity of the model.

In addition, the root-mean-squared error (RMSE) was reported, which is given by $RMSE = \sqrt{SSE/n}$, where SEE is the sum of the squared errors and n is the sample size.

5.4.6 Benefit measurement

As noted in (Section 3.1), the interpretation of implicit prices obtained from the first stage of the hedonic model as the willingness to pay (WTP) can be established under certain conditions. These relate to (a) the size of the change in the environmental characteristic and (b) the extent of its impact (Bishop et al., 2020; L. O. Taylor, 2008).

Arguably, the introduction of a new rapid or semi-rapid transit corridor can produce changes in accessibility that are neither marginal nor localised. If this is not the case, theoretical literature suggests as appropriate the use of alternative approaches to those presented above, for instance, Rosen's second-step demand analysis or the equilibrium sorting model (Freeman et al., 2014; L. O. Taylor, 2017). However, few empirical studies opt for these theoretically correct alternatives given the level of complexity and the additional information required.

While theory urges the adoption of such approaches (e.g. Freeman et al., 2014; L. O. Taylor, 2017), empirical literature provides limited references in terms of the expected degree of uncertainty following further consideration. With regard to the net benefits, Gjestland et al. (2014) concluded that a simplistic hedonic approach (i.e Rosen's first step) provides *reasonable* results and is 'very similar' in the case of a variable of an ex post hedonic model when compared to a travel demand model as the benchmark for benefits associated to changes in accessibility introduced by the transport network. In addition, an empirical study assumes changes in accessibility derived from the expansion of the rail network in London, UK, as marginal changes from the perspective of a partial equilibrium (Ahlfeldt, 2013). The study predicted a set of out-of-sample housing prices and validated them using observed data. The conclusions suggest that the approach produced 'satisfactory' predictions. Although these experiences provide general references, these types of studies have limitations and results may still be context specific. For example, relocation costs have been shown to be substantially different between the US and China at the regional level (Liang et al., 2021).

For the purpose of illustrating the magnitude of the net benefits using public transport location-based accessibility measures as a starting point, a simplified approach using the estimated hedonic price function has been provided. This has been developed from a partial equilibrium view. Accordingly, the WTP can be predicted from for a change in accessibility as following (using the notation of the source) (Ibid, p. 265):

$$\hat{\mathrm{WTP}} = \Delta z_c \frac{\partial h(z)}{\partial z_c}. \tag{5.14}$$

In the equation above, z_c is a vector denoting the characteristic of interest, e.g. public transport accessibility; Δz_c is given following the difference between z_c at the original time-point and a new/improved condition; $h(z)$ is the estimated hedonic function, and; $\partial h(z)/\partial z_c$ is the partial derivative with respect to MPTN accessibility. As the functional form of the hedonic function implemented in this work is the semi-log, L. O. Taylor (2017) suggests that the implicit price must be assessed in relation to a price level P as

$$\frac{\partial h(z)}{\partial z_c} = \beta_c \times P. \tag{5.15}$$

In these empirical estimates, β_c is the regression coefficient obtained in the final model which accounts for MPTN accessibility.

According to Freeman et al. (2014), the total willingness to pay is formally expressed as

$$w_q = \sum_{i=1}^{N} WTP. \tag{5.16}$$

Following the notation of the source, w_q is the aggregated marginal welfare change (or total willingness to pay -TWTP) and WTP is the individual marginal willingness to pay.[12]

[12]The notation in the source uses b^{*i} to represent the individual marginal willingness to pay. Here, the original notation is adapted for consistency with previous expressions.

This estimate does not explicitly consider transaction costs. However, it is not expected that these costs can radically bias the estimates.[13] This is because of the structure of the National Housing System and the policies in place seek to reduce the frictions of the housing market, as detailed in Chapter 4. Still, the figures represent an upper bound estimate (Palmquist, 2005; L. O. Taylor, 2017), as discussed in Chapter 3. The specific implications and interpretations for the present case are discussed along with the results

5.5 Environmental measures

This section provides the details of the data used to account for environmental measures in the hedonic models. It also describes the procedures and references followed for their computation.

5.5.1 Neighbourhood measures

According to hedonic property value literature and preliminary modelling, the following neighbourhood attributes are identified as appropriate controls in the context of GMC: availability of educational facility, presence of a public administration establishment, percentage of streets that do not have streetlights, percentage of street with trees, having a park, crime, and percentage of households with a private car available.

The availability of an *educational facility* and the presence of a *public administration* establishment is based on the 2015 National Directory of Economic Units (*Directorio Estadístico Nacional de Unidades Económicas 2015*, DENUE) (INEGI, 2015c).[14] The DENUE is an open-access spatially rich dataset which is organized at the point-level for each individual establishment that provides any kind of service. This data collection includes: (1) location (geographic co-ordinates); (2) number of occupied personnel in ordinal categories (i.e. '0 to 5,' '6 to 10,' and other factors, up to 'More than 251'); (3) and code corresponding to a hierarchical classification according to the 2012 North American Classification System (NAICS) by sector (first two digits of the code), sub-sector (following 3-digit code), and other relevant details. While this data cover all the national territory only establishments within GMC are considered.

An education facility is identified from the DENUE as an establishment in which its sector is equal to 'Educational Services,' including 'Elementary and Secondary Schools,' 'Junior Colleges,' and 'Colleges, Universities, and Professional Schools' at

[13]This is the case in Liang et al. (2021), which shows that the strict migration policy in China bias the direction of the estimates for the willingness to pay for air quality.

[14]Downloaded on the April the 14th, 2020, through the public URL available in https://www.inegi.org.mx/app/descarga/default.html.

the sub-sector level. The respective codes are '6111,' '6112,' and '6113.' From these, only those that are large are considered, i.e. 'More than 251.' This choice is aimed at reducing potential collinearity with accessibility to employment measures in the final models since these latter measures use similar information as their input, i.e. employment. According to the selection criteria, there are 176 major education facility locations in GMC.

An education facility is considered as being available for a property if the travel time by walking from the population weighted centroid of a SAS zone to the exact location of the service is twenty minutes or less. This threshold is based on the figures reported in the 2017 Travel Survey which indicate that the duration of non-working trips is 16.4 minutes on average. This figure is rounded up to allow some degree of tolerance given the spatial aggregation of the origin. The walking travel time was estimated using the `r5r` package (Saraiva, Pereira, Herszenhut, Braga, & Conway, 2021) for the `R` programming language (R Core Team, 2021). The main input to compute the estimated travel time walking is the road network, which was accessed from *OpenStreetMap* (OSM).[15]

A *public administration establishment* was identified as a DENUE record classed as 'Public Administration' at the sector level (2-digit code '93') and its size is equal to 'More than 251.' The criteria for the size of the establishment and the distance to be considered as available are the same as for educational facilities. There are 496 establishments of this type in GMC.

Two proxy variables are used to account for the overall quality of the neighbourhood characteristics; namely percentage of streets that do not have street lights (*Street light NA (%)*) and percentage of streets with trees (*Street with trees (%)*). These were chosen from a variety of characteristics obtained from the 2016 National Housing Inventory (*Inventario Nacional de Viviendas 2016*, INV)[16] (INEGI, 2016) based on preliminary modelling. The criteria considered the correlation levels with the dependent variable and the models' assumptions discussed in the Model assessment subsection.

The process to measure the percentage of block fronts including streetlights and streets with trees is as follows; first, the raw information from the INV is organized in spatial data files (`.SHP`) as vector objects which represent the front of an urban block where each front is classified according to several discrete characteristics. Then, the vectors were spatially intersected by the SAS zones. Through this process the segments of the block (including the characteristics) are assigned the overlapping SAS unit. Finally, the percentage is computed from the ratio of the length of fronts that

[15] Manually downloaded on October the 30th 2019 through web export tool available in `https://export.hotosm.org/es/v3/`.

[16] Manually downloaded from "`https://www.inegi.org.mx/contenidos/masiva/indicadores/inv/%5BENT%5D_Frentes_INV2016_shp.zip`" on the 07/02/2020.

match the characteristic of interest over the sum of urban block fronts within each zone.

The input data to account for the availability of a park (*Park*) use OSM. This source was preferred as the official information is not standardized across the different states or municipalities in this area of study. The data were accessed from OMS via the `R` package `osmdata` (Padgham, Rudis, Lovelace, & Salmon, 2021).[17] The key used in the API query to download the information was equal to 'leisure' and the value was set to 'park,' according to the documentation available from the source (https://wiki.openstreetmap.org/wiki/Category:Keys). The data were spatially subset to match only those polygons within Greater Mexico City. In addition, only polygons that are equal or larger than 5,000 m^2 (or 0.5 hectares) were considered, following the minimum area recommended by the World Health Organization (WHO) (WHO, 2017) and the figure employed in previous empirical studies (e.g. Cömertler, 2017; Ståhle, 2010). Furthermore, median strips were excluded (long green areas along roads) based on the criteria suggested by the Ministry of Agriculture, Land and Urban Development (*Secretaría de Desarrollo Agrario, Territorial y Urbano*, SEDATU) (DOF, 2020), which indicates that the strip 'must be at least 20 m wide on its shortest side to be considered as a park.' Given the ambiguity involved in implementing this rule, and the lack of further details regarding its practical application, this is operationalised as whether a 10 m radius circle drawn from the centroid of each geometry does fit within the corresponding polygon represeting a green area. Those that do not pass the test were excluded.

The walking distance to a park was estimated from the geometric centroid of the park[18] to the population weighted centroid of a SAS zone. The method and sources required to estimate the travel time by walking was the same as for educational facilities or public administration establishments. A park was assumed to be available if it was within fifteen minutes walking distance following the threshold adopted by the European Environment Agency (EEA, 2002)[19] and in common with the median travel time to leisure activities reported in the 2017 Travel Survey.

Crime was measured based on open data issued by the National Security System (*Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública*) (SESNSP, 2021). The information covers the period from 2011 to 2017 and records reported crimes aggregated at the municipal level by month.[20] Although the spatial aggregation level

[17]The data was downloaded on the 16/03/2021.

[18]Or a grid point within the park uniformly distributed for large polygons > 100 000 sqm. This is to account the proximity of houses located near to the edge of a park. Otherwise, the distance to the geometric centroid of a large park could misleading.

[19]The report suggest the accessibility concepts as 'within 15 minutes' walking distance and assumes as equivalent a 300 m straight-line distance. We adopt the 15 minutes walking distance on the road network as the criteria.

[20]The data was manually download from the https://www.gob.mx/sesnsp/acciones-y-

is relatively high (municipality), this source is preferred as it offers a standardized collection method for the whole area of study. Here, the offences are recorded in a diversity of discrete classes, types and subtypes. As most of the classes are correlated to each other (Pearson correlation coefficient ≥ 0.80), only homicides were chosen as a representative type of crime. The measure used in the models is the standardized net sum of homicides for the period reported (2011-2017). This information was imputed to observations according to the municipality where a valuation record is located. In a preliminary modelling stage, a normalized crime rate by population was also tried. However, the raw figure produced more robust results. This makes sense, as the total number of these types of high impact crimes can affect home buyers' perception.

A measure indicating the proportion of households having a car is used as a control variable for accessibility (*HH with car(%)*. For instance, this measure could control the estimated number of opportunities that can be accessed by car only. In addition, this can account for neighbourhood dis-amenities, such as road congestion as suggested in previous empirical studies (Ahlfeldt, 2011). The source is the 2010 Population Census collected by the INEGI and this was accessed through the Census Information Consultation System (*Sistema de Consulta de Información Censal*, SCINCE) (INEGI, 2012). The original data present the information aggregated at the urban block level. To create this measure, the information at the SAS zone was aggregated and a ratio of the number of households having at least one car available over the total number of households in the zone was computed.

5.5.2 Accessibility measures

As the main interest lies in public transport infrastructure, the analysis paid special attention to the possible measures used to account for *general accessibility* as further control variables. The benchmark is the distance to the traditional central business district (CBD), as it has been frequently used in the literature (Heyman et al., 2019). This was estimated as the Euclidean distance in kilometres from each population weighted centroid of a SAS zone to the location of traditional CBD (*Zócalo*, the main square in Mexico City).

Following Ahlfeldt (2011), a gravity-type measure accounting for accessibility to employment by car was introduced according to $A_j^{\text{CAR}} = \sum_{j=1}^{J} E_k \exp(-\beta_1 d_{jk})$.[21] Here, d_{jk} is the estimated travel time by car in minutes from the post code of origin j (where the property i is located) to all potential destination zones k, E is the number of employment opportunities at k and β_1 is a distance deterrence parameter. The

`programas/datos-abiertos-de-incidencia-delictiva?state=published` on the 20/03/2021.

[21] For consistency with the accessibility literature, I keep the usual notation of the distance deterrence parameter using β. The accessibility parameters are differentiated from the regression coefficients by assign numeric sub-indexes to the former.

estimation method of this parameter is detailed in the next section. These measures (distance to the CBD or potential accessibility to employment by car) can either replace or complement each other. This possibility is tested in Chapter 7.

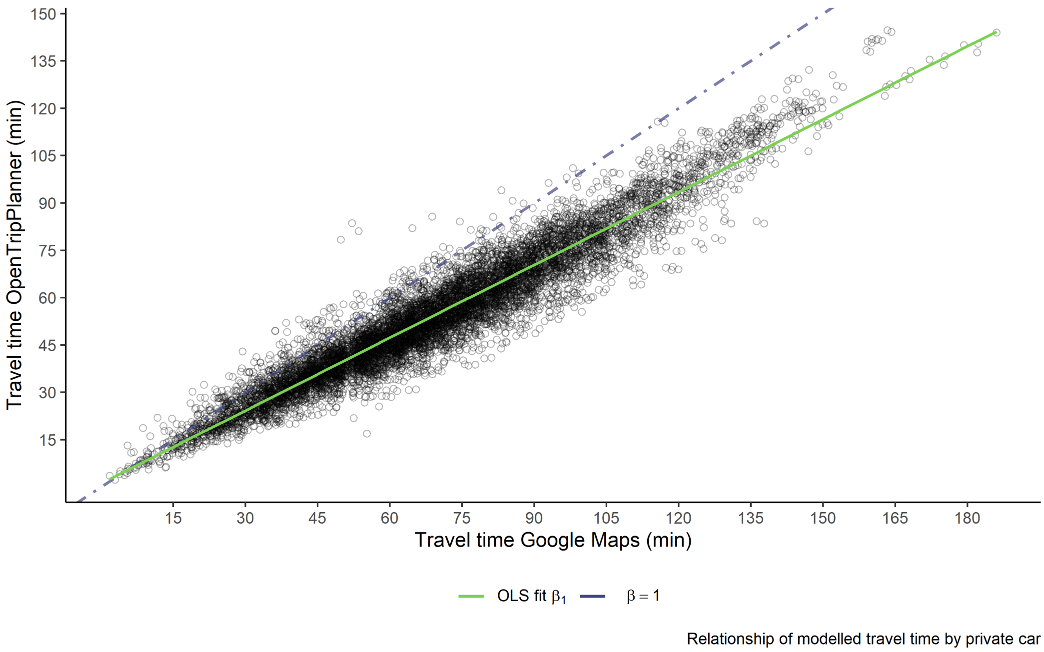
Travel time by private car is modelled for one fixed scenario departing at 10 a.m. on 8th August 2018 in OTP. The initial estimate is a free-flow model based on the road network obtained from OpenStreetMap (OSM) data in 2019. Although it is acknowledged that an ideal private car model should estimate the link average speed according to several time periods (e.g. a.m. peak, afternoon, p.m. peak, evening and night) accounting for differences in travel speeds, such information was not available for the scenarios studied. However, if this information existed for the area studied, it is understood that this is not owned by public agencies. Therefore, acquiring it would be costly and out of the budget of the present project. In addition, possible travel time inaccuracies can be controlled through the flexible parameters considered in the proposed accessibility measures (i.e. β_1) assuming generalized delays in the metropolitan area.

Therefore, the initial free-flow model was adjusted to account for congestion and other unobserved factors based on a sample of 9,264 random OD routes estimated by Google Maps[22] (GM) and accessed through the `gmapsdistance` package (Demetrio Rodriguez T & David Zarruk, 2018) for `R`. This estimate considers the typical conditions for different times of the day taking road congestion into account. The estimates were queried for a single departure time of 10 a.m. on Wednesday, 30th March 2021. It is worth highlighting that one of the limitations of the private routing services such as GM, is the restriction on formulating alternative scenarios. This includes the request of routings only on future dates. Therefore, it is not possible to query the routes for a temporal period corresponding to the property data available for this study. Having this in mind, the difference between the the routing estimates (OPT vs GM) was evaluated in a regression framework following Eq. (5.17).

$$d_{jk\text{GM}} = \alpha + \beta_1 d_{jk\text{OTP}} + \epsilon_{jk} \tag{5.17}$$

In the Eq. above, d_{ij} is the estimated travel time between each OD pair for its respective routing engine, where $i \neq j$. The proposed model is visualized in Figure 5.3. The plot shows a strong linear relationship. In addition, a visual comparison between the fitted line (OLS β_1) and the dashed line which represents equality between two estimates (where $\alpha = 0$ and $\beta = 1$), shows that the original free-flow model is consistently underestimated. This implies feasibility to correct the initial free-flow model.

[22]www.google.com/maps



Source: the Author based on own estimates and Google API.

Figure 5.3: Modelled travel time by car in Google Maps and OpenTripPlanner.

Alternative specifications to Eq. (5.17) were tested in a multivariate framework adding possible control variables; for example, a binary variable capturing whether the journey starts or ends in a core location of GMC. However, these controls did not add considerable explanatory power but added unnecessary complexity. Appendix B presents the details of the chosen parameters to correct the models and the alternative specifications.

In addition, a search time for parking was added following previous empirical work (Salonen & Toivonen, 2013). Here this corresponds to 8.1 minutes based on an international average estimated by Shoup (2006).

Internal travel estimates follow the same specification described earlier plus the constant accounting for the time needed to search for a parking space. The average speed was computed for short trips ($\leq$ 3 km) from the GM sample where $i \neq j$.

Therefore, the final model by private car is given by Eq. (5.18).

$$\hat{d_{jk\text{CAR}}} \begin{cases} 4.54 + 1.19 * d_{ij\text{OTP}} + 8.1 & \text{if } j \neq k \\ \frac{\frac{1}{2}\sqrt{\frac{area_i}{\pi}}}{\text{mean}(d_{jk\text{GMspeed}\leq 3km, j\neq k})} + 8.1 & \text{if } j = k \end{cases} \tag{5.18}$$

The public transport accessibility measures used in the subsequent analyses are computed following the location-based specification described in Chapter 2. The details of the specifications employed are provided in the following chapters. The reason for adopting such a format is for clarity and simplicity as these measures require a considerable level of discussion and technicalities that are context-specific to each of the Research Questions guiding the present work.

5.6 Summary and final remarks

LVC can benefit from additional knowledge about the distribution of the economic benefits derived from public actions. The incorporation of methods and tools enabled under the Open Science framework can provide further insights for the application of LVC. This chapter outlined the research strategy adopted to answer the Research Questions under such a framework. In addition, it provided the details of the different spatial analytical schemes that will be used to address the Research Questions, based on the knowledge and gaps identified in Chapter 3. In addition, it details the sources, data processes and sample selection criteria of the property transaction records which constitute a key piece of information for this work. An important element of the sample is the geolocation process which is supported by information obtained from the state land registry.

The modelling framework employed to empirically assess the willingness to pay for the benefits of the MPTN has been presented. This approach is based on a number of considerations guided by the theoretical framework, the hedonic methodology, existing studies, the specific characteristics of the sample and the area of study. Finally, the information sources and methods for computing the environmental amenity measures used in the hedonic models are detailed. Before proceeding to the estimation of the hedonic model in Chapter 7, the next chapter explores the role of the main public transport network shaping accessibility to employment in Greater Mexico City between 2009 and 2019.

Chapter 6

The role of the main public transport network in shaping accessibility to employment. An exploratory spatial data analysis

6.1 Introduction and Research Question

Location-based accessibility measures have been used to account for the level of service offered by the transport infrastructure in a limited number of property value studies (e.g. Ahlfeldt, 2013; Gjestland et al., 2014). At the same time, the literature looking at the theory of accessibility is extensive and it elucidates the key aspects involved in the concept (Geurs & van Wee, 2004; Levinson & Wu, 2020). Still, the outcomes generated by the public transport network have yet to be explored in empirical studies which consider some of the key features involved in the estimation of location-based measures. These range from the spatial definition of the area of study and its subsequent partition, to the various aspects of people's perception. Altogether, these considerations can offer useful guidance for the environmental measure of accessibility from the empirical perspective. In turn, this can support the implementation of different land value capture (LVC) instruments in terms of additional knowledge about the distribution of the benefits of public transport infrastructure.
From the above, the objective of this chapter is to:

- Explore the distribution of accessibility to employment enabled by the main public transit network in Greater Mexico City between 2009 and 2019.

Accordingly, the specific Research Question is:

- *What is the role of public transport infrastructure in shaping accessibility to employment in Greater Mexico City?*

This chapter offers two contributions to this research. First, it identifies and digitally reconstructs in a common format (general transit feed specification, GTFS) a series of temporal stages of the MPTN in GMC for a ten-year period (2009-2019). This considers the different services at the metropolitan level for the first time in academic literature for this area of study and is an addition to publicly available data since it is currently limited to *Ciudad de México*. Secondly, it illustrates the spatial distribution of general MPTN accessibility benefits, which are relative to previous temporal stages (in terms of environmental changes), by relaxing the assumption that all employment opportunities are equally attractive to all the population considering the MAUP. Sensitivity parameters (spatial or distance decay) are informed from the observation of commuting flows.

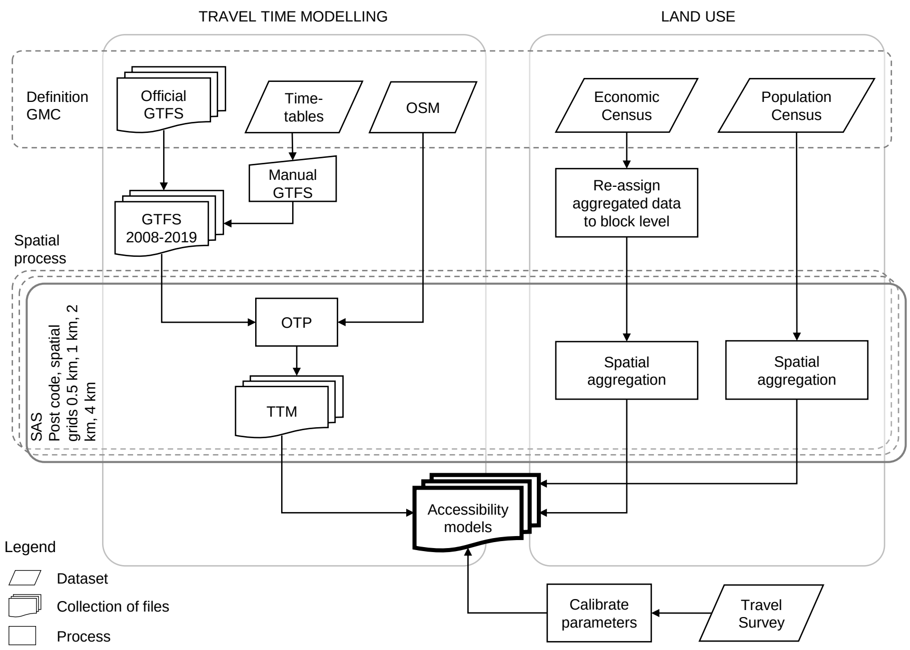
The key findings suggest that, by taking into account the different components of location-based measures, accessibility is far from symmetrical. The results demonstrate that the role of the MPTN depends on both internal (operational and physical) and external (land use) characteristics, which is consistent with expectations based on theory. The contribution of this study lies in the explicit examination of the extent and magnitude of an empirical case, which highlights the importance of incorporating these concepts into environmental valuation. In addition, the assumptions adopted about the perception of accessibility (i.e. whether all opportunities are equally attractive or are restricted according to the education level of residents) produce further heterogeneous results. The MAUP considerations represent an additional aspect related to perception. The results imply different patterns according to different ways in which the area of study is partitioned. These findings also have important implications for environmental measures in non-market valuation studies (e.g. hedonic property value). For instance, Chapter 3 highlighted that many studies still rely on simple metrics that reflect incomplete characterization of accessibility, e.g. equal-distance buffers or distance to stations. Thus, these tend to overlook key concepts in land-rent theory and several contributions in the accessibility literature reviewed in Chapter 2.

This chapter is structured as follows Section 6.2 details the methods employed to examine accessibility to employment. These include the description of sources and inputs employed, the procedures followed to model travel time by public transport and the approach adopted to calibrate the parameters framing the measures. Section 6.3 presents a comprehensive univariate exploratory analysis of the accessibility to employment measures. This is comprised of an overview, an examination of aspects relating to spatial aggregation and the temporal variability of access. Section 6.4 discusses the results and acknowledges the limitations of the analyses presented. Section 6.5 concludes the chapter.

6.2 Research strategy and estimation

6.2.1 Overview

Figure 6.1 depicts the conceptual route-map followed for the estimation of the accessibility models. The diagram can be read employing the four horizontal levels and two columns. The top horizontal level highlights the raw data sources. The second shows the preliminary data processes required to supplement or format the raw data. The next level represents the processes required to organize or produce the information according to the five different SASs proposed, namely the hexagonal grids of 0.5 km, 1 km, 2 km and 4 km, as well as the post code zones, as detailed in Section 5.2 in Chapter 5. The bottom level of the illustration shows how the information is assembled to produce the accessibility models. The box on the left-hand side refers to information about the transport network, while the box on the right-hand side accommodates the land use data.



Note: GTFS = general transit feed specification; OSM = OpenStreetMap; OTP = OpenTripPlanner; TTM = Travel time matrix; GMC = Greater Mexico City.

Figure 6.1: Conceptual route-map for the estimation of accessibility models.

The specific sources employed and procedures followed are detailed in the subsequent sub-sections and are structured as follows; first, the definition of the accessibility measures adopted, secondly, the sources of the data, and, thirdly, the estimation processes including the selection of a spatial decaying function along with the calibration of the corresponding parameters.

6.2.2 Accessibility measures

The general form of accessibility expressed in Eq. (6.1) is the form which has been adopted for the subsequent analyses.

$$A_{jt}^{\mathrm{PT}} = \sum_{j=1}^{J} E_k f(d_{jkt}) \tag{6.1}$$

This specification considers the level of MPTN accessibility at origin j at temporal stage t and impedance is given by a function of travel time between j and k (d_{jk}). Specifically, travel time is represented by a modelled estimate in minutes using a combination of walking and the MPTN at temporal stage t; the opportunities at destinations are expressed by E, which is the total employment in destinations k.

In Eq. (6.1), households and opportunities are assumed to be homogeneous, i.e., all types of employment are equally attractive for all types of residents or potential house-buyers in the context of the hedonic model. Based on theoretical contributions (Geurs & van Wee, 2004; Levinson & Wu, 2020), an additional accessibility measure takes into consideration the characteristics of the demand for each origin and the characteristics of supply at potential destinations. These characteristics are expressed as type of person m at j and type of employment m at k. According to this suggestion, the measure can be specified as shown in Eq. (6.2):

$$A_{jtm}^{\mathrm{PT}} = \sum_{j=1}^{J} E_{km} f(d_{jkt}), \tag{6.2}$$

where m at j is categorized according to the median education level of its residents by quintile and m at k is classified based on the average paid salary by quintile. Here, only employment opportunities E matching the education level quantile between j and k are considered (when $j \neq k$).[1] This assumes that residents would find attractive only employment opportunities that match their education level, i.e. those in the same quintile.

The approach is based on the Human Capital Theory (Becker, 1994; Schultz,

[1]An exception occurs when $j = k$. In this case, employment opportunities in the origin are considered as available. This is to avoid the assumption implying that some zones have accessibility values equal to 0

1961) and its operationalization is guided by existing empirical accessibility studies (Pereira, 2019). Its implementation is adapted according to the information available and some preliminary findings. For instance, Pereira (2019) classifies residents according to their income level by decile.[2] However, the households' income was not available in the Population Census data at the disaggregation level required. Therefore, the households' education level (i.e., median number of years of education) is employed assuming correspondence with the level of salary paid in the employment market. This is further supported by the positive relationship found between education level and wages in empirical studies in the context of Mexico (Nuñez et al., 2017; M. Rojas, Angulo, & Velázquez, 2000). The data sources and specific procedures to process the data are described in the next sub-section.

The proposed MPTN accessibility measures can be viewed at two extremes. At one end Eq. (6.1) does not impose any restriction in terms of the attractiveness of opportunities relative to the type of population at the origin. At the other end, the measure produced by Eq. (6.2) does not allow mobility across different employment or education levels. These views are adopted here as a starting point to verify whether there are substantial differences between these measures.

6.2.3 Data

Road network and pedestrian infrastructure

The road and pedestrian infrastructure network represents the basic layout for modelling origin-destination journeys. The network data were obtained from *OpenStreetMap* (OSM).[3] This information is required in multi-modal journeys to realistically represent the access/egress stage to a public transport service according to the characteristics of the network when combining walking and public transport modes.

Although OSM data is not complete or flawless, it offers considerable levels of detail, e.g., main pedestrian crossings or bridges, pathways crossing green areas or parks, or restricted roads for pedestrians. This source was preferred to official data as the level of detail in core areas is higher, e.g. unofficial paths, pedestrian crossings or in some cases over or under-ground metro passages. In addition, this is readily available for use in common open-source routing engines (the basic tool required to estimate the travel time matrices).

[2]Different quantiles were tested in an early modelling stage and categorizing the data in five breaks of equal size instead of ten produced better results in terms of model fit.

[3]Manually downloaded on 30th October 2019 through a web export tool available in https://export.hotosm.org/es/v3/

Public transport itineraries and temporal aggregation

For the present work, impedance is based on travel time. This is regarded as a superior measure over metric distances in literature (as in straight-line distance or network path-based) (Miller, 2018). This is further supported in that travel time can reflect many characteristics of public transport systems, e.g. topology of the network or operational characteristics of the modes as discussed in Chapter 2 and Chapter 5.

The estimated travel time by public transport is based on timetables shown in *general transit feed specification* (GTFS) files. The Transport Authority in Mexico City (*Secretaría de Movilidad de la Ciudad de México*, Semovi) published its first set of GTFS files in 2013. Since then, these have been updated approximately once or twice a year. GTFS files were accessed from the website `https://openmobilitydata.org/` via uniform resource locator (URL) on 14th August, 2020.[4] These sets of data include the agencies which operate most of the MPTN services in GMC, namely: *Metrobús* (MB), *Sistema de Transporte Colectivo Metro* (METRO), *Servicio de Transportes Eléctricos* (STE), *Ferrocarriles Suburbanos* (SUB). However, the data presents the following limitations:

1. The geographic coverage includes services that are primarily located within the state of Ciudad de México with some exceptions, i.e. the suburban rail network (*Tren Suburbano Buenavista-Cuautitlán*, SUB-L1) and Line B of the metro system (METRO-LB). This implies that the services provided by the BRT agency MXB are excluded from this data.
2. The temporal coverage includes data from the year 2013 onwards.

To overcome the first limitation, the geographic coverage of the official GTFS data was augmented to include three BRT corridors operated by MXB. This was carried out *manually* in `R` programming environment following the best practices for structuring GTFS data available in the platform `http://gtfs.org/`. The detailed information for these three corridors was obtained directly from the Transport Agency of the State of Mexico (*Sistema de Transporte Masivo y Teleferico del Estado de México*, Sitramytem) through two Freedom of Information requests. The information accessed considered the valid services for 2019 only (when the information was requested). The data provided a collection of non-standardized files (e.g. stops and corridors were shared in `.kml` files and the rest of the information as text or figures in `.pdf` files). The details of these MXB services are presented in Appendix C.

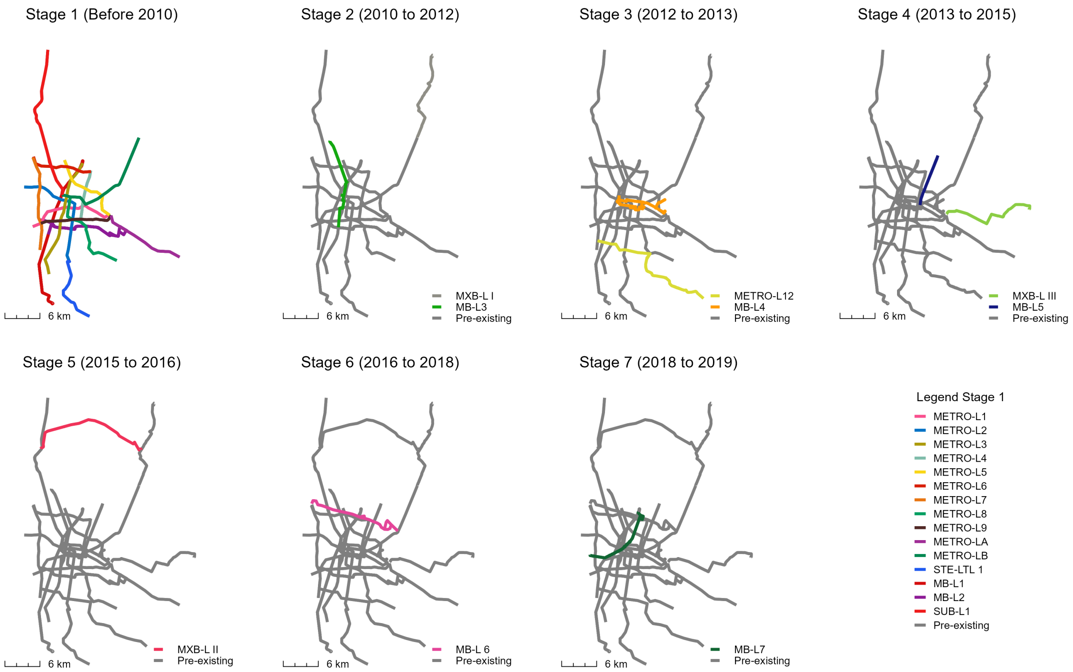
With regard to the second limitation, the MPTN's services were reconstructed as GTFS data for the period between late 2008 and 2019, in line with the time period

[4]The GTFS files downloaded correspond to the following dates of publication: 2013-10-21, 2014-12-30, 2015-01-28, 2017-09-02, 2019-01-09.

covering the residential property data. The basic input required to reconstruct each of the stages was the closest available in time official GTFS data reflecting these changes to the MPTN. With regard to the MXB services, the information employed was for 2019, as the details of the services, prior to the date the data was requested, were not available. This assumes that the level and type of services in the MXB corridors included in the different stages are fixed. Similarly, the level of service for the rest of the corridors included in the early stages were assumed to be as they were in 2013, as the details of the services before this year are not available. All the services that do not form part of the MPTN (as defined in Chapter 4) were excluded as these are incomplete (i.e. they do not consider informal services). In addition, the geographic extension to the state of Mexico would not be feasible given that the required data do not exist either.

The different states of the network during the period studied were grouped in seven temporal stages shown in Figure 6.2. This was to ease the computational burden required to estimate the various accessibility model scenarios while accounting for the main changes in the level of service through time. The key criteria to define a temporal stage are based on the physical changes in the MPTN network, i.e. the opening date of an entire new corridor or a partial extension of a existing one. The decision to consider the changes in the network according to the opening date (and not the announcement or consolidation stage, for example), follows the empirical findings of capitalization literature (i.e. Ahlfeldt, 2013; Gibbons & Machin, 2005).[5]

[5]Empirical findings report that important capitalization effects occur close in time to the opening date of the public transport infrastructure in areas with high owner-occupancy (Ahlfeldt, 2013; Gibbons & Machin, 2005). These studies do not report considerable 'anticipation effects,' which may occur after the announcement of a project. In Greater Mexico City the proportion of rented dwellings is relatively low, i.e. 20%.



Source: The author based on Semovi (n.d.) and SITRAMyTEM (n.d.).

Figure 6.2: Temporal stages of the main public transport network. Greater Mexico City, 2010 to 2019.

Once all the individual changes in the MPTN are identified, the difference in months to the immediately previous change is computed. If a modification occurred within six months, this is grouped in the same temporal stage. Otherwise, a new stage is created. This follows the assumption that accessibility benefits are perceived as simultaneous changes if they occur within relatively short time windows, as previously shown in existing literature (Ahlfeldt, 2013; Gibbons & Machin, 2005).[6] Table 6.1 shows all the individual changes in the network and the resulting temporal stages defined for the MPTN between 2010 and 2019.

Demographics

The demographic data employed to account for the individual components in accessibility measures come from the 2010 Population Census collected by INEGI and were accessed through the Census Information Consultation System (*Sistema de Consulta de Información Censal*, SCINCE) (INEGI, 2012). These data were chosen because the level of disaggregation is rich, namely the city block level (see Section 5.2 in Chapter 5 for more details about the characteristics of this spatial unit). In addition, the

[6] Both Ahlfeldt (2013) and Gibbons & Machin (2005) studied the overall changes in London's public transport network following three extensions which opened in a span of six months, namely: the Jubilee Underground (Stage 1) in May 1999; the Jubilee Underground (Stage 2) in November 1999; and the Docklands Light Railway (DLR) in November 1999.

Agency	Corridor	From station	To station	Opening date	Previous change*	GTFS source
Stage 7: 2018 to 2019						
MB	7	Indio Verdes	Campo Marte	2018-03-01	26	2019-01-09
Stage 6: 2016 to 2018						
MB	6	El Rosario	Martín Carrera	2016-01-01	12	2017-09-02
Stage 5: 2015 to 2016						
MXB	2	La Quebrada	Las Americas	2015-01-01	14	2015-01-28
Stage 4: 2013 to 2015						
MB	5	Río de los Remedios	San Lázaro	2013-11-01	6	2014-12-30
MXB	3	Chimalhuacán	Pantitlán	2013-05-01	7	2014-12-30
Stage 3: 2012 to 2013						
Metro	12	Mixcoac	Tlahuac	2012-10-01	6	2013-10-21
MB	4	Buenavista	San Lazaro Aeropuerto	2012-04-01	14	2013-10-21
Stage 2: 2010 to 2012						
MB	3	Tenacayuca	Etiopia	2011-02-01	4	2013-10-21
MXB	1	Ciudad Azteca	Ojo de Agua	2010-10-01	10	2013-10-21
Stage 1: Before 2010						
MB	2	Tepalcates	Tacubaya	2009-12-01	NA	2013-10-21

Note:
* Expressed in months relative to the previous modification of the main public transport network.

Table 6.1: Main public transport network extensions between 2010 and 2019 in Greater Mexico City grouped by temporal stage.

next census was not conducted until 2020 and the results published in early 2021 are aggregated at a higher level than the city block (e.g. census tract). Using the data from 2010 assumes that the changes in the composition of the population in the period studied are not substantially different enough to undermine the results of the proposed accessibility measures. This is a limitation imposed by the sources of information.

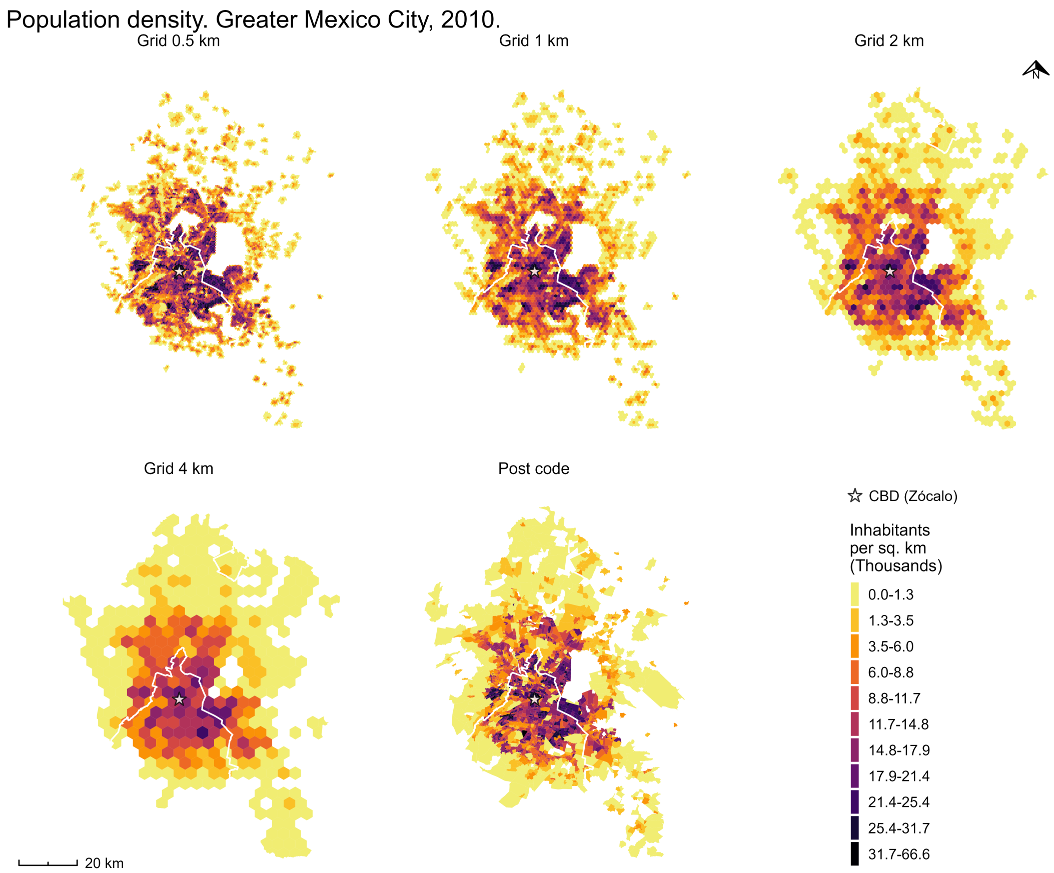
The key fields of this source include the total population, number of dwellings and the average level of education. The latter refers to the number of completed years of education in the 'National Education System' by individuals who are 15 or older.

The count data included in the 2010 Population Census (e.g. total population) originally aggregated at the block level (referred to as *source areas*) were transferred to the SAS (referred to as *target areas*) using the *areal weighting method* (Saporito, Chavers, Nixon, & McQuiddy, 2007). Mean and measures of centrality (i.e. average education level) variables were assigned using the *population weighting method* (both addressed in Saporito et al. (2007)). These methods were required because the nature of the target measures is different, in that one describes the total count and other an aggregated measure describing the population in the source areas.

The former method reassigns the data according to the proportional overlay of the source and target areas, assuming a uniform distribution of the source areas. For instance, if a quarter of a source area lies within a target area, then 25% of the source's population is assigned to the latter. Given that the post code polygons do not cover the totality of the territory of GMC (as detailed in Section 5.2 in Chapter 5), the following considerations were taken into account:

1. if a source area overlaps only one target area by 70% (and the other 30% of the surface does not overlap any other target area), the total population is assigned to the intersected target area;

2. on the other end, if the sum of the source area overlaps one target area only by less than 5% or 60 m^2, the source is considered to be out of the relevant post code coverage and is therefore excluded. The latter criteria are applicable for eight blocks only which contain a total of 767 inhabitants. Figure 6.3 shows the population density aggregated at the different SAS for the year 2010.



Source: The author based on Censo de Población y Vivienda 2010 (INEGI, n.d.)

Figure 6.3: Population density aggregated at various spatial analytical schemes. Greater Mexico City, 2010.

The second approach, the population weighting method, takes into consideration the proportional overlay of the areas but also the number of inhabitants and their type.

This is one mock example where there are two source areas overlapping a single target area. One of the source areas is transferring ten inhabitants with an average education level of 6 years to the target area and the second is transferring twenty inhabitants with an average education level of 8 years. In this example, the median education level is 8 years. One limitation is that this approach assumes homogeneity in the characteristics of the population in the source areas. However, because information is originally aggregated at low-scale spatial units (i.e. the block level) biases can be expected to be low, according to the literature. Figure 6.4 shows the level of education aggregated at the various SASs as the median number of years in education by quintile. In this case, the spatial patterns are fairly consistent across the different units of aggregation. For instance, it can be seen that the south-west quadrant is mostly populated by high-educated residents. In the meantime, low-educated residents locate mostly in peripheral areas, especially to the south-east and north.

Source: The author based on Censo de Población y Vivienda 2010 (INEGI, n.d.)

Figure 6.4: Level of education aggregated at various spatial analytical schemes. Greater Mexico City, 2010.

Employment opportunities

The main source of information accounting for employment opportunities, including their location and characteristics, comes from official data collected by the 2014 Economic Census (*Censos Economicos 2014*) (INEGI, 2015a). The information was accessed from the INEGI upon request via the *Microdata lab and remote processing* (*Laboratorio de microdatos y procesamiento remoto*) service (https://www.inegi.org.mx/microdatos/). The application was submitted on 21st February 2020. The data were manually downloaded via a temporary private URL. The key information considered includes the following fields:

1. The number of establishments: This definition corresponds to that of the *North American Classification System* (NAICS) and usually refers to a single physically fixed location producing goods or offering some type of service;[7]
2. Employed people:[8] These are not strictly necessarily formal employment positions but include everyone associated with an establishment performing an activity. This includes volunteers or family members, for example. Henceforth, this is referred to simply as 'employment.'
3. Total gross remuneration paid.

Although the 2014 Economic Census was originally collected at the establishment level, the data accessed were aggregated at the city block level. In some cases, the information was aggregated at a higher geographic level because of the need to maintain confidentiality; namely, 92% of the employment in GMC are referenced at the block level and the remaining 8% is aggregated as follows: ≈ 70% at the census tract; ≈ 1% at the rural locality, and; ≈ 30% at the municipal level.[9]

The figures referenced in geographic areas other than the block level were reallocated to the city blocks in GMC based on a proportional estimate which adjusts the figures according the spatial distribution and composition of the 2015 DENUE (INEGI, 2015c).[10] The DENUE is an open-access spatially-rich data set since it is organized at the point-level of each individual establishment, as described in Subsection 5.5 in Chapter 5. The details of the re-allocation process are presented in Appendix D.

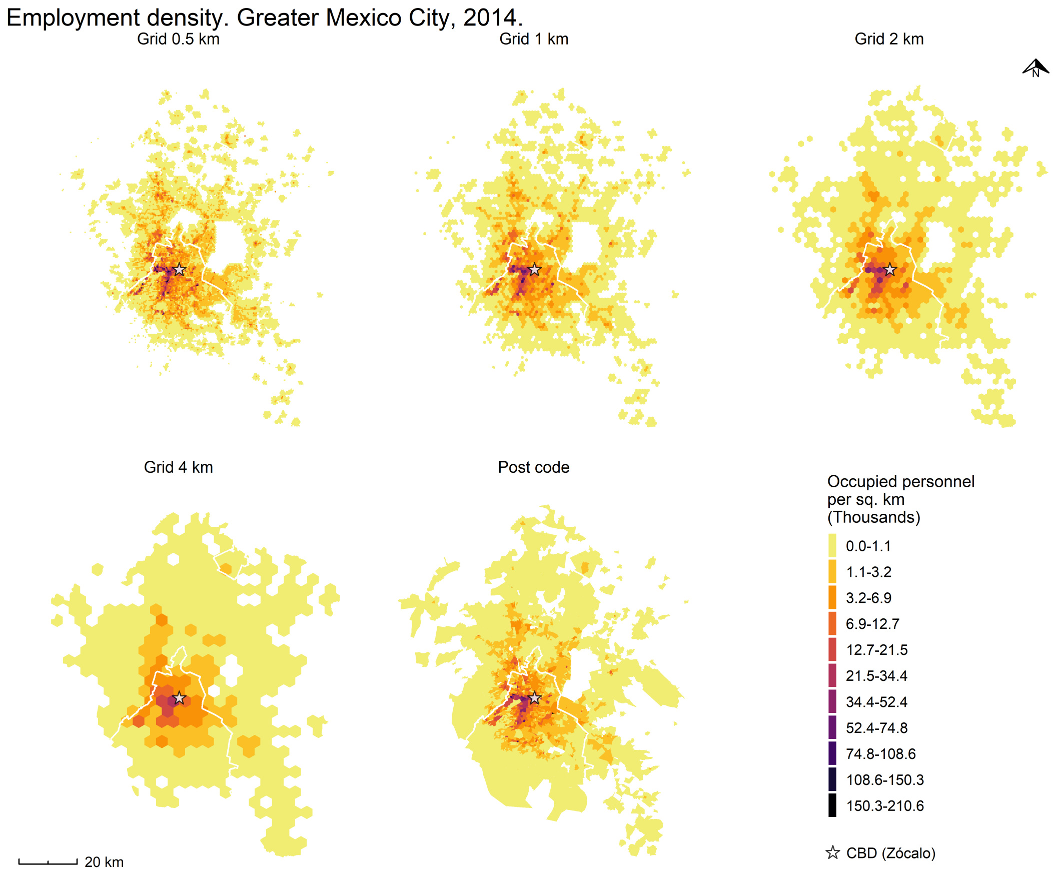
The use of the economic activity for a fixed point in time (i.e. a middle-point in time circa 2014) assumes that there are no substantial changes in the spatial structure of the labour market and its composition. This may be partially supported in that the

[7] More details are available in the following technical note: https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825075330.
[8] The original term in the source is "personal ocupado."
[9] The sum adds up to 101% as a result of rounding up.
[10] Downloaded on the April the 14th, 2020, through the public URL available in https://www.inegi.org.mx/app/descarga/default.html.

difference in total employment is -7% in 2008 and 21% in 2018, compared to the data collected for the 2014 Economic Census.[11]

This approach to reassign the data from city blocks (i.e. source areas) to SASs (target areas) is the same as the approach employed for the demographic data (described in the previous subsection), according to the areal weighting method. Figure 6.5 presents the employment density in GMC aggregated at the various SASs. The spatial distribution of employment looks less dispersed than the population. A general overview reveals that the areas with higher concentration of employment are located to the south-west of the CBD. A closer examination shows that the high-resolution schemes (e.g. 0.5 km or 1 km) reflect two types of agglomeration patterns, namely (1) centralities or nodes and (2) corridors. For instance, the nodes include *Polanco*, *Cuauhtémoc*, *Santa Fé* in a first order, and *Tlanepantla de Baz* or *Central de Abasto CDMX* in a second order. The corridors include *Av. Insurgentes* in a first-rank hierarchy and *Periferico Sur*, *Carretera México-Pachuca*, or *Av. López Portillo*. Meanwhile, low resolution schemes (e.g. 2 km or 4 km) denote an structure similar to a concentric-like pattern. The post code, however, reveals a mix of both. Here, some employment nodes are easy to identify, while some corridors are still salient.

[11]The figures include the total employment in all the municipalities in the state of Mexico and the State of Mexico City.



Source: The author based on Censos Economicos 2014 (INEGI, n.d.).

Figure 6.5: Employment density aggregated at various spatial analytical schemes. Greater Mexico City, 2010.

The average income was computed based on the total number of employment and the total gross remuneration paid allocated in the target areas. Although technically not all employed people are remunerated in reality (according to the statistical term used in the economic census), this is considered to be an indicator of the overall quality of jobs in a particular location. This is supported by consideration of the local context where informal employment is common, i.e. the contribution of the informal economy to the national GDP in the period studied was between 23% and 24% (INEGI, 2021).

6.2.4 Public transport travel time modelling

A series of travel time matrices has been calculated for each origin to all possible destinations for each of the SASs using a combination of walking and the public transport modes within the main public transport network. The origin-destination (OD) points j and k, respectively, are represented by the population weighted centroids of each spatial analytical unit that contains some economic activity (at least one economic unit according to the 2014 Economic Census) or demographic data (according to the 2010 Population Census). The total travel time is restricted to a maximum of 180 minutes based on the observed travel patterns in the 2017 Travel Survey (i.e. Figure 4.2).

The travel time for journeys using the MPTN are modelled according to a door-to-door approach using OpenTripPlanner (OTP) version '1.4.0' (OTP, n.d.),[12] a multi-modal open-source routing software platform. This considers the following steps in a journey:

1. Access time, i.e. walking from the point of origin to a public transport station;
2. Waiting time at the stop;
3. Time on board on vehicle to the next stop;
4. Transfer time if required, including waiting time and walking to the subsequent (second or third) station, and;
5. Egress time, i.e. walking from the last public transport station to the final destination.

A series of all-to-all travel time matrices were estimated for each of the SASs proposed for the seven temporal stages of the main public transport network defined above in OTP via the implementation of an open-access routine written in the `Python` programming language for parallel computing (Pereira et al., 2019). This uses, as primary inputs, the reconstructed GTFS files and the road and pedestrian networks. In addition, because it has been found that the specific time of departure at different times in the day can affect the estimated travel time in public transport because of the variability in the service (Conway et al., 2018; Owen & Levinson, 2015), eight random departure times were estimated for a typical business day within a time-window from 10 a.m. to 2 p.m. for every 30 minutes. From the different travel time estimates produced according to the different times of departure, the median travel time was used as a representative estimate in constructing the accessibility models, following Conway, Byrd, & van der Linden (2017).

The approach referred to above is computationally expensive, especially for high resolution SASs as the number of potential routes in a system grows exponentially for each additional origin considered. For instance, the theoretical size of full travel time matrix for the highest resolution SAS (0.5 km) results in 9.9 billion routes (considering that this contains 13,297 valid OD points, i.e. $n \times n \times 7$ *temporal stages* $\times$ 8 *departure times*). In reality, not all combinations are available as not all destinations can be accessed by public transport from every origin given the availability of infrastructure and the restriction of journeys equal to, or shorter than, 180 minutes.

[12] https://www.opentripplanner.org/

Working at this or other high resolutions represent technical challenges. Considering the above, an estimated total of 2.1 billion routes are required to model all the temporal stages of the main public transport network for the different spatial analytical frameworks. This procedure is also time-consuming even when taking advantage of parallel computing (the simultaneous processing of information). For instance, the effective running time required was 40.4 hours using an 'ordinary' 4-core processor CPU with a base clock speed of 3.4 GHz. Similarly, the limited capacity of virtual memory to 16 GB was often overtaken by the size of the matrices. This required the design of a programming strategy to break down the queries into smaller tasks which were implemented in an automated repetitive routine executed from `R` (R Core Team, 2021).

Later, the matrices were post-processed to merge the travel time estimates originally organised in multiple files and to calculate the median travel time for each OD pair from the different times of departure estimated in their respective scenarios. In addition, the total walking distance when using public transport (including all journey stages) was limited to 3 km and walking only to ≈ 4.8 km (or equivalently 60 minutes, according to OTP's default walking speed to 3 mph).

The internal-zone travel-times were defined based on the surface area of each polygon associated to an origin as follows: $d_{j=k} = \frac{1}{2}\sqrt{\frac{area_j}{\pi}}$, which has been widely used in this type of analysis (Frost & Spence, 1995; Stepniak & Rosik, 2015). The metric distance was transformed to travel time considering the average speed corresponding to modelled journeys shorter than 3 km.

6.2.5 Spatial decaying function and calibration of accessibility parameters

For the purposes of this chapter, the spatial decaying function and the respective parameters used for the accessibility measures have been chosen according to estimates produced by the trip distribution model, as suggested in Handy & Niemeier (1997). This model is frequently used in the classic four-step transportation modelling process for forecasting purposes (Ortúzar & Willumsen, 2011). However, here it is estimated for transport-behavioural purposes, i.e. to reveal the way in which commuters in GMC discount the attractiveness of employment opportunities as a function of travel time. Therefore, a series of alternative specifications have been estimated according to a doubly constrained gravity model framework.

The generic doubly-constrained specification is shown in Eq. (6.3) (Wilson, 1971). The notation presented here differs from the usual in literature in order to maintain consistency with the rest of the sections where origins have been denoted by

j and destinations by k.

$$T_{jk} = A_j O_j B_k D_k f(d_{jk}) \tag{6.3}$$

where

$$A_j = \left[\sum B_i D_i f(d_{jk})\right]^{-1} \tag{6.4}$$

and

$$B_k = \left[\sum A_k O_k f(d_{jk})\right]^{-1} \tag{6.5}$$

Here, T_{jk} is the size of the flow represented by the number of commuters between an origin j and a destination k. O_j is the origin marginal constraint (which restricts the number of journeys originated to those observed), and D_k is the constraint at destinations (restricting the number of attracted journeys to those observed). $f(d_{jk})$ is a spatial deterrence or decaying function between j and k. A_j and B_k are balancing factors which are estimated in an iterative routine (Batty, 1976; Dennett, 2012).

To avoid confusion between the balancing factor for origins (A_j) and accessibility (A_{PTj} or A_{CARj}), it should be noted that the latter specifies PT or CAR in the sub-index which refers to the corresponding mode for which accessibility is calculated.

These analyses draw on empirical data collected by the 2017 Travel Survey (INEGI, 2018). For this purpose, only journeys to work by public transport and walking in a business day are considered (i.e. Mon-Fri). Unobserved flows are assumed to be 0, producing a full square matrix (N = 37,636). The 2017 Travel Survey data is aggregated at transport analysis zones (TAZ). Given this condition of the source of information, the TAZ is the spatial analytical scheme used to estimate the SIM models.

Distance decay functions

The following two decaying functions have been tested in the context of the previously introduced gravity model (Oshan, 2021):

$$\text{Inverse Power: } f(d_{jk}) = d_{jk}^{-\beta} \tag{6.6}$$

$$\text{Negative Exponential: } f(d_{jk}) = e^{(-\beta d_{jk})} \tag{6.7}$$

In both expressions, β represents an estimated distance deterrence parameter. Furthermore, a modified specification proposed by Thorsen & Gitlesen (1998) is considered. This includes a parameter μ which affects only the diagonal elements of the OD matrix, i.e. T_{jk}. This is intended to capture an add-up cost for inter-zone journeys or analogously a benefit of residing and working within the same zone. The term is given by the following expression:

$$\mu\delta_{jk} \tag{6.8}$$

where Kronecker delta takes the following value:

$$\delta_{ij} \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k \end{cases} \tag{6.9}$$

Parameter estimates

To evaluate the different alternatives Knudsen & Fotheringham (1986) suggests that the standardized root mean square error (SRMSE) is the most accurate measure. This is given by Eq. (6.10) as following:

$$\text{SRMSE} = \frac{\sqrt{\frac{\Sigma_j \Sigma_k (\hat{T}_{jk} - T_{jk})^2}{JK}}}{\frac{T}{JK}} \tag{6.10}$$

The simultaneous calibration of the model parameters β and μ is achieved via optimization by minimizing the SRMSE using the *Nelder-Mead's simplex algorithm* (Nelder & Mead, 1965) as implemented in base `R` (R Core Team, 2021). The iterative routine to estimate the balancing factors A_j and B_k follows the routine proposed in Dennett (2012).

The estimates for the alternative doubly constrained gravity models are summarised in Table 6.2 where it is shown that :

- The inverse power function produces better results compared to the negative exponential when μ is set equal to 0, as shown by the estimates of Model 1 (M1) and Model 2 (M2);
- The inclusion of the $\mu\delta_{jk}$ term substantially improves the performance of both decaying functions, as shown in Model 3 (M3) and Model 4 (M4) compared to their counterparts M1 and M2, respectively. It is also noticeable that M3 and M3 produce lower β values;
- When the $\mu\delta_{jk}$ term is allowed to vary, the negative exponential function outperforms the inverse power, as shown in M3 and M4.

This preliminary analysis shows that the use of the within-zone parameter μ improves the estimates. Specifically, the negative exponential function provided more accurate estimates compared to the inverse power μ. Therefore, the selected specification is the modified negative exponential including μ.

The parameters employed to estimate accessibility are shown in Table 6.3. These have been selected according to the model parameter and decaying function, which is closer to observed flows in the region, namely M4. The parameter beta for the accessibility measures by private car is set equal to 0.085, drawing on previous empirical

	Doubly constrained gravity model			
	M1	M2	M3	M4
$\hat{\beta}$	-3.274	-0.07	-2.276	-0.044
$\hat{\mu}$	0	0	0.443	1.654
Decaying function	Inv. power	Neg. exponential	Inv. power	Neg. exponential
Observations	37 636	37 636	37 636	37 636
SRMSE	2.544	2.713	1.984	1.764
Log-likelihood	-2 380 345	-2 665 957	-2 335 166	-2 064 904

Table 6.2: Trip distribution model estimates

Accessibility model index	Specification (Eq.)	β	μ	Mode
1	6.1	0.044	1.654	Public transport
2	6.2	0.044	1.654	Public transport
3	6.1	0.031	1.662	Public transport
4	6.2	0.031	1.662	Public transport

Table 6.3: Parameters employed in accessibility models

studies (Osland & Thorsen, 2008). This value is in line with empirical data in the region collected by the 2017 Travel Survey, which shows that travellers who use public transport are more tolerant to long distances compared to commuters who drive. In addition, this decision simplifies the complexity in the methods involved for the analyses, given that this measure is intended to work as an overall benchmark.

6.3 Exploratory analysis

6.3.1 Overview of accessibility to employment

Table 6.4 presents the descriptive statistics of the accessibility to employment estimates for all MPTN temporal stages and for each SAS by type of measure, i.e. *all* and *matching* opportunities; the latter according to Eq. (6.1) and Eq. (6.2), respectively.

Table 6.4: Summary statistics of accessibility models.

	Stage	Grid 0.5 km (N=13 297)	Grid 1 km (N=3 972)	Grid 2 km (N=1 264)	Grid 4 km (N=413)	Post code (N=2 377)
All						
	Before 2010	35.5 (84.3)	33.0 (84.4)	33.3 (93.2)	38.9 (115.5)	66.9 (108.2)
		[1.0, 15.7]	[0.8, 14.9]	[0.7, 17.4]	[0.6, 21.4]	[4.8, 73.0]
	2010 to 2012	36.5 (84.9)	34.0 (85.4)	34.0 (94.1)	40.1 (117.1)	68.6 (108.9)
		[1.0, 17.2]	[0.8, 15.9]	[0.7, 17.9]	[0.6, 21.5]	[4.9, 78.2]

(Continued on next page...)

Table 6.4: Summary statistics of accessibility models. *(continued)*

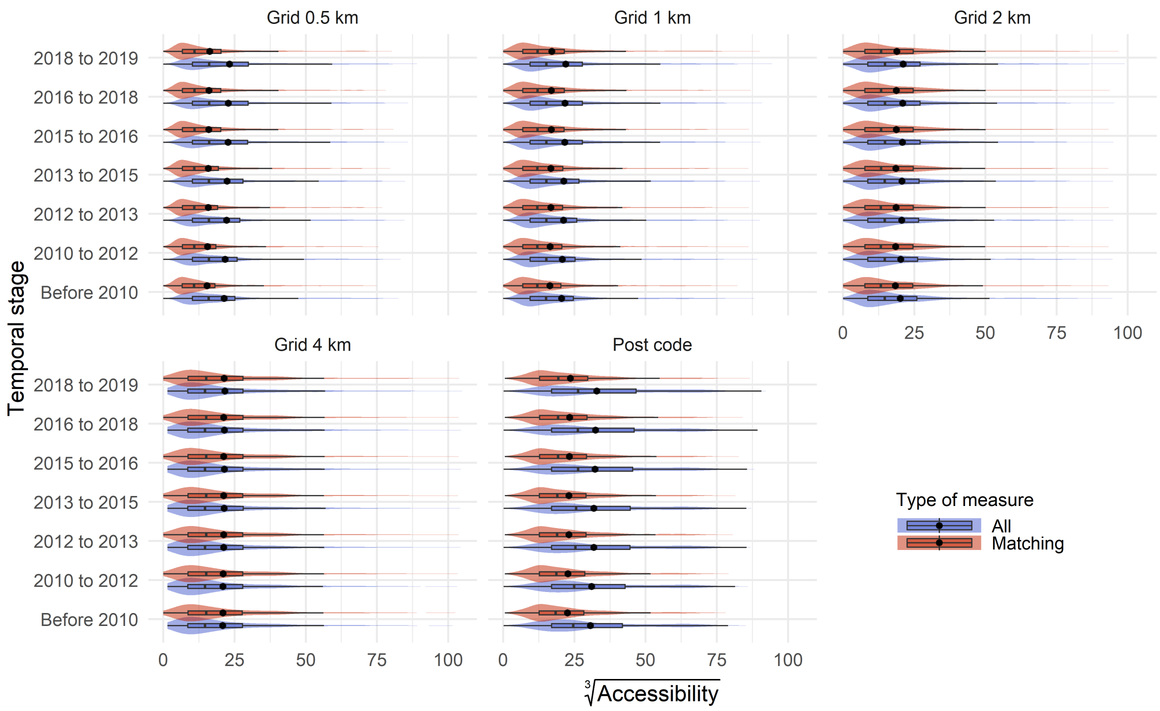
Type	Stage	Grid 0.5 km (N=13 297)	Grid 1 km (N=3 972)	Grid 2 km (N=1 264)	Grid 4 km (N=413)	Post code (N=2 377)
	2012 to 2013	39.7 (89.6)	36.7 (89.5)	36.3 (98.1)	42.3 (121.5)	73.5 (113.7)
		[1.0, 19.2]	[0.8, 17.1]	[0.7, 18.6]	[0.6, 21.7]	[4.9, 87.8]
	2013 to 2015	39.2 (86.4)	36.4 (86.6)	36.1 (96.0)	42.7 (120.5)	72.1 (110.2)
		[1.0, 21.7]	[0.8, 18.7]	[0.7, 19.1]	[0.6, 22.0]	[4.9, 88.6]
	2015 to 2016	40.8 (88.5)	37.7 (88.7)	37.0 (97.5)	43.1 (121.7)	74.3 (112.1)
		[1.0, 25.8]	[0.8, 21.3]	[0.7, 19.7]	[0.6, 21.7]	[4.9, 93.5]
	2016 to 2018	41.4 (89.5)	38.3 (89.8)	37.4 (98.2)	43.4 (121.8)	75.7 (113.5)
		[1.0, 26.2]	[0.8, 21.4]	[0.7, 19.7]	[0.6, 21.7]	[4.9, 97.2]
	2018 to 2019	45.9 (102.2)	41.9 (100.8)	40.3 (108.3)	46.1 (131.0)	81.4 (124.2)
		[1.0, 26.5]	[0.8, 21.4]	[0.7, 19.7]	[0.6, 21.7]	[4.9, 101.5]
Matching						
	Before 2010	17.5 (52.7)	20.3 (62.3)	27.4 (83.2)	39.2 (115.8)	30.0 (65.3)
		[0.3, 5.9]	[0.3, 8.3]	[0.4, 14.6]	[0.6, 21.2]	[2.0, 22.7]
	2010 to 2012	17.8 (52.9)	20.7 (62.8)	27.8 (83.8)	40.2 (117.3)	30.4 (65.5)
		[0.3, 6.2]	[0.3, 8.7]	[0.4, 14.8]	[0.6, 21.5]	[2.0, 23.5]
	2012 to 2013	19.0 (55.6)	21.8 (65.3)	29.0 (86.9)	41.5 (120.9)	32.1 (68.8)
		[0.3, 6.8]	[0.3, 9.1]	[0.4, 14.9]	[0.6, 21.5]	[2.0, 24.6]
	2013 to 2015	18.5 (53.7)	21.3 (63.3)	28.6 (85.2)	41.2 (119.4)	31.4 (66.8)
		[0.3, 7.1]	[0.3, 9.1]	[0.4, 14.8]	[0.6, 21.5]	[2.0, 24.6]
	2015 to 2016	19.0 (54.8)	21.8 (64.5)	29.0 (86.0)	41.5 (120.5)	31.9 (67.5)
		[0.3, 8.1]	[0.3, 9.8]	[0.4, 15.0]	[0.6, 21.7]	[2.1, 25.0]
	2016 to 2018	19.3 (55.3)	22.0 (65.1)	29.2 (86.4)	41.5 (120.4)	32.2 (68.1)
		[0.3, 8.2]	[0.3, 9.8]	[0.4, 15.0]	[0.6, 21.7]	[2.1, 25.5]
	2018 to 2019	21.4 (63.2)	24.0 (72.6)	31.1 (93.8)	43.8 (128.0)	34.3 (73.5)
		[0.3, 8.2]	[0.3, 9.8]	[0.4, 15.0]	[0.6, 21.7]	[2.1, 26.0]

Note:
Descriptive statistics: Mean (SD) [1st-Q, 3rd-Q]. Figures shown in thousands.

The top panel in Table 6.4 presents a summary of the measure that considers *all* employment opportunities. Initially, it can be seen that the mean does not display a clear systematic variation according to the scale of the SAS. For instance, the mean of Stage 1 ('Before 2010') in Grid 0.5 km is approximately 36 and decreases to approximately 33 in both the 1 km and 2 km grids. Meanwhile, the average for the lowest resolution is 39. By contrast, the dispersion of accessibility does display a systematic variability in relation to the scale of the SAS; namely, the corresponding relative standard deviation (computed as the standard deviation over the mean) is 2.4, 2.5, 2.8, 2.9, for the 0.5 km, 1 km, 2 km, and 4 km grid, respectively (this is not shown in the table). A similar pattern is observed for the subsequent temporal stages. In addition, it can be seen that the mean is higher than the respective third quartile for all hexagonal grids in all temporal stages, which suggests a skewed distribution. It should be noted that the mean accessibility for post code zones falls within the first and third quartile. With regard to the temporal stages, all SASs reflect upward trends in the period studied. The increase on the average accessibility ranges between an additional 29% in the 0.5 km Grid and 19% in the 4 km Grid when the latest stage (i.e. 2018 to 2019) is compared to the earliest ('Before 2010'). Section 5.3.3 presents the temporal trends in greater detail.

The second panel in Table 6.4 shows the accessibility estimates for measures considering *matching* employment opportunities. In contrast to the previous type of measure, the mean values do display a systematic pattern according to the scale of the SAS, namely the larger the scale of the SAS, the higher the mean accessibility. For instance, in the first temporal stage the mean gradually increases from 18 in the 0.5 km Grid to 39 in the 4 km Grid. In addition, the relative standard deviation displays a different pattern compared to the alternative accessibility measure. Here, this figure remains stable (around 3.0) for all hexagonal grids. With regard to the type of distribution, it can be seen that the mean is substantially larger than the third quartile in all SASs (including the post code scheme). This implies a strong skewed distribution. The figures according to temporal stages suggest upward trends, coinciding with the previous measure. However, these trends are not as marked as they are for the previous accessibility specification. Specifically, it can be seen that the relative increase in the mean ranges from an additional 12% to 22% for the 4 km Grid and 0.5 km grid, respectively.

The series of plots in Figure 6.6 illustrates the distribution of the accessibility estimates by SAS and temporal stages, directly comparing the type of employment opportunity considered. The horizontal axis shows accessibility (in the cube root scale) and the vertical axis shows the temporal stage. In addition to the descriptive measures presented above, there are several aspects worth noting. First, it can be seen that the measure considering matching opportunities tends to be more skewed than the one considering all types of employment. This is expected as the former is stricter than the latter and will therefore assign a lower accessibility index to a larger portion of SAS units. Secondly, skewness is more pronounced in high resolution grids than it is in low resolution grids. The post code scheme generates the less skewed distribution for both types of accessibility. Thirdly, the difference between the alternative measures used is more obvious in low resolution grids and the post code scheme than it is in high resolution schemes (e.g. 2 km or 4 km grid). Fourthly, the measure considering all type of employment displays changes according to the MPTN temporal stage more clearly than the one including only matching opportunities. Furthermore, this becomes evident in high resolution grids and post code schemes.



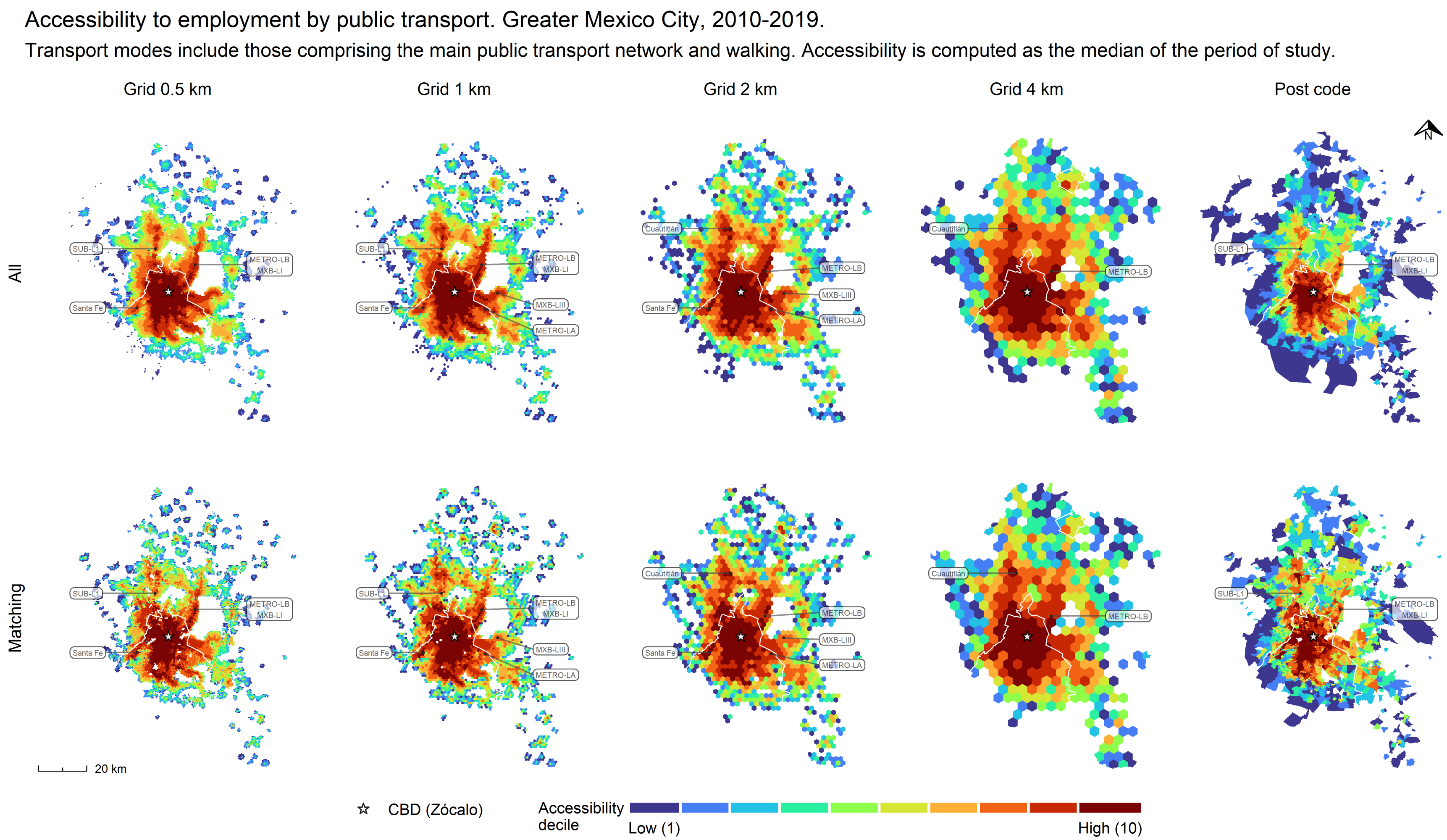
Source: The author based on own calculations.

Figure 6.6: Distribution of accessibility measure by temporal stage and SAS.

The following sub-section provides an in-depth discussion in relation to spatial aggregation effects and a subsequent discussion about temporal variability of accessibility.

6.3.2 Spatial aggregation of accessibility

The series of maps in Figure 6.7 show the modelled accessibility to employment enabled by the modes included in the MPTN in combination with walking aggregated at various SASs (columns) by type of specification (rows). The accessibility measure shown in the maps is summarised by the median of all the temporal stages and this is presented in decile breaks. This is designed to control possible temporal variations at this analytical stage and to focus on the spatial changes. The top row illustrates accessibility to employment considering all opportunities. Here, high resolution grids (e.g. 0.5 km or 1 km grids) clearly highlight the relevance of both central areas and radial public transport axes or corridors. This reflects an overall star-shaped structure. The post code SAS also partially reflects some of these patterns. Relevant examples of corridors connecting non-central areas include the services running to the north; namely, the regional rail Line 1 *Buenavista-Cuautitlán* (SUB-L1) (to the north-west with respect of the CBD) and the corridor comprised by the Metro line B (METRO-LB) in combination with the Mexibús BRT Line I (MXB-LI) to the north-east.



Source: The author based on own calculations, Censos Economicos 2014 (INEGI, n.d.), and Censo de Población y Vivienda 2010 (INEGI, n.d.).

Figure 6.7: Accessibility generated by the main public transport network aggregated at various spatial analytical schemes.

Although both of these public transport corridors tend to form continuous patterns along a path, there are some differences; namely the SUB-L1 generates accessibility benefits forming high-access islands around stations even at relatively distant locations from the CBD. Meanwhile, the north-east corridor introduces the most noticeable effects in the segment corresponding to the METRO-LB and diminishes as locations situate further away from the CBD as well as in those serviced by the MXB-L1 (to the north-end).

These differences result from a combination of factors. Arguably, the most relevant are the physical and operational characteristics of the distinct modes as well as land use characteristics. Specifically, at one end the SUB-L1 operates at a higher speed given the separation between stations, the high degree of segregation of the right of way (ROW-A), and the type of vehicles used. At the other end, the operational speed of the MXB-LI (a BRT system) is undermined by the high frequency of stops and the use of semi-segregated right of way (i.e. ROW-B). The METRO-LB is an intermediate case between the previous two. This line operates on a completely segregated ROW (A). Still, the separation between stops is shorter than it is in the SUB-L1. In addition, METRO-LB operates at a closer proximity to the CBD than both the SUB-L1 and MXB-L1. As a result, the METRO-LB expands the highest accessibility decile into a narrow corridor to the north-east.

Whilst high-resolution SASs accentuate the patterns of accessibility generated by the MPTN as in corridor-like shapes, the low-resolution patterns (e.g. 2 km or 4 km grid) dissolve many of these details into polygons or zones. At this scale, the previous star-shaped structure develops into what can be described as concentric bands or rings of accessibility. This is particularly noticeable in the 4 km Grid, where there is a clear, highest-decile accessibility zone around the CBD. This primary zone is surrounded by n gradual bands. Despite much of the structure identified in small-scale SASs being abstracted, some corridor-like patterns prevail even in the lowest-resolution SAS. Some examples are the METRO-LB as well as some sub-centres such as *Cuautitlán* in the north-west.

Although most of the spatial patterns noted earlier hold in both types of accessibility measures others differ (upper panel versus lower panel in Figure 6.7). One of the main distinctions is that the matching opportunities measure tends to emphasize land use patterns, attenuating the role of the MPTN. This is illustrated by the relevance of some areas which are not covered by the MPTN. For instance, *Santa Fe*, a sub-centre oriented to services in the south-west of GMC, appears in a secondary or tertiary level in the upper row maps. By contrast, this acquires a first order role when the characteristics of residents and employment are taken into account. This is chiefly the result of co-location of matching employment and residents in close proximity (as the resident/employment matches are enabled by walking only).

A second indication of the emphasis attributed to land use, rather than the role of the MPTN by the matching opportunities measure, is observed along transport corridors. Whilst transport corridors are regarded with high importance when all types of employment are included, these lose relevance in the matching opportunities models. Some examples are observed along the BRT MXB-LIII and the METRO-LA both running to the east, or the METRO-L12 servicing the south-east of GMC. In all of these cases, the level of accessibility benefits is mixed and discontinuous. This can be appreciated more clearly in the 0.5 km and 1 km grids than in the low resolution SASs. This difference reflects that, despite the residents adjacent to this transport infrastructure, employment opportunities which may be of relevance to them are still a considerable travel time distance away.

A third sign of the differentiated role of the MPTN captured by the alternative accessibility measures employed is the shape of the core accessibility area around the CBD; namely, this characteristic can be appreciated in the asymmetric distribution of the higher decile accessibility region around the CBD. Specifically, while the central highest accessibility region according to the measure, including all type of employment (upper row in Figure 6.7), is fairly circular around the CBD, in the alternative measure this tends to deviate to the south-west. This pattern resembles that of the highest level of education previously shown in Figure 6.4. This spatial trend can be observed in all the SASs.

A further aspect emerging from the aggregation of accessibility at various SASs is the degree of spatial heterogeneity. Table 6.5 presents the results of a series of spatial autocorrelation tests conducted for each SAS and the two alternative accessibility measures. According to the global Moran's I test, it can be seen that all MPTN accessibility estimates display spatial autocorrelation at a significant level (P-value <0.01). This is expected because, by conception, the level of accessibility in zone j depends on neighbouring areas k. Although the spatial correlation is high in all cases (>0.7), this systematically varies according to the SAS employed. For instance, the correlation coefficient gradually diminishes from 0.91 in the smallest grid to 0.68 in the largest (i.e. 4 km) for the type of measure 'All.' Here, the post code scheme is spatially correlated to a very similar level as the smallest grid (0.92). This implies that, although this SAS has fewer units than the high-resolution grids, the zoning produces spatially smooth accessibility values. The 'Matching' opportunity measure also displays a systematic variation in the strength of spatial correlation according to the scale of the SAS.

An additional quality explaining the role of MPTN in determining the level of accessibility is the degree of spatial heterogeneity produced by the alternative measures. Specifically, Table 6.5 shows that the strength of the spatial correlation is weaker when

	SAS				
Type	Grid 0.5 km	Grid 1km	Grid 2km	Grid 4km	Post code
All	0.912***	0.866***	0.789***	0.675***	0.919***
Matching	0.853***	0.793***	0.761***	0.688***	0.738***

Signif. Codes: ***: >0.01, **: >0.05, *: >0.1
SAS = Spatial analytical scheme.

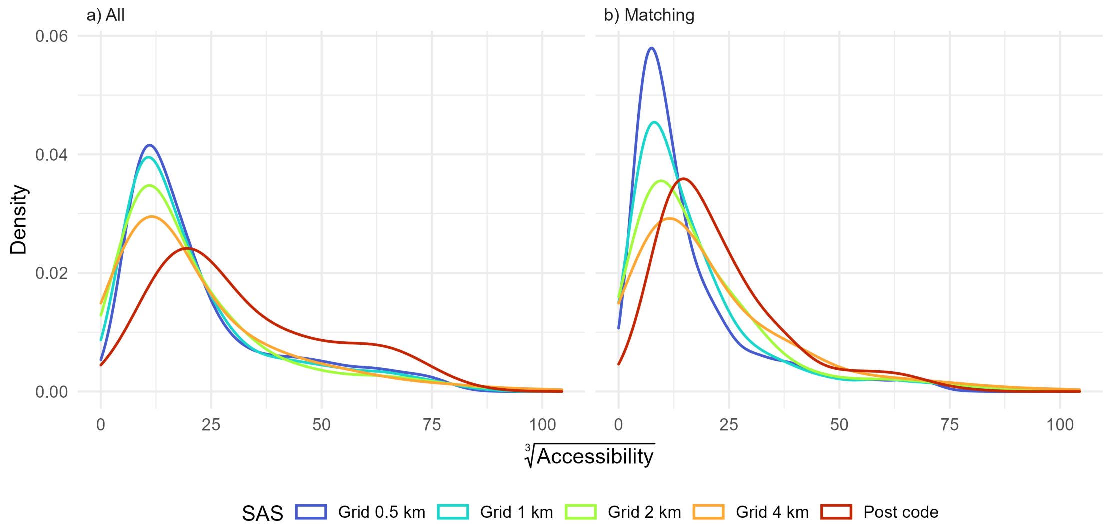
Table 6.5: Spatial correlation over various SAS according to Moran's I test.

only matching employment opportunities are considered rather than in the measure considering all (except for the Grid 4 km where it is very similar). This implies that the role of the MPTN is lessened in the former measure given that MPTN infrastructure is physically continuous in space.

Some MAUP effects can be observed in the distribution density of accessibility measures. Figure 6.8 shows the distribution of both accessibility for all and matching opportunities in the left-hand side panel (a) and right-hand side panel (b), respectively. The measures are aggregated in various SASs according to the median of all the temporal stages. Panel a) shows that the overall distribution of uniform hexagonal is similar between them. This is particularly noticeable between the 0.5 km and 1 km grids. Yet, a key difference is that the higher the resolution, the larger the relative number of observations containing low accessibility values (between 0 to 25 in the cube root scale). In other words, the density of observations with low accessibility gradually diminishes as the scale of the SAS is increased. In addition, the distribution of post code units substantially differs from the previous distribution. Here, the proportional number of observations including low accessibility values is considerably smaller than it is in any of the other SASs. Furthermore, the portion of units having middle and high accessibility levels (between 25 and 75) is consistently larger for post codes. The latter difference is a result of the zoning effects referred to in the MAUP literature. Specifically, zones on the periphery tend to acquire low accessibility values while central zones have generally high values. At the same time, the scale of post code zones tends to be larger on the periphery than near to the CBD. Therefore, there is a relatively smaller number of units with low accessibility values in the post code SAS than in the uniform grid SASs (where space is partitioned in equal bins). This observation is confirmed by a significant correlation coefficient between accessibility and the surface area in the post code scheme (r = -0.18, p = <0.001, the full results are shown in the Appendix E).[13] This effect produced by zoning influences the average accessibility and explains the considerable high figure for the postcode SAS compared to grids, as shown

[13]Although the correlation coefficient is also significant for other SASs, the magnitude of the post code coefficient according to the 'All' specification is considerably higher than other SASs.

in Table 6.4.



Source: The author based on won calculations, Censos Economicos 2014 (INEGI, n.d.), and Censo de Población y vivienda 2010 (INEGI, n.d.).

Figure 6.8: Distribution density of accessibility measures over various SAS.

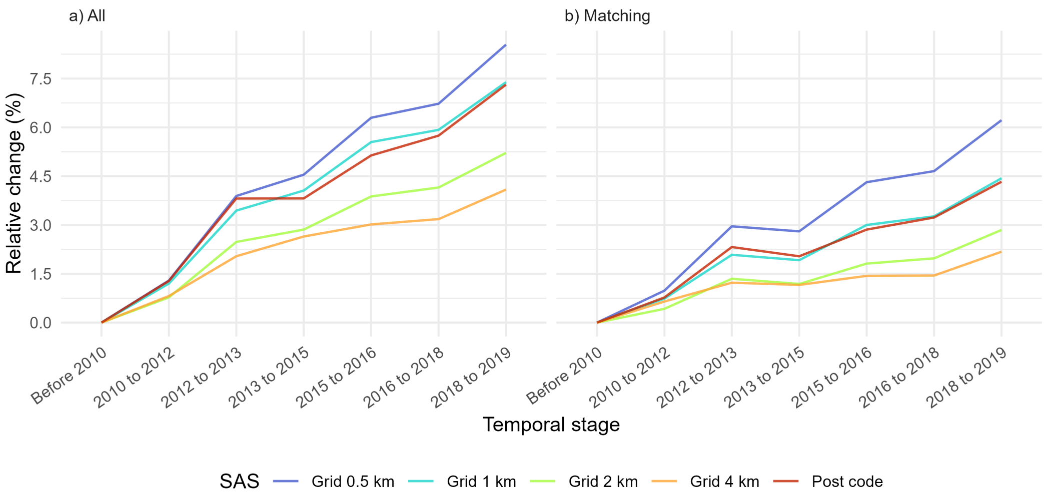
Panel b) in Figure 6.8 shows the distribution of the accessibility measure that considers matching opportunities. Here, the differences between the SASs resulting from scale effects referred to above are emphasized, while those related to zoning are attenuated. With regard to the first aspect, it can be seen that the relative density of zones with low accessibility values (0-25) is considerably higher in the 0.5 km and 1 km grid than it is in Panel a). This is affected by at least two interrelated factors: (1) the strict character of the former measures which ignores nearby opportunities if these are assumed to be unattractive for residents at the origin, and; (2) the consequent stronger reflection of the spatial distribution of employment, which tends to concentrate in small areas (as illustrated in Figure 6.5).

Thus, this results in a relatively large number of observations with low accessibility values, as illustrated in the density plots. These aspects also relate to the weakened effect of zoning. Specifically, the distribution for accessibility aggregated in post codes In Panel b) is closer to the uniform grids than it is in Panel a). The overall increased skewness also affects the mean values lowering the average in the summary statistics shown in Table 6.4.

6.3.3 Temporal variability of accessibility

Figure 6.9 illustrates the relative change in mean accessibility over the different temporal stages of the MPTN by SASs taking Stage 1 ('Before 2010') as the reference. Panel

a) corresponds to the measure including all employment opportunities and Panel b) only matching opportunities. Overall, all the SASs and the alternative measures reflect upward trends over time. This can be expected as the extension of the public transport network should reduce journeys' travel time and consequently enable additional employment opportunities in some locations. There is also coincidence in the degree of the change reflected according to the SAS. For instance, the largest changes are displayed by the 0.5 km SAS and the smallest by the 4 km grid. Another agreement shown between Panel a) and Panel b) and the different SASs is that the largest changes are observed in temporal Stage 4 (2012 to 2015) and Stage 7 (2018 to 2019), with the introduction of the Metro-L12 and MB-L4 in the former and MB-L7 in the latter. The close similarities in the patterns shown between the post code scheme and the 1 km Grid are very noticeable in both panels.



Source: The author based on own calculations.

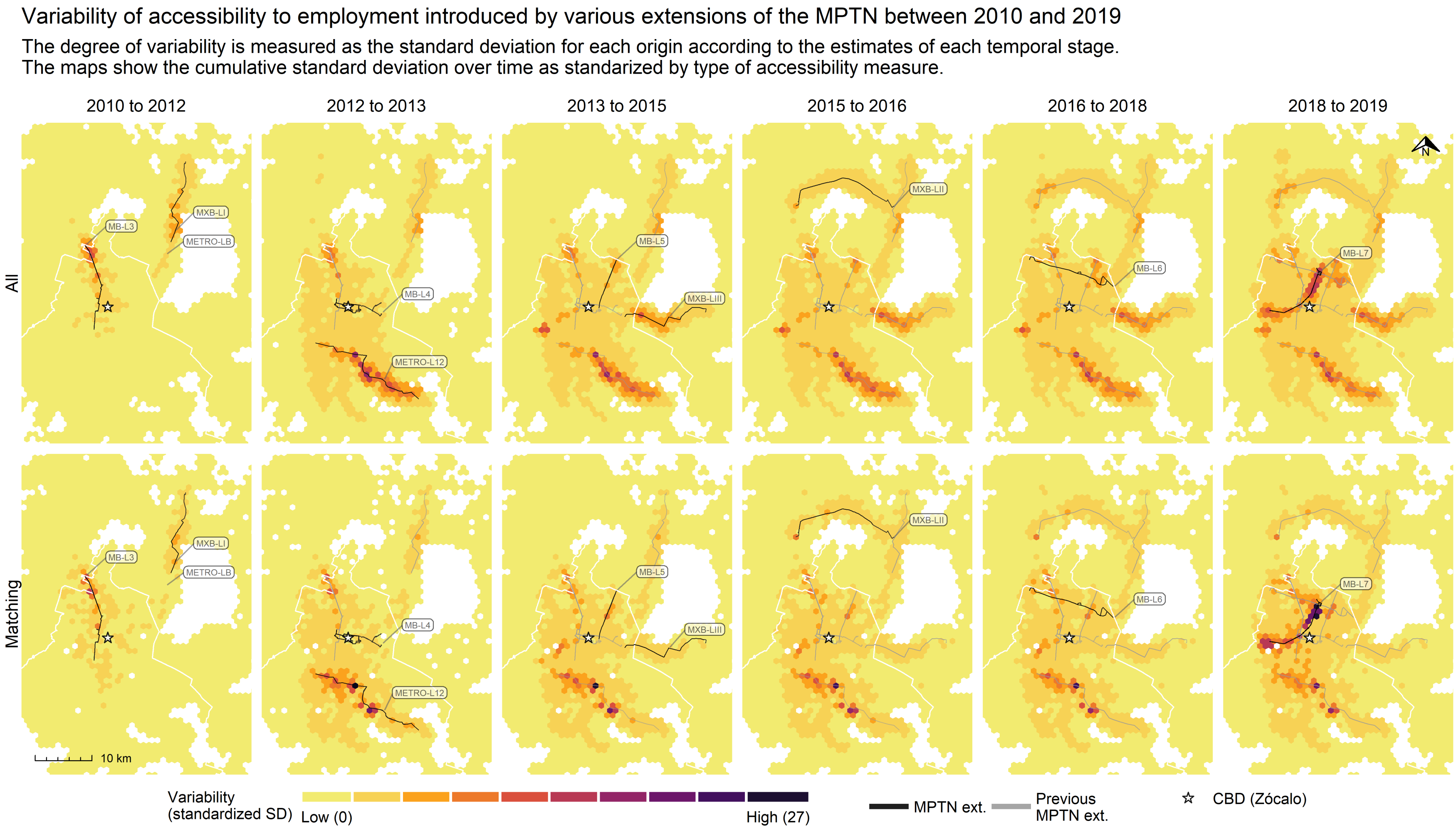
Figure 6.9: Relative change of mean accessibility to employment over various temporal stages (reference is 'Before 2010') according to all and matching employment opportunities measures.

There are some relevant particularities in the fluctuations shown in Figure 6.9.

First, the measure considering all opportunities is more sensitive to changes in the MPTN than the one including only matches. For instance, the additional accessibility in Stage 7 is approximately 7.5% in the 1 km grid and post code scheme in Panel a), while this is about 4.5% in Panel b) for both SASs. In addition, the variability captured by the 4 km in Panel b) is modest, being practically stagnated between Stage 3 and Stage 6. Secondly, there are some neutral and negative changes in Stage 4 (2013 to 2015). Those can be appreciated more clearly when only matching opportunities are considered. This pattern is counter to the general expectation. However, this can be explained by the temporary closure of some stations of the METRO-L12 in 2014 as a result of maintenance works. High resolution grids and the post code scheme are more sensitive in capturing these changes than other SASs regardless of the measures used. Thirdly, while the 0.5 km grid mirrors similar trends to those produced by the 1 km grid and the post code scheme when all opportunities are included, the 0.5 km grid substantially differs from these SASs when opportunities are constrained.

Figure 6.10 illustrates the spatial distribution of the changes in accessibility to employment introduced by the modifications in the MPTN according to the temporal stage and type of measure. Here, the variability is represented for each origin as the cumulative standard deviation of accessibility according to the estimate in each stage and it is shown as the standardized value for each type of measure in order to allow comparison.[14] The estimates are presented for a unique SAS (i.e. 0.5 km Grid, the highest resolution) to limit the effects of spatial aggregation at this analytical stage. From the maps presented, one of the most relevant aspects is that the benefits are not distributed homogeneously along the newly introduced public transport infrastructure. The magnitude and distribution of the patterns reflect several characteristics of the urban layout as well as those of the MPTN; for instance, distance to employment agglomerations (e.g. CBD), coverage of pre-exiting public transport services, transport mode, or the topology of the MPTN. These are discussed in detail in the following section.

[14] Stage 1 (Before 2010) is not shown as there is no data preceding it.



Source: The author based on own calculations.

Figure 6.10: Variability of accessibility to employment over various temporal stages. Greater Mexico City, 2010-2019.

The first column from left to right in Figure 6.10 clearly illustrates the incidence of some of the characteristics already highlighted. This shows the changes introduced by two different BRT corridors in Stage 2 ('2010 to 2012'), namely MXB-LI and MB-L3, respectively. In the first case, this infrastructure is connected to the rest of the network only via the METRO-LB (as shown in Figure 6.2) and the areas it serves are relatively distant to the CBD. The largest benefits are concentrated in units the closer they are to the CBD. By contrast, the changes introduced by the second case are concentrated towards the north end of the line. In this example, the difference can be explained by the proximity to important employment agglomerations and pre-existing infrastructure. In the first case, it can be argued that employment opportunities in the north end of MXB-LI are still distant. In the second case, areas close to the CBD were already served by the MPTN, while north end locations were not and are located at a reasonable distance necessary to realize relevant accessibility gains. It is also interesting to note that the benefits are not limited to zones adjacent to the newly built infrastructure. For instance, the upper row ('All') shows how some benefits introduced by the MXB-LI expand along a segment of the pre-existing METRO-LB, according to the network effect.

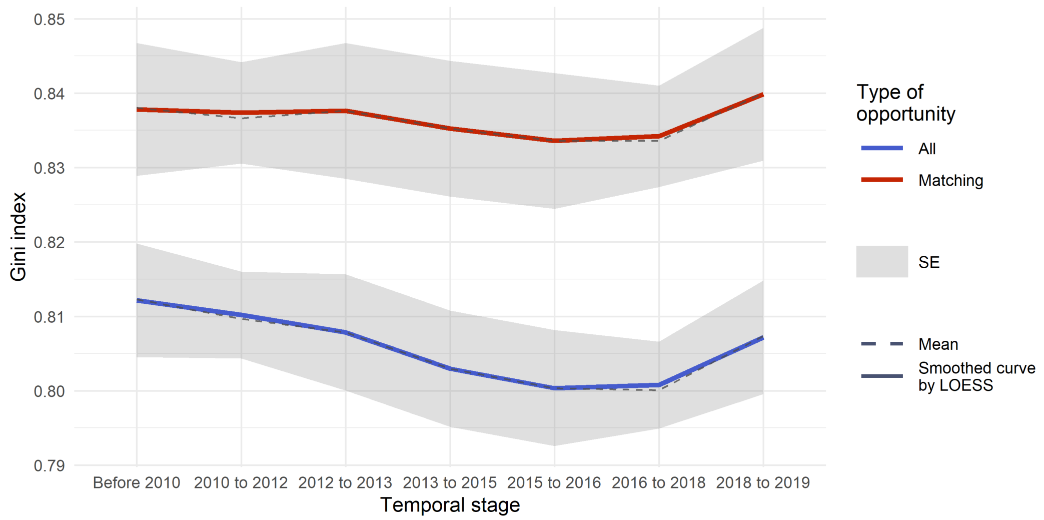
The variability shown in Stage 3 ('2012 to 2013') in Figure 6.10 is a further example of how the characteristics of both the land use and infrastructure influence the impacts of accessibility. There are two extensions occurring in this stage, namely MB-L4 (a BRT corridor) and the METRO-L12. One feature to note is that the gains are more extensive than in the previous stage, expanding to an important extent of the network. A second feature is the marked difference in the magnitude of the effects between the two innovations included in this stage. In the case of the MB-L4 the benefits are modest, extending along a short path around the METRO-LB (to the north-east from the CBD). Conversely, both the extent and magnitude of the changes induced by the METRO-L12 are worth noting. Specifically, some moderate effects extend to other perpendicular lines of the MPTN, such as the LRT-L1 and the MB-L1 to the south, as well as the METRO-L7 to the west.

Furthermore, the impacts adjacent to METRO-L12 are considerably larger than those to the MB-L4. The most important concentrate around the middle segments of the METRO-L12, i.e., in areas that are not too far from employment agglomerations and that were not previously served by the MPTN (e.g. *Lomas Estrella*, *Granjas Estrella I*, and *San Andrés Tomatlán* in the municipality of *Iztapalapa*, Ciudad de México, or around the *Hospital Pediátrico Iztapalapa* in the same municipality). Moreover, it can be noted that the magnitude of the effects gradually diminishes along the METRO-L12 the further the locations are from the CBD. By contrast, the locations to the west end of the line (relatively near to the CBD) reflect small benefits given the prior *good* coverage of the network. These patterns follow the same notions discussed earlier, which are influenced mostly by the pre-existing conditions of the MPTN and land-use. Still, the type of transport mode is a further component that appears to shape the magnitude of accessibility gains. Whereas the mode of most of the extensions in the period studied are semi-rapid transit (i.e. BRT), the METRO-L12 is a rapid-transit system. Therefore, its characteristics allow higher operational speeds and producing stronger and more far-reaching changes even in distant locations with respect to employment agglomerations than semi-rapid modes.

Many of the changes in Stage 4 ('2013 to 2015') and those subsequently shown in Figure 6.10 are moderate and follow similar spatial distribution patterns as those noted earlier. However, the final temporal stage (Stage 7, '2018 to 2019') is worth discussion. This stage consists of only one extension, namely the BRT MB-L7, which runs along one of the most emblematic business corridors in the city, i.e. *Av. Paseo de la Reforma* (continuing to the north on *Calzada de los Misterios*). As shown earlier in Figure 6.5, the employment density in this corridor is high especially in the segment between the intersection with *Circuito Interior* and *Avenida Juárez* (the latter near the traditional CBD, *Zócalo*). The configuration of the corridor establishes a direct link between the high-density employment zone and both ends of the MB-L7. This results in strong gains on the north-end (e.g. *Tepeyac Insurgentes*, *Industrial I*, *Estrella*) and the west-end (e.g. *Lomas De Chapultepec I*, *Lomas De Chapultepec II*, or *Palmitas (Polanco)*, all of these highly affluent zones) of the corridor. Meanwhile, it is worth noting that the gains around the middle segment of the MB-L7 are modest only (considering the pre-existing high employment density).

These temporal variations are reflected by both accessibility measures to a good extent. Still, there are some aspects that are worth discussing. A noticeable aspect in all temporal stages shown in Figure 6.10 is that the spatial patterns reflected by the measure considering matching opportunities (lower row) are more heterogeneous than the measure considering all types of employment (upper row). The latter tends to express changes as a smoother surface, while the former concentrates high peaks in small clusters. This is the case in Stage 3, for example. Here, the upper panel suggests that some moderate benefits extend up to the south-east end of the METRO-L12. However, the lower panel suggests that the gains in these distant areas are modest. Another example illustrating relevant differences between these measures are the changes produced by the MXB-LI in Stage 2. The upper panel suggests that some benefits continue to the south along the METRO-LB. However, this is not shown in the lower panel. This observation is interesting not only because it shows how the benefits extend to different locations of the MPTN (not only in adjacent zones to the innovations), but also because it shows the importance of the assumptions made in the alternative accessibility measures. While the 'All' measure reflects benefits as a result of the increased connectivity to employment opportunities located to the north of the METRO-LB (chiefly in the municipality of *Ecatepec de Morelos*, in the state of México), the 'matching' measure neglects these, given that residents located along METRO-LB would not find attractive those employment opportunities.

In a further approach, the Gini index is used to examine the variability of accessibility to employment across the different temporal stages of the MPTN. This is a global measure of distribution ranging from 0 to 1, where 0 implies the distribution is perfectly equal and 1 is perfectly unequal. This index is estimated in each temporal stage for each SAS and the full results are shown in the Appendix F. Figure 6.11 summarises the Gini index estimates according to the mean by temporal stage and type of measure in a longitudinal plot. This shows that inequality is higher in all stages according to the measure accounting for matching employment opportunities than it is for the one considering all types. Both measures show general downward trends as the MPTN expands. However, the 'Matching' measure is less sensitive to modifications of the MPTN. This is in agreement with the qualitative observations pointed out earlier. For instance, while the 'Matching' type reflects negligible changes from Stage 1 to Stage 3, the 'All' type captures slight improvements in the distribution of accessibility during these stages.



Source: The author based on own calculations. *Note:* LOESS = Local Polynomial Regression Fitting. SE = Standard error.

Figure 6.11: Mean Gini index for each MPTN temporal stage between 2010-2019.

Despite some of the differences already highlighted, there is coincidence between both measures in pointing out the largest improvements in Stage 4 and Stage 5 and a small deterioration followed by a considerable reversal in Stage 6 and Stage 7, respectively. These trends can be understood given that the extensions in Stage 5 and Stage 6 connect non-central municipalities of GMC to the MPTN (e.g. the municipality *Chimalhuacán* in the former and *Tultitlán* and *Coacalco de Berriozábal* in the latter), slightly balancing disadvantaged areas in terms of accessibility. By contrast, the extension in Stage 7 (MB-L7) operates at the core of both the metropolitan area and the MPTN and connects to zones consisting of highly remunerated employment opportunities. In this case, the 'Matching' measure is more sensitive than the one that considers all opportunities. This is reflected in the rebound of the index to even a higher point than the baseline ('Before 2010').

6.4 Discussion and limitations

The results which were presented in the previous section can be discussed from various angles. The following points are related to some spatial considerations:

- First, the findings can be discussed with regard to some key aspects of the spatial aggregation of accessibility on simple univariate statistics; namely scale and zoning. With regard to scale, the results show a mix in terms of the systematic variations of accessibility according to the size of the SAS. For instance, the average values do not display a specific pattern when a measure considering all employment opportunities is used. However, they do when the alternative measure is used. By contrast, the dispersion between observations (as measured by the relative standard deviation) gradually increases in higher resolution schemes when all opportunities are considered. However, this remains stable (around 3.0) in all grids when only *matching* opportunities are considered. These observations may respond to the increased uncertainty on univariate statistics in the presence of high positive spatial autocorrelation (S.-I. Lee et al., 2019).
- Furthermore, the post code SAS clearly illustrates how the accessibility is affected by zoning. In this case, it can be seen that the shape of post code zones induces the correlation between accessibility and surface area. This results in inflated mean and median values compared to uniform grids. Therefore, direct comparisons of accessibility measures generated by alternative zoning schemes should be made with care. The existing literature has identified this issue arising from the unevenness of units in terms of surface area and has recommended weighting methods to mitigate these effects (Arbia, 1989; e.g. K.-Y. Kim, 2011). Therefore, if the aim is to compare simple univariate statistics produced by different units of aggregations the previous suggestions should be considered.
- The results also show that the definition of scale and/or zoning can result in a differentiated representation of the overall spatial structure of accessibility to

employment at the city level. While high-resolution SASs (including the post code) portray a clear star-shaped structure, the low-resolution SASs depict what can be better described as concentric bands. However, some patterns introduced by the MPTN prevail even in the most abstract representations (e.g. METRO-LB or the asymmetry depicted by the matching measure).

- The high-resolution SASs show that the spatial patterns of accessibility differ according to the type of infrastructure and land use. For instance, while SUB-L1 is better represented by accessibility islands, BRT and metro form almost continuous corridors.

The role of public transport infrastructure from the temporal view is worth discussing too. The following features can be noted:

- Not all extensions produce global increases and some produce more than others, e.g. MB-L7 or METRO-L12. Some effects may be negligible at the global level. This is particularly noticeable in matching opportunities measures (where there are no improvements perceived).
- The ex post effects show heterogeneity in terms of extension and magnitude. Indeed, some effects can expand over the network to areas that are not necessarily adjacent to new public transport infrastructure. In other words, some indirect accessibility effects can be observed around the existing extension as a result of the effects of the network, in line with existing literature (Ahlfeldt, 2013).
- Changes in accessibility produced by the network clearly reflect two important components of accessibility measures; namely land use and infrastructure characteristics. There are various examples provided in the results. One of the clearest is the accessibility effects produced by the introduction of the MB-L7. Here, it is shown that the largest gains locate towards both ends of the corridor. By contrast, gains were low in middle segments where local employment density is very high.

  This finding can clarify the argument provided in some empirical studies regarding the lack of capitalization effects for BRT infrastructure (Guzman, Enríquez, & Hessel, 2021). Specifically, the method of the study assumes homogeneous accessibility improvements around stations. However, this lack of relationship is attributed to high employment density and good overall accessibility before the implementation of the infrastructure. These findings illustrate well those effects under these conditions.
- From the point of view of environmental benefit measurement (discussed in Chapter 3), the magnitude of the changes thus depend on all of the characteristics of the intervention highlighted in the points above. Therefore, whether the changes

in accessibility can be regarded as marginal and/or localized depends on whether these are estimated globally or locally. The latter also implies the consideration of the spatial scale, as it is shown that a small-scale geographic unit is more sensitive to changes in terms of both magnitude and extent. Thus, it remains difficult for an analyst to characterise such shocks in economic environmental analyses.

Furthermore, the results using alternative specifications for accessibility to employment, ‘All’ vs. ‘Matching’ opportunities, motivate the following considerations:

- The matching opportunities measure is less sensitive in capturing the role of the MPTN. This is illustrated in various examples, including *Santa Fe*, mixed patters of accessibility along transport corridors, and a close proximity to the education decile, also spatial autocorrelation denoting higher level of spatial heterogeneity.
- The average accessibility values are affected by a stronger skewed distribution produced by the implementation of matching opportunities. Therefore, there is a stronger reflection of spatial distribution of employment.
- As a side result, it can be seen that accounting for matching opportunities is key from the perspective of social equity.
- Related to the above, it can be seen that MB-L7 reinforces the concentration of accessibility in the most advantaged areas.
- This measure shows the potential to account for heterogeneous household hedonic models by relaxing the assumption that all employment is equally attractive implicitly made in metrics such as simple distance to the CBD or general location-based measures.

This analysis comes with limitations. First, the empirical data used to model commuters’ behaviour in the trip distribution model are aggregated at the TAZ. However, the aggregation of data at different spatial analytical schemes could affect this estimate, as previously suggested by empirical literature in the context of gravity modelling (e.g. K.-Y. Kim, 2011; Stillwell, Daras, & Bell, 2018). In a further chapter the stability of the spatial deterrence parameter is examined in the context of MAUP (i.e. Chapter 8). Secondly, information related to public transport services should ideally consider street-transit modes (e.g. *microbus* or regular buses) explicitly as a complement to semi-rapid and rapid transit modes. However, this information is not available and it would be highly time- and budget-consuming to assemble. Thirdly, the changes in the MPTN are aggregated in temporal stages which may include the introduction of more than one substantial change in the MPTN. This decision was aimed at simplifying the process and reducing the computational costs. Still, this diffuses the attribution of some effects to a single intervention.

There are also many strengths that are worth mentioning. The results draw on multi-dimensional analyses; namely the spatial and temporal effects of PT infrastructure. In addition, the personal component in accessibility measures is explored via the assumption of matching opportunities. Finally, the analyses employ empirical information processed in a trip-distribution model to account for the degree of spatial deterrence in accessibility measures.

Considering the above, there are relevant open avenues for future studies. First, it would be valuable to confirm and measure the degree of factors influencing changes in accessibility empirically using formal statistical techniques, e.g. topology of the network (e.g. connectivity and centrality), proximity to CBD or sub-centres, local employment density, pre-existing PT infrastructure, or the public transport mode being introduced and its characteristics (e.g. separation between stations, degree of segregation). Secondly, some important questions remain open. For instance, those derived from the adequacy, or otherwise, of the spatial scale at which homebuyers perceive accessibility benefits and the sensibility of results to the scale of the spatial unit of aggregation. In addition, researchers may consider whether homebuyers make any distinction in employment opportunities. Thirdly, comparative analyses of the results between hedonic methods and the equilibrium sorting models (according to a general equilibrium approach) can help to clarify the above questions. In addition, they can help in clarifying the appropriate characterization of accessibility changes (marginal and localized) in the context of environmental valuation.

6.5 Conclusions

This chapter has shown how accessibility to employment is shaped by the main public transport infrastructure network in Greater Mexico City drawing on a comprehensive exploratory spatial data analysis. An important empirical contribution refers to the characterization of public transport accessibility in GMC thanks to the identification and digital reconstruction of the MPTN at the metropolitan level in GTFS format over several temporal stages for a ten-year period (2009-2019). Furthermore, the results consistently show the heterogeneous spatial distribution of accessibility benefits. These are reflected in a variety of patterns which result from several characteristics of the land use and public transport infrastructure, e.g. proximity to employment, pre-existing infrastructure, and operational characteristics of the infrastructure. As a side result, these findings raise concerns about the equitable spatial distribution of public investments in public transport projects in the area of study.

The results highlight the importance of comprehensive measures which can capture the variations discussed above in the study of the capitalization of public transport benefits on land value. This is relevant in examining the extent of the potential of LVC as they provide further insights about the spatial and temporal distribution of benefits. For instance, the distribution of the charges in LVC instruments, such as betterment contributions, can be more equitable than under the assumption of constant catchment areas adopted in some research or local regulation, e.g. the Tax Code of Ciudad de México discussed in Chapter 4. In addition, the identification of the benefited areas could support the development of tools for prioritizing the implementation of land-based LVC instruments, such as joint developments or sale/lease of development rights (or the intermodal transport hub programme, CETRAM, in Ciudad de México, discussed in Chapter 4). Accordingly, the analysis in the next chapter adopts location-based measures in the context of the hedonic model to assess the potential of LVC as a financing tool supporting public transport.

Chapter 7

The value of public transport accessibility

7.1 Introduction and Research Question

The previous chapter illustrated the heterogeneous spatial distribution of accessibility to employment shaped by the main public transport network (MPTN) , drawing on the case of Greater Mexico City (GMC). The results set the ground for the potential role of location-based measures in the context of the hedonic model. Specifically, it was demonstrated that the accessibility estimates differ depending on the assumptions made regarding house-buyers' perceptions. As discussed in a review of the empirical literature (Chapter 3), the number of studies using these accessibility measures to examine the relationship between transport infrastructure and the property market has increased in recent years. However, knowledge in this area remains limited.

The main contribution of this chapter is the evaluation of the capitalization of public transport into residential land value while taking into account the perception of house-buyers. This consideration builds on environmental valuation literature, which assumes that there is a consistent link between measures of the environment (such as accessibility) and how services are perceived (Klaiber & Smith, 2011). Moreover, the accessibility literature suggests that "an accessibility measure is only appropriate as a performance measure if it is consistent with how residents perceive and evaluate their community" (Handy & Niemeier, 1997, p. 1176). In line with this perspective, it is suggested that specific population subgroups (such as age, gender, income, and education) evaluate opportunities differently (such as position, wages, or educational requirements in the supply side of the labour market) (Geurs & van Wee, 2004; Kwan & Weber, 2008; Levinson & Wu, 2020; Páez et al., 2012; M. Ryan et al., 2016; Thériault et al., 2013). The analyses presented in this chapter assess the willingness to pay (WTP) for public transport accessibility, taking into consideration the adequacy of

location-based measures, while explicitly incorporating the characteristics of both the offer (salaries paid at potential destinations) and the demand (education level at the origin) in the labour market. This is the first time that such a view has been integrated into the literature on the hedonic model.
The specific objective of this chapter is to:

- Evaluate the willingness to pay for the accessibility benefits derived from the MPTN in the residential land market while considering the adequacy of location-based measures. It will do this by acknowledging the salaries paid at the potential destination and the education level at the origins, drawing on the case of Greater Mexico City.

In line with the above, the specific Research Question (RQ) is:

- *What is the willingness to pay for the accessibility generated by the main public transport network in the residential land market between 2009 and 2019?*

The results suggest a robust capitalization of public transport accessibility into residential land value, as represented by location-based measures. This is confirmed by a positive and significant relationship across a variety of regression techniques for all the temporal stages of the MPTN studied. In addition, this association was positive in a variety of accessibility specifications tested. The pertinence of a location-based measure that distinguishes the characteristics of residents and employment population is evident when the analytical technique does not explicitly incorporate local heterogeneity, e.g. multilevel or spatial models. This is shown in two ways. First, the hedonic models performed better when the parameters estimated in a spatial interaction model were incorporated into an accessibility measure that matched residents' level of education and employment salaries, as opposed to when they were used in a generic one. Second, the incorporation of these aspects into location-based measures also improved hedonic models that were estimated simultaneously with accessibility parameters. This improvement was seen in the explanatory power, reduced heteroskedasticity, and possible misspecification issues. Furthermore, an estimate of the benefits of one of the MPTN's temporal stages from a partial equilibrium approach suggests that the size of the capitalization of public transport into residential land is considerable in GMC. The estimates for the bus rapid transit (BRT) Line 7 of the *Metrobús* system (MB-L7) ranged between 0.7 times the equivalent of the capital cost of in a partial scenario to 1.5 in a full scenario.

The present chapter is organised as follows: Section 7.2 sets out the research strategy adopted to address the Research Question. It provides specific information about the data and measures used. It also presents the model specifications and the estimation procedures used. Section 7.3 presents the results. First, it outlines the descriptive statistics of the sample. Following this, it presents the results of the estimated hedonic functions. It goes on to present the estimated aggregate willingness to pay for accessibility improvements associated with the MB-L7. Section 7.4 discusses the results in relation to the literature reviewed in Chapter 2 and Chapter 3. This section also outlines some of the limitations of this work. Section 7.5 provides a conclusion for this chapter.

7.2 Research strategy and estimation

To answer the Research Question, the analytical process works through the following steps:

1. Establish a parsimonious hedonic property value model to identify the house value determinants, including structural property and locational attributes.
2. Examine the adequacy of alternative gravity-type accessibility specifications for assessing the transport service level generated by the MPTN.
3. Evaluate the hedonic prices estimated in consideration of the spatial structure of the data across the various temporal stages of the MPTN between 2009 and 2019.
4. Illustrate the potential for LVC in association with improvements of the MPTN, according to the most recent transport extension in the period studied.

A cross-sectional design is adopted. The standard hedonic model assumes 'stability over time' (Parmeter & Pope, 2013) which implies that home-buyers' preferences were fixed over the period studied. This assumption is relaxed in step 3 by fitting independent hedonic functions according to the temporal stages of the MPTN, as identified in Chapter 6.

7.2.1 Data and measures

The data set employed in this chapter includes 781,898 housing valuation records obtained from the Federal Mexican Society (SHF), as detailed in Chapter 5. In addition to the observations discarded in the initial data cleaning process, 16,216 additional observations were excluded due to missing information relating to educational levels or because the *class* of the construction was 'Minimal.' The latter were not considered because the criterion for assigning this category after April 2015 (i.e. in SHIF-DB2). In addition, 145 records were removed due to reporting '0' bathrooms.

The structural attributes of dwellings were obtained directly from the SHF records, while the neighbourhood and locational characteristics were estimated from a variety of sources, as detailed in Chapter 5. In this chapter, the spatial analytical scheme is restricted to the post code zone as the lowest location reference for the properties in this data set after April 2015 was only available at the post code level. However, a thorough sensitivity analysis was conducted to account for the modifiable areal unit problem (MAUP) discussed in Chapter 3. The implications of the analysis are addressed in Section 7.4 and the specific results are presented in Appendix G.

MPTN accessibility measures

Two alternative location-based accessibility specifications accounting for the services enabled by the MPTN will be considered. The first is written as follows (Mcarthur, Kleppe, Thorsen, & Ubøe, 2013; Osland & Thorsen, 2008):

$$A_{jt}^{\mathrm{PT}} = \sum_{j=1}^{J} E_k^{\gamma} e^{-\beta_2 d_{jkt} + \mu\delta_{jk}}. \tag{7.1}$$

Here, MPTN accessibility at post code j at temporal stage t is given by the sum of the product of two functions. The first considers total employment, E, in post code k to the power of γ. The latter is a weighting parameter directly affecting the attractiveness of employment opportunities independent from their location. This has been shown to produce adequate results in the context of the hedonic property model (Osland & Pryce, 2012; Osland & Thorsen, 2008). The product of the second function discounts the attractiveness of employment opportunities exponentially, according to the travel time in between j and k via the MPTN in temporal stage t, d_{jkt}. The MPTN travel time is modelled according to the timetables in each t, as detailed in Chapter 6. The relevance of the travel time is regulated by β_2. μ is a parameter denoting additional benefits of residing and working in the same location or, analogously, it can be viewed as a start-up cost when considering employment opportunities in a different post code from the origin (Thorsen & Gitlesen, 1998). μ affects only potential employment opportunities where the origin and destination are the same, as indicated by the Kronecker delta product δ_{jk} which takes a value of 1 when $j = K$ and 0 otherwise.

The second MPTN accessibility specification adds the characteristics of both the offer (salaries paid at potential destination) and the demand (education level at the origin) of the labour market to Eq. (7.1). This formulation draws on the discussion presented in Chapter 6 and it is written as follows:

$$A_{jtm}^{\mathrm{PT}} = \sum_{j=1}^{J} E_{km}^{\gamma} e^{-\beta_2 d_{jkt} + \mu\delta_{jk}}. \tag{7.2}$$

Here, the notation follows the same as in Eq. (7.1) except that m at j is categorized according the median education level of its residents by quintile and m at k is classified based on the average paid salary by quintile. Thus, only employment opportunities E matching the education level quantile between j and k are considered (when $j \neq k$), as discussed in Chapter 6.

7.2.2 Model specification and estimation

The hedonic function discussed in Chapter 3 and Chapter 5 is estimated in a regression framework as shown below.

$$\ln(P_i) \sim N(\alpha_+ \beta_S^\intercal \mathbf{X}_{iS} + \beta_N^\intercal \mathbf{X}_{iN} + \beta_A^\intercal \mathbf{X}_{iA} + \beta_t Year_i + \beta_s State_i, \sigma_y^2) \text{ for } i = 1, ..., n \tag{7.3}$$

P_i is the value of each property (in MXN) and it is entered in the logarithmic form; $\mathbf{X}_S$ denotes the structural controls of the property (i.e. saleable area, class, number of bathrooms, number of parking spaces, and age of the building); $\mathbf{X}_N$ is the neighbourhood attributes (i.e. availability of a major educational facility, presence of a major public administration establishment, percentage of streets that do not have street light, percentage of street with trees, availability of a park, crime, and percentage of households with a private car available); $\mathbf{X}_A$ is a matrix including locational attributes interpreted in accessibility measures estimated at the post code level (i.e. distance to the CBD, accessibility to employment by car, and main public transport network accessibility to employment); β_S, β_N, and β_A are vectors including the respective regression coefficients of the covariates; $Year$ is a time control for the specific year when the value of the property was assessed which enters as a dummy variable and β_t is its regression coefficient (the base year is 2009); $State$ is a further control variable indicating the state of the metropolitan area where the property is located (i.e. Ciudad de México, Hidalgo, México) which is entered as a dummy variable and β_s is its respective regression coefficient (Ciudad de México is the reference); α is a constant term; and σ_y^2 is the error term variance associated to each observation i. The error term is assumed to be independent and to follow a normal distribution. The continuous variables in $\mathbf{X}_S$ enter in the log form. Additionally, age is entered in the logarithmic as well as in the log-squared form as $\log(age)^2$, following Osland & Thorsen (2008) and McArthur et al. (2012). This helped to improve the model parsimony, as found in a preliminary modelling phase.

Accessibility to employment by car at j, A_j^{CAR} is also represented by a location-based measure which is discussed at length in Chapter 5.[1] This measure takes its own distance deterrence parameter, β_1 which was estimated beforehand by minimizing the sum of the squared errors (SSE) of Eq. (7.3) (excluding MPTN accessibility). This procedure was implemented using Brent's optimization algorithm (Brent, 1973) in R programming (R Core Team, 2021). The estimated optimal β_1 parameter is 0.084 (as introduced in the log-linear form). This value is very consistent when compared to previous empirical studies that employed a similar measure. For instance, this was 0.086 in Rogaland, Norway (Osland & Thorsen, 2008), and 0.10 in Berlin, Germany (Ahlfeldt, 2011).

Two different methods are used for the estimation of MPTN accessibility parameters in Eq. (7.1) and Eq. (7.2): one exogenous and other endogenous. The exogenous estimates β_2 and μ a priori in a trip-distribution model (within a doubly-constrained gravity framework) using observed origin-destination (OD) flows from the 2017 Travel Survey (INEGI, 2018), as detailed in Chapter 6. Here, γ equals to 1 . This procedure is similar to the one followed in Adair et al. (2000) and Cordera et al. (2018). The calibrated accessibility parameter values according to this method are the following:

$$\beta_2 = 0.044,$$

and

$$\mu = 1.654.$$

The endogenous method is a simultaneous estimation using the observed property data available combing Eq. (7.3) and Eq. (7.1) or Eq. (7.2). As these measures include parameters to be estimated, they are fitted by non-linear methods by the maximum likelihood, as in McArthur et al. (2012). For example, Eq. (7.1) enters to Eq. (7.3) as following:

$$\ln(P_i) = \alpha + \sum_{c=1}^{C} \beta_c \boldsymbol{x}_{ci} + \beta_{\text{APT}} \left(\sum_{j=1}^{J} E_k^{\gamma} \exp(-\beta_2 d_{jkt} + \mu \delta_{jk}) \right) + \epsilon_i. \tag{7.4}$$

In Eq. (7.4), α is the intercept; $\boldsymbol{x}$ is the design matrix which collapses the c covariates as specified in the hedonic model in Eq. (7.3) (i.e. structural controls X_S, X_N neighbourhood attributes, X_A accessibility, $Year$, and $State$) with β_c being a vector denoting their corresponding linear effect; β_{APT} is the regression coefficient for the MPTN accessibility (e.g. A_{jt}^{PT}) ; and ϵ is an independent and normally distributed

[1] Accessibility to employment by car is given by

$$A_j^{\text{CAR}} = \sum_{j=1}^{J} E_k \exp(-\beta_1 d_{jk})$$

.

error term. The implementation of the nonlinear models (NMLs) is achieved using the `fixest` software version '0.9.0' for `R` (Berge, 2021). In the first set of models, γ is fixed as equal to 1 for comparability with the exogenous estimates. In the rest, all accessibility parameters are allowed to vary.

The combination of MPTN accessibility specification and the calibration method of parameters results in a set (Set 2, S2) of the following six models :

1. M1:S2 introduces MPTN accessibility to the hedonic model. Here, households and employment opportunities are assumed to be homogeneous and the parameters are estimated exogenously using data from the 2017 Travel Survey only.
2. M2:S2 is similar to M1:S2 except that the MPTN accessibility to employment acknowledges the type of household in j and employment opportunities at k.
3. M3:S2 estimates the parameters β_2 and μ of Eq. (7.1) by simultaneously fixing γ as equal to 1 and assuming home-buyers and employment to be homogeneous (as in M1:S2).
4. M4:S2 also estimates parameters β_2 and μ of Eq. (7.2) simultaneously holding γ fixed equal to 1. As in M2:S2, this specification acknowledges the level of education of residents at j and the characteristics of employment at k.
5. M5:S2 is analogous to M3:S2, with the exception that here γ is allowed to vary too.
6. M6:S2 is analogous to M4:S2, with the exception that the former allows γ to vary too.

The linearised versions of the models from Eq. (7.4) are fitted holding the accessibility parameters $\{\rho, \gamma, \beta\}$ fixed to calculate the standard diagnostic statistical tests, as done in similar studies (McArthur et al., 2012; e.g. Osland & Thorsen, 2008).

Multilevel models and spatial multilevel models

In a next step, the full sample is split according to the respective temporal stages of the MPTN of each observation i and estimated in set of independent models to account for possible temporal variations. In addition, the hierarchical and spatial structure of the data is acknowledged in multilevel and spatial multilevel models. These aspects are discussed in detail in Chapter 5. Here, the structural characteristics of the dwellings are considered as lower-level attributes, while the neighbourhood and locational characteristics aggregated by post code zone are considered as upper-level attributes.

The specification of the multilevel model (MLM) is as follows (Gelman & Hill, 2007, pp. 264–265):

$$\begin{aligned}
&\ln(P_i) \sim N(\alpha + \beta_S^\intercal \mathbf{X}_{iS} + \beta_N^\intercal \mathbf{X}_{iN} + \beta_A^\intercal \mathbf{X}_{iA} + \beta_t Year_i + \beta_s State_i + \nu_{j[i]}, \sigma_\nu^2), \text{ for } i = 1, ..., n \\
&\quad \text{and} \\
&\quad \nu_j \sim N(0, \sigma_\nu^2).
\end{aligned} \tag{7.5}$$

This is the usual formula style used to represent the 'mixed effects' model and it is adopted here because it facilitates the upcoming discussion. In this formula, the notation follows that of Eq. (7.3). $\boldsymbol{\nu} = (\nu_1, ..., \nu_J)$ can be thought as the upper-level error term, which is assumed to be independent and identically distributed (iid) , and; σ_ν^2 represents the variance at the upper level (i.e. between post codes zones).

The multilevel model above is extended to incorporate a spatially structured random effect for zone j, $\boldsymbol{\upsilon} = (\upsilon_1, ..., \upsilon_J)$, according to the Besag-York-Mollie model (BYM) detailed in Chapter 5 and is written as follows (Besag et al., 1991):

$$\begin{aligned}
\ln(P_i) &\sim N(\alpha + \beta_S^\intercal \mathbf{X}_{iS} + \beta_N^\intercal \mathbf{X}_{iN} + \beta_A^\intercal \mathbf{X}_{iA} + \beta_t Year_i + \beta_s State_i + \xi_{j[i]}, \sigma_e^2), \text{ for } i = 1, ..., n, \\
\xi_{j[i]} &= \nu_j + \upsilon_j, \\
\nu_j &\sim N(0, \sigma_\nu^2), \\
\upsilon_j \mid \boldsymbol{\upsilon}_{-j} &\sim N\left(\frac{\sum_{j=1}^J w_{jk}\upsilon_k}{\sum_{j=1}^J w_{jk}}, \frac{\sigma_\upsilon^2}{\sum_{j=1}^J w_{jk}}\right), \\
\sigma_\nu^2, \sigma_\upsilon^2 &\sim \text{Gamma}(a, b),
\end{aligned} \tag{7.6}$$

where a and b are the shape and scale parameters of the variance hyperparameters, respectively. Due to the complexity of the spatially structured random effects, the spatial multilevel model is estimated in a Bayesian setting (Lawson, 2009), as detailed in Chapter 5.

According to the INLA's framework notation (Blangiardo et al., 2013), the spatial multilevel model in Eq. (7.6) can be re-written as:

$$\eta_{ij} = log(P_{ij}) = \alpha + \sum_{c=1}^{C} \beta_c x_{qij} + \nu_j + \upsilon_j + \epsilon_{ij}, \tag{7.7}$$

where $\boldsymbol{x}$ is a matrix that includes the c characteristics and covariates specified in the hedonic model in Eq. (7.3) (i.e. X_S structural controls, X_N neighbourhood attributes, X_A accessibility, $Year$, and $State$), with $\boldsymbol{\beta}$ being a vector denoting their corresponding linear fixed effects; the random effects at the upper level enter to the INLA framework as $\xi_j = \nu_j + \upsilon_j$ and υ_j (the first term denotes the sum of the structured and unstructured upper-level components, while the second is only the spatial random effect). ϵ_{ij} is the individual observation error term which follows a normal distribution with variance σ_y^2.

Component	Type	Shape	Rate	Source
Set 1				
Unstructured τ_{ν}^2	logGamma	1.0	0.0005	Rue et al. (2012)
Spatial τ_{υ}^2	logGamma	1.0	0.0005	
Set 2				
Unstructured τ_{ν}^2	logGamma	1.0	0.001	Bivand et al. (2015)
Spatial τ_{υ}^2	logGamma	1.0	0.001	
Set 3				
Unstructured τ_{ν}^2	logGamma	1.0	0.01	Bivand et al. (2017)
Spatial τ_{υ}^2	logGamma	1.0	0.01	
Set 4				
Unstructured τ_{ν}^2	logGamma	1.0	0.001	Blangiardo et al. (2013)
Spatial τ_{υ}^2	logGamma	1.0	0.01	

Table 7.1: Prior parameters used for the sensitivity analysis.

Here, the vector parameters estimated in the INLA framework are $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\upsilon}\}$ and the hyperparameters are $\boldsymbol{\psi} = \{\tau_y^2, \tau_\nu^2, \tau_\upsilon^2\}$ (note that within the INLA's framework the variance is expressed by the precision as $\tau = 1/\sigma^2$).

For the empirical analysis, 'vague' prior distributions are used which resemble a noninformative with a flat shape. This choice is because: (1) there is not enough information nor enough strong preferences for a priori distribution, and (2) the sample size is large (>50K observations in each subset), which means that the data can *speak for itself*, as suggested by several authors (Blangiardo & Cameletti, 2015; Lawson, 2009; LeSage & Pace, 2009). The specific parameters are taken from Bivand et al. (2017) and are the following: on the log of both the unstructured and structured precision is $\tau_\nu^2, \tau_\upsilon^2 \sim \text{logGamma}(1, 0.01)$,[2] which specify the shape and inverse-scale parameter, respectively; and on the fixed effects part of the model (for the regression coefficients and the intercept) is $\alpha, \beta \sim N(0.0, 0.001)$, the mean and precision, respectively, where the precision is the inverse of the variance.

Sensitivity analysis was conducted in one subset defined by the latest temporal stage of the data (Stage 7) exploring four combinations of prior parameter specifications for the hyperparameters presented in Table 7.1. The results of the preliminary sensitivity analysis show that the estimates are very consistent with minor differences on the model fit, as displayed in Table G.1 in Appendix H.

[2]This is on the log form as the computing method employed internally transforms the scale of the model for computational ease

7.3 Results

7.3.1 Descriptive statistics

Table 7.2 provides a summary of the descriptive statistics of the total sample and by sub-sample according to the MPTN temporal stage. The overall mean value of the properties is MXN$1.29 million (GBP£46,705)[3] at current prices. The value of the properties consistently increases in each of the temporal stages. For instance, the mean value practically trebles from Stage 1 to Stage 7. While the market value substantially increases in the period studied, the average saleable area remains fairly constant at around 83 m^2 in almost all temporal stages, with the exception of the first, where the area is smaller (74 m^2). About half of the sample is constituted by 'Low-Cost/Social' class housing (53%). The proportion of this class consistently falls with time. This type is followed by 'Middle'-class properties which make up almost two in every five units (38%). The 'Semi-luxury' units are considerably fewer than other classes, representing less than one tenth (8%). Only less than 2% of the observations are 'Residential/Luxury.' About seven in every ten houses has only one bathroom (69%). The proportion of this type falls from about three quarters in Stage 1 (77%) to nearly two thirds in Stage 7 (64%), while the number of houses with two or more bathrooms increases. Around two thirds of the dwellings in the sample have only one parking space (69%).

Houses without and with tow parking slots are equally frequent in the total sample (14%) and only a few have three or more than three (1% and 2%, respectively). The average age of the construction is almost nine years. This has consistently increased from seven years in Stage 1 to eleven years in Stage 7. The temporal patterns observed in the structural characteristics of houses are in line with the shift identified in the region where the housing market in newly developed horizontal areas on the periphery (e.g., small one- or two-story houses) decelerated, and a focus on central areas increased (i.e. Mexico City) (Flores, 2019). This is likely the result of an urban development agenda established by the state of Ciudad de México through the Mexico City Development and Investment Agency (*Agencia de Promoción de Inversiones y Desarrollo para la Ciudad de México*).

[3]Using an exchange rate of 27.60 MXN per one GBP, according to the currency exchange rate published by the Central Bank of Mexico (*Banxico*) on the 15/09/2021 on the website https://www.banxico.org.mx.

Table 7.2: Descriptive statistics

	Stage 1 (N=146779)	Stage 2 (N=116866)	Stage 3 (N=79200)	Stage 4 (N=118206)	Stage 5 (N=73903)	Stage 6 (N=142375)	Stage 7 (N=104569)	Total (N=781898)
Value (MXN/1000)								
Mean (SD)	723.69 (1333.88)	863.47 (1120.26)	1 011.60 (1374.62)	1 262.91 (1685.07)	1 458.08 (2010.79)	1 728.77 (2788.34)	2 067.00 (3804.14)	1 287.34 (2255.94)
Saleable area (sqm.)								
Mean (SD)	73.86 (57.76)	79.00 (63.27)	82.70 (69.07)	88.15 (73.96)	87.03 (70.74)	86.93 (72.80)	86.75 (77.05)	83.03 (69.25)
Class								
Low-Cost/Social	97 823 (66.6%)	71 571 (61.2%)	44 237 (55.9%)	56 754 (48.0%)	34 469 (46.6%)	62 740 (44.1%)	43 747 (41.8%)	411 341 (52.6%)
Middle	40 037 (27.3%)	36 242 (31.0%)	28 220 (35.6%)	49 120 (41.6%)	31 107 (42.1%)	62 083 (43.6%)	46 925 (44.9%)	293 734 (37.6%)
Semi-luxury	7 189 (4.9%)	7 230 (6.2%)	5 099 (6.4%)	9 994 (8.5%)	6 936 (9.4%)	15 071 (10.6%)	11 514 (11.0%)	63 033 (8.1%)
Residential/Luxury	1 730 (1.2%)	1 823 (1.6%)	1 644 (2.1%)	2 338 (2.0%)	1 391 (1.9%)	2 481 (1.7%)	2 383 (2.3%)	13 790 (1.8%)
N. of bathrooms								
1	113 553 (77.4%)	87 034 (74.5%)	56 418 (71.2%)	77 531 (65.6%)	47 663 (64.5%)	89 659 (63.0%)	66 375 (63.5%)	538 233 (68.8%)
2	25 652 (17.5%)	21 588 (18.5%)	16 372 (20.7%)	30 400 (25.7%)	19 489 (26.4%)	39 071 (27.4%)	28 405 (27.2%)	180 977 (23.1%)
3	4 992 (3.4%)	5 742 (4.9%)	4 347 (5.5%)	6 648 (5.6%)	4 206 (5.7%)	8 753 (6.1%)	6 124 (5.9%)	40 812 (5.2%)
3+	2 582 (1.8%)	2 502 (2.1%)	2 063 (2.6%)	3 627 (3.1%)	2 545 (3.4%)	4 892 (3.4%)	3 665 (3.5%)	21 876 (2.8%)
N. of parking spaces								
0	16 335 (11.1%)	13 114 (11.2%)	8 849 (11.2%)	13 206 (11.2%)	13 980 (18.9%)	27 503 (19.3%)	19 315 (18.5%)	112 302 (14.4%)
1	114 167 (77.8%)	88 365 (75.6%)	57 999 (73.2%)	83 252 (70.4%)	44 719 (60.5%)	84 860 (59.6%)	63 325 (60.6%)	536 687 (68.6%)
2	13 192 (9.0%)	12 386 (10.6%)	9 726 (12.3%)	17 236 (14.6%)	12 266 (16.6%)	24 503 (17.2%)	17 968 (17.2%)	107 277 (13.7%)
3	2 034 (1.4%)	1 927 (1.6%)	1 666 (2.1%)	2 818 (2.4%)	1 895 (2.6%)	3 564 (2.5%)	2 418 (2.3%)	16 322 (2.1%)
3+	1 051 (0.7%)	1 074 (0.9%)	960 (1.2%)	1 694 (1.4%)	1 043 (1.4%)	1 945 (1.4%)	1 543 (1.5%)	9 310 (1.2%)
Age (years)								
Mean (SD)	6.63 (10.71)	7.56 (11.32)	8.17 (11.72)	9.32 (12.15)	9.42 (12.22)	9.73 (12.35)	10.87 (12.83)	8.73 (11.95)
Edu. facility (20 min.)	18 364 (12.5%)	14 093 (12.1%)	10 953 (13.8%)	16 935 (14.3%)	11 832 (16.0%)	21 426 (15.0%)	15 826 (15.1%)	109 429 (14.0%)
Public admin. (20 min.)	33 232 (22.6%)	27 447 (23.5%)	19 999 (25.3%)	33 829 (28.6%)	22 002 (29.8%)	43 740 (30.7%)	29 765 (28.5%)	210 014 (26.9%)
Street light NA (%)								
Mean (SD)	9.84 (9.60)	9.14 (9.08)	8.91 (9.29)	7.59 (8.41)	7.33 (8.54)	7.66 (8.69)	7.39 (8.12)	8.34 (8.91)
Streets with trees (%)								
Mean (SD)	52.94 (26.16)	54.51 (25.36)	55.70 (25.31)	57.30 (27.27)	56.50 (29.60)	58.01 (27.84)	57.08 (28.61)	55.93 (27.18)

Table 7.2: Descriptive statistics *(continued)*

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7	Total
	(N=146779)	(N=116866)	(N=79200)	(N=118206)	(N=73903)	(N=142375)	(N=104569)	(N=781898)
Park (15 min.)	83 920 (57.2%)	68 727 (58.8%)	46 746 (59.0%)	72 225 (61.1%)	46 238 (62.6%)	89 247 (62.7%)	61 583 (58.9%)	468 686 (59.9%)
Crime (std. homicides)								
Mean (SD)	1.25 (1.12)	1.23 (1.11)	1.26 (1.15)	1.34 (1.23)	1.26 (1.18)	1.30 (1.22)	1.30 (1.20)	1.28 (1.18)
HH with car (%)								
Mean (SD)	55.16 (16.09)	53.80 (16.39)	52.49 (16.36)	51.78 (17.81)	54.00 (19.19)	53.49 (18.80)	53.37 (18.05)	53.52 (17.54)
State								
Ciudad de México	45 165 (30.8%)	37 290 (31.9%)	27 853 (35.2%)	48 514 (41.0%)	30 970 (41.9%)	58 137 (40.8%)	42 361 (40.5%)	290 290 (37.1%)
Hidalgo	5 343 (3.6%)	4 191 (3.6%)	2 587 (3.3%)	7 894 (6.7%)	7 295 (9.9%)	12 159 (8.5%)	7 729 (7.4%)	47 198 (6.0%)
México	96 271 (65.6%)	75 385 (64.5%)	48 760 (61.6%)	61 798 (52.3%)	35 638 (48.2%)	72 079 (50.6%)	54 479 (52.1%)	444 410 (56.8%)
Year								
2009	86 053 (58.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	86 053 (11.0%)
2010	60 726 (41.4%)	20 315 (17.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	81 041 (10.4%)
2011	0 (0.0%)	75 663 (64.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	75 663 (9.7%)
2012	0 (0.0%)	20 888 (17.9%)	56 675 (71.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	77 563 (9.9%)
2013	0 (0.0%)	0 (0.0%)	22 525 (28.4%)	45 577 (38.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	68 102 (8.7%)
2014	0 (0.0%)	0 (0.0%)	0 (0.0%)	72 629 (61.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	72 629 (9.3%)
2015	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	73 903 (100.0%)	0 (0.0%)	0 (0.0%)	73 903 (9.5%)
2016	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	67 641 (47.5%)	0 (0.0%)	67 641 (8.7%)
2017	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	64 716 (45.5%)	0 (0.0%)	64 716 (8.3%)
2018	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10 018 (7.0%)	52 814 (50.5%)	62 832 (8.0%)
2019	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	51 755 (49.5%)	51 755 (6.6%)
Dist. to CBD (Km)								
Mean (SD)	25.22 (14.41)	24.43 (14.00)	23.62 (14.17)	21.82 (14.28)	22.52 (15.16)	22.67 (15.20)	22.69 (15.21)	23.37 (14.68)
Accessibility car								
Mean (SD)	1.02 (0.98)	1.05 (0.96)	1.11 (0.99)	1.22 (1.00)	1.22 (1.01)	1.21 (1.02)	1.20 (1.02)	1.14 (1.00)
Accessibility PT_{All}								
Mean (SD)	0.53 (0.93)	0.54 (0.92)	0.63 (0.99)	0.69 (1.00)	0.72 (1.01)	0.73 (1.02)	0.78 (1.11)	0.65 (1.00)
Accessibility PT_m								
Mean (SD)	0.45 (0.92)	0.46 (0.94)	0.53 (1.00)	0.57 (1.02)	0.57 (1.03)	0.58 (1.03)	0.60 (1.07)	0.53 (1.00)

Note:

Desciptive statistic values: numeric variables = M(SD): Categorical variables = N(%). MPTN = Main public transport network.

Regarding the neighbourhood characteristics shown in Table 7.2, 14% of the houses had a major educational facility available within 20 minutes of walking. This proportion is below the overall figure in the first three stages (between 12.1% and 13.8%), while it is slightly above in the rest. A public administration establishment can be reached by about one quarter of the dwellings in 20 minutes or less. On average, 8% of the streets in the post code zone where the house is located do not have street lights available. Also, a little more than half of the streets have trees. Furthermore, nearly three in every five observations in the sample has a park available within 15 minutes of walking. This proportion is fairly consistent within the sub-samples. The standardized crime mean for the total sample is 1.28. The houses in the first three stages were transacted in areas with slightly lower crime incidence (between 1.25 and 1.23), whilst the ones sold in the last two are somewhat above the overall mean (1.30). Concerning car ownership, a little more than half of the households have at least one car. This remains practically unchanged in the sub-samples.

A little more than half of the houses were transacted in the state of Mexico (57%), this is followed by Ciudad de México, with 37%, whereas Hidalgo had a substantially lower proportion (6%), as shown in Table 7.2. The composition of the sub-samples by state shows important fluctuations. For instance, while houses in Hidalgo represent only 3.6% in Stage 1, this proportion peaks in Stage 5 at 9.9%, decreasing to 7.4% in the more recent stage. In addition, the market in the state of Mexico has a larger presence in the early stages (up to 66% in Stage 1) and steadily decreases to stabilize at around 50% from Stage 4 onwards. Conversely, the share in Ciudad de México is less than one third in Stage 1. This figure grows to about 40% in Stage 4 and remains practically unchanged in later stages. The number of observations by year remains roughly stable with close to one tenth each year between 2009 and 2018. One exception is the year 2019, which represents only about 7% of the total sample. These fluctuations correspond to the observed urban dynamics discussed previously.

As seen in Table 7.2, the observations in the total sample are located 23.4 kilometres away from the CBD on average. The location of houses in the first three stages is at a longer distance from the CBD than the total sample. This is also reflected in the accessibility by car. For instance, whilst the global average is 1.14, the first three sub-samples display lower values (between 1.02 and 1.11) than the overall value. This means that in early stages there were proportionally more houses located in low-accessibility zones. The mean potential accessibility to employment generated by the MPTN, which considers all types of employment as available, is 0.65. As expected, this increases in each of the temporal sub-samples, going from 0.53 in the first to 0.78 in the last. Meanwhile, the average value of the alternative accessibility measure, which distinguishes the labour market offer and demand (as denoted in Eq. (7.2)), is 0.53. Similarly, this steadily rises across the temporal stages. It should be noted that these two measures related to the public transport service were estimated a priori. The following section details the estimation procedure for the endogenous accessibility measures used.

7.3.2 The base hedonic property value model

In the first modelling stage, an adequate base hedonic property value model following Eq. (7.3) is identified. For this purpose, a set of six alternative model specifications (Set 1) are fitted and the performance of different *general* accessibility measures is evaluated, namely: distance to the CBD, accessibility by car, or a mix of both, according to different functional forms (i.e. log-linear and log-log). The full results of Set 1 are summarised in Table H.1 and presented in Appendix I.

Considering the different diagnostic tests and goodness-of-fit measures, the preferred specification is given by the model that conceptualizes general accessibility as the potentiality of access to employment by car A_j^{CAR} (Model 3 (M3:S1) in Table H.1). This specification replaces the more abstract measure computed as the Euclidean distance to the traditional CBD (*Zócalo*). The preferred model provides good explanatory power while reducing misspecification, collinearity, and heterogeneity issues. This choice is also supported by neoclassic urban economic theory (Alonso, 1964), as gravity-type accessibility measures are flexible enough to capture the spatially dispersed employment opportunities, or analogously commuting cost, compared to the traditional measures that restrict the location of opportunities to a single location. This is in coincidence with previous empirical results (Ahlfeldt, 2011).

7.3.3 Incorporating a measure of public transport accessibility to employment to the hedonic property value model

Having identified a base hedonic property value model, the variables accounting for the benefits generated by the MPTN to Eq. (7.3) are entered in the second set of models (Set 2). The results are summarised in Table 7.3. To improve visibility, only the estimated coefficients for neighbourhood and locational attributes are included. Structural and year dummy controls are significant, and in the direction expected. In Set 2, the MPTN's accessibility to employment is entered as specified in Eq. (7.1) or Eq. (7.2), referred to as *All* or m in the type of employment opportunity row in Table 7.3, respectively. These are entered in the log-linear form. This choice is based on preliminary modelling specifications tests.

Regarding the neighbourhood characteristics, the results show that the esti-

	M1:S2	M2:S2	M3:S2	M4:S2	M5:S2	M6:S2
	(OLS)	(OLS)	(NLM)	(NLM)	(NLM)	(NLM)
Intercept	9.67***	9.70***	9.69***	9.70***	9.66***	9.66***
	(0.064)	(0.062)	(0.062)	(0.062)	(0.070)	(0.074)
Edu. facility: Yes	0.043***	0.037***	0.044***	0.039***	0.041***	0.037***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Public admin.: Yes	0.023*	0.016	0.020*	0.016	0.013	0.008
	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)
Street with trees (%)	0.053*	0.059**	0.061**	0.060**	0.033	0.028
	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)	(0.030)
Street light NA (%)	-0.213***	-0.228***	-0.222***	-0.230***	-0.260***	-0.272***
	(0.056)	(0.057)	(0.056)	(0.057)	(0.069)	(0.069)
Park: Yes	-0.039***	-0.040***	-0.036***	-0.040***	-0.034***	-0.037***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
Crime	-0.010***	-0.008***	-0.009***	-0.008***	-0.011***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ln HH with car	-0.238***	-0.225***	-0.232***	-0.226***	-0.232***	-0.230***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
State: Hidalgo (ref: Cd. de México)	-0.325***	-0.337***	-0.340***	-0.339***	-0.321***	-0.321***
	(0.027)	(0.027)	(0.026)	(0.027)	(0.029)	(0.030)
State: Mexico	-0.289***	-0.297***	-0.301***	-0.299***	-0.283***	-0.284***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Accessibility car	0.108***	0.103***	0.107***	0.102***	0.101***	0.093***
	(0.012)	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)
Accessibility PT	0.033***	0.045***	0.036***	0.044***	0.056***	0.066***
	(0.011)	(0.006)	(0.007)	(0.006)	(0.011)	(0.011)
$\hat{\beta}_2$			0.201***	0.053**	0.130***	0.099***
			(0.058)	(0.025)	(0.034)	(0.034)
$\hat{\mu}$			0.186	1.84*	3.54*	2.26**
			(1.09)	(1.01)	(1.86)	(1.01)
$\hat{\gamma}$					0.387***	0.337***
					(0.109)	(0.102)
Employment opportunity	All	m	All	m	All	m
Structural controls	Yes	Yes	Yes	Yes	Yes	Yes
Dummy year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	781,898	781,898	781,898	781,898	781,898	781,898
Log-Likelihood	158,423.3	163,774.0	162,065.2	164,015.3	164,776.5	166,176.0
AIC	-316,776.7	-327,478.0	-324,056.4	-327,956.6	-329,477.0	-332,275.9
RMSE	0.1976	0.1962	0.1967	0.1962	0.1960	0.1956
RESET, statistic	13572.6	11635.4	12688.6	11700.2	14445.3	13501.3
VIF, mean	3.597	3.296	3.270	3.288	3.371	3.439
B-P test, statistic	62268.8	60962.5	63304.0	61230.0	61119.6	60196.9

Note:
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. Standard errors clustered by post code ID. OLS = Ordinary least squares. NML = Non-linear maximum likelihood. RMSE = Root mean squared error. RESET = Ramsey reset test. VIF = Variance of inflation factor. B-P = Breusch-Pagan test.

Table 7.3: Main public transport network (MPTN) accessibility to employment in the hedonic property value model.

mated coefficient of having an educational establishment within 20 minutes , the proportion of streets without public light, crime, and the proportion of households with a car, denote a relationship and association with property value in the expected direction and are consistently significant across all alternative models. The presence of a public administration establishment is only significant in Model 2 (M2:S2) and Model 3 (M3:S2), showing a positive sign. The proportion of streets with trees is significant and positively associated with the value of properties from M1:S2 to Model 4 (M4:S2). Having a park within 15 minutes walking distance is suggested to have a consistent significant and negative impact in all alternative specifications of Set 2. The direction of this coefficient is not as expected according to the hedonic property literature (Loret de Mola et al., 2017). Accessibility to employment by car is significant and in the direction expected, implying an increase in house values of between 9.3% and 10.8% for an additional standard deviation.

In M1:S2, MPTN accessibility shows a positive and significant relationship. The magnitude of the coefficient suggests that the value of land is expected to raise by 3.3% for an increase in MPTN accessibility equivalent to one standard deviation. In M2:S2, the relationship between the MPTN accessibility and land value is also positive and significant. The coefficient indicates that an improvement by one standard deviation is related to a 4.5% increase in the value of residential land. It is worthwhile noticing that M2:S2 substantially improves the goodness-of-fit (Log-Likelihood/AIC/RMSE) and all the model's diagnostic measures (RESET/VIF/B-P test) compared to M1:S2.

In M3:S2, the magnitude of the coefficient is slightly larger than in M1:S2, suggesting an increase of 3.6% in the value of land for every additional standard deviation in MPTN accessibility. It is worthwhile noting that the distance deterrence parameters $\hat{\beta}_2$ and $\hat{\mu}$ in the accessibility measure substantially deviate from those estimated exogenously in the gravity model (0.201 in M3:S2 versus 0.044, and 0.186 in M3:S2 versus 1.654, respectively). Furthermore, M3:S2 considerably improves the goodness-of-fit and practically all the diagnostic measures compared to its counterpart M1:S2 (except for the B-P test statistic). The results of M4:S2 show that the magnitude of the relationship is very similar to that estimated in M2:S2. In M4:S2, each accessibility unit increase generated by the MPTN is associated to a 4.4% gain in land value. The goodness-of-fit, RESET statistic and Breusch-Pagan test statistic improve compared to the generic accessibility specification used in M3:S2 whilst the VIF remains practically unchanged.

Interestingly, the results of M4:S2 and M2:S2 are very similar. Specifically, this time the accessibility parameters β_2 and μ are considerably closer to those estimated in the gravity transport model (0.053 in M4:S2 versus 0.044, and 1.84 in M4:S2 versus 1.65) than those estimated in M3:S2. It is worthwhile remembering that the parameters used to construct the accessibility measure used in M2:S2 are completely exogenous. This is because they are based on commuting flow data only and estimated in a different modelling framework (a gravity model). Bearing this in mind, it is remarkable the accuracy level and relative stability obtained when the exogenous accessibility parameters are entered in an accessibility measure which acknowledges the characteristics of the labour market on the side of both the supply and demand (M2:S2) and benchmarked against an endogenous method that optimizes these parameters using detailed information from the residential land market (M4:S2).

The coefficient for MPTN accessibility in M5:S2 is considerably higher than in M1:S2 and M3:S2, suggesting an increment of 5.6% in the value of land for one standard deviation increase. Moreover, the goodness-of-fit measures and the heterogeneity of the residuals are more favourable than in the previous equivalent models (M1:S2 and M3:S2). However, there are not improvements in the model specification (RESET) or collinearity (VIF). As in M5:S2, in M6:S2 the coefficient of accessibility to employment provided by the MPTN is larger than its counterparts (M2:S2 and M4:S2). One additional standard deviation increase is expected to have a positive effect on land value of about 6.6%. Also, this model specification substantially improves the explanatory power and reduces heteroskedasticity compared to M2:S2 and M4:S2. Nonetheless, there are no improvements in the model specification or collinearity.

It should be noted that the coefficient estimates of some of the neighbourhood variables change according to the MPTN accessibility measure employed. For instance, the presence of a major public administration establishment is significant only in M1:S2 and M3:S2, that is, where MPTN accessibility considers all types of employment. Yet, this becomes insignificant when γ is allowed to vary. By contrast, this variable is insignificant in the presence of the accessibility measure that restricts employment to matching opportunities in all cases. These observations make sense, as the availability of a major public administration establishment can also add up to accessibility. Therefore, a comprehensive accessibility measure would capture the effect of this neighbourhood characteristic as well. A further similar effect occurs with the proportion of trees on streets. While this variable is significant in the first four models (M1:S2 to M4:S2), this is not significant when γ varies in M5:S2 and M6:S2. This could be explained by the fact that the proportion of trees follows a spatial structure which is correlated to the location of houses with respect to distance to the CBD or employment opportunities (Pearson correlation coefficient is -0.74 for distance to the CBD and > 0.65 for accessibility PT in M5:S2 and M6:S2). Therefore, this measure could be capturing some accessibility effects in the initial model specifications which were later reflected in more comprehensive measures.

Overall, the results in Table 7.3 show that the models that use accessibility measure parameters estimated simultaneously in NLMs (M3:S2 to M6:S2) perform better than the models where the accessibility parameters were estimated a priori in terms of goodness-of-fit (M1:S2 to M2:S2). It is interesting to note, that considering the type of household and employment in accessibility measures increases the explanatory power, and reduces heteroskedasticity and possible misspecification issues in all cases when compared to models including accessibility measures that assume the homogeneity of these elements. Considering the results presented in Table 7.3, the preferred specification is given by M6:S2 at this step. This is because it produces the best goodness-of-fit, the lowest heterogeneity of error variance, and a good balance of RESET/VIF/B-P test values. In a further modelling stage, the performance of this accessibility model is contrasted to the specification in M5:S2 introducing further spatial and temporal considerations.

7.3.4 Multilevel models over time

With a hedonic model which incorporates adequate controls and accounts for MPTN accessibility, the selected specifications in M5:S2 and M6:S2 are further examined by considering possible temporal fluctuations and spatial effects.

Before proceeding to the estimation of all temporal stages in multilevel and spatial multilevel models, the most recent stage (Stage 7) is fitted in different regression frameworks in a third set of models (Set 3, S3). OLS models, MLMs, and BYM models are fitted for each alternative MPTN accessibility formulation (as in model M5:S2 and M6:S2). The purpose of this is twofold. First, it is useful to verify the assumption of independence of the random effects in the 'unstructured' multilevel model at the upper-level (ν_j in Eq. (7.5)). This is required to confirm or reject the need for a statistical technique that explicitly models spatial autocorrelation at this level. Second, this intermediate step allows for a comparison of the estimated coefficients between the different regression techniques. This can illustrate some possible implications of ignoring the structure of the data in relation to the research objective.

The results of Set 3 are shown in Table 7.4. A first aspect to note is that the neighbourhood variables that were not significant in M5:S2 and M6:S2 (of the previous modelling stage), namely the presence of a public administration establishment, and relative number of streets with trees, remain insignificant in all modelling frameworks and for both MPTN accessibility specifications (i.e. 'Accessibility: All' and 'Accessibility: m'). Second, the availability of a park maintains a negative sign in all frameworks. This is significant in MLMs and in only one of the OLS models (M4:S3). Yet, it is non-significant in the BYM models. This means that the apparent negative effects of having a park in nonspatial models, in fact, resemble some characteristics of the spatial structure.

	Access PT: All			Access PT: m		
	M1:S3 (OLS)	M2:S3 (MLM)	M3:S3 (BYM)	M4:S3 (OLS)	M5:S3 (MLM)	M6:S3 (BYM)
Edu. facility: Yes	0.046***	0.063***	0.028**	0.040***	0.058***	0.026**
	(0.016, 0.075)	(0.034, 0.093)	(0.005, 0.051)	(0.011, 0.070)	(0.028, 0.087)	(0.004, 0.049)
Public admin.: Yes	0.006	-0.007	0.007	0.002	-0.010	0.006
	(-0.028, 0.040)	(-0.033, 0.019)	(-0.014, 0.027)	(-0.031, 0.035)	(-0.037, 0.016)	(-0.015, 0.027)
Street with trees (%)	-0.011	-0.034	-0.017	-0.017	-0.039	-0.022
	(-0.088, 0.066)	(-0.085, 0.017)	(-0.057, 0.022)	(-0.095, 0.061)	(-0.090, 0.012)	(-0.062, 0.018)
Street light NA (%)	-0.367***	-0.250***	-0.072	-0.387***	-0.262***	-0.076
	(-0.570, -0.164)	(-0.378, -0.123)	(-0.172, 0.028)	(-0.595, -0.179)	(-0.389, -0.134)	(-0.176, 0.025)
Park: Yes	-0.022	-0.056***	-0.001	-0.027*	-0.059***	-0.002
	(-0.050, 0.006)	(-0.079, -0.034)	(-0.018, 0.016)	(-0.055, 0.001)	(-0.081, -0.037)	(-0.019, 0.015)
Crime	-0.018***	-0.027***	-0.014**	-0.016***	-0.026***	-0.014**
	(-0.025, -0.011)	(-0.034, -0.020)	(-0.026, -0.002)	(-0.023, -0.009)	(-0.033, -0.019)	(-0.026, -0.002)
ln HH with car	-0.269***	-0.384***	-0.198***	-0.266***	-0.382***	-0.198***
	(-0.308, -0.230)	(-0.405, -0.363)	(-0.218, -0.178)	(-0.305, -0.227)	(-0.403, -0.361)	(-0.218, -0.178)
State: Hidalgo (ref: Cd. de México)	-0.346***	-0.470***	-0.127	-0.352***	-0.476***	-0.136
	(-0.424, -0.269)	(-0.585, -0.356)	(-0.303, 0.048)	(-0.431, -0.273)	(-0.590, -0.361)	(-0.309, 0.037)
State: México	-0.335***	-0.321***	-0.144***	-0.338***	-0.321***	-0.145***
	(-0.379, -0.291)	(-0.347, -0.295)	(-0.198, -0.089)	(-0.382, -0.295)	(-0.347, -0.295)	(-0.199, -0.092)
Accessibility car	0.129***	0.166***	0.034**	0.123***	0.160***	0.030**
	(0.104, 0.155)	(0.148, 0.184)	(0.005, 0.062)	(0.096, 0.149)	(0.141, 0.178)	(0.002, 0.058)
Accessibility PT	0.082***	0.073***	0.046***	0.087***	0.082***	0.067***
	(0.058, 0.106)	(0.055, 0.092)	(0.032, 0.060)	(0.064, 0.110)	(0.063, 0.102)	(0.050, 0.084)
$\hat{\sigma}_y^2$	0.0438	0.0231	0.0231	0.0437	0.0231	0.0231
$\hat{\sigma}_v^2$	-	0.0431	0.0058	-	0.0429	0.0066
$\hat{\sigma}_\upsilon^2$	-	-	0.0406	-	-	0.0387
Observations	104569	104569	104569	104569	104569	104569
Groups	1706	1706	1706	1706	1706	1706
AIC	-30377	-91723	-	-30559	-91727	-
DIC	-	-	-95873	-	-	-95873
RMSE	0.2092	0.1508	0.1509	0.2090	0.1508	0.1509
Moran's I	-	0.3613***	0.8475***	-	0.3666***	0.8433***

Note:
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1; In Bayesian models 99%, 95% or 90% of the highest posterior density (HPD) credible interval does not include zero, respectively. The regression coefficient reported in Bayesian models is the posterior marginal mean with corresponding credible intervals in parenthesis (2.5%, 97.5%). All models include structural and year controls. OLS = Ordinary least square. MLM = Multilevel model. BYM = Besag-York-Mollié model. Moran's I estimated on the upper-level random effects.

Table 7.4: Regression estimate comparison across various models for Stage 7 (2018 to 2019) for PT accessibility.

There are further relevant observations across the different modelling frameworks presented in Table 7.4. The availability of an education facility shows a positive and significant coefficient in all the models regardless of the modelling technique or the accessibility specification used. Yet, the magnitude of this variable differs, being larger in MLMs than in OLS and BYM models. Furthermore, the estimate is higher in the presence of the MPTN accessibility measure that considers all opportunities than the alternative formulation comparing corresponding models on the left- and right-hand side panels. The proportion of streets without public light shows a negative sign in all models but it is significant only in OLS models and MLMs. This time, the magnitude is higher in OLS models than in MLMs. For this subset, crime is negative and significant in all regressions. As in the case of education facility, the estimate is greater in MLMs than in OLS and BYM models. This size of the coefficient is consistent in the presence of either accessibility measure.

Similarly, the proportion of households with a car is negative and significant in all cases. Also, the coefficients are substantially larger in OLS models than in the alternative techniques, and the estimates are steady in the presence of alternative accessibility formulations. Houses in Hidalgo show a significant and negative sign in OLS models and MLMs. Yet, this is not significant in BYM models. This can be understood by the fact that the simple locational differences are captured by the spatial random effects at the upper-level. An additional aspect to note is that the magnitude of the estimate in MLMs is very high. For example, M2:S3 suggests that residential land in Hidalgo is almost 50% cheaper than in Ciudad de Mexico, with everything else being equal. Meanwhile, OLS models suggest a difference of about 35% less. The estimate for the state of Mexico is negative and significant in all techniques. The estimates are more or less similar between OLS models and MLMs regardless of the accessibility specification. Yet, the spatial random effects capture a considerable portion of these coefficients. This is shown in Model 3 (M3:S3) and Model 6 (M6:S3), where the magnitude shrinks to approximately half.

Accessibility variables display similar patterns as those observed previously across the various frameworks tested, as shown in Table 7.4. While accessibility to employment by car is positive and significant in all models, the magnitude is higher in OLS models and MLMs (especially in the latter). However, most of the expected changes on land value are captured by the spatial random effects at the upper level in M3:S3 and M6:S3. For instance, while the association in both MLMs is suggested to be in the order of 16% to 17% for one standard deviation increase, this is around 3.4% to 3.0% in BYM models.

The estimate for the key variable—MPTN accessibility—is positive and significant and it does not show fluctuations as dramatic as the previously discussed variable across the different techniques. Still, the variations are relevant. The largest estimate in both of the alternative measures is produced by models fitted by OLS. These models suggest gains of 8.2% and 8.7% for one unit increase in MPTN accessibility considering all employment and only matching opportunities, respectively. The magnitude of these effects is reduced to 7.3% and 8.2% in MLMs, accordingly. While the size of the coefficient shrinks for both accessibility measures in BYM models, the proportional change differs. For instance, in the first case (all opportunities), the coefficient practically halves compared to OLS , from 8.2% to 4.6%, whilst in the second case (matching opportunities) it falls from 8.7% to 6.7%. This implies that in the former case the spatially structured random effects capture more of the variance at the upper level than the in the latter case. This is also reflected in the estimated variance of the spatial component (σ_{υ}^2), since the estimate value is higher in M3:S3 than in M6:S3. In sum, it is shown that both OLS models and MLMs tend to overestimate the willingness to pay for some locational characteristics. Yet, it is not possible to identify a systematic inflation pattern.

Regarding the fit statistics and diagnostic tests, the models display substantial differences as presented at the bottom section of Table 7.4. The variance at the individual level is higher in OLS models, at around 0.044. This considerably decreases in both MLMs and BYM models to approximately 0.023. Concerning the variance of the upper-level elements, both MLMs display very similar values (≈ 0.043). In BYM models the upper-level variance is largely absorbed by the spatially structured elements since σ_{ν}^2 shrinks, while σ_{υ}^2 takes a larger value. In terms of the explanatory power, OLS models produce much higher AIC values than MLMs. Considering the analogous measure in BYM models, the DIC value is the highest compared to the corresponding AIC values. In terms of accuracy, OLS models result in substantially higher RMSEs than MLMs and BYM models. These values are slightly lower for MLMs than BYM models in both cases, and the value is very similar for both accessibility measures. The Moran's I statistic for MLMs is near 0.36 and the p-value is lower than 0.05. These results confirm the presence of spatial autocorrelation between the random effects in MLMs. Therefore, it is required to explicitly model autocorrelation at the upper level. This is in line with expectations since many characteristics at the zonal level in the residential land market are shared between neighbouring areas. As expected, the autocorrelation statistic is significant and much larger in BYM models.

Spatial multilevel models

Table 7.5 presents a summary of the results for each of the MPTN temporal stages according to the BYM spatial multilevel model denoted in Eq. (7.7). Each stage includes the two selected model specification resulting from the previous section, namely one including the MPTN accessibility to employment measure which considers all types of employment (columns under ‘All’); and another that restricts matching employment opportunities (columns under ‘m’). To increase visibility, structural and year controls are not shown.

	Stage 1 (Before 2010)		Stage 2 (2010 to 2012)		Stage 3 (2012 to 2013)		Stage 4 (2013 to 2015)		Stage 5 (2015 to 2016)		Stage 6 (2016 to 2018)		Stage 7 (2018 to 2019)	
Accessibility:	All	m	All	m	All	m	All	m	All	m	All	m	All	m
Edu. facility: Yes	0.017*	0.015*	0.021**	0.019**	0.017*	0.014	0.021**	0.018*	0.022**	0.018*	0.027**	0.023**	0.028**	0.026**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.010)	(0.012)	(0.012)
Public admin.: Yes	-0.004	-0.005	0.006	0.005	0.002	0.001	0.010	0.009	-0.003	-0.004	0.010	0.008	0.007	0.006
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
Street with trees (%)	0.010	0.007	-0.004	-0.007	-0.005	-0.009	-0.006	-0.009	0.006	0.003	-0.032*	-0.035*	-0.017	-0.022
	(0.015)	(0.015)	(0.016)	(0.016)	(0.018)	(0.018)	(0.017)	(0.017)	(0.020)	(0.019)	(0.019)	(0.018)	(0.020)	(0.020)
Street light NA (%)	0.058	0.058	-0.001	-0.000	0.059	0.059	0.041	0.040	0.035	0.035	-0.001	-0.002	-0.072	-0.076
	(0.038)	(0.038)	(0.041)	(0.041)	(0.045)	(0.045)	(0.042)	(0.043)	(0.048)	(0.047)	(0.047)	(0.046)	(0.051)	(0.051)
Park: Yes	0.009	0.009	-0.003	-0.003	0.009	0.008	0.004	0.003	0.005	0.004	0.007	0.006	-0.001	-0.002
	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
Crime	-0.003	-0.003	-0.012***	-0.012***	0.002	0.002	0.004	0.004	-0.003	-0.003	-0.001	-0.001	-0.014**	-0.014**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)
ln HH with car	-0.183***	-0.182***	-0.173***	-0.172***	-0.193***	-0.191***	-0.198***	-0.197***	-0.191***	-0.187***	-0.208***	-0.205***	-0.198***	-0.198***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)
State: Hidalgo														
(ref: Cd. de México)	-0.100	-0.104	-0.100	-0.105	-0.070	-0.072	-0.140	-0.144	-0.123	-0.123	-0.128	-0.129	-0.127	-0.136
	(0.068)	(0.067)	(0.079)	(0.077)	(0.090)	(0.089)	(0.095)	(0.095)	(0.081)	(0.079)	(0.101)	(0.098)	(0.089)	(0.088)
State: México	-0.098***	-0.099***	-0.062***	-0.064***	-0.090***	-0.091***	-0.118***	-0.119***	-0.137***	-0.137***	-0.122***	-0.124***	-0.144***	-0.145***
	(0.018)	(0.018)	(0.019)	(0.019)	(0.021)	(0.021)	(0.020)	(0.020)	(0.025)	(0.025)	(0.026)	(0.025)	(0.028)	(0.027)
Accessibility car	0.036***	0.033***	0.031***	0.027**	0.031**	0.026**	0.033***	0.029**	0.034**	0.028**	0.023*	0.018	0.034**	0.030**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)	(0.014)
Accessibility PT	0.036***	0.049***	0.046***	0.065***	0.044***	0.060***	0.048***	0.065***	0.052***	0.070***	0.057***	0.075***	0.046***	0.067***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.009)
$\hat{\sigma}_y^2$	0.0184	0.0184	0.0182	0.0183	0.0186	0.0186	0.0213	0.0213	0.0200	0.0200	0.0224	0.0223	0.0231	0.0231
$\hat{\sigma}_\nu^2$	0.0034	0.0028	0.0032	0.0041	0.0042	0.0041	0.0038	0.0040	0.0054	0.0035	0.0056	0.0056	0.0058	0.0066
$\hat{\sigma}_\upsilon^2$	0.0235	0.0241	0.0296	0.0262	0.0319	0.0312	0.0310	0.0303	0.0338	0.0382	0.0411	0.0371	0.0406	0.0387
Observations	146779	146779	116866	116866	79200	79200	118206	118206	73903	73903	142375	142375	104569	104569
Groups	1784	1784	1788	1788	1699	1699	1790	1790	1656	1656	1799	1799	1706	1706
Log-Likelihood	83239	83245	66184	66195	43701	43707	57736	57745	37973	37984	66474	66485	46744	46753
DIC	-168836	-168833	-134692	-134694	-89382	-89379	-117856	-117854	-77957	-77944	-135677	-135669	-95873	-95873
RMSE	0.1349	0.1349	0.1343	0.1343	0.1353	0.1353	0.1452	0.1452	0.1402	0.1402	0.1487	0.1487	0.1509	0.1509
Moran's I^a	0.7694***	0.7670***	0.7894***	0.7831***	0.7792***	0.7772***	0.8138***	0.8105***	0.8053***	0.8094***	0.8475***	0.8489***	0.8475***	0.8433***

Note:
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1; in Bayesian models 99%, 95%, or 90% of the highest posterior density (HPD) credible interval does not include zero, respectively. Posterior marginal mean reported with corresponding standard deviation in parentheses. All models include structural and year controls. DIC = Deviance information criterion. RMSE = Root mean squared error. Moran's I statistic on the upper-level random effect.

Table 7.5: Spatial multilevel model results for the Besag-York-Mollié (BYM) specification.

One of the most visible aspects in this modelling set is that many of the neighbourhood attributes are not significant. The variables that were not significant in M5:S2 and M6:S2 (of the previous modelling subsection), namely the presence of a public administration establishment and the proportion of streets with trees, remain insignificant in all of the MPTN stages. The explanation offered above regarding the correlation with accessibility may be applicable for these models too. One exception to this observation is Stage 6, where the proportion of streets with trees is negative and significant. The proportion of block fronts without streetlights or a park become insignificant in all temporal stages, while these were both negative and significant in regression frameworks that do not include spatial effects (presented in Table 7.4).

The presence of an educational facility remains significant in practically all stages (except for model 'm' in Stage 3) and the direction of it is as expected in all cases. Crime, as measured in this study, is not a robust covariate in the presence of spatial effects. This variable is significant only in Stage 2 and Stage 7 (in both specifications 'All' and 'm'). The implicit indifference to crime in most of the stages is in line with previous studies in the context of Mexico, which report this variable as neutral (or even positive in some specific states outside of GMC) (Nuñez et al., 2017). The authors suggest that illegal earning opportunities can outweigh crime as a disamenity. The proportion of households having at least one car displays a negative and significant coefficient in all stages. Furthermore, the magnitude of the estimate is within a small range. One percent increase in car ownership is associated with a decline of between 17% and 21% in the value of land. Also, a modest downward trend in time can be observed. This is more visible when comparing the first two stages against the last two. In the former, the size of the raw coefficient is between 0.17 and 0.18, while in the latter it is between 0.20 and 0.21. As the negative effect can be related to road congestion, this trend can indicate a growing concern of house-buyers regarding this issue.

The estimate for dwellings located in Hidalgo compared to those in *Ciudad de México* is not significant in any of the stages. As before, the interpretation of this is that the differences are entirely explained by the spatial random component at the upper level. By contrast, the estimate for the value of residential land in the state of Mexico is negative and significant compared to that in Ciudad de México. This is as expected. The reading of this result in light of the lack of significance for Hidalgo is that, in addition to the spatial local heterogeneities accounted in the zonal elements, there is a generalizable discount in the value associated to the value of land in the state of Mexico. This could be explained by unobserved locational characteristics, for example by the quality of the public services, quality of urban amenities, availability of shopping centres, or air quality. Furthermore, while the difference in the first two stages is between 6% and 10%, this is between 12% and 15% in the last two. This shift may indicate growing divergences between these two states in the urban terms previously discussed. It should be noted that these differences are very consistent in the presence of either MPTN accessibility specification. This implies that the observed disparities are not related to MPTN accessibility to employment.

In contrast to many of the neighbourhood attributes discussed above, the transport-related variables are robust after modelling the spatial structure over the various temporal stages examined. Accessibility to employment by car is positive and significant in practically all temporal stages, except for model m in Stage 6. The estimates suggest a rise in land value of between 2.3% (in Stage 6) and 3.6% (in Stage 1) for one standard deviation increase. In general, the coefficients are similar between them and do not display a clear temporal tendency. It is worthwhile noting that the estimate is consistently smaller in models that consider matching employment opportunities (‘m’) in MPTN accessibility measures than in models using a more generic measure (‘All’).

MPTN accessibility shows a positive and significant relationship with land value in all temporal stages. In contrast to the relatively small variability for accessibility by car, these estimates display some fluctuations.[4] Considering all employment opportunities, MPTN accessibility is associated to a value increment of 3.6% in Stage 1 for one standard deviation improvement, gradually increasing up to 5.7% in Stage 6. From Stage 2 to Stage 4, the association is fairly stable at around 4.5% (in Stage 3), going up in Stage 5 to 5.2%. In the final stage, the estimate falls back to 4.6%. Considering one standard deviation gain in the ‘matching’ MPTN accessibility specification (‘m’), the magnitude of the relationship increases from 5.0% in Stage 1 to 6.5% in Stage 2. Then, it remains roughly stable in the next two stages (Stage 3 and Stage 4) only to increase again to 7.0% and 7.5% in Stage 5 and Stage 6, respectively. Lastly, it goes back to 6.7% in the last period observed. Overall, there is a minor upward temporal trend which suggests that the accessibility benefits of public transport infrastructure are valued at a higher rate in later stages.

Regarding the models’ goodness-of-fit and diagnostic measures, there is not a substantial difference between the MPTN accessibility specifications used. Firstly, the upper-level elements’ variance is discussed given the interest in the between-zone differences. In all models it is observed that this is largely captured by the spatially structured elements since the values for σ^2_ν are considerably smaller than σ^2_υ. This is in line with the contrasts observed in Table 7.4 between MLMs and BYM models.

Although the variance of spatially structured random errors is similar between

[4]It should be noted that it is not appropriate to make direct comparisons between the magnitude of the coefficients since they follow different specifications and even though they are scaled by the standard deviation , the scale maintains differences.

'All' and 'm' models, this is lower for the latter in most of the stages (five out of seven). This implies that the matching accessibility specification explains slightly more of the spatial heterogeneity than its counterpart. Concerning the goodness-of-fit of the models, the estimated marginal log-likelihood is slightly higher for all models 'm.' However, this improvement is not consistent when the DIC is considered. When this is the case, this measure favours the 'All' models. In terms of accuracy, both models produce the same RMSE. The Moran I statistic over the upper-level elements may reflect the extent to which each of the alternative MPTN accessibility specifications reduce spatial heterogeneity. It is observed that spatial autocorrelation is slightly lower for 'm' models only in the first four stages and the last one.

All in all, the results produced by both MPTN accessibility formulations are not radically different in the spatial multilevel framework. Since the performance of the alternative accessibility measures are not consistently superior across the temporal stages examined, the measure that reduces complexity and demands less data should be preferred for now. Therefore, in the remainder of this section, further observations are based on accessibility measures that do not constrain employment opportunities according to the MPTN accessibility measure specified in Eq. (7.1).

7.3.5 The potential for LVC

This section illustrates the potential for LVC derived from improvements in accessibility to employment resulting from the expansion of the MPTN. Here, the focus is on the changes introduced by the implementation of the bus rapid transit (BRT) Line 7 of the *Metrobus* system (MB-L7): the most recent stage (Stage 7). For this purpose, two scenarios are estimated. The first is a full scenario which considers all post code zones that received MPTN accessibility benefits. The second is a partial scenario that considers only the 15 most benefited post code zones. The aim is to show the degree of uncertainty between a best-case scenario and a conservative alternative.

The partial scenario also incorporates elements from LVC literature. Specifically, it acknowledges that the introduction of a LVC strategy in a context with limited prior experience would be more feasible if it initially affects a small number of areas, according to the proposed gradual approach suggested by international organizations (Suzuki et al., 2015). In addition, the link between the benefits for the infrastructure and a potential recovery strategy is localized and is, therefore, more evident, which would be more likely to be supported by households if "they understand what they are receiving in return" (Vadali et al., 2018, pp. C–4).

The change in MPTN accessibility is calculated as $\Delta A_j^{\mathrm{PT}} = A_j^{\mathrm{PT}_{after}} - A_j^{\mathrm{PT}_{before}}$. This accounts for the benefits introduced in Stage 7 (after) compared to Stage 6 (before). In line with the results reported in the previous section, the accessibility specification used considers all employment opportunities (Eq. (7.1)). The full scenario identifies 447 post code zones that received MPTN accessibility benefits ($\Delta A_j^{\mathrm{PT}} > 0$). The magnitude of these is on average equivalent to a 0.02 standard deviation increase with a maximum increase of 0.43. This change is equivalent to a 1.1% mean gain with a maximum mean gain of 22.0%. Figure 7.1 illustrates the spatial distribution of the accessibility benefits.

Source: the Author based on own calculations. *Note:* Accessibility measure considers all employment opportunities.

Figure 7.1: Absolute accessibility to employment gains after the MB-L7 extension in Stage 7 featuring top benefited post code zones (reference is Stage 6).

The observed accessibility effects should be interpreted according to the land use context around the transport infrastructure. The MB-L7 is located at the core of the MPTN and runs along the main business boulevard (Av. *Paseo de la Reforma*). From Figure 7.1, it is important to note that the access gains are not homogeneous. For instance, it is shown that the areas at the north end of the of the line received the largest benefits, e.g. in *Tepeyac Insurgentes*, *Industrial*, and *Estrella*. This is because the transport infrastructure generated a direct link between these areas (which are not dense in employment) to zones with a high employment density. The transport implementation also brought important additional benefits to areas that already had an important number of employment opportunities, e.g. *Polanco III*, *Juárez*, and

Cuauhtémoc. The post code zones considered in the partial scenario are highlighted by a white boundary in Figure 7.1. By definition, these received the largest MPTN accessibility benefits.

The marginal willingness to pay (WTP) is recovered from the individual implicit prices for a change in accessibility according to Equation (5.14), as detailed in Chapter 5. The total willingness to pay (TWTP) can be obtained from the sum of the individual WTP of each affected individual (Freeman et al., 2014). This estimate does not consider transactions costs. From the discussion in Chapter 4, it was observed that the institutional housing system in Mexico operates as a financial enable r smoothing potential frictions in the housing market, for example by enabling access to loans to different segments of the population. Thus, there were no major conditions undermining the overall efficiency of the market. Accordingly, the estimate is an upper bound, as discussed in Chapter 3 and Chapter 5.

The TWTP should be considered as the individual characteristics of each of the dwellings affected (Gjestland et al., 2014). Since such information is not available for all the housing stock, as is usual in practice, L. O. Taylor (2017) references the use of mean or median as a common practice. Here, the estimates follow a similar approach to those employed in Feng & Humphreys (2018) and assume a representative house value according to the median of the SHF records in the jth post code zone in Stage 6, median($P_{i,j}^{\text{Stage 6}}$). Following this, the census data (2010 Population Census) is used to account for the size of the housing stock affected within each post code zone.

The results for the full scenario indicate that the individual WTP ranges from MXN$25 to MXN$143,000 with a mean value of MXN$3,000. In this scenario, the TWTP is MXN$4,321 million (GBP£157 million) at current prices. This amount is equivalent to 1.5 times the total initial cost of the MB-L7 (the cost of the infrastructure is MXN$2,800 million, including civil works and the bus fleet). In the partial scenario, the mean MPTN accessibility change is 0.25 standard deviations which represents an average gain of 13%. The housing stock included in this scheme is made up of 59,928 units. Under this scenario, the individual mean WTP is MXN$38,000, the minimum is MXN$7,000, and the maximum is MXN$143,000. Here, the TWTP is MXN$1,900 million (GBP£69 million). In the partial scenario, the number properties considered are relatively small. Yet, the estimated increase represents about two thirds of the total initial costs of the MB-L7 (69%).

7.4 Discussion

7.4.1 Location-based accessibility in the hedonic model

In the first instance, the results indicate a measure of accessibility to employment by car as adequate, and it is able to replace the commonly used measure of distance to the CBD. This can be explained from various angles. For instance, some authors suggest that the location of a focal point in cities can change over time (Plaut & Plaut, 1998). Therefore, the location of the CBD may not be the optimal surrogate measure of *general* accessibility in hedonic models. This may be related to the findings of empirical studies for GMC which suggest a shift from a dominant mono-centric structure to a poly-centric one (Fernández-Maldonado et al., 2014; Montejano-Escamilla, 2015). In either case, a location-based measure has been shown to be more appropriate than the distance to the CBD. This is in line with previous empirical findings (Ahlfeldt, 2011).

The analysis of the adequacy of the location-based accessibility measures suggests that the observed travel behaviour in the MPTN captured in the 2017 Travel Survey corresponds to a perception of accessibility where employment opportunities are evaluated according to the matching characteristics of the supply and demand of the labour market, in line with theory (Geurs & van Wee, 2004; Levinson & Wu, 2020; Páez et al., 2012). This is based on the result which shows that the accessibility parameters estimated a priori in a trip-distribution model substantially improve the performance of the hedonic property value model when they are entered into an accessibility specification that distinguishes the characteristics of the labour market (i.e. M1:S2 versus M2:S2). This idea is further supported by the surprisingly similar results obtained between the hedonic model that employs accessibility parameters estimated in the gravity-model and the model that optimises these parameters endogenously (i.e. M2:S2 versus M4:S2).

These results suggest that the transferability of accessibility parameters from the gravity-transport model to the hedonic model is more adequate if employment opportunities are matched to the characteristics of the residents at the origin. These results are partially in line with previous empirical literature. For example, Ahlfeldt & Wendland (2016) reports that the spatial decay parameter (referred to as β_2 here) estimated in either a land value- or commuting flow-based model is equivalent in the context of Germany even when considering all types of employment in gravity-type accessibility measures. These contrasting results may be due to a wider range of inequality in terms of social mobility in the context of Mexico than in the German context. These contrasting results may reflect the low social mobility in the context of Mexico compared to that in Germany. In the latter case, the expectations of accessing employment in a higher band may be higher than in the latter. This implies that making a distinction in accessibility measures is more relevant in economies with high economic and social inequality rates.

Taking into account the characteristics of the labour market addressed here, the accessibility measure improves the OLS models and NLMs but do not improve those which include local heterogeneity, namely MLMs and BYM models so clearly. For example, the models which constrain the employment supply to those matching the education level of the demand at the residential location increase the explanatory power, reduce heteroskedasticity, and reduce possible misspecification issues in comparison to those assuming homogeneity of opportunities in OLS models (as shown in the second set of results presented in Table 7.3). Yet, the predictive ability of models which consider matching employment opportunities in accessibility measures is not consistently superior to a more generic alternative formulation in models accounting for local heterogeneity (i.e. both MLMs and BYM models). Still, it is possible to observe a moderate reduction in the variance of the spatial heterogeneity reflected in a lower Morans' I statistic in models type m from Stage 1 to Stage 4 (shown in Table 7.5). One reason for this result may be that models which include spatially structured elements (i.e. BYM models) capture omitted variables via neighbouring areas. Therefore, multilevel and spatial multilevel models can compensate the performance of models which do not consider matching employment opportunities.

These are relevant implications regarding the temporal considerations of the analysis. While the results suggest that the coefficients produced through accessibly by car remain fairly stable across the temporal stages, MPTN accessibility displays a minor upward trend in the period studied. This implies that the service generated by the MPTN infrastructure becomes more valuable in the residential land market over time. First, this finding calls for increased attention regarding the classic hedonic model assumption of identical preferences of home-buyers over time usually adopted in empirical studies. Therefore, case by case sensitivity analyses to assess this assumption are important. This finding is also in line with the criticism of quasi-experimental hedonic models for environmental valuation (Klaiber & Kuminoff, 2014), which rely on this assumption.

Concerning the econometric aspects of these models , it is observed that many of the neighbourhood variables intended to function as controls at the local level are rendered nonsignificant when modelling spatial effects. This finding is similar to previous empirical studies, where it was reported that these types of variables become less relevant in spatial models than in OLS models (G. Lee, Cho, & Kim, 2016). This could be attributed to the fact that spatially structured elements carry out information about the characteristics of the neighbourhood. This suggests that control variables play an uncritical role in spatial hedonic models. Therefore, this may readdress the crucial role of these variables in traditional hedonic methods as omitted bias (Parmeter & Pope, 2013).

7.4.2 The implications of the MPTN in the housing market system

There are some considerations regarding the net benefit estimates. First, it could be argued whether the changes in accessibility between Stage 6 and Stage 7 are localized or not. In the full scenario, it was found that 20% of the housing stock is affected to some degree. Yet, as shown by the results, at least half of the estimated changes are very small, e.g. the median WTP in the full scenario is MXN$500 (equivalent to only GBP£18). Therefore, these may not represent a 'market wide' shift in the housing supply, as cautioned by L. O. Taylor (2017) for the case of nonlocalized amenities. Furthermore, in the partial scenario, the additional MPTN accessibility can be argued to be localized with more certainty as the affected dwellings represent only 1% of the total stock in GMC (considering a total of 5.96 million units in 2010). The identification of these types of changes is still contended. For instance, similar studies accounting for the benefits derived from (public) transport infrastructure follow the assumption of marginal and localized changes (e.g. Diao, 2015; McIntosh et al., 2015, 2014), while others make further considerations (e.g. Gjestland et al., 2014). The estimate provided in the previous section is from the perspective of a partial equilibrium approach which does not consider transaction cost. Thus, it should be considered as an upper bound.

Still, the differentiated potential accessibility to employment measure by population sub-groups in the housing market could be discussed from the point of view of relocation effects. This alternative view from a general equilibrium perspective is suitable if the size of the accessibility change and extension is perceived as *large* enough to incentivise residents to move. Following neoclassic economic theory, the house-buyer maximises its utility subject to a fixed income considering the various attributes of a house including accessibility. Considering this, it could be hypothesized that if the expansion of the main public transport network particularly favours a sector of the population with higher income than the current residents benefited, the former would outbid the latter (e.g. by offering a higher rent and achieving equilibrium). Consequently, higher income households would displace the 'original' population that did not perceive additional accessibility benefits. This literature makes clear that the residential land value reflects the upper envelope of the bid-rent curve (Duranton & Puga, 2015). According to this view, a distinction by population sub-groups in accessibility measures becomes of low relevance. This alternative view has not been explicitly discussed in accessibility literature yet.

Furthermore, literature from an alternative perspective acknowledges some of the economic pressures mentioned here and, in addition, suggests that other factors such as local crime, ethnic/class diversity, or aesthetic preference, can play important roles in a reallocation process (Hamnett, 2003; Klaiber & Kuminoff, 2014; Kuminoff et al., 2013; Lees, Slater, & Wyly, 2008). However, it is not clear from empirical literature how, or to what extent, these factors catalyse/deter residential displacement effects derived from public transport investments (Zuk, Bierbaum, Chapple, Gorska, & Loukaitou-Sideris, 2018). Approaching these from the perspective of the equilibrium sorting model (ESM) can relax some of the classic assumptions of the hedonic model by estimating the benefits in a general equilibrium. The implications of overlooking these aspects in our results may range from differentiated willingness to pay for specific groups of households (Klaiber & Phaneuf, 2010) to biased coefficients if reallocation costs are high in reality (Liang et al., 2021). The ESM could incorporate further realism in terms of heterogeneous income and preferences, and, thus, can contribute to clarifying the processes involved in the perception of public transport accessibility and its role in the broader housing market.

7.4.3 Limitations and future research

As stated in the previous sub-section, there are several constraints introduced by the main analytical method adopted, namely the hedonic approach. The main refer to the possibility that the change in accessibility derived from the expansion of the MPTN represents a large shock in the housing market. If this occurs, it is challenging to account for the elasticity of demand using the hedonic method. This limits the analysis to account for the possibility of residents to re-sort according to their income, preferences, or 'homophily,' as discussed in Chapter 3. Furthermore, if transaction costs are high there is the potential for bias in the WTP estimates (Liang et al., 2021). However, it is believed that these may not be as extreme as in other contexts (e.g. China in Liang et al. (2021)) as the national housing system in GMC actively seeks to lower these costs. These issues could be mitigated by adopting the ESM approach. This represents an avenue with enormous potential to advance the discussion of public transport accessibility valuation. Unfortunately, as discussed in Chapter 5, there are aspects that prevented its use for this research.

Another aspect is that the precise location of the lower-level units within the post code areas is unknown. Therefore, the spatial analytical unit is constrained to these geographic definitions. This represents potential concerns related to the modifiable areal unit problem (MAUP). Responding to this concern, Appendix I presents a detailed sensitivity analysis including alternative spatial units of aggregation for one the temporal stages of the data where the location was available at a disaggregated level.

The results of the MAUP sensitivity analysis in Appendix I indicate that all estimates are significant and go in the direction expected. This includes the key variable, namely MPTN accessibility. Thus, it is shown that the overall relationship between the potential accessibility to employment generated by the MPTN and land value is not simply an artefact of the way the information is partitioned in space, as cautioned for in literature (Fotheringham, Brunsdon, & Charlton, 2000; Páez & Scott, 2004). The MAUP effects are further examined in terms of the individual WTP and TWTP for changes in the MPTN which are equivalent to those introduced in the MPTN Stage 3 . Overall, it is shown that both individual and aggregate WTP estimates are very sensitive to scale effects. In terms of zoning and the WTP and TWTP, the results show that the post code scheme produces the closest results relative to the highest resolution and the most accurate SAS, i.e. the 0.5 km Grid.

This analysis comes with the following additional limitations:

- The main data source (residential value records) is constrained to the formal property market and houses with a mortgage. However, this data covers an important share of the transactions in the market (≈ 50%), as discussed in Chapter 4. In addition, this property data is used to build the national House Price Index in Mexico.
- Even if the precise location was available for all lower-level units, there are important technical challenges in terms of the computing power required to implement the processes required in standard spatial econometric approaches, e.g. neighbour matrices at the individual level. This type of limitation has been reported previously (Bala, Peeters, & Thomas, 2014), narrowing the spatial models to a couple of specific sectors. In contrast, the spatial multilevel model techniques adopted in this chapter help to overcome this limitation. Therefore, the aggregation of individual observations in multilevel models is still needed in the presence of large datasets, as the one used here.
- There are a few aspects concerning the spatial data sources of the post code zones, namely that they are not always consistent in terms of their technical definition of the boundaries (e.g. the vertices are not snapped); the polygons do not cover the total of the metropolitan surface, therefore some employment opportunities or properties may not be considered in the analysis due to the lack of spatial representation, or, in a few cases, one post code may refer to more than one geographic area. To overcome some of these limitations, the post code of the observation is validated by matching the name of the settlement or the municipality included in the property data. In addition, locations are represented as the population weighted centroids which may reduce some of the associated

problems.

- Most locational controls are estimated at one point in time due to a lack of more frequent data updates from the official sources, in addition to time and resource constraints. However, this issue can affect results in OLS models more severely than those which include spatially structured elements (i.e. BYM models) as the latter have been found to be robust regarding omitted variables (Bivand et al., 2017; G. Lee et al., 2016; Osland et al., 2016).

In light of the findings and limitations presented, the following points are recommended to be addressed in future research:

- There is an important opportunity to adopt the ESM approach in the study of the capitalization of accessibility into land value. This is mainly due to its capacity to incorporate many of the complexities of the housing market and home-buyers' preferences, income, and heterogeneous characteristics. This is possible given the ability to simulate a new equilibrium of the housing market where residents are allowed to re-sort according to the new conditions, as discussed in Chapter 3. Future studies are encouraged to consider this alternative as a way of alleviating some of the limitations of this research and the overall hedonic model.
- It would be useful to examine whether the consideration of surrogate employment data (e.g. the National Directory of Economic Units, DENUE) is as consistent as the disaggregated and detailed information from the Economic Census, used here for the construction of accessibility measures. This would be useful since it would allow the use of more up-to-date information.
- Further analysis in relation to the MAUP is recommended with a special focus on confirming the preliminary findings of the sensitivity analysis presented in Appendix I, which suggest that zoning can play a role in mitigating some of the scale effects.

7.5 Conclusions

The present study offers an in-depth analysis of the relevance of location-based accessibility measures in assessing the role of public transport infrastructure on the residential land market between the period of 2009 and 2019, drawing on the case of GMC. The results show that the relationship found between the residential land value and accessibility to employment generated by the MPTN is positive, robust, and consistent across the various temporal stages examined here, even after modelling the spatial structure at the local level (post code zones). These findings support a transition towards comprehensive accessibility measures to account for the level of service provided by transportation infrastructure, acknowledging the heterogeneity of the network and spatial distribution of opportunities to assess the benefits capitalized in the residential land market, in line with literature (Ahlfeldt, 2013; Higgins & Kanaroglou, 2016; Rodríguez & Mojica, 2009). These considerations may be useful to clarify apparent negative or neutral associations previously reported in the Latin American context (e.g. D'Elia et al., 2020; Flores Dewey, 2011).

The positive relationship found is also relevant for LVC policy. In the illustration concerning the expansion of the BRT corridor L-7 of the Metrobus system, a partial scenario suggests capitalizations equivalent to two thirds of the capital costs (including civil works and fleet). Even in a limited scheme, these benefits are attractive. For instance, sharing these benefits between landowners and the public by half would considerably aid the expansion of the transport network. In terms of policy implementation, the possible strategies for capturing all or a portion of this increase in value are very diverse. They depend on multiple criteria, e.g. whether it is voluntary, negotiated, or compulsory, or would be implemented before or after the development of the infrastructure, or at the time of the land development, or after, and other elements discussed in Chapter 2 (Medda, 2012; Suzuki, Cervero, & Luchi, 2013; Zhao et al., 2012). An in-depth LVC feasibility analysis is beyond the scope of this study. However, the findings reported here are encouraging and should be useful for informing policy analyses.

The next chapter concludes the thesis and includes further reflections about the methodological, empirical, and policy implications.

Chapter 8

Conclusion

8.1 Introduction and review of Research Questions

The delivery and effective operation of public transport is fundamental for a variety of reasons, including environmental and social equity aspects (Banister, 2018; Holden et al., 2020; Pereira et al., 2019), yet it remains difficult to fund (Falcocchio et al., 2018; Mathur, 2019). This is especially relevant in developing contexts with both rapid growth rates and public financial constraints (Gómez-Ibáñez et al., 2022). Land value capture (LVC) could alleviate part of this financial pressure from the bottom-up (Alterman, 2012; Gómez-Ibáñez et al., 2022). A core argument for LVC rests on the economic benefits generated by public actions that capitalize on land value. Thus, in order to support and extend the use of such mechanisms, it is essential to advance the discussion surrounding the empirical assessment of the potential benefits of public transport investments on residential land.

The present thesis aims at revisiting the implicit willingness to pay for the benefits of public transport infrastructure in the residential land market. This is done, primarily, by integrating key and novel elements, which are chiefly addressed in the fields of urban and transport geography looking at the concept of accessibility and drawing on Open Science resources.

Section 8.2 presents the conclusions reached, and focuses on the findings provided in the empirical chapters according to the following Research Questions:

- RQ1. *What is the role of the main public transport network in shaping accessibility to employment in Greater Mexico City?*
- RQ2. *What is the willingness to pay for the accessibility generated by the main public transport network in the residential land market between 2009 and 2019?*

Section 8.3 provides a reflection on the contributions of this thesis to knowledge and their implications for methodological, empirical, and policy perspectives. Section 8.4

acknowledges the main limitations of the analyses presented. Section 8.5 provides recommendations for future work. Section 8.6 offers some final thoughts.

8.2 Addressing the Research Questions

This research project draws on the case of Greater Mexico City (GMC), Mexico. The core object of study is the *main public transport network* (MPTN). This is comprised of *rapid transit* and *semi-rapid transit* modes, including the following systems: the Metro system, two bus rapid transit (BRT) systems (Metrobús, MB, and Mexibús, MB), one light rail transit line (LRT-L1), and one suburban/regional rail (RGR, denominated SUB-L1) (see Chapter 4 for details).

8.2.1 What is the role of the main public transport network in shaping accessibility to employment in Greater Mexico City?

The large-scale origin-destination (OD) route modelling phase was a key element of this analysis since it reflects the various detailed elements of the MPTN in location-based accessibility measures. A total of ≈ 2.1 billion door-to-door routes were modelled to estimate the travel time for each origin to every possible destination, for eight departure times in a day, for five spatial analytical schemes (SASs), and for seven MPTN temporal stages.

The results of the analysis are evidenced the heterogeneous distribution of accessibility. From a spatial perspective, it is possible to observe various patterns associated with the different characteristics of the transport modes studied. For instance, the accessibility generated by the services of the SUB-L1 tended to be reflected as islands around stations. In contrast, those enabled by the BRT corridors and Metro system reflected a path-like structure. This is a consequence of the special characteristics of the modes discussed in Chapter 4. Specifically, the SUB-L1 operates at a higher average speed with a longer distance between stations than the Metro and BRT modes. In addition, the SUB-L1 generates benefits in more distant locations from the station than the Metro or BRT corridors since the reduction in time in vehicles can be balanced-out by longer walking distances to the stations. These patterns are dissolved as the scale of the SAS increases. At the city level, this structure changes from a star-shaped form to a mono-centric like form.

The analyses from the temporal view unveiled additional aspects about the role of the MPTN in shaping accessibility. The ex-post comparison showed that extensions of the MPTN can play different roles according to the contextual characteristics. This is illustrated by the implementation of the BRT MB-L7 in Stage 7 where the benefits were concentrated on both ends of the corridor. This is because the employment density was high in central areas of the corridor. Thus, the connection to other (low- or moderate-density) zones was perceived as a marginal improvement. Similarly, the implementation of the BRT MXB-L1 in Stage 2 shows that the largest benefits are located towards the end of the line closest to the CBD (and the rest of the network too). The benefits gradually diminish in locations along the corridor that are further away from the CBD. In contrast, the most significant benefits of the BRT MB-L3 can be seen towards the end of the corridor which is the furthest from the CBD since sites in proximity to the main employment centre already enjoyed high employment densities. The outcomes of these examples can be explained by the overall location of the metropolitan area with respect of the spatial distribution of employment and connectivity with the rest of the network. A further observation from these analyses was the network effect, where even certain non-adjacent locations to a MPTN extension received benefits. This partially confirms the coarse assumption made in some empirical studies examining network effects in public transport systems (e.g. Mulley et al., 2017; Rodríguez & Mojica, 2009). However, this assumption is conditioned by the set of characteristics discussed above.

Most of the observations above hold for both accessibility specifications employed in this analysis, i.e. 'All' and 'matching.' However, there are a few differences worth noting. An interesting example at the local level is the network effect produced by the implementation of the BRT MXB-LI in Stage 2, which extends 'downstream' along the pre-existing METRO-LB. This effect assumes that residents close to the pre-existing metro line find attractive the potential employment opportunities which have been made available in the north (in the opposite direction of the CBD), now accessible by the BRT MXB-LI extension. In contrast, the matching accessibility specification did not reflect this network effect. This implies that the employment characteristics adjacent to MXB-LI do not match the education level of the residents located along the METRO-LB. At the global level, the matching accessibility specification captures more modest changes than its counterpart. This is expected since opportunities which may be valuable for some groups of people continue to be restricted by the travel time needed to reach them, even after the improvement of the network. These examples depict some of the consequences of the assumptions adopted in accessibility measures.

To summarize, the analysis clearly illustrates the marked spatial heterogeneities of accessibility and how the MTPN interacts with the multiple components of accessibility. In practical terms, it can be said that the outcome is influenced by the concentration of employment (e.g . proximity to important employment centres), operational characteristics of the transport mode/infrastructure, and the availability of pre-exiting

MPTN services, for example. These findings highlight the variety of outcomes which result from the expansion of the network, contrasting with the standard assumption of homogeneous benefits represented as buffers around stations or an entire transport corridor.

8.2.2 What is the willingness to pay for the accessibility generated by the main public transport network in the residential land market between 2009 and 2019?

Here, the methodological approach was aimed at covering a long temporal period (between 2009 and 2019), while holding the spatial aggregation scheme fixed for the computation of neighbourhood and accessibility measures, i.e. in post code zones. The analysis offers a thorough examination including various modelling methods (i.e. OLS models, nonlinear models, multilevel models, and spatial multilevel models), two accessibility specifications and two calibration methods (one exogenous and one endogenous), for seven temporal stages over a relatively wide time span. The analysis use a property value sample obtained from administrative records collected by the Federal Mortgage Society (SHF) ($N \approx 800,000$).

The results show that the relationship between MPTN accessibility and residential land value is very consistent. This was positive and significant in all of the specific analyses even after controlling for a wide range of neighbourhood characteristics, competing modes, and structured spatial effects. However, the magnitude of the coefficients was found to differ, especially when the spatial structure of the data was not modelled. A comparison showed that the largest estimates are produced by OLS models. The coefficient for the MPTN measure considering all opportunities practically halved in the Besag-York-Mollié (BYM) models compared to OLS models, falling from 8.2% to 4.6%. For the matching accessibility measure , the estimate varied from 8.7% to 6.7%, respectively. The coefficients slightly decreased in MLMs compared to OLS models. In the final set of models (BYM, spatial multilevel model), the residential land value was found to increase between 3.6% and 5.7% for one additional standard deviation in access to all types of employment opportunities depending on the MPTN temporal stage.

The role of the type of accessibility specification in hedonic models varied according to the inferential method used. Considering only matching employment opportunities in accessibility measures showed additional explanatory power and reduced possible coefficient bias in OLS and nonlinear models. Yet, the difference became irrelevant in multilevel models and spatial multilevel models. A possible explanation of this is that the latter methods could reduce potential bias introduced by unobserved variables given the information obtained from neighbouring zones (e.g. Kuminoff, Parmeter, & Pope (2010) show how bias can be reduced by employing spatial fixed effects). This finding can also be discussed from the perspective of urban economic theory. Specifically, the AMM model suggests that the land-rent bid-rent curve is given by upper envelope bid-rent (Duranton & Puga, 2015). Accordingly, if accessibility is not perceived by a lower-education group of residents as being improved, but is perceived as such by a group of residents in a higher-education quantile, the latter are expected to bid higher for residential land than the former (following the assumed correlation between education and income). Thus, if residents sort themselves to achieve spatial equilibrium right after an accessibility improvement, the distinction of employment opportunities in relation to salary in location-based measures would become irrelevant.

The analysis also provided an illustration of the potential for LVC by assessing the aggregated willingness to pay for accessibility improvements of the magnitude introduced by the MB-L7 BRT corridor in Stage 7. Two scenarios were developed to establish a possible range of uncertainty. A full scenario considers all the benefited housing stock and a second considers only those properties within the 15 most benefited post code zones. The results show that the total willingness to pay is considerable. This ranged from 0.7 to 1.5 times the equivalent of the capital costs of the transport infrastructure in the partial and full scenario, respectively. This is within the range of previous analyses (e.g. Gupta et al., 2022; McIntosh et al., 2015; TfL, 2017), revealing considerable potential for LVC.

8.3 Implications and contributions to knowledge

The implications and contributions to knowledge of the present work are discussed from a empirical, policy, and methodological point of view.

8.3.1 From the empirical point of view

The aggregate result of this thesis fills an important gap in the literature where the relationship between public transport infrastructure and residential land value is heavily understudied in GMC. The review of empirical literature shows that, although there are a few contributions from *grey* literature sources, contributions from peer-reviewed sources which explicitly address this topic are lacking. This is surprising given the size and relevance of the metropolitan area at the national and regional level. In addition, the public transportation system is complex and vast. The metro system alone is the largest in the region. Thus, its study, in the academic arena, represents a relevant contribution which can inform policy makers in GMC working in various fields such as housing, land use and planning, and transportation.

Specifically, the analysis carried out to answer RQ1 assessed and illustrated the role of the MPTN in the spatial distribution of accessibility over the past decade for the whole of GMC. In addition to the academic goals accomplished and discussed above, these findings are useful for reflection on the decisions made at the city level with respect to the transport network. For instance, the output can support in identifying areas where the infrastructure has had the largest impact, which areas have been left behind, and whether an intervention had the expected outcome. An example of the latter is the small accessibility effects in high-employment density areas such as the middle segment of the MB-L7 (in Stage 7) or the overall small effects of the MB-L4 (in Stage 3). Furthermore, the strategy employed can constitute the basis of an analytical method or tool informing local programmes such as the redevelopment of the multimodal transport hubs in *Ciduad de México* (CETRAMs) (discussed in Chapter 4).

In addition, the consistent findings (in terms of the positive relationship and the magnitude of the potential for LVC, addressing RQ2) provide relevant information for local bodies regarding the expectations and the range of the willingness to pay for these benefits. The results of this work can help to clarify previous findings in the same area which were approached from the *classic* perspective. These findings have reported neutral or even negative effects on residential land value for the announcement of public transport infrastructure (Flores Dewey, 2011) or for the construction of the METRO-L12 (Velandia Naranjo, 2013).

Additionally, the sensitivity analysis in Appendix I adds robustness to the positive relationship found between location-based measures accounting for accessibility to employment and land value in light of the MAUP effects. Specifically, the findings confirm positive capitalizations attributed, chiefly, to the accessibility generated by the Metro-L12 regardless of the spatial aggregation unit employed. This clearly illustrates how one project (MB-L4) can reflect only minor improvements compared to the more evident benefits of others (METRO-L12). This is according to the components embodied in the accessibility concept as discussed when addressing RQ1.

8.3.2 From the public policy point of view

The findings for RQ2 showed that the magnitude of the aggregated economic benefits is considerable, even in a conservative scenario. The magnitude was consistent with the estimate of other studies pursuing similar objectives. Yet, the main difference between the previous estimates and the present estimates lies in the spatial distribution of the assessed value increments according to location-based measures, which is in line with the theoretical definition of accessibility. This is relevant for the design of fee-based LVC instruments for at least two reasons, e.g . betterment charges, or tax increment financing (TIF). The first is related to aspects of fairness. Based on the view adopted for this work, benefited households would be charged according to the *effective* benefits received by the implementation of a transport project. This contrasts with the coarse assumption of the classic approach where all stations are assumed to produce benefits of the same order. This view is also stipulated in the prevalent legal tax code of Ciudad de México for the application of betterment contributions related to public transport infrastructure. Therefore, households could be over or undercharged. Second, LVC literature suggests that benefits should be clear to households in order to mitigate any potential resistance from the public to the implementation of LVC instruments (Vadali et al., 2018). Hence, the approach adopted, and the assumptions made, in measuring accessibility benefits can influence the success in implementing these types of strategies.

LVC fee-based instruments are affected by the assessment of the benefits, as discussed in Chapter 2. Thus, the MAUP is relevant as the sensitivity analysis in Appendix I showed that the estimated willingness to pay is sensitive to the selected SAS. Specifically, it was observed that small areas tend to display a wider range of accessibility and, consequently, these are more likely to contain extreme values. In the context of LVC, this implies that if the assessment is intended for the design of fee-based instruments, just few households would be expected to pay heavy contributions which may complicate the implementation. In such cases, the policy maker could consider a trade-off between accuracy and the smoothing process produced by the scale effects. The latter implies the loss of information, but it can be argued that the overall spatial pattern follows the spatial process of MPTN accessibility benefits. The sensitivity analysis also showed the relevance of zoning. Specifically, middle-size administrative zones (post code zones) are more appropriate than middle-size uniform zones (> 1 km). In practical applications, the use of equal area zones may run the risk of assessing a group of households equally even if there is a major physical barrier that prevents the use of public transport infrastructure for some of the population within the same zone.

In contrast to these observations, the trade-off referred to above may not be relevant for guiding the design and implementation of development-based LVC instruments, e.g. land readjustments, sale/lease of air development rights, or joint developments. Here, the policy designer may prefer accuracy. In such a case, it is appropriate to use a high-resolution spatial scheme as the objective of these instruments entails the identification of the sites receiving the largest benefits and, thus, those with the largest potential for the recovery of land value increments. As shown, the incorporation of location-based measures in the assessment of the WTP represents a compelling approach for guiding the implementation of development-based instruments also.

The adoption of location-based accessibility measures has collateral implications which are relevant for policy. For example, the results addressing RQ1, in terms of the spatial distribution of the accessibility benefits assessed by the Gini index, are concerning. These show how the MB-L7, a centrally located intervention in the network, clearly re-concentrated accessibility advantages at the city level. Meanwhile, projects connecting non-central areas of the metropolis to the network (e.g. MXB-LIII, MB-L5, MXB-LII) support the de-concentration of these benefits.

8.3.3 From the methodological view

Answering the Research Questions illustrates the spatial heterogeneity of MPTN accessibility well. Location-based measures are shown to capture not only characteristics of the transport network and its different modes, but also the spatial dispersion of opportunities. One of the most evident methodological implications is that regular buffers around stations or public transport corridors are not always appropriate. The spatial extent, magnitude and shape of the accessibility benefits depend on a number of aspects on the ground, e.g. operational characteristics of the mode, local and overall spatial distribution of employment (local and global land use), perception of spatial decay, physical barriers, pedestrian network, or the characteristics of residents and employment opportunities, as suggested in the literature (Levinson & Wu, 2020). This observation adds to the current literature discussing the definition of public transport *areas of influence* by providing additional insights into the intervening factors (Guerra et al., 2012; Yen et al., 2019).

Additionally, the heterogeneity of MPTN accessibility outcomes can help to clarify a lack of correlation between some empirical studies in the region. For example, the results reported for Bogotá, Colombia (using catchment areas based on distance to public transport facilities) suggest that in one of the corridors where there was no correlation with the proximity to BRT stations also presented with a high concentration of employment (Guzman et al., 2021). As shown in the results for RQ1, these types of areas tend to receive low accessibility benefits which, in turn, can be expected to reflect low, or a lack of, capitalizations. Thus, it is strongly recommended to move from oversimplifying measures to location-based ones in order to assess the transport benefits, particularly in the context of LVC (as previously discussed). This illustrates well the argument in Higgins & Kanaroglou (2016) regarding the mixed outcomes in the empirical studies discussed in Chapter 3.

The role of the characteristics of residents and employment opportunities in accessibility measures was found to be relevant in some cases only. For example, an examination of the roles of the MPTN using both measures, including ‘All’ and only ‘matching’ opportunities, was demonstrated to be relevant if the purpose of the research comes from a social equity perspective. The previous section addressing RQ1 describes an empirical situation where this may play a determining role (MXB-LI). In addition, this distinction is relevant in global Gini index measures when highlighting distributive aspects. This is in line with the conclusions of previous empirical research aimed at the aforementioned purposes (Pereira, 2018; Pereira et al., 2019). However, this differentiation was not essential in the assessment of residential land value. The results of the models that acknowledge the structure of the data (multilevel models and spatial multilevel models) do not show clear improvements in the reduction of spatial heterogeneity, explanatory power, or mitigation of possible bias. This is supported from the perspective of urban economic theory as discussed in the previous section. Thus, it is concluded that such distinction is not indispensable for the environmental assessment of public transport accessibility as long as the spatial structure of data is acknowledged. This is an overlapping aspect between urban economic theory and accessibility theory which has not been clearly discussed as of yet.

The use and comparison of various multivariate techniques to estimate the hedonic function confirms and extends previous notions which suggest that overlooking the spatial structure of the data can produce bias, inefficiency, and uncertainty in parameter estimates (LeSage, 2015; LeSage & Pace, 2009). Specifically, although the literature has suggested the (unstructured) multilevel model as a way to model spatial heterogeneity (Habib & Miller, 2008; Páez & Scott, 2004), the results of RQ2 show that this technique produces similar coefficients to the OLS models. Furthermore, in this type of model the upper-level component is assumed to be independent and identically distributed. A test for spatial autocorrelation shows that this is not the case for our sample and, most likely, not the case in the context of property value data either. Thus, researchers are encouraged to consider the assumptions of the multilevel model (random intercept model) in this context. The use of a spatially structured component at the upper level proves to be necessary for the adequate modelling of the data. The results of the latter provided substantially lower parameter estimates for MPTN accessibility.

Answering RQ2 also offers practical recommendations for the implementation of spatial inferential methods for large datasets. The current capacity of computational resources proves to be insufficient for the implementation of 'standard' spatial econometric methods in two related aspects. First, the computation of the spatial matrices at the point level (both using the full sample ($N \approx 800,000$) and temporal stage ($N \approx 100,000$)) is infeasible even in a CPU of considerable capacity (virtual memory of up to 120 GB RAM). A spatial multilevel model is thus appropriate to account for such effects in large samples, particularly if the within-zone location of the observations is unknown.

Second, the frequently used simulation-based approach for the implementation of spatial models in a Bayesian setting was impractical for the purpose of this work (i.e. Markov Chain Monte Carlo simulations). For instance, the estimation of one model, which inclued 80,000 to 150,000 observations, required between three to five days to reach a convergence (for 10,000 to 15,000 simulations) for only one temporal stage using the `HSAR` software version '0.5.1' (Dong et al., 2020).[1] The use of an approximate Bayesian inference method implemented via the `INLA` open-source software for `R` proved to be an excellent alternative for the fast implementation of both complex models and large samples, reducing the estimation of the models to minutes. Thus, the use of the `INLA` is recommended for spatial multivariate models for large datasets (for a technical comparison between the INLA and MCMC methods see B. M. Taylor & Diggle (2014)). This is expected to be increasingly relevant since large samples are more common in the digital era (Birkin, Clarke, Corcoran, & Stimson, 2021).

8.4 Limitations

As discussed throughout various chapters, the hedonic approach is constrained in various aspects. First, it is challenging to account for the heterogeneity of residents in terms of income and preferences (Klaiber & Smith, 2011). There are further challenges for the estimation of the net benefits produced by public transport improvements. These primarily arise from the potential re-sorting effects of residents in response of this type of infrastructure (Kuminoff et al., 2013). This process may be affected by complexities such as 'homophily' (Bakens & Pryce, 2019; Dean & Pryce, 2017), ethnic mix preferences (Bayer et al., 2014; McPherson et al., 2001), quality of other public services (Bayer & McMillan, 2012), or residents' lifestyles which may affect the heterogeneity in preferences for accessibility across different groups. Those considerations are particularly relevant when changes in the urban landscape are large (L. O. Taylor, 2017). The equilibrium sorting model (ESM) can help to address some of these complexities. In addition, it can introduce considerations such as reallocation costs (Liang et al., 2021), which are not included in this analysis. Although the use of the ESM remains limited for this research, the next section discusses the feasibility and potential for the study of land value capitalizations in future studies.

In addition to the limitation of the main analytical approach employed, there are further specific considerations. For instance, the accessibility measures used fall into the ecological fallacy (Openshaw, 1984) as they assume homogeneity within the spatial analytical unit (i.e. that all residents or employments are of the same type). However, socio-economic or financial information of people and establishments is protected by

[1] `CARBayes` version '0.5.2' allowed faster computation for considerably more simulations (~100,000 to 200,000) (D. Lee, 2016). Yet, the spatial multilevel model is limited to the CAR model proposed in Leroux, Lei, & Breslow (2000). In addition, the implementation was still time-consuming, i.e. around 12 hours to reach convergence.

the Data Protection Act (*Ley Federal de Protección de Datos Personales en Posesión de Particulares*). Therefore, the data was accessed aggregated at the lowest possible level (i.e. the urban block) including statistical disclosure controls with further aggregations in some cases (for more details see Chapter 6 and Appendix D).

Although this research offers important advancements in modelling public transport travel costs relative to existing literature in the field, these could be improved in various aspects. First, the street transit modes (e.g. regular bus and other modes detailed in Chapter 4) are not modelled explicitly due to limitations of data. Partial information is available for Mexico City, but the incompleteness of information is severe for the state of México. Due to the magnitude of the resources required to collect such information, it is not expected that this information will be available for GMC any time soon. The assumption made in the present study is that street transit services are comparable across the zones studied. This is based on data from the 2017 Travel Survey. In addition, the effectiveness of such services is mostly a function of the *overall* location in the metropolis, which can be accounted for by the measure of accessibility by car and the local heterogeneity effects in multilevel models. Second, the MPTN travel time modelled does not account for delays or changes with respect to the timetables. Possible mitigation strategies are discussed in the next section. Third, the fare costs are not considered. However, these are not expected to substantially affect the estimates since the cost of a single ticket is affordable, as suggested in the literature (Crotte, Graham, & Noland, 2011).

In addition, the property value data used in this analysis is restricted to dwellings that were acquired through a mortgage in the formal market. The characteristics of the stock reported in Chapter 4 show that this type of acquisition represents a considerable proportion of the total stock. Thus, the sample is expected to be representative. This could be confirmed by other sources in future research, as will be discussed in the next section.

There are further considerations in the context of LVC. The TWTP estimates focus on the benefits flowing from travel to work. Still, there are other potential benefits of the network, e.g. access to other urban amenities/services (although at least some of these may be captured by the accessibility measures).

8.5 Future work

In light of some of the limitations due to travel costs, as suggested above, future work could extend the analysis by incorporating GTFS data including street transit services which are partially available. As mentioned above, this is unlikely to be included for the whole metro area. Still, it would be possible to conduct analyses limited to *Ciudad* *de México*. Also, this can include real-time GTFS data which are now available for the MB BRT system. This would enable sensitivity analyses in relation to possible operational delays.

The present analysis is focused on the residential land market only. Still, other types of land use are likely to respond to accessibility resulting from improvements to the public transport network as well. Future research could examine other types of property, such as retail or offices. This could supplement the estimates presented in the results. Other sources of information could be useful in confirming the findings of this analysis, provide further insights for the residential market, and incorporate additional land uses. For example, administrative data collected for taxing purposes at the local level records all the transactions in the formal market regardless of the financing scheme or their use. In addition, offering value data collected by web platforms could be used to examine possible detailed timing variations in the market as a response to public transport projects. Although these alternative sources could provide additional insights, they may also represent some limitations. For instance, the offered value may differ from the transaction, or the values reported for taxing purposes may be underestimated.

The following methodological aspects that were not explored due to time constraints could be considered in future research. For example, literature in the region has suggested the interaction between residents' socio-economic status and the valuation of public transport services (Crôtte, Noland, & Graham, 2009). Future work can extend the analysis by identifying submarkets according to such characteristics. In addition, difference-in-difference (DID) research designs are suggested to present advantages in the hedonic modelling, e.g. mitigate potential omitted variable bias (Bishop et al., 2020; Parmeter & Pope, 2013). The approach adopted in this thesis could be easily extended by designating the treatment groups based on the level of accessibility benefits received.

Further work is recommended in relation to the MAUP to examine other administrative-based zone schemes. This may be useful in confirming the implications discussed previously. For example, this would help to address the following questions: is it just post code zones that provide reasonable results? What is the acceptable scale of administrative-based schemes? It would also be worth extending the analysis to test the suggestion in the literature that spatial econometric methods can represent a potential solution to mitigate some of the MAUP effects as compared to classic multilevel or OLS models (Wong, 2009a).

The approach could also be expanded to incorporate additional realism in accessibility measures. These could consider penalties for transfers from one public transport system to another (given the lack of fare integration) or the uncertainty of the public transport travel time estimates (Conway et al., 2018). Another avenue for future research could be the clarification of whether the capacity or saturation (offer/demand in the transport component) of different modes (metro vs BRT) can affect the conclusions provided in the present work, as suggested in previous reviews (Mohammad et al., 2013).

A useful contribution in the future could be the expansion of the discussion regarding the time and computational limitations in implementing inferential methods for large samples, possibly f rom a comparative perspective using different techniques. Although there are some comparative studies in the literature, these tend to focus on the results rather than on the process (e.g. Bivand et al., 2017; Comber, Arribas-Bel, Singleton, Dong, & Dolega, 2020). Such work could provide practical guidance to future researchers who may face challenges related to the scalability of the analysis, as presented in this work and in the reported literature (e.g. Bala et al., 2014).

From the perspective of theory or policy design, LVC literature can be extended by drawing on the results of this work. For example, qualitative literature could assess further the implications of the findings presented here in the governance, planning, or legal arena. For instance, these could examine the role of the institutional or regulatory framework, or discuss further areas of opportunity for the implementation of LVC instruments while having the range of figures and methods developed in this work at hand. In addition, studies from the perspective of urban planning could elucidate whether LVC instruments supporting public transport infrastructure financing can simultaneously mitigate the rise of housing prices, as suggested in the literature (Borrero, 2013). This could also be approached from the point of view of the consequential displacement of local residents relating to gentrification effects (Zuk et al., 2018).

Macro events or global shocks could represent further implications for this analysis. For instance, a considerable volume of emerging work is documenting changes in travel behaviour (van Wee & Witlox, 2021) and the property market (Balemi, Füss, & Weigand, 2021) in light of the COVID-19 global pandemic. This major event broke out right after the data acquisition phase, in early 2020, and is, therefore, outside of the temporal scope of the study. Still, there are relevant potential implications that could be built on top of the work presented here. For example, a study of the temporality of the effects, in both the short- and long-term, in the relationship studied could generate important insights for LVC in terms of the certainty of such instruments. In addition, the rise of teleworking triggered by lockdowns may affect commuting patterns (Campisi, Tesoriere, Trouva, Papas, & Basbas, 2022) which in turn is expected to have implications on the residential land market (Brueckner, 2011; Elldér, 2017). These results may also question or confirm the robustness of LVC instruments for public transport infrastructure. The expected effects also expand to office and retail property (Hoesli & Malle, 2021). This segment of the land market could be addressed as a cross-cutting aspect in relation to the recommended work above, given its previously discussed relevance. All in all, the emerging evidence suggests important implications related to the COVID-19 pandemic that should be considered in future research.

Considering many of the points mentioned in this section, adopting the ESM approach may provide further insights into the study of public transport. This is because it can address some of the limitations of the hedonic model in a general equilibrium while incorporating key complexities of the housing market. The review of the literature has identified that the ESM has not yet been implemented in order to address these objectives. While this may reflect some practical limitations, this also represents an important avenue for future studies. For this research, there were technical and data availability constraints preventing the use of this method at the time of choosing the main analytical framework. However, computational resources have become more accessible in recent years. Additionally, the prosperity of the digital era could offer the information needed about residents and the housing market using digital footprint data collected for administrative or private purposes. These expected shifts may lower the barriers for the generalization of this approach. Taken altogether, exploiting the full potential of the ESM approach represents an extraordinary opportunity to advance the knowledge of this area in future research.

8.6 Final thoughts

The present work has advanced discussion on the assessment of the economic benefits generated by public transport infrastructure which capitalize on the value of residential land. It has done this by taking advantage of the resources developed and used according to the Open Science framework. These enabled a thorough examination of public transport accessibility and its incorporation into the hedonic framework from various relevant perspectives and may guide the decision-making processes involved in LVC.

Altogether, the findings of the present research are hoped to encourage and provide useful resources and methods to, at least, kick-off and progress the discussion for the implementation of LVC instruments for expanding, operating, and maintaining the public transport network in GMC. These actions are very much needed in order to address multiple urban development objectives, including the decarbonization of the transport sector.

At a higher level, the contributions presented in this work represent an example of an integrated perspective of the theory of land rent and land value capture literature drawing on an empirical example. Such a perspective can contribute for framing a "more complex united theory of land value capture" (Vejchodská et al., 2022, p. 8).

Appendix A

List of source code repositories

All the source code used for this PhD thesis is organized in a series of `RStudio` projects hosted eight individual public GitHub repositories (https://github.com). The structure within the projects follows some general conventions. Namely, the `data` folder contains the raw input information, the `R` folder includes the scripts to process the information, and the `output` or `plots`/`maps` the products of the process, e.g. modelling results, visualizations, post-processed data, as well as temporary outputs.

The repositories follow a logic sequence indicated by the two digits that follow the common identification `phd_...` of their name in ascending order. Similarly, the flow of the `R` scripts within each project is defined by the numeric sequence indicated by first two digits of the file's name, e.g. `01...`, `02...`, ..., `08...`. The `.R` files starting with `00` are preliminary and do not follow a sequence.

1. Name: `phd_01_census_access`.
    - URL: https://github.com/rafavdz/phd_01_census_access
    - Description: This repository focuses on organising the essential input data. It contains the source code used for the the acquisition, processing, and restructuring or aggregation of cartographic, administrative, sociodemographic, and transport information (road network, GTFS files, and routing modelling). It is relevant for Chapter 4, Chapter 5, Chapter 6, Appendix B, Appendix C, and Appendix D.
2. Name: `phd_02_housing`.
    - URL: https://github.com/rafavdz/phd_02_housing
    - Description: This repository contains the source code describing the housing stock both in Mexico and Greater Mexico City setting the context of the present thesis. It draws on data from the Population Census (2010 and 2020), National Household Income Survey (Encuesta Nacional de Ingresos y Gastos de los Hogares, version 2010 and 2020), and the System for Na-

tional Housing Statistics (Sistema Nacional de Información e Indicadores de Vivienda). It is relevant for Chapter 4.

3. Name: `phd_03_sim_mexico`.
   - URL: https://github.com/rafavdz/phd_03_sim_mexico
   - Description: This repository chiefly includes the code used to estimate the spatial interaction models for Greater Mexico City. It employs data from the 2017 Travel Survey (INEGI, n.d.-b). It also includes code describing the general mobility trends in the metropolis. It is relevant for Chapter 4 and Chapter 5.
4. Name: `phd_04_shif_data`.
   - URL: https://github.com/rafavdz/phd_04_shif_data
   - Description: This repository processes the raw administrative property value data obtained and collected by the Federal Mortgage Society (SHF). It also geolocates the records by linking the observation's address or other administrative field to local cadastre data. It is relevant to the sample selection process described in Chapter 5.
5. Name: `phd_05_amenities_osm`.
   - URL: https://github.com/rafavdz/phd_05_amenities_osm
   - Description: This repository organises most of the environmental data according to different spatial analytical schemes. It processes information obtained from National Directory of Economic Units (DENUE, Directorio Estadístico Nacional de Unidades Económicas) (INEGI, n.d.-b) and OpenStreetMap. It is relevant for the process in Chapter 5 describing the computation of neighbourhood measures.
6. Name: `phd_06_crime_snsp`.
   - URL: https://github.com/rafavdz/phd_06_crime_snsp
   - Description: This repository processes the crime data obtained from the National Security System (*Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública*) (SESNSP, 2021) further describing the neighbourhood attributes of a dwelling. The information is aggregated at the municipality level. It is relevant for the process described in Chapter 5.
7. Name: `phd_07_pop_statanalyses`.
   - URL: https://github.com/rafavdz/phd_07_pop_statanalyses
   - Description: This repository contains the code source used to run the statistical analyses of the present thesis. It uses the outputs produced in previous repositories. It is relevant for Chapter 7, Chapter 8, Appendix F, Appendix

G, and Appendix H (N.B. the code under the folder named as `RQ1` refers to the analyses in Chapter 7 while `RQ2` refers to those in Chapter 8).

8. Name: `phd_08_thesis`.
   - URL: https://github.com/rafavdz/phd_08_thesis
   - Description: This repository contains the source code to create the main thesis document in `.pdf` format. It draws on the packages `thesisdown` (Ismay & Solomon, 2021) and `bookdown` (Xie, 2021a) for `R` language. It is relevant for all the contents of the present thesis.

Appendix B

Model specifications to adjust private car travel times

Table B.1 presents a summary of the various OLS specifications that were tested to adjust the private car free-flow model originally produced in OTP. The models are fitted using the estimates generated by Google Maps as the dependent variable in a regression framework. As shown, Model 1 (M1) is a simple linear model where β_1 is the travel time estimated in OTP. Model 2 (M2) includes a binary variable indicating whether the origin or the destination of the estimated route is one of the core municipalities.[1] The last two models, Model 3 (M3) and Model 4 (M3), include a categorical variable for the state of origin or the state of destination, respectively. As shown, the performance of the models is similar (indicated by the Adjusted R^2 and the results of the Ramsey Regression equation Specification Error Test, RESET). We therefore opted for M1 since it performs good when compared to more complex specifications while maintaining simplicity.

[1]Coyoacán, Iztacalco, Iztapalapa, Álvaro Obregón, Benito Juárez, Cuauhtémoc, Miguel Hidalgo, Venustiano Carranza.

	Dependent variable:			
	GM travel time (min)			
	(1)	(2)	(3)	(4)
OTP travel time (min)	1.190***	1.193***	1.195***	1.196***
	(0.004)	(0.004)	(0.004)	(0.004)
Core area		1.135***		
		(0.209)		
Origin state: Hidalgo (ref: Ciudad de México)				−3.505***
				(0.590)
Origin state: Mexico (ref: Ciudad de México)				−0.925***
				(0.190)
Destination state: Hidalgo (ref: Ciudad de México)			−1.483*	−1.497*
			(0.810)	(0.808)
Destination state: Mexico (ref: Ciudad de México)			−1.822***	−1.833***
			(0.168)	(0.168)
Intercept	4.540***	4.150***	5.370***	6.042***
	(0.221)	(0.232)	(0.232)	(0.266)
Observations	9,264	9,264	9,264	9,264
R^2	0.916	0.916	0.917	0.917
Adjusted R^2	0.916	0.916	0.917	0.917
Reset statistic	90.271	88.027	81.408	85.135
Reset p-value	1.5e-39	1.3e-38	9.0e-36	2.3e-37
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B.1: OLS summary. Dependent variable is Google Maps (GM) travel time by car in minutes

Appendix C

Mexibus (MXB) services

Table C.1 summarises the details of the Mexibus (MxB) services employed to reconstruct the GTFS data. These are valid for the year 2019 in a regular working day (Monday to Friday), and were assumed as fixed for the corridors included in the different temporal scenarios of the main public transport network.

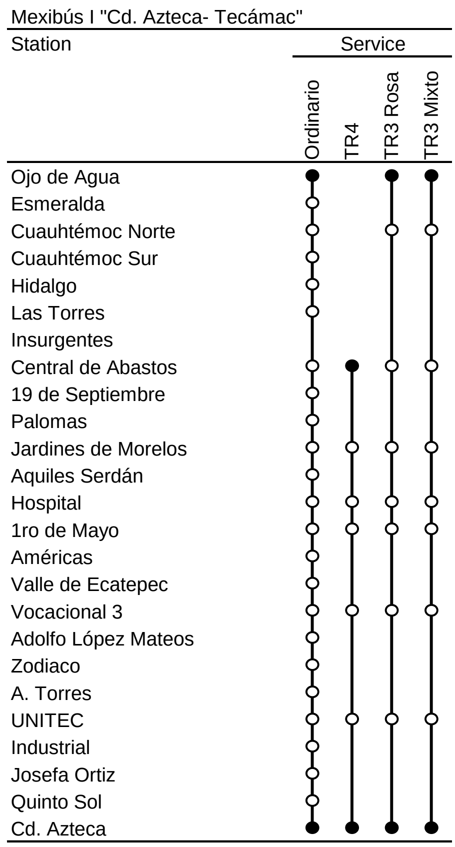
Corridor/Route	Route service name	Total length (Km)	Travel time (minutes)	Average speed (Km/h)	Headway (seconds)
Mexibús I	Ordinario (Cd. Azteca - Tecámac)	33.6	78	25.85	300
Mexibús I	TR3 Exprés Rosa (Cd. Azteca - Tecámac)	33.6	60	33.60	540
Mexibús I	TR3 Exprés Mixto(Cd. Azteca - Tecámac)	33.6	60	33.60	360
Mexibús I	TR4 Exprés (C. Azteca - Central de Abastos)	22.5	40	33.75	180
Mexibús II	Ordinario (Lechería - Plaza las Amerícas)	46.1	120	23.05	240*
Mexibús II	Ordinario Rosa (Lechería - Plaza las Amerícas)	46.1	120	23.05	1200
Mexibús II	Exprés Mixto (Lechería -Coacalco)	34.0	70	29.14	600
Mexibús II	Exprés Rosa (Lechería -Coacalco)	34.0	70	29.14	1800
Mexibús III	Ordinario (Chimalhuacán - Pantitlán)	36.0	100	21.60	300*
Mexibús III	Exprés 1 (Chimalhuacán - Pantitlán)	34.0	80	25.50	480*
Mexibús III	Exprés 2 (Chimalhuacán - Pantitlán)	36.0	80	27.00	360*
Mexibús III	Exprés 2 Rosa (Chimalhuacán - Pantitlán)	36.0	80	27.00	1260
Mexibús III	Exprés 3 (Nezahualcóyolt - Pantitlán)	16.6	44	22.64	660

Note:
Source: Sistema de Transporte Masivo y Teleferico del Estado de México
Total length and travel time and are expressed as a full cycle wich includes the return trip to the starting terminal;
Average speed is based the total length and travel time reported;
* Headway considered is the largest reported (max. difference of 120 secs), since the specific frequency time windows are not defined. All other services are constant, as reported.

Table C.1: Mexibus corridor services, 2019

The following three figures illustrate the operation of various Mexibús BRT corridors. These are drawn based on the information provided by the Sistema de Transporte Masivo y Teleférico (SITRAMyTEM) of the state Mexico between 2019 and 2020. The 'TR' prefix is used to identify express services. The term 'Rosa' is used to identify exclusive services for women.

Figure C.1 presents an schematic representation of the operation of the different services of the Mexibús Line I, namely: 'Ordinario,' 'TR4,' 'TR3 Rosa' and 'TR3 Mixto.' The filled dots represent terminals while the empty dots the station where a service operates.



Source: The author based on information provided by the Sistema de Transporte Masivo y Teleférico (SITRAMyTEM) del Estado de México accessed via freedom of information requests.

Figure C.1: Stations serviced in the corridor Mexibús I.

Figure C.2 presents an schematic representation of the operation of the different services of the Line II, namely: 'Ordinario,' 'Ordinario Rosa,' 'Exprés mixto' and 'Exprés Rosa.'

Mexibús II "Lechería - Coacalco - Plaza las Américas"

Station	Ordinario	Ordinario Rosa	Exprés Mixto	Exprés Rosa
Las Américas	●	●		
1ro de Mayo	○	○		
San Martín	○	○		
Puente de Fierro	○	○		
Casa de Morelos	○	○		
UPE	○	○		
San Cristóbal	○	○		
Agricultura	○	○		
ISSEMYM	○	○		
El Carmen	○	○		
Ecatepec	○	○	●	●
DIF	○	○	○	○
Guadalupe Victoria	○	○	○	○
Venustiano Carranza	○	○		
FOVISSSTE	○	○	○	○
San Carlos	○	○		
La Laguna	○	○	○	○
Parque Residencial	○	○	○	○
Eje 8	○	○	○	○
1a de Villa	○	○	○	○
Las Flores Zacuatitla	○	○	○	○
San Francisco	○	○	○	○
Héroes Canosas	○	○	○	○
Coacalco Tultepec	○	○	○	○
Ex Hacienda San Felipe	○	○		
Bosques del Valle	○	○	○	○
Coacalco Berrizábal	○	○	○	○
Santa María	○	○		
Villas de San José	○	○	○	○
Mariscala Real del Bosque	○	○	○	○
Fuentes del Valle	○	○	○	○
De la Cruz San Mateo	○	○		
Cartagena	○	○	○	○
Bello Horizonte	○	○		
Bandera Tultitlán	○	○	○	○
Buenavista	○	○		
COCEM	○	○		
Recursos Hidráulicos	○	○		
Chilpan	○	○	○	○
Cd. Labor	○	○		
Vidrera	○	○		
Lechería	○	○		
ERO			●	●
La Quebrada	●	●		

Source: The author based on information provided by the Sistema de Transporte Masivo y Teleférico (SITRAMyTEM) del Estado de México accessed via freedom of information requests.

Figure C.2: Stations serviced in the corridor Mexibus II.

Figure C.3 presents an schematic representation of the operation of the different services of the BRT system Mexibús Line III, namely: 'Ordinario,' 'Exprés 1,' 'Exprés 2,' and 'Exprés 2 Rosa,' and 'Exprés 3.'

Mexibús III "Chimalhuacán - Nezahualcóyotl - Pantitlán"					
Station	Service				
	Ordinary	Exprés 1	Exprés 2	Exprés 2 Rosa	Exprés 3
Pantitán	●	●	●	●	●
Calle 6	○				○
El Barquito	○	○			○
Maravillas	○				
Vicente Riva Palacio	○	○	○	○	○
Virgencitas	○				
Nezahualcóyotl		○			
Lago de Chapala	○				
Adolfo López Mateos	○	○	○	○	○
Palacio Municipal	○	○			○
Sor Juana I. de la Cruz	○				
El Castillito	○				
General Vicente Villada	○	○	○	○	○
Rayito de Sol	○	○			●
Las Mañanitas	○				
Rancho Grande	○				
Bordo de Xochiaca	○				
Las Torres	○	○	○	○	
Guerrero Chimalli	○				
Las Flores	○	○			
Canteros	○				
La Presa	○	○	○	○	
Embarcadero	○				
Santa Elena	○				
Ignacio M. Altamirano	○				
San Pablo	○				
Los Patos	○	○	○	○	
Refugio	○				
Acuitlapilco	○	●			
Chimalhuacán	●		●	●	

Source: The author based on information provided by the Sistema de Transporte Masivo y Teleférico (SITRAMyTEM) del Estado de México accessed via freedom of information requests.

Figure C.3: Stations serviced in the corridor Mexibus III.

Appendix D

Estimation procedure of economic activity spatial reallocation

The process to re-allocate the aggregated data at a higher level than the block is based on a proportional estimate which adjusts the figures following the spatial and classification structure of the 2015 DENUE through the next steps:

1. Summarise point data from the DENUE at the block level;
2. Associate the Economic Census and DENUE data in a common analytical framework (blocks);
3. Identify the blocks that that did not contain data of the Census but the did from the DENUE. These are assumed to be the location of the aggregated data;
4. Proportionally assign the aggregated figures in the Economic Census to their respective blocks based on the number and size of establishments recorded in the DENUE.

The previous presents a limitation related to a mismatch in the collection period. The Economic Census data was collected in early 2014, whilst the DENUE was collected about a year later. Therefore we could have allocated data in areas where at the time of the Census data collection did not have economic activity. Conversely, we could have omitted data that had activity at the time of the Census but stopped at the time of the data collection for the DENUE. However, this mismatch is expected to be minor, since the vast majority of information is located at the block or census tract level (97%) and there was not a considerable shock that could have substantially altered the spatial distribution of economic activity.

Another limitation is that the data collection methodology of the Economic Census appends some figures to the geographic location of its headquarters for industries where some personnel do not have a fixed working location (e.g. construction or energy production). Unfortunately, we do not have access to the type of industry at the block level and we cannot systematically exclude those potential cases. To limit possible overestimation, the occupied personal density was limited to 30,000 per hectare and a total of 50,000 per block. This rule excludes rural localities, since the block geometries are not available. Therefore, it is not possible to compute the density. These type of localities contain a minor proportion of the total economic activity figures. This resulted in the exclusion of 6 blocks that are equivalent to about 9% of the total occupied personnel originally referenced within the metropolitan area.

The respective cartographic data for the information described above is available in the INEGI's public URL.[1] It should be noted that blocks are represented differently in urban and rural settlements[2] Rural localities are represented by spatial points and urban by polygons. To standardize the data in a single spatial data frame, we buffered a representative square shape around each of these points. Measures relative to the area of blocks in rural areas (e.g. employment density) in rural settlements are disregarded.

[1] `https://datos.gob.mx/busca/dataset/cartografia-geoestadistica-urbana-cierre-de-los-censos-economicos-2014-denue-01-2015`.

[2] An urban settlement is considered when the total number of inhabitants is more than 2,500, which is the case for the vast majority in of areas in Greater Mexico City.

Appendix E

Correlation of surface area and accessibility

Table E.1 presents the results of the Pearson correlation test for each SAS for both accessibility measures presented in Chapter 5, namely *all* and *matching* employment opportunities. Surface area is entered in the logarithmic scale while accessibility is in the cube root. This is aimed at meeting the assumption of the test referring to linear relationship.

	SAS				
Type	Grid 0.5 km	Grid 1km	Grid 2km	Grid 4km	Post code
All	0.012***	0.020***	0.041***	0.080***	-0.182***
Matching	0.007**	0.013**	0.036***	0.075***	-0.044***

Signif. Codes: ***: >0.01, **: >0.05, *: >0.1
SAS = Spatial analytical scheme.

Table E.1: Bivariate correlation between accessibility and surface area in various SAS according to the Pearson correlation coeficient.

Appendix F

Distribution of accessibility according to Gini index

Table F.1 shows the full results of the Gini index for both matching and all accessibility to employment measures in all temporal stages for each Grid. Overall, it is shown that the estimates are larger for the 'Matching' measure than the alternative measure. Also, both panels show downward trends in time, except for the last temporal stage. This trend is consistent in all SAS and both measures.

	Stage	*Grid* 0.5 km	1 km	2 km	4 km	Post code	Overall (mean)
All							
	Before 2010	0.823	0.833	0.843	0.848	0.714	0.833
	2010 to 2012	0.819	0.831	0.842	0.848	0.709	0.831
	2012 to 2013	0.816	0.829	0.842	0.849	0.703	0.829
	2013 to 2015	0.810	0.823	0.839	0.846	0.697	0.823
	2015 to 2016	0.806	0.821	0.837	0.846	0.693	0.821
	2016 to 2018	0.806	0.821	0.837	0.845	0.692	0.821
	2018 to 2019	0.814	0.828	0.844	0.851	0.700	0.828
Matching							
	Before 2010	0.867	0.860	0.852	0.846	0.765	0.860
	2010 to 2012	0.865	0.859	0.851	0.846	0.762	0.859
	2012 to 2013	0.864	0.860	0.853	0.848	0.763	0.860
	2013 to 2015	0.861	0.857	0.852	0.847	0.759	0.857
	2015 to 2016	0.859	0.855	0.851	0.846	0.757	0.855
	2016 to 2018	0.859	0.855	0.851	0.846	0.757	0.855
	2018 to 2019	0.866	0.862	0.856	0.851	0.764	0.862

Table F.1: Gini index for all temporal stages of the MPTN between 2010 and 2019 for measures of accessibility considerin all and matching opportunities.

Appendix G

Prior sensitivity analysis

Table G.1: Prior sensitivity analysis. Stage 1 (Before 2010).

	Set1 (BYM)	Set2 (BYM)	Set3 (BYM)	Set4 (BYM)
Intercept	9.809	9.810	9.808	9.811
	(0.021)	(0.021)	(0.021)	(0.021)
ln Saleable area	0.780	0.780	0.780	0.780
	(0.002)	(0.002)	(0.002)	(0.002)
Class: Middle (ref:	0.114	0.114	0.114	0.114
Low-cost/Social)	(0.001)	(0.001)	(0.001)	(0.001)
Class: Semi-Luxury	0.205	0.205	0.205	0.205
	(0.002)	(0.002)	(0.002)	(0.002)
Class: Residential/Luxury	0.279	0.279	0.279	0.279
	(0.004)	(0.004)	(0.004)	(0.004)
N. of bathrooms: 2 (ref: 1)	0.042	0.042	0.042	0.042
	(0.001)	(0.001)	(0.001)	(0.001)
N. of bathrooms: 3	0.100	0.099	0.100	0.099
	(0.003)	(0.003)	(0.003)	(0.003)
N. of bathrooms: 3+	0.126	0.126	0.126	0.126
	(0.004)	(0.004)	(0.004)	(0.004)
N. of parking: 1 (ref:0)	0.077	0.077	0.077	0.077
	(0.001)	(0.001)	(0.001)	(0.001)
N. of parking: 2	0.177	0.177	0.177	0.177
	(0.002)	(0.002)	(0.002)	(0.002)
N. of parking: 3	0.208	0.208	0.208	0.208
	(0.004)	(0.004)	(0.004)	(0.004)
N. of parking: 3+	0.263	0.263	0.263	0.263
	(0.005)	(0.005)	(0.005)	(0.005)
$\ln age$	-0.031	-0.031	-0.031	-0.031
	(0.001)	(0.001)	(0.001)	(0.001)
$\ln age^2$	-0.015	-0.015	-0.015	-0.015
	(0.000)	(0.000)	(0.000)	(0.000)
Edu. facility: Yes	0.016	0.015	0.016	0.015
	(0.009)	(0.009)	(0.009)	(0.009)

(Continued on next page...)

Table G.1: Prior sensitivity analysis. Stage 1 (Before 2010). *(continued)*

	Set1 (BYM)	Set2 (BYM)	Set3 (BYM)	Set4 (BYM)
Public admin.: Yes	-0.004	-0.004	-0.004	-0.004
	(0.008)	(0.008)	(0.008)	(0.008)
Street with trees (%)	0.007	0.007	0.007	0.007
	(0.015)	(0.015)	(0.015)	(0.015)
Street light NA (%)	0.055	0.055	0.055	0.055
	(0.038)	(0.038)	(0.037)	(0.037)
Park (20 min): Yes	0.010	0.010	0.010	0.010
	(0.006)	(0.007)	(0.006)	(0.006)
Crime (homicides)	-0.003	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
ln HH with car	-0.183	-0.183	-0.184	-0.182
	(0.007)	(0.007)	(0.007)	(0.007)
Year: 2010 (ref:2009)	0.049	0.049	0.049	0.049
	(0.001)	(0.001)	(0.001)	(0.001)
State: Hidalgo (ref: Cd. de	-0.103	-0.101	-0.104	-0.100
México)	(0.067)	(0.067)	(0.067)	(0.068)
State: Mexico	-0.100	-0.100	-0.100	-0.099
	(0.018)	(0.018)	(0.018)	(0.018)
Accessibility car	0.035	0.034	0.035	0.034
	(0.011)	(0.011)	(0.011)	(0.011)
Accessibility PT_m	0.046	0.046	0.047	0.046
	(0.006)	(0.006)	(0.006)	(0.006)
σ^2_y	0.0184	0.0184	0.0184	0.0184
σ^2_ν	0.0036	0.0030	0.0034	0.0028
σ^2_υ	0.0220	0.0232	0.0226	0.0244
DIC	-168824	-168822	-168824	-168824
RMSE	0.1349	0.1349	0.1349	0.1349

Note:

Posterior marginal mean reported and standard deviation in parentheses. DIC = Deviance information criterion. RMSE = Root mean squared error.

Appendix H

Full base hedonic property value model results

The full results of the alternative base hedonic property models are shown in Table H.1. The results show that the structural variables, namely saleable area, class of the house quality, number of bathrooms, number of parking spaces, and age, are in the direction expected and significant in all alternative model specifications. The signs of the neighbourhood characteristics, namely availability of a major education establishment, presence of public administration building, proportion of streets with trees, lack of street light, crime, and car ownership, are as expected, and consistent across almost all models. One exception is the availability of a park, which shows a negative and significant coefficient. The control variables for year and state are significant the sign is in the direction expected. The dummy year variable shows a consistent increase for every additional year compared to the year of reference (2009). This is in line with the Housing Price Index for Greater Mexico City (SHF, 2021).

Model 1 (M1:AA) and Model 2 (M2:AA) of Table H.1 show the results for the linear and logarithmic form of distance to the CBD, respectively. M1:AA shows that the value of houses fall by 0.9% for every additional kilometre to the CBD. Similarly, M2:AA suggests that one percent increase in distance to the CBD is associated with a decrease in the value by 0.14%. It is worthwhile noting that while M1:AA results in higher explanatory power than M2:AA, the latter improves aspects related to mis-specification and collinearity issues, but not those related to heterogeneity of error variance. Model 3 (M3:AA) and Model 4 (M4:AA) introduce the measure of potential accessibility to employment by car $\{A_{\mathrm{CARj}}\}$ to Eq. (7.3) in the log-linear and log-log form, respectively. The relationship between A_{CARj} and house value in M3:AA results in an increase of 14.2% for an for each additional standard deviation. Likewise, the introduction accessibility by car in the log form $\{\log(A_{\mathrm{CARi}})\}$ in M4:AA shows an additional 0.09% in the value of houses for increase equivalent to one per cent. Potential accessibility to employment performs better in both of its two functional forms in terms of explanatory power than distance to the CBD. Furthermore, the log-linear form in M3:AA produces the highest R^2 and reduces the most possible misspecification and collinearity, while it generates the second lowest B-P test statistic.

In Model 5 (M5:AA) we include both distance to CBD in the log form and accessibility by car in the linear form, considering that they produced relatively good results in the previous models (from M1:AA to M4:AA). The results of M5:AA show that the goodness-of-fit does not improve, whilst worsening all the diagnostic measures (RESET/VIF/B-P statistic) compared to M3:AA. Additionally, distance to the CBD becomes non-significant. M6:AA is similar to M5:AA, exception that it enters distance to the CBD in the linear form. This time, the explanatory power does increase and both accessibility by car and distance to the CBD are significant. However, the possible misspecification and collinearity problems are considerably higher than in M3:AA, whereas the heterogeneity of error variance slightly improves.

Table H.1: House value determinants. Greater Mexico City, 2009-2019.

	M1:S1 (OLS)	M2:S1 (OLS)	M3:S1 (OLS)	M4:S1 (OLS)	M5:S1 (OLS)	M6:S1 (OLS)
Intercept	10.2***	10.2***	9.68***	10.0***	9.74***	9.87***
	(0.062)	(0.071)	(0.065)	(0.063)	(0.088)	(0.071)
ln Saleable area	0.758***	0.780***	0.785***	0.773***	0.785***	0.774***
	(0.017)	(0.016)	(0.014)	(0.017)	(0.015)	(0.016)
Class: Middle (ref:	0.185***	0.191***	0.187***	0.188***	0.185***	0.177***
Low-cost/Social)	(0.012)	(0.011)	(0.011)	(0.012)	(0.010)	(0.011)
Class:	0.430***	0.436***	0.410***	0.426***	0.410***	0.404***
Semi-Luxury	(0.021)	(0.019)	(0.020)	(0.021)	(0.019)	(0.019)
Class: Residen-	0.606***	0.609***	0.584***	0.597***	0.583***	0.577***
tial/Luxury	(0.030)	(0.030)	(0.031)	(0.031)	(0.030)	(0.030)
N. of bathrooms: 2 (ref: 1)	0.079***	0.078***	0.071***	0.078***	0.072***	0.073***
	(0.009)	(0.009)	(0.008)	(0.009)	(0.008)	(0.008)
N. of bathrooms: 3	0.126***	0.117***	0.117***	0.125***	0.117***	0.123***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
N. of bathrooms:	0.170***	0.150***	0.160***	0.175***	0.159***	0.167***
3+	(0.022)	(0.022)	(0.022)	(0.022)	(0.021)	(0.022)
N. of parking: 1 (ref:0)	0.087***	0.088***	0.085***	0.079***	0.087***	0.092***
	(0.009)	(0.008)	(0.008)	(0.010)	(0.008)	(0.008)
N. of parking: 2	0.244***	0.247***	0.242***	0.235***	0.244***	0.249***
	(0.012)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)
N. of parking: 3	0.277***	0.272***	0.272***	0.270***	0.273***	0.282***
	(0.014)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)
N. of parking: 3+	0.332***	0.317***	0.317***	0.319***	0.319***	0.334***
	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)	(0.020)
ln *age*	-0.031***	-0.024***	-0.022***	-0.027***	-0.022***	-0.025***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)

(Continued on next page...)

Table H.1: House value determinants. Greater Mexico City, 2009-2019. *(continued)*

	M1:S1 (OLS)	M2:S1 (OLS)	M3:S1 (OLS)	M4:S1 (OLS)	M5:S1 (OLS)	M6:S1 (OLS)
ln age^2	-0.013***	-0.014***	-0.014***	-0.014***	-0.014***	-0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Edu. facility: Yes	0.063***	0.064***	0.044***	0.051***	0.045***	0.048***
	(0.013)	(0.013)	(0.011)	(0.012)	(0.011)	(0.012)
Public admin.: Yes	0.070***	0.054***	0.033**	0.044**	0.033**	0.038**
	(0.014)	(0.013)	(0.012)	(0.014)	(0.012)	(0.013)
Street with trees (%)	0.088**	0.093***	0.065*	0.058*	0.060*	0.037
	(0.030)	(0.028)	(0.028)	(0.029)	(0.028)	(0.030)
Street light NA (%)	-0.162*	-0.202***	-0.195***	-0.283***	-0.191***	-0.154*
	(0.078)	(0.056)	(0.055)	(0.074)	(0.055)	(0.061)
Park: Yes	-0.037**	-0.034**	-0.036**	-0.034*	-0.037**	-0.041**
	(0.014)	(0.012)	(0.011)	(0.015)	(0.011)	(0.012)
Crime	-0.025***	-0.019***	-0.011***	-0.014***	-0.012***	-0.019***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
ln HH with car	-0.222***	-0.250***	-0.232***	-0.216***	-0.234***	-0.226***
	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)
Year: 2010 (ref:2009)	0.060***	0.059***	0.059***	0.059***	0.059***	0.060***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Year: 2011	0.106***	0.106***	0.107***	0.107***	0.107***	0.107***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Year: 2012	0.147***	0.145***	0.145***	0.145***	0.145***	0.146***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Year: 2013	0.214***	0.213***	0.213***	0.214***	0.213***	0.214***
	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Year: 2014	0.285***	0.287***	0.288***	0.289***	0.287***	0.286***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Year: 2015	0.378***	0.379***	0.379***	0.381***	0.379***	0.380***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Year: 2016	0.454***	0.453***	0.454***	0.455***	0.455***	0.456***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Year: 2017	0.536***	0.532***	0.535***	0.536***	0.535***	0.538***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Year: 2018	0.611***	0.606***	0.608***	0.611***	0.608***	0.611***
	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
Year: 2019	0.678***	0.674***	0.677***	0.679***	0.677***	0.679***
	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)
State: Hidalgo (ref: Cd. de México)	-0.207***	-0.298***	-0.328***	-0.327***	-0.315***	-0.230***
	(0.042)	(0.030)	(0.026)	(0.039)	(0.028)	(0.037)
State: Mexico	-0.285***	-0.291***	-0.294***	-0.313***	-0.287***	-0.260***
	(0.020)	(0.018)	(0.015)	(0.017)	(0.017)	(0.018)
Dist. to CBD (km)	-0.009***					-0.005***
	(0.001)					(0.001)
ln Dist. to CBD		-0.136***			-0.019	
		(0.014)			(0.018)	
Accessibility car			0.133***		0.123***	0.103***
			(0.009)		(0.014)	(0.011)

(Continued on next page...)

Table H.1: House value determinants. Greater Mexico City, 2009-2019. *(continued)*

	M1:S1 (OLS)	M2:S1 (OLS)	M3:S1 (OLS)	M4:S1 (OLS)	M5:S1 (OLS)	M6:S1 (OLS)
ln Accessibility car				0.087***		
				(0.010)		
Observations	781,898	781,898	781,898	781,898	781,898	781,898
R2	0.9436	0.9429	0.9455	0.9438	0.9455	0.9462
Adj. R2	0.9436	0.9429	0.9455	0.9438	0.9455	0.9462
RMSE	0.2014	0.2027	0.1981	0.2010	0.1980	0.1966
RESET, statistic[a]	28855.1	16611.3	14543.3	24922.2	15553.6	24230.5
VIF, mean[b]	3.473	3.348	3.232	3.274	3.933	3.793
B-P test, statistic[c]	62688.4	67887.0	62113.5	57928.6	62422.0	60408.1

Note:

Clustered standard-errors by post code in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

[a] RESET = Ramsey reset test.

[b] VIF = Variance of inflation factor.

[c] B-P = Breusch-Pagan test.

Appendix I

MAUP sensitivity analysis

I.1 Introduction and objective

The objective of this sensitivity analysis is to examine the effects of the modifiable areal unit problem (MAUP) in the study of capitalizations of the main public transport network (MPTN) on residential land value in the context of land value capture.

The general approach is to hold the individual level data and analytical methods constant while employing five alternative spatial aggregation schemes (SAS) to account for neighbourhood and locational attributes associated to the dwelling. The SAS employed are four hexagonal grids of 0.5 km, 1 km, 2 km and 4 km, and the post codes zone system, as defined in Section 5.2 in Chapter 5. Specifically, the analytical steps for each of the SAS defined are the following:

1. Calibrate the spatial deterrence parameter for MPTN accessibility measures.
2. Estimate a hedonic function using a fixed model specification.
3. Estimate the individual and aggregated willingness to pay for MPTN accessibility.

I.1.1 Data and measures

The property value data employed in the present analysis is a subset of the main dataset obtained from the Federal Mortgage Society (SHF, for its acronym in Spanish) referred to in Chapter 5. The observations included here consider dwellings valuated between April 2012 and April 2014. This timespan is defined based on two main considerations. First, the exact location (geocoded at the parcel or block level) is only available for records registered between the year 2009 and early-2015. This level of geographic detail is important to freely study the effects of aggregating neighbourhood and locational attributes according to different spatial schemes. Therefore, this study is constrained to the availability of such information.

The analysis uses five datasets which are constructed by spatially linking the locational characteristics resulting from the aggregation or computation of upper-level measures at each SAS to the individual observations. The dependent variable (assessed value of a dwelling, P_i) and the structural characteristics of the property are directly obtained from the SHF records. The data for neighbourhood and locational characteristics comes from a variety of sources, as detailed in Chapter 5.

The accessibility benefits enabled by the MPTN are assessed using a location-based measure to employment. A simplified accessibility specification is chosen with the aim to focus on the effects of MAUP, making comparisons as clear as possible. This is given by the following formula:

$$A_{jt}^{\mathrm{PT}} = \sum_{j=1}^{J} E_k e^{-\beta_2 d_{jkt}}. \tag{I.1}$$

MPTN accessibility at location j at the MPTN temporal stage t is given by the sum of total employment at potential destination E_k which is affected by the product of a negative exponential function of travel time from the origin at stage t, d_{jkt}. The distance deterrence is regulated by the parameter β_2. The estimation details for β_2 are provided in the next section. Travel time, d_{jkt}, considers a multi-stage door-to-door routing approach using a combination of walking and the modes of the MPTN, as detailed in Chapter 6. This is modelled according to the services available in the respective temporal stage.

I.1.2 Model specifications

Spatial decay in the MPTN

The first analytical step estimates the spatial deterrence parameter β_2 in Eq. (I.1) for each SAS. A simple model specification is used which includes structural controls and minimal locational controls as following:[1]

$$\ln(P_i) = \alpha + \sum_{s=1}^{S} \beta_s \boldsymbol{x}_{is} + \sum_{l=1}^{L} \beta_l \boldsymbol{x}_{il} + \beta_{\mathrm{PT}} \ln\left(\sum_{j=1}^{J} E_k \exp(-\beta_2 d_{jk})\right) + \beta_{year} Year + \epsilon_i. \tag{I.2}$$

In Eq. (I.2), P_i is the assessed value of dwelling i in MXN at current prices; α is a constant; $\boldsymbol{x}_{is}$ is a vector including individual structural characteristics s for property i (i.e. saleable area, class, number of bathrooms, number of parking spaces, and age of the building); $\boldsymbol{x}_{il}$ is a vector of the locational characteristics l for property i (i.e. distance to CBD, and HH with car). MPTN accessibility (A_j^{PT}) enters the equation in the log form; $Year$ is a time control dummy variable indicating the year in which the property

[1] Here, structural controls are required to account for the improvements on residential land (the structure of the dwelling), in contrast to Ahlfeldt & Wendland (2016) which use land value directly.

value was assessed (reference is the year 2012). $\beta_s, \beta_l, \beta_{\mathrm{PT}}, \beta_{year}$ are the linear regression coefficients to be estimated; α is a constant, and; ϵ_i is assumed to be an independent and normally distributed error term. The continuous variables enter in the log form with the exception of distance to CBD, which was found to be more adequate in the linear form in Chapter 7. Additionally, age is entered both in the logarithmic as well as in the log-squared form as $\log(age)^2$, following Osland & Thorsen (2008). This helps to improve the model parsimony, as found in a preliminary modelling phase.

Spatial multilevel model

The specification and method employed to estimate the hedonic function used in this analysis draws on the results of Chapter 7. The hedonic spatial multilevel model is written as following:

$$\begin{aligned} \mathrm{E}(\ln P_{ij}) = \eta_{ij} &= \alpha + \sum_{s=1}^{S} \beta_s x_{ijs} + \sum_{l=1}^{L} \beta_l x_{ijl} + \beta_{year} Year + \nu_j + \upsilon_j + \epsilon_{ij}, \\ \epsilon_{ij} &\sim N(0, \sigma_y^2), \\ \nu_j &\sim N(0, \sigma_\nu^2), \\ \upsilon_j \mid \boldsymbol{\upsilon}_{-j} &\sim N\left(\frac{\sum_{j=1}^{J} w_{jk} \upsilon_k}{\sum_{j=1}^{J} w_{jk}}, \frac{\sigma_\upsilon^2}{\sum_{j=1}^{J} w_{jk}} \right), \\ \sigma_y^2, \sigma_\nu^2, \sigma_\upsilon^2 &\sim \mathrm{Gamma}(a, b). \end{aligned} \tag{I.3}$$

Here, $\boldsymbol{x}_{ijs}$ is a vector of structural characteristic s of dwelling i in j (i.e. saleable area, class, number of bathrooms, number of parking spaces, and age of the building); $\boldsymbol{x}_{ijl}$ is a vector denoting locational characteristic l for property i in j (i.e. HH with car, accessibility by car, accessibility by the MPTN, and state); $Year$ denotes a set of dummy variables indicating the year in which the property value was assessed (the reference year is 2012); βs are the linear regression effects for the aforementioned vectors and α is a scalar denoting the intercept; ν_j is unstructured random effect (independent and identically distributed, iid) and υ_j is an spatially structured random effect (intrinsic conditional autoregressive, iCAR), according to the Besag-York-Mollie model (BYM) (Besag et al., 1991) discussed in Chapter 5; ϵ_{ij} is the individual lower-level error term with variance σ_y^2.

Eq. (I.3) is estimated in a Bayesian setting by the integrated nested Laplace Approximation (INLA) method (presented in Chapter 5). The lower- and upper-level variance parameters follow a Gamma distribution. These parameters enter to the INLA framework in the logarithmic form (as logGamma) and are expressed as the precision by the following hyperparameters $\boldsymbol{\psi} = \{\tau_y^2, \tau_\nu^2, \tau_\upsilon^2\}$.[2]

[2]Within INLA's framework the variance is expressed by the precision as $\tau = 1/\sigma^2$.

I.1.3 Implicit prices and willingess to pay

To further examine the potential implications resulting from aggregating the upper-level characteristics using alternative SAS, the willingness to pay for MPTN accessibility is estimated for the equivalent to the MPTN accessibility change (ΔA_j^{PT}) introduced in temporal Stage 3 with respect to Stage 2. The change in accessibility is computed as following:

$$\Delta A_j^{\mathrm{PT}} = A_{jt_{\mathrm{S3}}}^{\mathrm{PT}} - A_{jt_{\mathrm{S2}}}^{\mathrm{PT}} \tag{I.4}$$

In a third analytical step, the individual willingness to pay WTP and aggregate WTP (total willingness to pay, TWTP) for MPTN accessibility is estimated in line with Subsection 5.4.6 in Chapter 5.

The TWTP is estimated for two scenarios with the purpose of illustrating the possible degree of uncertainty introduced by the MAUP. The first employs all the observations in the sample examined here while the second considers only observations in the highest WTP decile for each SAS. The focus on the latter is to represent a reasonable fraction of the affected stock that may be included in a LVC intervention, according to some of the elements that may support their feasibility in policy implementation as discussed in Chapter 2 and Chapter 7. It should be noted that TWTP estimates are intended for internal comparative purposes only using a sample at hand. These do not represent comprehensive assessment of the welfare change.

I.2 Results

I.2.1 Descriptive statistics

Property data

Table I.1 presents a descriptive statistics summary for the datasets assembled according to different SAS examined. The mean property value of the sample selected is MXN$1.3 million (GBP£47,100)[3] at current prices. The average saleable area is 96 m^2. Half of the properties' construction class is middle (50%). This is followed by low-cost or social houses which represent about a third of the sample. A little less than one tenth of the properties is semi-luxury (9%) and only a small fraction (3%) are luxury (or 'Residential') class dwellings. Most of the properties has only one bathroom (62%) and more than a quarter has two (26%). The remaining observations (12%) has three or more bathrooms. Most of the houses have one parking space (61%) while about one in every six houses has

[3]Considering an exchange rate of MXN27.60 per one GBP, according currency exchange published by the Central Bank of Mexico (*Banxico*) on the 15/09/2021 on the website https://www.banxico.org.mx.

	Grid 0.5 km	Grid 1 km	Grid 2 km	Grid 4 km	Post code
	(N=50 381)	(N=50 381)	(N=50 381)	(N=50 381)	(N=50 381)
Value (MXN/1000)					
Mean (SD)	1 300.22 (1580.18)	1 300.22 (1580.18)	1 300.22 (1580.18)	1 300.22 (1580.18)	1 300.22 (1580.18)
Saleable area (sqm.)					
Mean (SD)	95.67 (79.40)	95.67 (79.40)	95.67 (79.40)	95.67 (79.40)	95.67 (79.40)
Class					
Low-Cost/Social	18 869 (37.5%)	18 869 (37.5%)	18 869 (37.5%)	18 869 (37.5%)	18 869 (37.5%)
Middle	25 333 (50.3%)	25 333 (50.3%)	25 333 (50.3%)	25 333 (50.3%)	25 333 (50.3%)
Semi-luxury	4 718 (9.4%)	4 718 (9.4%)	4 718 (9.4%)	4 718 (9.4%)	4 718 (9.4%)
Residential/Luxury	1 461 (2.9%)	1 461 (2.9%)	1 461 (2.9%)	1 461 (2.9%)	1 461 (2.9%)
N. of bathrooms					
1	31 235 (62.0%)	31 235 (62.0%)	31 235 (62.0%)	31 235 (62.0%)	31 235 (62.0%)
2	13 157 (26.1%)	13 157 (26.1%)	13 157 (26.1%)	13 157 (26.1%)	13 157 (26.1%)
3	4 094 (8.1%)	4 094 (8.1%)	4 094 (8.1%)	4 094 (8.1%)	4 094 (8.1%)
3+	1 895 (3.8%)	1 895 (3.8%)	1 895 (3.8%)	1 895 (3.8%)	1 895 (3.8%)
N. of parking spaces					
0	8 384 (16.6%)	8 384 (16.6%)	8 384 (16.6%)	8 384 (16.6%)	8 384 (16.6%)
1	30 746 (61.0%)	30 746 (61.0%)	30 746 (61.0%)	30 746 (61.0%)	30 746 (61.0%)
2	8 819 (17.5%)	8 819 (17.5%)	8 819 (17.5%)	8 819 (17.5%)	8 819 (17.5%)
3	1 553 (3.1%)	1 553 (3.1%)	1 553 (3.1%)	1 553 (3.1%)	1 553 (3.1%)
3+	879 (1.7%)	879 (1.7%)	879 (1.7%)	879 (1.7%)	879 (1.7%)
Age (years)					
Mean (SD)	11.93 (12.69)	11.93 (12.69)	11.93 (12.69)	11.93 (12.69)	11.93 (12.69)
HH with car (%)					
Mean (SD)	47.84 (18.14)	48.60 (16.11)	50.63 (13.94)	52.19 (12.02)	47.95 (16.43)
Distance to CBD (Km)					
Mean (SD)	16.66 (11.29)	16.67 (11.27)	16.67 (11.31)	16.68 (11.24)	16.66 (11.22)
Accessibility car					
Mean (SD)	1.96 (1.33)	2.01 (1.38)	2.14 (1.52)	2.32 (1.57)	1.72 (1.16)
State					
Ciudad de México	27 213 (54.0%)	27 213 (54.0%)	27 213 (54.0%)	27 213 (54.0%)	27 213 (54.0%)
México	23 168 (46.0%)	23 168 (46.0%)	23 168 (46.0%)	23 168 (46.0%)	23 168 (46.0%)
Year					
2012	35 878 (71.2%)	35 878 (71.2%)	35 878 (71.2%)	35 878 (71.2%)	35 878 (71.2%)
2013	14 503 (28.8%)	14 503 (28.8%)	14 503 (28.8%)	14 503 (28.8%)	14 503 (28.8%)

Note:
Note for desciptive statistic values: numeric variables = M(SD): Categorical variables = N(%).

Table I.1: Descriptive statistics

none or two (17%). The average age of the dwellings in the sample is 12 years.

The neighbourhood and locational measures present systematic differences according to the SAS employed. For example, the mean percentage of households owing at least one car gradually increases from 48% in the highest-resolution scheme to 52% in the lowest (4 km Grid). The figure is very similar between the 0.5 km Grid and the post-code zones (48%). Distance to the CBD is consistent across the different SASs at 17 km on average. The potential accessibility to employment by car displays an upward trend as the scale of the grid increases. While the average accessibility by car is 1.96 standard deviation units in the 0.5 km Grid, this is 2.3 in the 4 km Grid. The post code scheme presents the lowest average (1.7). The sample is roughly split by half between the state of México and Ciudad de México. The value of most of the

	Grid 0.5 km	Grid 1 km	Grid 2 km	Grid 4 km	Post code
	(NLM)	(NLM)	(NLM)	(NLM)	(NLM)
ln Accessibility MPTN	0.091***	0.092***	0.093***	0.079***	0.077***
	(0.006)	(0.008)	(0.009)	(0.014)	(0.013)
MPTN spat. decay ($\hat{\beta_2}$)	0.123***	0.110***	0.096***	0.091***	0.107***
	(0.014)	(0.017)	(0.016)	(0.023)	(0.016)
Structural controls	Yes	Yes	Yes	Yes	Yes
Year controls	Yes	Yes	Yes	Yes	Yes
Observations	50 381	50 381	50 381	50 381	50 381
Log-Likelihood	5 798	5 338	5 252	3 832	5 107
AIC	-11 559	-10 638	-10 466	-7 626	-10 175

Note:
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1; Standard error clustered at the respective SAS. AIC= Akaike information criterion.

Table I.2: MPTN spatial decay parameter across various spatial analytical schemes.

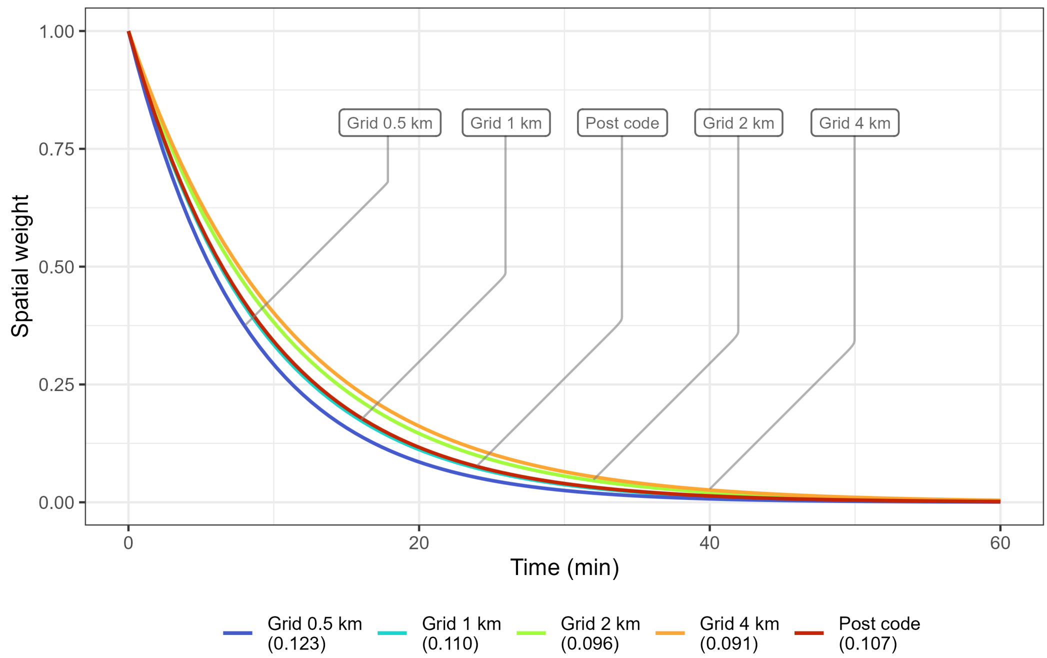
observations (71%) was assessed in the year 2012 and the rest (29%) in the year 2013.

I.2.2 MAUP and spatial decay in the main public transport network

Table I.2 shows the results of the specification in Eq. (I.2). To increase visibility, the estimates for the structural and year controls are not shown. All regression coefficients for these covariates are significant and in the expected direction. The coefficient for MPTN accessibility is significant and positive in all schemes. The magnitude of these coefficients is within a relative small range. The smallest estimate is given by the post code scheme, where it is implied that a doubling MPTN accessibility increase the value of residential land by 7.7% . The coefficients estimated using a 0.5 km, 1 km, and 2 km grids are very similar to each other, implying increase in the land value of between 9.1% and 9.3% percent for an increase in MPTN accessibility of the same order. The estimate using the 4 km grid is lowest among the hexagonal grids, suggesting a 7.9% increase in the value of land for and equivalent improvement in accessibility to employment.

The focus in this analytical step is the spatial decay parameter in MPTN accessibility, β_2. Since the model specification Eq. (I.2) assumes this parameter takes a negative value, the reported results should show a positive sign. The results are in line with the expectation in all models shown in Table I.2. The estimated parameters using an alternative SAS ranges from 0.91 for the 4 km Grid to 0.12 in the 0.5 km Grid. A further interesting finding is that the parameter changes systematically according to the SAS used. The estimated strength at which people discount the attractiveness of employment opportunities as a function of distance declines as the resolution of the grid is reduced.

This trend is illustrated in Figure I.1. The attractiveness of opportunities decline in a negative exponential form by 0.12 in the highest resolution grid. This gradually changes to 0.09 in the lowest resolution grid. The 1 km Grid and the post code zones are in between and produce almost identical results (0.11 and 0.107, respectively). Meanwhile, the estimate for the 2 km Grid is close to the lowest resolution scheme. This results can be associated to the fact that travel time is computed from/to weighted centroids. This implies that the highest resolution scheme can capture *enough* employment opportunities at relatively short distances, while the coarser resolutions consider opportunities at further locations.



Source: The author based on own calculations.

Figure I.1: Estimated spatial decay parameter for various grids.

I.2.3 MAUP and implicit prices in the spatial multilevel hedonic model

The results for spatial multilevel hedonic models according to the specification in Eq. (I.3) for the five SASs are shown in Table I.3. The locational variables are all significant and in the expected direction in all models. Although not the focus of the the present work, it is observed that the coefficients estimated for the structural components vary systematically according to the aggregation level of the locational characteristics. Except for age in the log squared form $\{\ln(age)^2\}$, all structural characteristics show upward trends as the resolution of the grid decreases. The saleable area shows only small variations. However, other characteristics such as the number of parking spaces imply substantial differences. For example, having two bathrooms is associated to an additional 5% in the value compared to properties that have only one using a 0.5 km Grid, while this is nearly 8% in the 4 km Grid. In general, the estimates for the structural characteristics in the post code scheme are close to the 1 km grid or in between the 0.5 km and the 1 km Grid. These variations in the structural estimates are somehow unexpected, since these are characteristics entered at the individual-level. Therefore, these are exactly the same for all datasets.

As the structural covariates' coefficients, all the locational ones are significant, in the direction expected, and substantially vary according to the SAS employed as shown in Table I.3. Specifically, the proportion of households with a car shows a gradual upward trend according to an increase in the size of the gird used. Yet, the lowest resolution grid present a value slightly smaller than the 2 km Grid. Also, the coefficient for this characteristic in the post code model is very similar that in the the 4 km Grid. Regarding accessibility by car, the estimates produced by the 0.5 km Grid and 1 km Grid are the highest and are similar to each other. These suggest that one standard deviation increase in access to employment by car is associated to an additional value of land between 0.10% and 0.11%. The size of these coefficients is followed by the 2 km Grid and the 4 km Grid. These display similar values to each other. The coefficients imply increments in the value of land of approximately 0.06% for a rise in car accessibility of the same order. This time, the coefficient produced by the post code zones is not within the range of approximations produced by the hexagonal grids and it is in fact the smallest. The results suggest that for one standard deviation increase in accessibility by car the value of land is expected to rise by 0.03%.

In contrast to the considerable variability of the estimates for accessibility by car, the MPTN accessibility coefficients are within a small range. The percentage increase in residential land value for the estimates produced by the hexagonal grids is around 3.0% for a doubling in MPTN accessibility. Interestingly, in three out of the four hexagonal grids the coefficient is consistently estimated around this figure, except for the 2 km Grid. According to the latter, the magnitude of these coefficients imply that a doubling in accessibility is associated to a 1.7% increment in the value of land. Furthermore, the estimate produced by the model using post code zones is close to range estimated produce for the hexagonal grids. In this SAS, the gain in residential land value is 3.4% under a doubling of MPTN accessibility.

The regression coefficients for the overall control variables also display some variability. First, the state where the property is located acts both as a location characteristic but also as an overall control for the quality of the public services. These

	Grid 0.5 km	Grid 1 km	Grid 2 km	Grid 4 km	Post code
	(BYM)	(BYM)	(BYM)	(BYM)	(BYM)
Intercept	10.017***	9.895***	9.826***	9.747***	10.053***
	(0.020)	(0.022)	(0.024)	(0.029)	(0.025)
ln Saleable area	0.795***	0.800***	0.799***	0.805***	0.787***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Class: Middle	0.088***	0.103***	0.127***	0.142***	0.114***
(ref: Low-cost/Social)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Class: Semi-Luxury	0.153***	0.195***	0.254***	0.301***	0.182***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Class:	0.183***	0.249***	0.329***	0.393***	0.227***
Residential/Luxury	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
N. of bathrooms: 2	0.049***	0.052***	0.067***	0.076***	0.054***
(ref: 1)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
N. of bathrooms: 3	0.069***	0.072***	0.093***	0.103***	0.096***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N. of bathrooms: 3+	0.069***	0.077***	0.095***	0.100***	0.090***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)
N. of parking: 1	0.084***	0.098***	0.101***	0.122***	0.090***
(ref: 0)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
N. of parking: 2	0.171***	0.192***	0.211***	0.237***	0.176***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
N. of parking: 3	0.209***	0.233***	0.248***	0.267***	0.213***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
N. of parking: 3+	0.258***	0.276***	0.296***	0.323***	0.266***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
$\ln age$	-0.021***	-0.025***	-0.026***	-0.026***	-0.029***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
$\ln age^2$	-0.017***	-0.017***	-0.016***	-0.014***	-0.016***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
ln HH with car	-0.162***	-0.179***	-0.215***	-0.205***	-0.202***
	(0.006)	(0.009)	(0.015)	(0.023)	(0.009)
Accessibility car	0.099***	0.107***	0.057***	0.064***	0.033***
	(0.009)	(0.013)	(0.015)	(0.021)	(0.013)
ln Accessibility MPTN	0.030***	0.030***	0.017***	0.028***	0.034***
	(0.004)	(0.004)	(0.005)	(0.007)	(0.005)
State: Mexico	-0.194***	-0.117***	-0.069***	-0.066***	-0.089***
(ref: Cd. de México)	(0.013)	(0.012)	(0.009)	(0.007)	(0.022)
Year: 2013	0.044***	0.046***	0.048***	0.049***	0.046***
(ref: 2012)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
MPTN spat. decay ($\hat{\beta_2}$, imputed)	0.1230	0.1096	0.0963	0.0912	0.1075
$\hat{\sigma_y^2}$	0.0184	0.0236	0.0282	0.0332	0.0226
$\hat{\sigma_\nu^2}$	0.0089	0.0058	0.0056	0.0041	0.0034
$\hat{\sigma_\upsilon^2}$	0.0332	0.0299	0.0123	0.0116	0.0315
Observations	50381	50381	50381	50381	50381
Groups	4961	1883	666	240	1629
Log-Likelihood	25446	21156	17592	13896	22697
DIC	-54707	-44311	-36191	-28389	-46813
RMSE	0.1304	0.1514	0.1671	0.1818	0.1483

Note:
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1; Correspond to 99%, 95% or 90% of the highest posterior density (HPD) credible interval not including zero, respectevely. Posterior marginal mean reported with corresponding standard deviation in parentheses. DIC = Deviance information criterion. RMSE = Root mean squared error.

Table I.3: Spatial multilevel Besag-York-Mollié (BYM) model results for various grid scales.

have been reported to be lower in the state of *México* than in *Ciudad de México*. This is reflected in the negative coefficient associated to observations located in the state of Mexico compared to Ciudad de México. The divergence in the value of land for being located in one of these states ranges from 7% in the 4 km Grid to 21% in the 0.5 km grid. Second, the temporal control entered in the year in which the value of the property was assessed shows small variability. This ranges from an additional 4.5% to 5% in land value for observations assessed in the year 2013 compared to the year 2012.

The summary statistics for the models show gradual variations according to the SAS used. For example, the individual variance parameter σ_y^2 increases in higher resolution grids. The individual-level variance for the 0.5 km Grid is 0.018 and this progressively increases up to 0.032 in the 4 km Grid. This implies that the individual residuals widen as spatial aggregation of neighbourhood and locational attributes increase. σ_y^2 is very similar between the post code scheme and the 1 km Grid. Regarding the upper-level variance, it is clear that the spatially structured component is capturing most of the the variance in all models (i.e. σ_υ^2 is consistently larger than σ_ν^2). This reflects the presence of strong spatial autocorrelation. The value of both σ_ν^2 and σ_υ^2 parameters decrease as the scale of the grid increases. For instance, σ_υ^2 diminishes from 0.03 in the 0.5 km Grid to 0.01 in the 4 km grid. The spatially structured variance for the post code zones is between the 0.5 and the 1 km grid.

The opposite direction of the trends between the lower-level and the upper-level variance is in line with the expectation. This makes sense if it considered that in high resolution schemes the number of observations within groups is smaller and the information about the locational characteristics is more detailed. Therefore, the within variance is reduced. In the latter case (upper-level parameters), the variance reflects that extreme values are less likely to occur in low resolution schemes due to the spatial smoothing process referred to in the MAUP literate (Wong, 2009b). Accordingly, the degree of heterogeneity between zones in fine-grained schemes is high while these differences flatten in large zones.

Concerning the goodness-of-fit measures, both the RMSE and DIC measures consistently identify the 0.5 km grid as the most satisfactory model. As expected, the RMSE gradually increases as the size of the grid increases. It is interesting to observe that despite the mean surface area (scale) of post code zones is larger than the 1 km Grid by two thirds, the goodness-of-fit measures favour the former.

	Grid 0.5 km	Grid 1 km	Grid 2 km	Grid 4 km	Post code
Full sample					
Negative obs.	8 155	9 587	2 220	3 961	9 769
	(16.19%)	(19.03%)	(4.41%)	(7.86%)	(19.39%)
Neutral obs.	8 854	11 718	16 119	20 163	9 306
	(17.57%)	(23.26%)	(31.99%)	(40.02%)	(18.47%)
Positive obs.	33 372	29 076	32 042	26 257	31 306
	(66.24%)	(57.71%)	(63.60%)	(52.12%)	(62.14%)
TWTP	46 961 692	33 580 654	15 800 128	29 415 761	40 220 745
	(4 470.77)	(3 446.90)	(1 709.40)	(2 315.38)	(3 643.57)
Selected sample					
Count	5 039	5 039	5 040	5 039	5 040
	(10.00%)	(10.00%)	(10.00%)	(10.00%)	(10.00%)
WTP median	4 280.09	3 424.32	1 374.98	3 710.78	4 194.09
WTP mean	8 393.45	6 056.78	2 655.43	5 118.73	7 143.69
TWTP	42 294 575	30 520 098	13 383 363	25 793 282	36 004 190
	(11 712.66)	(9 268.43)	(4 788.04)	(5 522.51)	(9 344.46)

Note:
In desciptive statistic values: numeric variables = M(SD); Categorical variables = N(%). Currency is MXN at current prices. WTP = Willingess to pay. TWTP = Total willingness to pay.

Table I.4: Total willingness to pay summary

I.2.4 MAUP and the willingness to pay for accessibility to employment

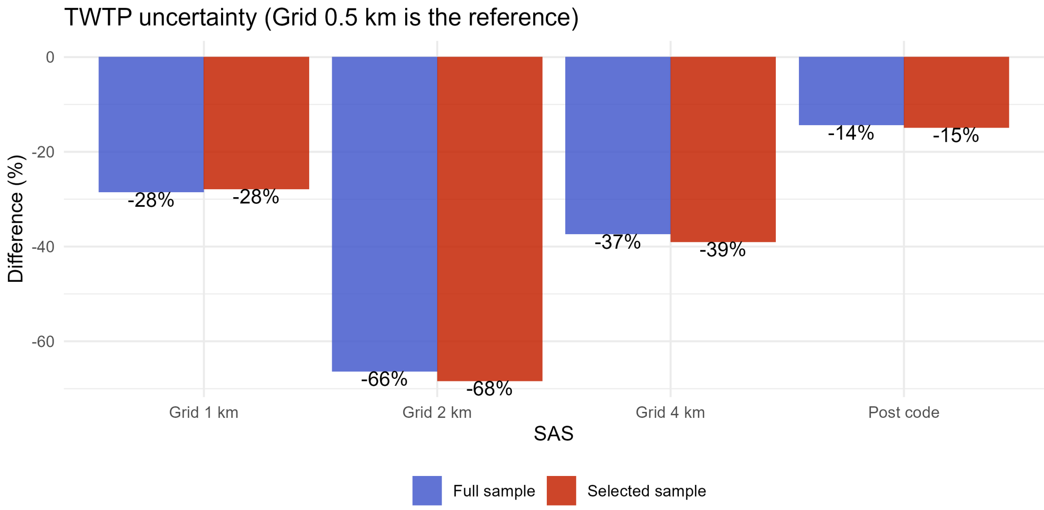
Having identified the changes in MPTN accessibility, the WTP for MPTN accessibility according to changes introduced in Stage 3 are estimated for each observations in the sample. This estimate reflects three elements: (1) changes in the level of MPTN accessibility; (2) implicit price of MPTN accessibility (from hedonic function), and; (3) the assessed value of the property (P_i). This section provides a summary for the individual WTP and the TWP. Rather than providing a full assessment of the aggregated value derived from the public transport interventions in Stage 3 (i.e. total welfare effects), this is intended to expose the proportional potential differences using an identical sample without introducing further factors, e.g. housing stock.

First, the number of observations that were estimated to receive negative, neutral, or positive impacts is presented in Table I.3. There are relevant contrasts. For instance, while only 4% of the observations are considered to be impacted negatively in the 2 km Grid, this is 19% in post code zones and the 1 km Grid. Similarly, the neutral effects range from 40% in the 4 km Grid to 18% in the 0.5 km Grid. Considering the positive effects, the 0.5 km Grid reflects the largest proportion, followed by the 2 km Grid and the post code scheme.

Table I.3 shows the aggregated WTP (TWTP) considering all observations in the sample. Here, the largest estimate is given by the 0.5 km Grid, which adds up to MXN$47 million (GBP£2 million). This is followed by the post code zones, which produces a TWTP of MXN$40 million. The estimate substantially decreases as the spatial resolution increases in the 1 km and 2 km Grid, summing MXN$34 million and MXN$16 million, respectively. Despite the lowest resolution SAS estimated the smallest number of positive observations, it produced a MXN$29 million TWTP, which is similar to that of the 1 km Grid. This implies that even the estimated coefficient for MPTN is similar to the 0.5 km, the accessibility gains show less variance. This can be observed in the standard deviation of the estimates in parenthesis. For example, the 0.5 km show a large standard deviation. This is because the WTP covers a smoother and broader range than the rest of SAS.

The lower panel in Table I.3 shows the results of the analysis including only the highest WTP decile for each SAS (selected sample). In the selected sample, the median individual WTP is more or less consistent across most of the SAS, which ranges between MXN$3,400 (GBP£120) in the 1 km Grid and MXN$4,300 (£155) in the 0.5 km Grid. This is with the exception of the 2 km Grid, where the estimate is as low as MXN$1,400 (GBP£50). This can be attributed to the relatively small estimated coefficient at this resolution. The mean WTP shows a similar pattern as the median. Yet, the mean is well above the median WTP, which is expected since this resembles the skewed distribution of the full value of the properties. This part of the analysis includes the TWTP for the selected sample. A first thing to note is that TWTP does not show substantial differences compared to the previous estimate using the total sample. This is due to the type of density distribution of the WTP, where many of the observations are expected to have only small effects. This also implies that few of the most benefited observations show important potential for the recovery of land value. Here, the TWTP for the highest resolution grid is MXN$42 million (compared to MXN$47 million in the full sample), this falls to MXN$30 million in the 1 km Grid and only MXN$13 million in the 2 km Grid. The TWTP for the post code zones is between the 0.5 km Grid and the 1 km Grid. The results of this analysis suggest that the closest coincidences are between the 0.5 km Grid and the post code scheme.

Figure I.2 illustrates the uncertainty in the TWTP introduced by the MAUP based on Table I.3. The plot shows the relative difference compared to the 0.5 km grid for both the full sample and the selected sample. The largest difference is generated by the 2 km grid, suggesting that the TWTP is almost 70% smaller than the reference. This is followed by the 4 km grid, resulting in an estimate of almost 40% smaller than the 0.5 km grid. The 1 km grid is about 30% below the reference. Interestingly, the post code is the closest to the reference halving the uncertainty compared to the 1 km grid, from approximately -30% to about -15%.



Source: The author based on own calculations. *Note:* Grid 0.5 km is the reference.

Figure I.2: TWTP uncertainty.

I.3 Final remarks

The analyses illustrate that the spatial decay parameter varies systematically according to scale. The strongest discount of employment attractiveness as a function of distance is given by the highest resolution grid (0.12). This parameter gradually relaxes as the resolution of the SAS increase.

The results indicate that all firs stage hedonic model coefficients are significant and in the direction expected. This includes MPTN accessibility. Thus, it is shown that the overall relationship between the potential accessibility to employment generated by the MPTN and land value is not simply because of the way the information is partitioned in space. Yet, the findings show that the magnitude of the coefficients vary in both systematic and unsystematic fashion depending of the type of covariate. Specifically, the magnitude of most of the structural attributes tend to increase as the resolution lowers.

The MAUP effects are further examined in terms of the individual WTP and TWTP for changes in the MPTN equivalent to the introduced in the MPTN Stage 3. It is shown that both individual and aggregate WTP estimates are very sensitive to scale effects. Considering the results produced using the selected sample, it is shown that both the mean WTP and TWTP are about 30% and 70% below the 0.5 km Grid for the 1 km and 2 km Grid, respectively. The estimates for the 4 km Grid lie between these two.

In terms of zoning and the WTP and TWTP, the results show that the post code scheme produces the closest results relative to the highest resolution and the most accurate SAS, i.e. the 0.5 km Grid. For example, both the WTP and TWTP are 15%

below the 0.5 km Grid. This scheme also produces the second best goodness-of-fit. This is despite the fact that the mean surface area of the post code zones is between the 1 km Grid and the 2 km Grid. This suggests that zoning compensates its relatively large scale the study case studied. This makes sense since the boundaries of post code zones often follow physical barriers of the urban context which can ultimately shape residents' accessibility perception. By contrast, regular grids, especially of medium and large scale, disregard the features of the urban layout and may introduce higher heterogeneity within the area that intends to represents all residents within j.

References

Abiad, A., Farrin, K., & Hale, C. (2019). *Sustaining Transit Investment in Asias's Cities* (No. 9789292615727). Manila: Asian Development Bank. Retrieved from https://dx.doi.org/10.22617/ARM190085-2

Adair, A., McGreal, S., Smyth, A., Cooper, J., & Ryley, T. (2000). House Prices and Accessibility: The Testing of Relationships within the Belfast Urban Area. *Housing Studies*, *15*(5, 5), 699–716. http://doi.org/10.1080/02673030050134565

Ahlfeldt, G. M. (2011). If Alonso was right: Modeling accessibility and explaining the residential land gradient. *Journal of Regional Science*, *51*(2, 2), 318–338. http://doi.org/10.1111/j.1467-9787.2010.00694.x

Ahlfeldt, G. M. (2012). Rail mega-projects in the realm of inter- and intra-city accessibility: Evidence and outlooks for Berlin. *Built Environment*, *38*(1, 1), 71–88. http://doi.org/10.2148/benv.38.1.71

Ahlfeldt, G. M. (2013). If we build it, will they pay? Predicting property price effects of transport innovations. *Environment and Planning A*, *45*(8, 8), 1977–1994. http://doi.org/10.1068/a4549

Ahlfeldt, G. M., & Wendland, N. (2016). The spatial decay in commuting probabilities: Employment potential vs. Commuting gravity. *Economics Letters*, *143*, 125–129. http://doi.org/10.1016/j.econlet.2016.04.004

Alexander, F. S. (2008). *Land Banking as Metropolitan Policy* (SSRN Scholarly Paper No. 1955584). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1955584

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press. http://doi.org/10.1126/science.11.277.620

Alterman, R. (2012). Land-Use Regulations and Property Values: The "Windfalls Capture" Idea Revisited. In *The Oxford Handbook of Urban Economics and Planning* (pp. 755–786). http://doi.org/10.1093/oxfordhb/9780195380620.013.0034

Anas, A., Arnott, R., & Small, K. (1998). Urban Spatial Structure. *Journal of Economic Literature*, *36*(3, 3), 1426–1464. Retrieved from https://econpapers.repec.org/RePEc:aea:jeclit:v:36:y:1998:i:3:p:1426-1464

Andelson, R. V. (2000). Land-Value Taxation around the World. *The American Journal of Economics and Sociology*, *59*(5, 5), i–490. Retrieved from https://www.jstor.org/stable/3487821

Anselin, L., & Lozano-Gracia, N. (2009). Spatial Hedonic Models. In T. C. Mills & K. Patterson (Eds.), *Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics* (pp. 1213–1250). London: Palgrave Macmillan UK. http://doi.org/10.1057/9780230244405_26

Arbia, G. (1989). *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems* (Vol. 14). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-009-2395-9

Aveline-Dubach, N., & Blandeau, G. (2019). The political economy of transit value capture: The changing business model of the MTRC in Hong Kong. *Urban Studies*, 0042098018821519–0042098018821519. http://doi.org/10.1177/0042098018821519

Bakens, J., & Pryce, G. (2019). Homophily horizons and ethnic mover flows among homeowners in Scotland. *Housing Studies*, *34*(6), 925–945. http://doi.org/10.1080/02673037.2018.1504007

Bala, A. P., Peeters, D., & Thomas, I. (2014). Spatial issues on a hedonic estimation of rents in Brussels. *Journal of Housing Economics*, *25*, 104–123. http://doi.org/10.1016/j.jhe.2014.05.002

Balchin, P., Isaac, D., & Chen, J. (2000). The spatial structure of Urban Areas. In *Urban Economics: A global perspective* (6th ed., pp. 80–136). Hampshire: Palgrave. http://doi.org/10.4337/9781849808057.00023

Balemi, N., Füss, R., & Weigand, A. (2021). COVID-19's impact on real estate markets: Review and outlook. *Financial Markets and Portfolio Management*, *35*(4, 4), 495–513. http://doi.org/10.1007/s11408-021-00384-6

Banister, D. (2018). *Inequality in transport*. Oxford: Alexandrine Press.

Banzhaf, H. S. (2020). Panel Data Hedonics: Rosen's First Stage as a "Sufficient Statistic." *International Economic Review*, *61*(2), 973–1000. http://doi.org/10.1111/iere.12446

Banzhaf, H. S. (2021). Difference-in-Differences Hedonics. *Journal of Political Economy*, *129*(8), 2385–2414. http://doi.org/10.1086/714442

Bartik, T. (2008). Measuring the Benefits of Amenity Improvements in Hedonic Price Models. In R. E. Just, D. L. Hueth, & A. Schmitz (Eds.), *Applied Welfare Economics* (pp. 643–654). Northampton, MA: Edward Elgar Publishing, Inc. Retrieved from https://research.upjohn.org/bookchapters/20

Bartosiewicz, B., & Marcińczak, S. (2020). Investigating polycentric urban regions: Different measures – Different results. *Cities*, *105*, 102855. http://doi.org/10.1016/j.cities.2020.102855

Batty, M. (1976). *Urban Modelling*. Cambridge: Cambridge University Press.

Bayer, P., Fang, H., & McMillan, R. (2014). Separate when equal? Racial inequality and residential segregation. *Journal of Urban Economics*, *82*, 32–48. http://doi.org/10.1016/j.jue.2014.05.002

Bayer, P., Ferreira, F., & McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, *115*(4), 588–638. http://doi.org/10.1086/522381

Bayer, P., & McMillan, R. (2012). Tiebout sorting and neighborhood stratification. *Journal of Public Economics*, *96*(11), 1129–1143. http://doi.org/10.1016/j.jpubeco.2012.0

Bayer, P., McMillan, R., & Rueben, K. (2004). *An Equilibrium Model of Sorting in an Urban Housing Market* (No. w10865) (p. w10865). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w10865.pdf

Becker, G. S. (1994). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* (3rd ed.). Chicago: The University of Chicago Press.

Berge, L. (2021). *Fixest: Fast fixed-effects estimations*. Retrieved from https://CRAN.R-project.org/package=fixest

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(2, 2), 192–236. Retrieved from https://www.jstor.org/stable/2984812

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*(1, 1), 1–20. http://doi.org/10.1007/BF00116466

Birch, C. P. D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, *206*(3-4, 3-4), 347–359. http://doi.org/10.1016/j.ecolmodel.2007.03.041

Birkin, M., Clarke, G., Corcoran, J., & Stimson, R. (2021). Introduction to Big Data Applications in Geography and Planning. In *Big Data Applications in Geography and Planning* (pp. 1–7). Cheltenham: Edward Elgar Publishing Limited. Retrieved from https://www.elgaronline.com/view/edcoll/9781789909784/9781789909784.00006.xml

Bishop, K. C., Kuminoff, N. V., Banzhaf, H. S., Boyle, K. J., von Gravenitz, K., Pope, J. C., … Timmins, C. D. (2020). Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality. *Review of Environmental Economics and Policy*, *14*(2, 2), 260–281. http://doi.org/10.1093/reep/reaa001

Bishop, K. C., & Timmins, C. (2018). Using Panel Data to Easily Estimate Hedonic Demand Functions. *Journal of the Association of Environmental and Resource Economists*. http://doi.org/10.1086/696981

Bishop, K. C., & Timmins, C. (2019). Estimating the marginal willingness to pay function without instrumental variables. *Journal of Urban Economics*, *109*, 66–83. http://doi.org/10.1016/j.jue.2018.11.006

Bivand, R. (2021). *Spdep: Spatial dependence: Weighting schemes, statistics*. Retrieved from https://github.com/r-spatial/spdep/

Bivand, R., Gómez-Rubio, V., & Rue, H. (2015). Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software*, *63*(20, 20), 1–31. http://doi.org/10.18637/jss.v063.i20

Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R* (2nd ed., Vol. 10). New York, NY: Springer. Retrieved from https://doi.org/10.1007/978-1-4614-7618-4

Bivand, R., Sha, Z., Osland, L., & Thorsen, I. S. (2017). A comparison of estimation methods for multilevel models of spatially structured data. *Spatial Statistics*, *21*, 440–459. http://doi.org/10.1016/j.spasta.2017.01.002

Blanco, A., Moreno, N., Vetter, D., & Vetter, M. (2016). *El potencial de la captura de plusvalías para la financiación de proyectos urbanos: consideraciones metodológicas y casos prácticos*. Banco Interamericano de Desarrollo. Retrieved from https://publications.iadb.org/publications/spanish/document/

El-potencial-de-la-captura-de-plusvalÃŋas-para-la-financiaciÃşn-de-proyectos-urbanos-Consideraciones-metodolÃşgicas-y-casos-prÃącticos.pdf

Blangiardo, M., & Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Chichester: Wiley.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, *4*(1, 1), 33–49. http://doi.org/10.1016/j.sste.2012.12.001

Blaug, M. (1997). *Economic Theory in Retrospect* (5th ed.). Cambridge: Cambridge University Press.

Bockstael, N. E., & McConnell, K. E. (2007). Hedonic Models of Heterogeneous Goods. In N. E. Bockstael & K. E. McConnell (Eds.), *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models* (pp. 151–187). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-1-4020-5318-4_6

Borrero, O. (2013). Plusvalías urbanas: Generación y administración. In D. A. Erba (Ed.), *Definicion de políticas de suelo urbano en America Latina: Teoría y práctica*. Cambridge, MA: Lincoln Institute of Land Policy.

Bourassa, S. C., Hoesli, M., Merlin, L., & Renne, J. (2021). Big data, accessibility and urban house prices. *Urban Studies*, *58*(15, 15), 3176–3195. http://doi.org/10.1177/0042098020982508

Bourassa, S. C., & Hong, Y.-H. (2003). *Leasing Public Land: Policy Debates and International Experiences*. Toronto: Webcom Ltd.

Brady, M., & Irwin, E. (2011). Accounting for Spatial Effects in Economic Models of Land Use: Recent Developments and Challenges Ahead. *Environmental and Resource Economics*, *48*(3, 3), 487–509. http://doi.org/10.1007/s10640-010-9446-6

Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs N.J.: Prentice-Hall.

Breslow, N., Leroux, B., & Platt, R. (1998). Approximate hierarchical modelling of discrete data in epidemiology. *Statistical Methods in Medical Research*, *7*(1, 1), 49–62. http://doi.org/10.1191/096228098675989345

Briant, A., Combes, P. P., & Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of*

Urban Economics, *67*(3, 3), 287–302. http://doi.org/10.1016/j.jue.2009.09.014

Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. In E. S. Mills (Ed.), *Handbook of Regional and Urban Economics* (Vol. 2, pp. 821–845). Elsevier. http://doi.org/10.1016/S1574-0080(87)80006-8

Brueckner, J. K. (2011). *Lectures on Urban Economics*. Cambridge, MA: MIT Press.

CAF. (n.d.). Ferrocarriles Suburbanos. Retrieved January 8, 2022, from http://fsuburbanos.com/

Campisi, T., Tesoriere, G., Trouva, M., Papas, T., & Basbas, S. (2022). Impact of Teleworking on Travel Behaviour During the COVID-19 Era: The Case Of Sicily, Italy. *Transportation Research Procedia*, *60*, 251–258. http://doi.org/10.1016/j.trpro.2021.12.033

Carr, D. B., Olsen, A. R., & White, D. (1992). Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. *Cartography and Geographic Information Systems*, *19*(4, 4), 228–236. http://doi.org/10.1559/152304092783721231

Cervero, R., & Murakami, J. (2008). *Rail + Property Development: A model of sustainable transit finance and urbanism* (Working paper). Berkeley, CA: University of California. Retrieved from https://escholarship.org/uc/item/6jx3k35x

Cervero, R., & Murakami, J. (2009). Rail and Property Development in Hong Kong: Experiences and Extensions. *Urban Studies*, *46*(10, 10), 2019–2043. http://doi.org/10.1177/0042098009339431

Cervero, R., Murphy, S., Ferrell, C. L., Goguts, N., Tsai, Y.-H., Arrington, G. B., ... Tryssenaar, J. (2004). *Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects* (No. 102). Washington, D.C.: Transport Research Board. Retrieved from http://www.trb.org/Publications/Blurbs/TransitOriented_Development_in_the_United_States_E_154989.aspx

Champ, P. A., Boyle, K. J., & Brown, T. C. (Eds.). (2017). *A Primer on Nonmarket Valuation* (Vol. 13). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-7104-8

Chapelle, G., Wasmer, E., & Bono, P.-H. (2021). An urban labor market with frictional housing markets: Theory and an application to the Paris urban area. *Journal of Economic Geography*, *21*(1, 1), 97–126. http://doi.org/10.1093/jeg/lbaa006

Chapman, J. I. (2016). Value Capture Taxation as an Infrastructure Funding Technique. *Public Works Management & Policy*, *22*(1, 1), 31–37. http://doi.org/10.1177/1087724X16670395

Chava, J., & Newman, P. (2016). Stakeholder Deliberation on Developing Affordable Housing Strategies: Towards Inclusive and Sustainable Transit-Oriented Developments. *Sustainability*, *8*(10, 10), 1024. http://doi.org/10.3390/su8101024

Christensen, G., & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, *56*(3, 3), 920–980. http://doi.org/10.1257/jel.20171350

Clemens, M. (2015). *The Meaning of Failed Replications: A Review and Proposal* (Working paper No. 399). Center for Global Development. Retrieved from https://www.cgdev.org/publication/meaning-failed-replications-review-and-proposal-working-paper-399

Comber, S., Arribas-Bel, D., Singleton, A., Dong, G., & Dolega, L. (2020). Building Hierarchies of Retail Centers Using Bayesian Multilevel Models. *Annals of the American Association of Geographers*, *110*(4, 4), 1150–1173. http://doi.org/10.1080/24694452.2019.1667219

Cömertler, S. (2017). Greens of the European Green Capitals. *IOP Conference Series: Materials Science and Engineering*, *245*(5, 5). http://doi.org/10.1088/1757-899X/245/5/052064

CONAVI. (2017). *Código de Edificación de Vivienda*. México, D.F.: Comisión Nacional de Vivienda. Retrieved from https://www.gob.mx/cms/uploads/attachment/file/320345/CEV_2017__FINAL_.pdf

Conway, M. W. (2021). *Eqsormo: Equilibrium sorting models in Python*. Retrieved from https://github.com/mattwigway/eqsormo (Original work published September 6, 2019)

Conway, M. W. (2021). *If You Zone It, Who Will Come, and How Will They Travel? The Effects of Relaxed Zoning Regulations on Travel Behavior* (PhD thesis). Arizona State University, Phoenix, AZ.

Conway, M. W., Byrd, A., & van der Linden, M. (2017). Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule and Headway-Based Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2653*(1, 1), 45–53. http://doi.org/10.3141/2653-06

Conway, M. W., Byrd, A., & Van Eggermond, M. (2018). Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning. *Journal of Transport and Land Use*, *11*(1, 1). http://doi.org/10.5198/jtlu.2018.1074

Cordera, R., Coppola, P., Dell'Olio, L., & Ibeas, Á. (2018). The impact of accessibility by public transport on real estate values: A comparison between the cities of Rome and Santander. *Transportation Research Part A: Policy and Practice*, (xxxx, xxxx), 1–12. http://doi.org/10.1016/j.tra.2018.07.015

Crane, R. (1996). The Influence of Uncertain Job Location on Urban Form and the Journey to Work. *Journal of Urban Economics*, *39*(3, 3), 342–356. http://doi.org/10.1006/juec.1996.0018

Crook, T., Henneberry, J., & Whitehead, C. M. E. (2016). *Planning gain: Providing infrastructure and affordable housing*. Chichester, West Sussex: John Wiley & Sons Ltd.

Crotte, A., Graham, D. J., & Noland, R. B. (2011). The Role of Metro Fares, Income, Metro Quality of Service and Fuel Prices for Sustainable Transportation in Mexico City. *International Journal of Sustainable Transportation*, *5*(1, 1), 1–24. http://doi.org/10.1080/15568310903050073

Crôtte, A., Noland, R. B., & Graham, D. J. (2009). Is the Mexico City metro an inferior good? *Transport Policy*, *16*(1, 1), 40–45. http://doi.org/10.1016/j.tranpol.2009.02.009

D'Elia, V. V., Grand, M. C., & León, S. (2020). Bus rapid transit and property values in Buenos Aires: Combined spatial hedonic pricing and propensity score techniques. *Research in Transportation Economics*, *80*(August 2018, August 2018), 1–13. http://doi.org/10.1016/j.retrec.2020.100814

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography: Earth and Environment*, *31*(5, 5), 471–479. http://doi.org/10.1177/0309133307083294

De Cesare, C. M. (2016). *Sistemas del impuesto predial en America Latina y el Caribe*. Cambridge, MA: Lincoln Institute of Land Policy.

Dean, N., & Pryce, G. (2017). Is the housing market blind to religion? A perceived substitutability approach to homophily and social integration. *Urban Studies*, *54*(13), 3058–3070. http://doi.org/10.1177/0042098016668779

Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *Journal of Real Estate Finance and Economics*, *35*(2, 2), 161–180. http://doi.org/10.1007/s11146-007-

9032-z

Demetrio Rodriguez T & David Zarruk, R. A. M. & &. (2018). *Gmapsdistance: Distance and travel time between two points from google maps*. Retrieved from https://github.com/rodazuero/gmapsdistance

Dennett, A. (2012). *Estimating flows between geographical locations: 'Get me started in' spatial interaction modelling* (Working paper No. 181). London: University College London. Retrieved from https://www.ucl.ac.uk/bartlett/casa/publications/2012/mar/casa-working-paper-181

Diao, M. (2015). Selectivity, spatial autocorrelation and the valuation of transit accessibility. *Urban Studies*, *52*(1, 1), 159–177. http://doi.org/10.1177/0042098014523686

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1, 1), 269–271. http://doi.org/10.1007/BF01386390

Dijst, M., de Jong, T., & van Eck, J. R. (2002). Opportunities for transport mode change: An exploration of a disaggregated approach. *Environment and Planning B: Planning and Design*, *29*(3, 3), 413–430. http://doi.org/10.1068/b12811

DiPasquale, D., & Wheaton, W. C. (1996). *Urban economics and real estate markets*. Englewood Cliffs, NJ: Prentice Hall.

DOF. Proyecto de Norma Oficial Mexicana PROY-NOM-001-SEDATU-2020, Espacios públicos en los asentamientos humanos. (2020). Ciudad de México: Diario Oficial de la Federación. Retrieved from https://www.dof.gob.mx/nota_detalle.php?codigo=5608336&fecha=21/12/2020

Dong, G., Harris, R., & Mimis, A. (2020). *HSAR: Hierarchical spatial autoregressive model*. Retrieved from https://CRAN.R-project.org/package=HSAR

Du, H., & Mulley, C. (2007). The short-term land value impacts of urban rail transit: Quantitative evidence from Sunderland, UK. *Land Use Policy*, *24*(1, 1), 223–233. http://doi.org/10.1016/j.landusepol.2005.12.003

Du, H., & Mulley, C. (2012). Understanding spatial variations in the impact of accessibility on land value using geographically weighted regression. *Journal of Transport and Land Use*, *5*(2, 2). http://doi.org/10.5198/jtlu.v5i2.225

Dubé, J., Thériault, M., & Des, F. (2013). Commuter rail accessibility and house values : The case of the Montreal South Shore, Canada, 1992 – 2009. *Transportation Research Part A Journal*, *54*, 49–66. http://doi.org/10.1016/j.tra.2013.07.015

Dumolard, P. (2013). Distance, Accessibility and Spatial Difussion. In M. Thériault & F. Des Rosiers (Eds.), *Modeling Urban Dynamics: Mobility, Accessibility and Real Estate Value* (pp. 189–204). London: ISTE Ltd and John Wiley & Sons, Inc.

Duranton, G., & Puga, D. (2015). Urban Land Use. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of Regional and Urban Economics* (Vol. 5, pp. 467–560). Elsevier. http://doi.org/10.1016/B978-0-444-59517-1.00008-8

Eberts, R. W. (1981). An empirical investigation of intraurban wage gradients. *Journal of Urban Economics*, *10*(1, 1), 50–60. http://doi.org/10.1016/0094-1190(81)90022-X

EEA. (2002). *Towards an urban atlas: Assessment of spatial data on 25 European cities and urban areas*. Copenhagen. Retrieved from https://www.eea.europa.eu/publications/environmental_issue_report_2002_30

Eibenschutz, R., & Rodríguez, R. (2013). Globalización, desarrollo y ciudades-¿Veinticinco años de política urbana en México? In *Reflexiones sobre política urbana*. Zinacatepec: El Colegio Mexiquence, A.C.

Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy*, *112*(S1, S1), S60–S109. http://doi.org/10.1086/379947

El-Geneidy, A. M., & Levinson, D. M. (2006). *Access to Destinations: Development of Accessibility Measures* (Policy report No. MN/RC-2006-16). Minneapolis, MN: Minnesota Department of Transportation. Retrieved from http://www.lrrb.org/PDF/200616.pdf

Elldér, E. (2017). Does telework weaken urban structure–travel relationships? *Journal of Transport and Land Use*, *10*(1, 1), 187–210. Retrieved from https://www.jstor.org/stable/26211727

Epple, D. (1987). Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products. *Journal of Political Economy*, *95*(1), 59–80. Retrieved from https://www.jstor.org/stable/1831299

Erba, D. A. (2013). *Definicion de políticas de suelo urbano en America Latina: Teoría y práctica*. Cambridge, MA: Lincoln Institute of Land Policy.

Ermagun, A. (2021). Transit access and urban space-time structure of American cities. *Journal of Transport Geography*, *93*. http://doi.org/10.1016/j.jtrangeo.2021.103066

Fainstein, S. S. (2010). *The just city*. Ithaca: Cornell University Press.

Fainstein, S. S. (2012). Land Value Capture and Justice. In G. K. Ingram & Y.-H. Hong (Eds.), *Value Capture and Land Policies* (pp. 21–40). Cambridge, MA: Lincoln Institute of Land Policy. Retrieved from http://www.envplan.com/epc/fulltext/c31/c471rev.pdf

Falcocchio, J. C., Malik, A., & Kontokosta, C. E. (2018). A data-driven methodology for equitable value-capture financing of public transit operations and maintenance. *Transport Policy*, *66*, 107–115. http://doi.org/10.1016/j.tranpol.2018.02.005

Feng, X., & Humphreys, B. (2018). Assessing the Economic Impact of Sports Facilities on Residential Property Values: A Spatial Hedonic Approach. *Journal of Sports Economics*, *19*(2, 2), 188–210. http://doi.org/10.1177/1527002515622318

Fernández-Maldonado, A. M., Romein, A., Verkoren, O., & Parente Paula Pessoa, R. (2014). Polycentric Structures in Latin American Metropolitan Areas: Identifying Employment Sub-centres. *Regional Studies*, *48*(12, 12), 1954–1971. http://doi.org/10.1080/00343404.2013.786827

Fischel, W. A. (2015). *Zoning Rules! The economics of land use regulation*. Cambridge, MA: Lincoln Institute of Land Policy.

Flores, A. (2019). Cartografía del tsunami inmobiliario: El movimiento urbano popular y los vínculos urbano-rurales del boom inmobiliario en la Ciudad de México. *Espiral, Revista de Geografías y Ciencias Sociales*, *1*(1, 1), 7–28. http://doi.org/10.15381/espiral.v1i1.15843

Flores Dewey, O. (2011). *The Value of a Promise: Housing Price Impacts of Plans to Build Mass Transit in Ecatepec, Mexico* (Working paper No. WP11OD1) (pp. 1–36). Cambridge, MA: Lincoln Institute of Land Policy.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: Perspectives on spatial data analysis*. London ; Thousand Oaks, Calif: Sage Publications.

Fotheringham, A. S., & Sachdeva, M. (2022). Scale and local modeling: New perspectives on the modifiable areal unit problem and Simpson's paradox. *Journal of Geographical Systems*. http://doi.org/10.1007/s10109-021-00371-5

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment & Planning A*, *23*(7, 7), 1025–1044. http://doi.org/10.1068/a231025

Frambach, H. (2012). Johann Heinrich von Thünen: A Founder of Modern Economics. In J. G. Backhaus (Ed.), *Handbook of the History of Economic Thought* (pp. 299–322). New York, NY: Springer. http://doi.org/10.1007/978-1-4419-8336-7

Freeman, A. M. I., Herriges, J. A., & Kling, C. L. (2014). *The Measurement of Environmental and Resource Values. Theory and Methods* (3rd ed.). Oxon: RFF Press. http://doi.org/10.4324/9781936331826-6

Frost, M. E., & Spence, N. A. (1995). The rediscovery of accessibility and economic potential: The critical issue of self-potential. *Environment & Planning A*, *27*(11, 11), 1833–1848. http://doi.org/10.1068/a271833

Fujita, M. (1989). *Urban Economic Theory: Land Use and City Size*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511625862

Fujita, M., & Thisse, J.-F. (2013). *Economics of Agglomeration Cities, Industrial Location, and Globalization* (2nd ed., p. 544). New York, NY: Cambridge University Press. http://doi.org/10.2277/0521801389

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research)*. New York, NY: Cambridge University Press.

George, H. (1934). *Progress and Poverty* (6th ed., p. 571). Tennessee: Kingsport press, Inc. Retrieved from `http://books.google.com/books/about/Progress_and_Poverty.html?id=3Ex_yT6hJtEC&pgis=1`

Gerber, F., & Furrer, R. (2015). Pitfalls in the Implementation of Bayesian Hierarchical Modeling of Areal Count Data: An Illustration Using BYM and Leroux Models. *Journal of Statistical Software*, *63*(1, 1), 1–32. http://doi.org/10.1002/wics.10

Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, *12*(2, 2), 127–140. http://doi.org/10.1016/j.jtrangeo.2003.10.005

Gibb, K., Osland, L., & Pryce, G. (2014). Describing Inequalities in Access to Employment and the Associated Geography of Wellbeing. *Urban Studies*, *51*(3), 596–613. http://doi.org/10.1177/0042098013498283

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, *57*(1, 1), 148–169. http://doi.org/10.1016/j.jue.2004.10.002

Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region. *Regional Science and Urban Economics*, *21*(2, 2), 163–182. http://doi.org/10.1016/0166-0462(91)90032-I

Gjestland, A., McArthur, D. P., Osland, L., & Thorsen, I. (2014). The suitability of hedonic models for cost-benefit analysis: Evidence from commuting flows. *Transportation Research Part A: Policy and Practice*, *61*, 136–151. http://doi.org/10.1016/j.tra.2014.01.006

Gjestland, A., Osland, L., & Thorsen, I. (2020). Measures of labour market accessibility. What can we learn from observed commuting patterns? *Region*, *7*(1, 1), 49–70. http://doi.org/10.18335/region.v7i1.261

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. Oxford, New York: Oxford University Press.

Gobillon, L., & Selod, H. (2021). Spatial Mismatch, Poverty, and Vulnerable Populations. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 573–588). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-662-60723-7_7

Gobillon, L., Selod, H., & Zenou, Y. (2007). The Mechanisms of Spatial Mismatch. *Urban Studies*, *44*(12, 12), 2401–2427. http://doi.org/10.1080/00420980701540937

Goldstein, H. (2003). *Multilevel Statistical Models* (3rd ed.). London: Hodder Arnold.

Gómez-Ibáñez, J. A., Hong, Y.-H., & Du, H. (2022). Infrastructure Finance Through Land Value Capture. In *Infrastructure Economics and Policy. International Perspectives*. Cambridge, MA: Lincoln Institute of Land Policy.

Gómez-Ibáñez, J. A., & Liu, Z. (Eds.). (2022). *Infrastructure economics and policy: International perspectives*. Cambridge, MA: Lincoln Institute of Land Policy.

Graizbord, B. G. (2008). *Geografia del transporte en el area metropolitana de la Ciudad de México*. México, D.F.: Colegio de México.

Greater London Authority. Crossrail Bussiness Rates Suplement: Aproval of Policies for 208-19 (2018). https://www.london.gov.uk/sites/default/files/md2225_crossrailbrs_ 19_signed_pdf.pdf.

Guerra, E. (2014). Mexico City's suburban land use and transit connection: The effects of the Line B Metro expansion. *Transport Policy*, *32*, 105–114. http://doi.org/10.1016/j.tranpol.2013.12.011

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-Mile Circle: Does It Best Represent Transit Station Catchments? *Transportation Research Record*, *2276*(1, 1), 101–109. http://doi.org/10.3141/2276-12

Gupta, A., Van Nieuwerburgh, S., & Kontokosta, C. (2022). Take the Q train: Value capture of public infrastructure projects. *Journal of Urban Economics*, *129*, 103422. http://doi.org/10.1016/j.jue.2021.103422

Guzman, L. A., Enríquez, H. D., & Hessel, P. (2021). BRT system in Bogotá and urban effects: More residential land premiums? *Research in Transportation Economics*, *90*, 101039. http://doi.org/10.1016/j.retrec.2021.101039

Habib, M. A., & Miller, E. J. (2008). Influence of transportation access and market dynamics on property values: Multilevel spatiotemporal models of housing price. *Transportation Research Record*, (2076, 2076), 183–191. http://doi.org/10.3141/2076-20

Haider, M., & Donaldson, L. (2016). *Can Tax Increment Financing Support Transportation Infrastructure Investment?* (No. 25). Toronto: Institute on Municipal Finance & Governance.

Haila, A. (2016). *Urban Land Rent: Singapore as a Property State*. London: Wiley Blackwell.

Hamilton, B. W., & Röell, A. (1982). Wasteful Commuting. *Journal of Political Economy*, *90*(5, 5), 1035–1053. Retrieved from https://www.jstor.org/stable/1837132

Hamnett, C. (2003). Gentrification and the middle-class remaking of inner London, 1961-2001. *Urban Studies*, *40*(12, 12), 2401–2426. http://doi.org/10.1080/0042098032000136138

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A*, *29*(7, 7), 1175–1194. http://doi.org/10.1068/a291175

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2, 2), 73–76.

Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. A. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science*, *7*(2, 2), 190806. http://doi.org/10.1098/rsos.190806

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, *5*(1, 1), 81–102. http://doi.org/10.1016/0095-0696(78)90006-2

Hernández-Moreno, A., & Mugica-Álvarez, V. (2013). Vehicular fleets forecasting to project pollutant emissions: Mexico city metropolitan area case. *Transport Policy*, *27*, 189–199. http://doi.org/10.1016/j.tranpol.2013.01.010

Heyman, A. V., Law, S., & Berghauser Pont, M. (2019). How is Location Measured in Housing Valuation? A Systematic Review of Accessibility Specifications in Hedonic Price Models. *Urban Science*, *3*(1, 1), 3. http://doi.org/10.3390/urbansci3010003

Higgins, C. D. (2019). Accessibility Toolbox for R and ArcGIS. *Findings*, 8416. http://doi.org/10.32866/8416

Higgins, C. D., & Kanaroglou, P. S. (2016). Forty years of modelling rapid transit's land value uplift in North America: Moving beyond the tip of the iceberg. *Transport Reviews*, *36*(5, 5), 610–634. http://doi.org/10.1080/01441647.2016.1174748

Higgins, C. D., Palm, M., DeJohn, A., Xi, L., Vaughan, J., Farber, S., … Miller, E. (2022). Calculating place-based transit accessibility: Methods, tools and algorithmic dependence. *Journal of Transport and Land Use*, *15*(1, 1). http://doi.org/10.5198/jtlu.2022.2012

Hillier, B. (2007). *Space is the machine* (e-edition). London: Space Syntax. http://doi.org/10.1016/S0142-694X(97)89854-7

Hillsman, E. L., & Rhoda, R. (1978). Errors in measuring distances from populations to service centers. *The Annals of Regional Science*, *12*(3, 3), 74–88. http://doi.org/10.1007/BF01286124

Hoesli, M., & Malle, R. (2021). Commercial real estate prices and COVID-19. *Journal of European Real Estate Research*, *ahead-of-print*(ahead-of-print, ahead-of-print). http://doi.org/10.1108/JERER-04-2021-0024

Holcombe, R. G. (1997). A theory of the theory of public goods. *The Review of Austrian Economics*, *10*(1, 1), 1–22. http://doi.org/10.1007/BF02538141

Holden, E., Banister, D., Gössling, S., Gilpin, G., & Linnerud, K. (2020). Grand Narratives for sustainable mobility: A conceptual review. *Energy Research & Social Science*, *65*, 101454. http://doi.org/10.1016/j.erss.2020.101454

Hong, J., Shen, Q., & Zhang, L. (2014). How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation*, *41*(3, 3), 419–440. http://doi.org/10.1007/s11116-013-9462-9

Hong, Y.-H., & Needham, B. (Eds.). (2007). *Analyzing land readjustment: Economics, law, and collective action*. Cambridge, MA: Lincoln Institute of Land Policy.

Horner, M. W., & Murray, A. T. (2004). Spatial representation and scale impacts in transit service assessment. *Environment and Planning B: Planning and Design*, *31*(5, 5), 785–797. http://doi.org/10.1068/b3046

Iacono, M., & Levinson, D. M. (2017). Accessibility dynamics and location premia: Do land values follow accessibility changes? *Urban Studies*, *54*(2, 2), 364–381. http://doi.org/10.1177/0042098015595012

INEGI. (n.d.-a). Finanzas públicas estatales y municipales. Retrieved September 14, 2022, from https://www.inegi.org.mx/programas/finanzas/#Tabulados

INEGI. (n.d.-b). Instituto Nacional de Estadística y Geografía (INEGI). Retrieved January 24, 2022, from https://inegi.org.mx/default.html

INEGI. (2012). Sistema para la Consulta de Información Censal (SCINCE). Retrieved from https://www.inegi.org.mx/app/descarga/?ti=13&ag=01

INEGI. (2015a). *Censos Económicos 2014*. Instituto Nacional de Estadística y Geografía. Retrieved from https://www.inegi.org.mx/programas/ce/2014/

INEGI. (2015b). Directorio Estadístico Nacional de Unidades Económicas 2015. Retrieved from https://www.inegi.org.mx/default.html

INEGI. (2015c). Encuesta Intercensal 2015. Retrieved January 19, 2022, from https://www.inegi.org.mx/programas/intercensal/2015/

INEGI. (2016). Inventario Nacional de Viviendas 2016. Retrieved from https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825097905

INEGI. (2018). Encuesta Origen Destino en Hogares de la Zona Metropolitana del Valle de México (EOD) 2017. Retrieved from https://www.inegi.org.mx/programas/eod/2017/#:~:text=Las%20Encuestas%20de%20Origen%20y,ciudad%20durante%20un%20tiempo%20determinado.

INEGI. (2021). *Actualización de la medición de la economía informal 2003-2020 preliminar* (Press release No. 776/21). Ciudad de Mexico: Instituto Nacional de Estadística y Geografía. Retrieved from https://www.inegi.org.mx/

contenidos/saladeprensa/boletines/2021/pibmed/pibmed2020.pdf

Ingram, G. K., & Hong, Y.-H. (2012). Land Value Capture: Types and Outcomes. *Value Capture and Land Policies*, 1–28.

Islas Rivera, V. (2000). *Llegando tarde al compromiso: La crisis del transporte en la Ciudad de México*. México, D.F.: El Colegio de México.

Ismay, C., & Solomon, N. (2021). *Thesisdown: An updated r markdown thesis template using the bookdown package*.

Isunza-Vizuet, G., Castro, E., & Munévar, C. (2021). La plusvalía como sistema de financiación urbana: estudio comparativo en Ciudad de México y Manizales, Colombia. *EURE*, *47*, 229–248. http://doi.org/10.7764/EURE.47.142.11

Janić, M. (2019). Accessibility. In *Landside Accessibility of Airports: Analysis, Modelling, Planning, and Design* (pp. 27–47). Cham: Springer. http://doi.org/10.1007/978-3-319-76150-3_2

Junge, J. R., & Levinson, D. M. (2012). Financing transportation with land value taxes: Effects on development intensity. *Journal of Transport and Land Use*, *5*(1, 1), 49–63. http://doi.org/10.5198/jtlu.v5i1.148

Kanaroglou, P. S., Higgins, C. D., & Chowdhury, T. A. (2015). Excess commuting: A critical review and comparative analysis of concepts, indices, and policy implications. *Journal of Transport Geography*, *44*, 13–23. http://doi.org/10.1016/j.jtrangeo.2015.02.009

Kestens, Y., Thériault, M., & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: Looking at buyers' household profiles. *Journal of Geographical Systems*, *8*(1), 61–96. http://doi.org/10.1007/s10109-005-0011-8

Kim, H., & Song, Y. (2018). An integrated measure of accessibility and reliability of mass transit systems. *Transportation*, *45*(4, 4), 1075–1100. http://doi.org/10.1007/s11116-018-9866-7

Kim, K.-Y. (2011). Effects of the modifiable areal unit problem (MAUP) on a spatial interaction model. *Journal of the Korean Geographical Society*, *46*(2, 2), 197–211.

King, J. E., & McLure, M. (2015). *Value: History of the Concept*. *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 25, p. 13). Elsevier.

Klaiber, H. A., & Kuminoff, N. V. (2014). Equilibrium Sorting Models of Land Use and Residential Choice. In J. M. Duke & J. Wu (Eds.), *The Oxford Handbook of Land Economics*. Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199763740.013.011

Klaiber, H. A., & Phaneuf, D. J. (2010). Valuing open space in a residential sorting model of the Twin Cities. *Journal of Environmental Economics and Management*, *60*(2), 57–77. http://doi.org/10.1016/j.jeem.2010.05.002

Klaiber, H. A., & Smith, V. K. (2011). Preference Heterogeneity and Non-market Benefits: The Roles of Structural Hedonic and Sorting Models. In J. Bennett (Ed.), *The International Handbook on Non-Market Environmental Valuation*. Cheltenham and Northampton, MA: Edward Elgar.

Knowles, R. D. (2012). Transit Oriented Development in Copenhagen, Denmark: From the Finger Plan to Ørestad. *Journal of Transport Geography*, *22*, 251–261. http://doi.org/10.1016/j.jtrangeo.2012.01.009

Knowles, R. D., & Ferbrache, F. (2016). Evaluation of wider economic impacts of light rail investment on cities. *Journal of Transport Geography*, *54*, 430–439. http://doi.org/10.1016/j.jtrangeo.2015.09.002

Knudsen, D. C., & Fotheringham, A. S. (1986). Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*, *10*(2, 2), 127–147. http://doi.org/10.1177/016001768601000203

Kumbhakar, S. C., & Parmeter, C. F. (2010). Estimation of hedonic price functions with incomplete information. *Empirical Economics*, *39*(1), 1–25. http://doi.org/10.1007/s00181-009-0292-8

Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *Journal of Environmental Economics and Management*, *60*(3, 3), 145–160. http://doi.org/10.1016/j.jeem.2010.06.001

Kuminoff, N. V., Smith, V. K., & Timmins, C. (2013). The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets. *Journal of Economic Literature*, *51*(4, 4), 1007–1062. http://doi.org/10.1257/jel.51.4.1007

Kwan, M.-P. (1998). Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis*, *30*(3, 3), 191–216. http://doi.org/10.1111/j.1538-4632.1998.tb00396.x

Kwan, M.-P., & Weber, J. (2008). Scale and accessibility: Implications for the analysis of land use-travel interaction. *Applied Geography*, *28*(2, 2), 110–123. http://doi.org/10.1016/j.apgeog.2007.07.002

Lámbarry Vilchis, F., Rivas Tovar, L. A., & Peña Cruz, M. del P. (2011). Planeación de los sistemas BRT y consensos entre transportistas y autoridades de gobierno durante su implementación. El caso de Metrobús y Mexibús. *Administración & Desarrollo*, *39*(54, 54), 133–150.

Lawson, A. B. (2009). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Boca Raton, FL: Chapman and Hall/CRC.

Lee, D. (2016). CARBayes version 5.2: An R Package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors. *Journal of Statistical Software*, *55*(February, February), 32–32.

Lee, G., Cho, D., & Kim, K. (2016). The modifiable areal unit problem in hedonic house-price models. *Urban Geography*, *37*(2, 2), 223–245. http://doi.org/10.1080/02723638.2015.1057397

Lee, S.-I., Lee, M., Chun, Y., & Griffith, D. A. (2019). Uncertainty in the effects of the modifiable areal unit problem under different levels of spatial autocorrelation: A simulation study. *International Journal of Geographical Information Science*, *33*(6, 6), 1135–1154. http://doi.org/10.1080/13658816.2018.1542699

Lees, L., Slater, T., & Wyly, E. K. (2008). *Gentrification*. New York, NY: Routledge/Taylor & Francis Group.

Lei, T. L., & Church, R. L. (2010). Mapping transit-based access: Integrating GIS, routes and schedules. *International Journal of Geographical Information Science*, *24*(2, 2), 283–304. http://doi.org/10.1080/13658810902835404

Leroux, B. G., Lei, X., & Breslow, N. (2000). Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence. In M. E. Halloran & D. Berry (Eds.), *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (pp. 179–191). New York, NY: Springer. http://doi.org/10.1007/978-1-4612-1284-3_4

LeSage, J. (2015). Spatial econometrics. In C. Karlsson, M. Andersson, & T. Norman (Eds.), *Handbook of Research Methods and Applications in Economic Geography*. Cheltenham, UK: Edward Elgar Publishing. http://doi.org/10.4337/9780857932679.00007

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics.* Boca Raton, FL: Chapman & Hall/CRC.

Levinson, D. M., & Istrate, E. (2011). *Access for Value: Financing Transportation Through Land Value Capture* (pp. 1–21). Washington, DC.

Levinson, D. M., & Wu, H. (2020). Towards a general theory of access. *Journal of Transport and Land Use*, *13*(1, 1), 129–158. http://doi.org/10.5198/jtlu.2020.1660

Levkovich, O., & de Graaff, T. (2017). *Sortingmod.* Retrieved from https://github.com/Thdegraaff/sortingmod

Liang, W., Song, R., & Timmins, C. (2021). The Role of Migration Costs in Residential Sorting. In G. Pryce, Y. P. Wang, Y. Chen, J. Shan, & H. Wei (Eds.), *Urban Inequality and Segregation in Europe and China* (pp. 251–283). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-030-74544-8_4

Liu, L., Porr, A., & Miller, H. J. (2022). Realizable accessibility: Evaluating the reliability of public transit accessibility using high-resolution real-time data. *Journal of Geographical Systems.* http://doi.org/10.1007/s10109-022-00382-w

Liu, W., Wang, Q., & Wang, J. (2018). Research on the mechanism of value creation and capture process for urban rail development. *Journal of Ambient Intelligence and Humanized Computing.* http://doi.org/10.1007/s12652-018-1162-z

Loomis, J. (2011). What's to Know About Hypothetical Bias in Stated Preference Valuation Studies? *Journal of Economic Surveys*, *25*(2), 363–370. http://doi.org/10.1111/j.1467-6419.2010.00675.x

Loret de Mola, U., Ladd, B., Duarte, S., Borchard, N., Anaya La Rosa, R., & Zutta, B. (2017). On the Use of Hedonic Price Indices to Understand Ecosystem Service Provision from Urban Green Space in Five Latin American Megacities. *Forests*, *8*(12, 12). http://doi.org/10.3390/f8120478

Ma, K., & Banister, D. (2006). Excess Commuting: A Critical Review. *Transport Reviews*, *26*(6, 6), 749–767. http://doi.org/10.1080/01441640600782609

Malamis, S., Katsou, E., Inglezakis, V. J., Kershaw, S., Venetis, D., & Folini, S. (2016). Chapter 5 - Urban Environment. In S. G. Poulopoulos & V. J. Inglezakis (Eds.), *Environment and Development* (pp. 287–362). Amsterdam: Elsevier. http://doi.org/10.1016/B978-0-444-62733-9.00005-8

Malekzadeh, A., & Chung, E. (2020). A review of transit accessibility models: Challenges in developing transit accessibility models. *Inter-*

national Journal of Sustainable Transportation, *14*(10, 10), 733–748. http://doi.org/10.1080/15568318.2019.1625087

Malpezzi, S. (2003). Hedonic Pricing Models: A Selective and Applied Review. In T. O'Sullivan & K. Gibb (Eds.), *Housing Economics and Public Policy* (pp. 67–89). Oxford: Blackwell. http://doi.org/10.1002/9780470690680.ch5

Martínez, F., & Araya, C. (2000). Transport and Land-Use Benefits under Location Externalities. *Environment and Planning A: Economy and Space*, *32*(9, 9), 1611–1624. http://doi.org/10.1068/a32131

Mathur, S. (2014). Funding Public Transportation Through Special Assessment Districts: Addressing the Equity Concerns. *Public Works Management & Policy*, *20*(2, 2), 127–145. http://doi.org/10.1177/1087724X14550252

Mathur, S. (2019). An evaluative framework for examining the use of land value capture to fund public transportation projects. *Land Use Policy*, *86*, 357–364. http://doi.org/10.1016/j.landusepol.2019.05.021

Mathur, S., & Gatdula, A. (2020). *Promoting Transit-Oriented Developments by Addressing Barriers Related to Land Use, Zoning, and Value Capture* (No. 20-35). Mineta Transportation Institute. Retrieved from `https://scholarworks.sjsu.edu/mti_publications/324`

Mcarthur, D. P., Kleppe, G., Thorsen, I., & Ubøe, J. (2013). The impact of monetary costs on commuting flows. *Papers in Regional Science*, *92*(1, 1), 69–86. http://doi.org/10.1111/j.1435-5957.2011.00405.x

McArthur, D. P., Osland, L., & Thorsen, I. (2012). Spatial Transferability of Hedonic House Price Functions. *Regional Studies*, *46*(5, 5), 597–610. http://doi.org/10.1080/00343404.2010.521147

McDonald, J. F. (2007). William Alonso, Richard Muth, Resources for the Future, and the Founding of Urban Economics. *Journal of the History of Economic Thought*, *29*(01, 01), 67–67. http://doi.org/10.1080/10427710601178310

McDonald, J. F. (2018). Mr. Ricardo's Theory of Land Rent. *Journal of Real Estate Practice and Education*, *21*(1, 1), 1–15. http://doi.org/10.1080/10835547.2018.12091774

McDonald, J. F., & McMillen, D. P. (2010). *Urban Economics and Real Estate: Theory and Policy*. John Wiley & Sons.

McIntosh, J., Newman, P., Trubka, R., & Kenworthy, J. (2015). Framework for land value capture from investments in transit in car-dependent cities. *Journal* *of Transport and Land Use*, 155–185. http://doi.org/10.5198/jtlu.2015.531

McIntosh, J., Trubka, R., & Newman, P. (2014). Can value capture work in a car dependent city? Willingness to pay for transit access in Perth, Western Australia. *Transportation Research Part A: Policy and Practice*, *67*, 320–339. http://doi.org/10.1016/j.tra.2014.07.008

McMillen, D. P., & Singell, L. D. (1992). Work location, residence location, and the intraurban wage gradient. *Journal of Urban Economics*, *32*(2, 2), 195–213. http://doi.org/10.1016/0094-1190(92)90005-6

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444. http://doi.org/10.1146/annurev.soc.27.1.415

McSpiritt, J. (2012). Developing Hudson Yards with a $2 Billion Bond Issue. In *The Handbook of Municipal Bonds* (pp. 1257–1262). John Wiley & Sons, Ltd. http://doi.org/10.1002/9781119198093.ch93

Medda, F. (2012). Land value capture finance for transport accessibility: A review. *Journal of Transport Geography*, *25*, 154–161. http://doi.org/10.1016/j.jtrangeo.2012.07.013

Medina Ciriaco, S., & Kunz Bolaños, I. (2013). Políticas de suelo y vivienda en México durante el sigo xx. In *Reflexiones sobre política urbana*. Zinacatepec: El Colegio Mexiquence, A.C.

Merton, R. K. (1979). *The Sociology of Science: Theoretical and Empirical Investigations*. (N. W. Storer, Ed.). Chicago, IL: University of Chicago Press. Retrieved from https://press.uchicago.edu/ucp/books/book/chicago/S/bo28451565.html

Metrobús. (2021). Metrobús. Retrieved December 30, 2021, from https://www.metrobus.cdmx.gob.mx

Mieszkowski, P., & Mills, E. S. (1993). The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives*, *7*(3), 135–147. http://doi.org/10.1257/jep.7.3.135

Mill, J. S. (1909). *The Principles of Political Economy with some of their applications to social philosophy*. (W. J. Ashley, Ed.). London: Longmans, Green and Co. Retrieved from https://www.econlib.org/library/Mill/mlP.html?

Miller, E. J. (2018). Accessibility: Measurement and application in transportation planning. *Transport Reviews*, *38*(5, 5), 551–555. http://doi.org/10.1080/01441647.2018.1492778

Mills, E. S. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *The American Economic Review*, *57*(2, 2), 197–210. Retrieved from https://www.jstor.org/stable/1821621

Mills, E. S. (1972). *Studies in the Structure of the Urban Economy.* Baltimore, MD: Johns Hopkins Press.

Mohammad, S. I., Graham, D. J., Melo, P. C., & Anderson, R. J. (2013). A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, *50*, 158–170. http://doi.org/10.1016/j.tra.2013.01.013

Monkkonen, P. (2008). Using Online Satellite Imagery as a Research Tool: Mapping Changing Patterns of Urbanization in Mexico. *Journal of Planning Education and Research*, *28*(2, 2), 225–236. http://doi.org/10.1177/0739456X08323771

Monkkonen, P. (2011). The Housing Transition in Mexico: Expanding Access to Housing Finance. *Urban Affairs Review*, *47*(5, 5), 672–695. http://doi.org/10.1177/1078087411400381

Montejano-Escamilla, J. A. (2015). ¿Es la ciudad de México policéntrica? Nuevos datos y algoritmos para la detección de centralidades urbanas. *Economia, Sociedad y Territorio*, *XV*(48, 48), 333–361.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., & DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. *Spatial and Spatio-Temporal Epidemiology*, *31*. http://doi.org/10.1016/j.sste.2019.100301

Mulalic, I., & Rouwendal, J. (2020). Does improving public transport decrease car ownership? Evidence from a residential sorting model for the Copenhagen metropolitan area. *Regional Science and Urban Economics*, *83*, 103543. http://doi.org/10.1016/j.regsciurbeco.2020.103543

Mulley, C. (2014). Accessibility and Residential Land Value Uplift: Identifying Spatial Variations in the Accessibility Impacts of a Bus Transitway. *Urban Studies*, *51*(8, 8), 1707–1724. http://doi.org/10.1177/0042098013499082

Mulley, C., Sampaio, B., & Ma, L. (2017). South Eastern Busway Network in Brisbane, Australia: Value of the Network Effect. *Transportation Research Record: Journal of the Transportation Research Board*, *2647*(1, 1), 41–49. http://doi.org/10.3141/2647-06

Mulley, C., & Tsai, C. (Patrick). (2016). When and how much does new transport infrastructure add to property values? Evidence from the bus rapid transit system in Sydney, Australia. *Transport Policy*, 1–9. http://doi.org/10.1016/j.tranpol.2016.01.011

Muñiz, I., Sánchez, V., & García-López, M. Á. (2015). Estructura espacial y densidad de población en la ZMVM 1995-2010: Evolución de un sistema urbano policéntrico. *Eure*, *41*(122, 122), 75–102.

Muñoz Gielen, D., & Lenferink, S. (2018). The role of negotiated developer obligations in financing large public infrastructure after the economic crisis in the Netherlands. *European Planning Studies*, *26*(4, 4), 768–791. http://doi.org/10.1080/09654313.2018.1425376

Muñoz Gielen, D., & Mualam, N. (2019). A framework for analyzing the effectiveness and efficiency of land readjustment regulations: Comparison of Germany, Spain and Israel. *Land Use Policy*, *87*, 104077–104077. http://doi.org/10.1016/j.landusepol.2019.104077

Muñoz Gielen, D., & van der Krabben, E. (2019). *Public Infrastructure, Private Finance Developer Obligations and Responsibilities.* (D. Muñoz Gielen & E. van der Krabben, Eds.). Abingdon and New York, NY: Routledge.

Munoz-Raskin, R. (2010). Walking accessibility to bus rapid transit: Does it affect property values? The case of Bogotá, Colombia. *Transport Policy*, *17*(2, 2), 72–84. http://doi.org/10.1016/j.tranpol.2009.11.002

Muth, R. F. (1969). *Cities and housing; the spatial pattern of urban residential land use. American Journal of Agricultural Economics* (Vol. 91, p. 41). http://doi.org/10.1111/j.1467-8276.2008.01187.x

Nassir, N., Hickman, M., Malekzadeh, A., & Irannezhad, E. (2016). A utility-based travel impedance measure for public transit network accessibility. *Transportation Research Part A: Policy and Practice*, *88*, 26–39. http://doi.org/10.1016/j.tra.2016.03.007

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4, 4), 308–313. http://doi.org/10.1093/comjnl/7.4.308

Niemeier, D. A. (1997). Accessibility: An evaluation using consumer welfare. *Transportation*, *24*(4, 4), 377–396. http://doi.org/10.1023/A:1004914803019

Nuñez, H. M., Paredes, D., & Garduño-Rivera, R. (2017). Is crime in Mexico a disamenity? Evidence from a hedonic valuation approach. *Annals of Regional*

Science, *59*(1, 1), 171–187. http://doi.org/10.1007/s00168-017-0823-8

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). Reproducible research and GIScience: An evaluation using AGILE conference papers. *PeerJ*, *6*, e5072. http://doi.org/10.7717/peerj.5072

Oates, W. E., & Schwab, R. M. (2009). The Simple Analytics of Land Value Taxation. In *Land Value Taxation : Theory, Evidence, and Practice* (p. 22). Cambridge, MA: Lincoln Institute of Land Policy.

OECD. (2015). *OECD Territorial Reviews: Valle de México, México* (No. 9789264245174) (pp. 32–32). Paris: OECD Publishing. Retrieved from https://www.oecd.org/regional/regional-policy/valle-de-mexico-highlights-spanish.pdf

Openshaw, S. (1984). Ecological Fallacies and the Analysis of Areal Census Data. *Environment and Planning A: Economy and Space*, *16*(1, 1), 17–31. http://doi.org/10.1068/a160017

Openshaw, S., & Rao, L. (1995). Algorithms for reengineering 1991 census geography. *Environment & Planning A*, *27*(3, 3), 425–446. http://doi.org/10.1068/a270425

Ortega, E., López, E., & Monzón, A. (2012). Territorial cohesion impacts of high-speed rail at different planning levels. *Journal of Transport Geography*, *24*, 130–141. http://doi.org/10.1016/j.jtrangeo.2011.10.008

Ortúzar, J. de D., & Willumsen, L. G. (2011). *Modelling Transport* (4th ed.). John Wiley & Sons. Retrieved from https://books.google.com?id=qWa5MyS4CiwC

Oshan, T. M. (2021). The spatial structure debate in spatial interaction modeling: 50 years on. *Progress in Human Geography*, *45*(5, 5), 925–950. http://doi.org/10.1177/0309132520968134

Osland, L. (2010a). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3, 3), 289–320. http://doi.org/10.1080/10835547.2010.12091282

Osland, L. (2010b). Spatial variation in job accessibility and gender: An intraregional analysis using hedonic house-price estimation. *Environment and Planning A*, *42*(9, 9), 2220–2237. http://doi.org/10.1068/a42422

Osland, L., & Pryce, G. (2012). Housing Prices and Multiple Employment Nodes: Is the Relationship Nonmonotonic? *Housing Studies*, *27*(8, 8), 1182–1208. http://doi.org/10.1080/02673037.2012.728571

Osland, L., & Thorsen, I. (2008). Effects on housing prices of urban attraction and labor-market accessibility. *Environment and Planning A*, *40*(10, 10), 2490–2509. http://doi.org/10.1068/a39305

Osland, L., & Thorsen, I. (2013). Spatial Impacts, Local Labour Market Characteristics and Housing Prices. *Urban Studies*, *50*(10, 10), 2063–2083. http://doi.org/10.1177/0042098012474699

Osland, L., Thorsen, I. S., & Thorsen, I. (2016). Accounting for Local Spatial Heterogeneities in Housing Market Studies. *Journal of Regional Science*, *56*(5, 5), 895–920. http://doi.org/10.1111/jors.12281

OTP. (n.d.). OpenTripPlanner. Retrieved October 10, 2019, from https://www.opentripplanner.org/

Owen, A., & Levinson, D. M. (2015). Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, *74*, 110–122. http://doi.org/10.1016/j.tra.2015.02.002

Padgham, M., Rudis, B., Lovelace, R., & Salmon, M. (2021). *Osmdata: Import OpenStreetMap data as simple features or spatial objects*. Retrieved from https://CRAN.R-project.org/package=osmdata

Páez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*(1, 1), 53–67. http://doi.org/10.1007/s10708-005-0877-5

Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, *25*, 141–153. http://doi.org/10.1016/j.jtrangeo.2012.03.016

Palmateer, C., Owen, A., Levinson, D. M., & Ermagun, A. (2021). The Role of Transit Service Area Definition for Access-based Evaluation. In *Applications of Access*. Sidney: The University of Sidney.

Palmquist, R. B. (2005). Property Value Models. In K.-G. Mler & J. R. Vincent (Eds.), *Handbook of Environmental Economics* (Vol. 2, pp. 763–819). Elsevier. http://doi.org/10.1016/S1574-0099(05)02016-4

Parmeter, C. F., & Pope, J. C. (2013). Quasi-experiments and hedonic property value methods. In J. A. List & M. K. Price (Eds.), *Handbook on Experimental Economics and the Environment* (pp. 3–66). Cheltenham: Edward Elgar Publishing Limited. http://doi.org/10.4337/9781781009079.00007

Pebesma, E., & Bivand, R. (2021). *Sp: Classes and methods for spatial data.* Retrieved from https://CRAN.R-project.org/package=sp

Perdomo Calvo, J. A. (2017). The effects of the bus rapid transit infrastructure on the property values in Colombia. *Travel Behaviour and Society*, *6*, 90–99. http://doi.org/10.1016/j.tbs.2016.08.002

Pereira, R. H. M. (2018). *Distributive Justice and Transportation Equity: Inequality in accessibility in Rio de Janeiro.* Retrieved from https://www.researchgate.net/publication/328969941%0ADistributive

Pereira, R. H. M. (2019). Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro. *Journal of Transport Geography*, *74*(October 2018, October 2018), 321–332. http://doi.org/10.1016/j.jtrangeo.2018.12.005

Pereira, R. H. M., Banister, D., Schwanen, T., & Wessel, N. (2019). Distributional effects of transport policies on inequalities in access to opportunities in Rio de Janeiro. *Journal of Transport and Land Use*, *12*(1 SE -, 1 SE -). http://doi.org/10.5198/jtlu.2019.1523

Pereira, R. H. M., Grégoire, L., & Karner, A. (2019). Tutorial with reproducible example to estimate a travel time matrix using OpenTripPlanner and Python. http://doi.org/10.5281/zenodo.3242134

Pérez Torres, D. E. (2021). *Intentos recientes de la aplicación de contribuciones de mejoras en la Ciudad de México: el caso del proyecto Masaryk.* Presented at the Contribución de Valorización en Colombia y América Latina: Fronteras del conocimiento sobre el instrumento, Boston, MA. Retrieved from https://www.lincolninst.edu/sites/default/files/sources/events/sesion_4_-_daniel_perez_torres_masaryk_cdmx.pdf

Peterson, G. E. (2009). *Unlocking Land Values to Finance Urban Infrastructure.* Washington, DC: World Bank. http://doi.org/10.1596/978-0-8213-7709-3

Petheram, S. J., Nelson, A. C., Miller, M., & Ewing, R. (2013). Use of the Real Estate Market to Establish Light Rail Station Catchment Areas: Case Study of Attached Residential Property Values in Salt Lake County, Utah, by Light Rail Station Distance. *Transportation Research Record*, *2357*(1, 1), 95–99. http://doi.org/10.3141/2357-11

Phaneuf, D. J., & Requate, T. (2017). *A course in environmental economics: Theory, policy, and practice.* Cambridge, United Kingdon ; New York, NY, USA:

Cambridge University Press.

Plassmann, F., & Tideman, T. N. (2000). A Markov Chain Monte Carlo Analysis of the Effect of Two-Rate Property Taxes on Construction. *Journal of Urban Economics*, *47*(2, 2), 216–247. http://doi.org/10.1006/juec.1999.2140

Plaut, P. O., & Plaut, S. E. (1998). Endogenous Identification of Multiple Housing Price Centers in Metropolitan Areas. *Journal of Housing Economics*, *7*(3, 3), 193–217. http://doi.org/10.1006/jhec.1998.0230

Proost, S., & Thisse, J.-F. (2019). What Can Be Learned from Spatial Economics? *Journal of Economic Literature*, 71.

Pryce, G. (2009). *Dwelling Substitutability and the Delineation of Submarkets* (Discussion paper No. 21). Centre for Public Policy Regions.

Pryce, G. (2013). Housing Submarkets and the Lattice of Substitution. *Urban Studies*, *50*(13), 2682–2699. http://doi.org/10.1177/0042098013482502

Pryce, G. (2021). Future directions for research on residential segregation and inequality in China. In *Urban Inequality and Segregation in Europe and China* (pp. 329–361). Cham: Springer.

Pryce, G., Wang, Y. P., Chen, Y., Shan, J., & Wei, H. (Eds.). (2021). *Urban Inequality and Segregation in Europe and China: Towards a New Dialogue*. Cham: Springer International Publishing. http://doi.org/10.1007/978-3-030-74544-8

PUEC-UNAM. (2012). *México, perfil del sector de la vivienda*. Ciudad de México: Programa Universitario de Estudios sobre la Ciudad (PUEC), Universidad Nacional Autónoma de México (UNAM). Retrieved from `https://www.puec.unam.mx/index.php/publicaciones/145-publicaciones-digitales/pd-instituciones/146-mexico-perfil-del-sector-de-la-vivienda`

PUEC-UNAM. (2015). *Habitabilidad y política de vivienda en México*. Ciudad de México: Universidad Nacional Autónoma de México.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `https://www.R-project.org/`

Reggiani, A., Bucci, P., & Russo, G. (2011). Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network. *International Regional Science Review*, *34*(2, 2), 230–252. http://doi.org/10.1177/0160017610387296

Rhee, H.-J. (2009). Telecommuting and urban sprawl. *Transportation Research Part D: Transport and Environment*, *14*(7, 7), 453–460. http://doi.org/10.1016/j.trd.2009.05.0

Ricardo, D. (2001). *Principles of Political Economy and Taxation* (Third). Ontario: Batoche Books.

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, *25*(4, 4), 1145–1165. http://doi.org/10.1177/0962280216660421

Rocha Chiu, A., & Gama, T. (2007). Costos preliminares en proyectos de edificación. *Anuario Administración y Tecnología Para El Diseño*, 142–158. Retrieved from https://administracionytecnologiaparaeldiseno.azc.uam.mx/publicaciones/2006/9_2006.pdf

Rodríguez, D. A., & Mojica, C. H. (2009). Capitalization of BRT network expansions effects into prices of non-expansion areas. *Transportation Research Part A: Policy and Practice*, *43*(5, 5), 560–571. http://doi.org/10.1016/j.tra.2009.02.003

Rojas, C., & Yojan, M. (2016). Una lectura del transporte público concesionado en Ciudad de México desde la visión del orden urbano. *Sociedad y Economía*, *31*, 71–94.

Rojas, M., Angulo, H., & Velázquez, I. (2000). Rentabilidad de la inversión en capital humano en México. *Economía Mexicana. Nueva Época*, *IX*(2, 2), 113–142. Retrieved from http://hdl.handle.net/11651/4177

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*, *82*(1, 1), 34–55.

Rosik, P., Stępniak, M., & Komornicki, T. (2015). The decade of the big push to roads in Poland: Impact on improvement in accessibility and territorial cohesion from a policy perspective. *Transport Policy*, *37*, 134–146. http://doi.org/10.1016/j.tranpol.2014.10.007

Roukoni, A., & Medda, F. (2012). Evaluation of Value Capture Mechanisms as a Funding Source for Urban Transport: The Case of London's Crossrail. *Procedia - Social and Behavioral Sciences*, *48*, 2393–2404. http://doi.org/10.1016/j.sbspro.2012.06.1210

Rouwendal, J. (2021). Commuting, Housing, and Labor Markets. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 497–514). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-662-60723-

7_11

Rue, Havard, Lindgren, F., & Teixeira Krainski, E. (2021). *INLA: Full bayesian analysis of latent gaussian models using integrated nested laplace approximations*.

Rue, Håvard, Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *71*(2, 2), 319–392. http://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, Havard, Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, *4*, 395–421. http://doi.org/10.1146/annurev-statistics-060116-054045

Ryan, M., Lin, T. G., Xia, J. C., & Robinson, T. (2016). Comparison of perceived and measured accessibility between different age groups and travel modes at Greenwood Station, Perth, Australia. *European Journal of Transport and Infrastructure Research*, *16*(2, 2), 406–423.

Ryan, S. (1999). Property values and transportation facilities: Finding the transportation-land use connection. *Journal of Planning Literature*, *13*(4, 4), 412–427. http://doi.org/10.1177/08854129922092487

Salon, D., Sclar, E., & Barone, R. (2019). Can Location Value Capture Pay for Transit? Organizational Challenges of Transforming Theory into Practice. *Urban Affairs Review*, *55*(3, 3), 743–771. http://doi.org/10.1177/1078087417715523

Salonen, M., & Toivonen, T. (2013). Modelling travel time in urban networks: Comparable measures for private car and public transport. *Journal of Transport Geography*, *31*, 143–153. http://doi.org/10.1016/j.jtrangeo.2013.06.011

Saporito, S., Chavers, J. M., Nixon, L. C., & McQuiddy, M. R. (2007). From here to there: Methods of allocating data between census geography and socially meaningful areas. *Social Science Research*, *36*(3, 3), 897–920. http://doi.org/10.1016/j.ssresearch.2006.05.004

Saraiva, M., Pereira, R. H. M., Herszenhut, D., Braga, C. K. V., & Conway, M. W. (2021). *r5r: Rapid realistic routing with R5*. Retrieved from `https://github.com/ipeaGIT/r5r`

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, *51*(1, 1), 1–17. Retrieved from `https://www.jstor.org/stable/1818907`

Sedatu, Conapo, & INEGI. (2018). *Delimitación de las zonas metropolitanas México 2015* (No. En trámite). Ciudad de Mexico: Secretaría de Desarrollo Agrario, Territorial y Urbano. Consejo Nacional de Población. Instituto Nacional de Estadística y Geografía. Retrieved from https://www.inegi.org.mx/contenido/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825006792.pdf

Seduvi. (n.d.). CETRAM home. Retrieved September 17, 2022, from http://seduvi.proyectosurbanos.cdmx.gob.mx/cetram/cetram_home.html#

Semovi. (2020). *Programa Integral de Movilidad 2020-2024: Diágnostico Técnico*. Ciudad de Mexico. Retrieved from https://www.semovi.cdmx.gob.mx/tramites-y-servicios/participa-en-proyectos-de-la-semovi/pim_2020_2024

SESNSP. (2021). Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública. E-mail. Retrieved from https://www.gob.mx/sesnsp

Severen, C., Costello, C., & Deschênes, O. (2018). A Forward-Looking Ricardian Approach: Do land markets capitalize climate change forecasts? *Journal of Environmental Economics and Management*, *89*, 235–254. http://doi.org/10.1016/j.jeem.2018.03.009

SHF. (2015). Manual de Usuario Sistema Maestro de Avalúos. Ciudad de México: Sociedad Hipotecaria Federal. Retrieved from http://doc.shf.gob.mx/sobreshf/unidval/Documents/Manual_Usuario_SMA_BajaCalidad.pdf

SHF. (2021). Sociedad Hipotecaria Federal. E-mail. Retrieved from https://www.gob.mx/shf/

Shoup, D. C. (2006). Cruising for parking. *Transport Policy*, *13*(6, 6), 479–486. http://doi.org/10.1016/j.tranpol.2006.05.005

SIE. (2019). *Balance Nacional de Energía*. Ciudad de México: Sistema de Información Energética. Retrieved from http://sie.energia.gob.mx/

Sieg, H., Smith, V. K., Banzhaf, H. S., & Walsh, R. (2004). Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods*. *International Economic Review*, *45*(4), 1047–1077. http://doi.org/10.1111/j.0020-6598.2004.00297.x

Silva, J. M. C. S., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, *88*(4, 4), 641–658. http://doi.org/10.1162/rest.88.4.641

Silva, J. M. C. S., & Tenreyro, S. (2010). On the existence of the maximum likelihood estimates in Poisson regression. *Economics Letters*, *107*(2, 2), 310–312. http://doi.org/10.1016/j.econlet.2010.02.020

SITRAMyTEM. (n.d.). SITRAMyTEM. Secretaría de Comunicaciones. Gobierno del Estado de México. Retrieved December 30, 2021, from https://comunicaciones.edomex.gob.mx/sistramytem

Smolka, M. O. (2013). *Implementing Value Capture in Latin America: Policies and Tools for Urban Development* (No. 9781558442849). Cambridge, MA.

Smolka, M. O., & Maleronka, C. (2018). Assessing the monetary relevance of land value capture : The case for charges for additional building rights in São Paulo , Brazil. *International Journal of Real Estate and Land Planning*, *1*. Retrieved from https://ejournals.lib.auth.gr/reland

SNIIV. (n.d.). Sistema Nacional de Información e Indicadores de Vivienda (SNIIV). Retrieved January 24, 2022, from https://sniiv.sedatu.gob.mx

Sorensen, A. (2000). Land readjustment and metropolitan growth: An examination of suburban land development and urban sprawl in the Tokyo metropolitan area. *Progress in Planning*, *53*(4, 4), 217–330. http://doi.org/10.1016/S0305-9006(00)00002-7

Srour, I. M., Kockelman, K. M., & Dunn, T. P. (2002). Accessibility Indices: Connection to Residential Land Prices and Location Choices. *Transportation Research Record*, *1805*(1, 1), 25–34. http://doi.org/10.3141/1805-04

Ståhle, A. (2010). More green space in a denser city: Critical relations between user experience and urban form. *Urban Design International*, *15*(1, 1), 47–67. http://doi.org/10.1057/udi.2009.27

STC. (2022). Metro CDMX. Retrieved January 8, 2022, from https://www.metro.cdmx.gob.mx

Stepniak, M., & Jacobs-Crisioni, C. (2017). Reducing the uncertainty induced by spatial aggregation in accessibility and spatial interaction applications. *Journal of Transport Geography*, *61*(April, April), 17–29. http://doi.org/10.1016/j.jtrangeo.2017.04.001

Stepniak, M., & Rosik, P. (2015). The Impact of Data Aggregation on Potential Accessibility Values. In I. Ivan, I. Benenson, B. Jiang, J. Horák, J. Haworth, & T. Inspektor (Eds.), *Geoinformatics for Intelligent Transportation* (pp. 227–240). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-

319-11463-7_16

Stępniak, M., & Rosik, P. (2018). The Role of Transport and Population Components in Change in Accessibility: The Influence of the Distance Decay Parameter. *Networks and Spatial Economics*, *18*(2, 2), 291–312. http://doi.org/10.1007/s11067-017-9376-8

Stewart, J. Q. (1947). Empirical Mathematical Rules concerning the Distribution and Equilibrium of Population. *Geographical Review*, *37*(3, 3), 461–485. http://doi.org/10.2307/211132

Stillwell, J., Daras, K., & Bell, M. (2018). Spatial Aggregation Methods for Investigating the MAUP Effects in Migration Analysis. *Applied Spatial Analysis and Policy*, *11*(4, 4), 693–711. http://doi.org/10.1007/s12061-018-9274-6

Stokenberga, A. (2014). Does Bus Rapid Transit Influence Urban Land Development and Property Values: A Review of the Literature. *Transport Reviews*, *34*(3, 3), 276–296. http://doi.org/10.1080/01441647.2014.902404

Stokenberga, A., & Schipper, L. (2012). Trends in transport activity, energy use, and carbon footprint in Mexico City, Mexico. *Transportation Research Record*, (2287, 2287), 105–112. http://doi.org/10.3141/2287-13

Surprenant-Legault, J., Patterson, Z., & El-Geneidy, A. M. (2013). Commuting trade-offs and distance reduction in two-worker households. *Transportation Research Part A: Policy and Practice*, *51*, 12–28. http://doi.org/10.1016/j.tra.2013.03.003

Suzuki, H., Cervero, R., & Luchi, K. (2013). *Transforming Cities with Transit*. Washington, DC: World Bank. http://doi.org/10.1017/CBO9781107415324.004

Suzuki, H., Murakami, J., Hong, Y.-H., & Tamayose, B. (2015). *Financing Transit-Oriented Development with Land Values*. Washington, DC: The World Bank.

Tahmasbi, B., & Haghshenas, H. (2019). Public transport accessibility measure based on weighted door to door travel time. *Computers, Environment and Urban Systems*, *76*, 163–177. http://doi.org/10.1016/j.compenvurbsys.2019.05.002

Tan, P. Y., & Samsudin, R. (2017). Effects of spatial scale on assessment of spatial equity of urban park provision. *Landscape and Urban Planning*, *158*, 139–154. http://doi.org/10.1016/j.landurbplan.2016.11.001

Taylor, B. M., & Diggle, P. J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, *84*(10, 10), 2266–2284.

http://doi.org/10.1080/00949655.2013.788653

Taylor, L. O. (2008). Theoretical Foundations and Empirical Developments in Hedonic Modeling. In A. Baranzini, J. Ramirez, C. Schaerer, & P. Thalmann (Eds.), *Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation* (pp. 15–37). New York, NY: Springer New York. http://doi.org/10.1007/978-0-387-76815-1_2

Taylor, L. O. (2017). Hedonics. In P. A. Champ, K. J. Boyle, & T. C. Brown (Eds.), *A Primer on Nonmarket Valuation* (2nd ed., pp. 235–292). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-7104-8

TfL. (2017). *Land value capture* (Policy report). London: Transport for London and Greater London Authority. Retrieved from https://www.london.gov.uk/sites/default/files/land_value_capture_report_transport_for_london.pdf

Thériault, M., Voisin, M., & Des Rosiers, F. (2013). Accessibility of Urban Services: Modeling Socio-spatial Differences and their Impacts on Residential Values. In M. Thériault & F. Des Rosiers (Eds.), *Modeling Urban Dynamics: Mobility, Accessibility and Real Estate Value* (pp. 225–254). London: ISTE Ltd adn John Wiley & Sons, Inc.

Thorsen, I., & Gitlesen, J. P. (1998). Empirical evaluation of alternative model specifications to predict commuting flows. *Journal of Regional Science*, *38*(2, 2), 273–292. http://doi.org/10.1111/1467-9787.00092

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, *64*(5), 416–424. Retrieved from https://www.jstor.org/stable/1826343

Tischler, S., & Mailer, M. (2019). Cable Propelled Transit Systems in Urban Areas. *Transportation Research Procedia*, *41*, 169–173. http://doi.org/10.1016/j.trpro.2019.09.032

Todaro, R. (1978). La renta de la tierra: Algunos antecedentes teoricos. *Revista de Estudios Urbano Regionales*, *5*(15, 15).

Torrens, P. M. (2003). Cellular Automata and Multi-agent Systems as Planning Support Tools. In S. Geertman & J. Stillwell (Eds.), *Planning Support Systems in Practice* (pp. 205–222). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-24795-1_12

Torrens, P. M. (2012). Urban Geosimulation. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-Based Models of Geographical Systems* (pp. 435–450). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-90-481-8927-4_20

Tra, C. I. (2010). A discrete choice equilibrium approach to valuing large environmental changes. *Journal of Public Economics*, *94*(1), 183–196. http://doi.org/10.1016/j.jpubeco.2009.10.006

Tra, C. I. (2013). Measuring the General Equilibrium Benefits of Air Quality Regulation in Small Urban Areas. *Land Economics*, *89*(2, 2).

Transportation Research Board*Case studies in bus rapid transit.* (2003). (Synthesis of case study No. 90). Washington, D.C. Retrieved from https://nap.nationalacademies.org/catalog/24726/bus-rapid-transit-volume-1-case-studies-in-bus-rapid-transit

United Nations. (2015). World Population Prospects. Retrieved from https://esa.un.org/unpd/wpp/

Vadali, S., Zmud, J., Carlson, T., DeMoors, K., Rybeck, R., Fitzroy, S., … Sieber, M. (2018). *Guidebook to Funding Transportation Through Land Value Return and Recycling* (Research report No. 873). Washington, D.C.: National Academies of Sciences, Engineering, and Medicine. Retrieved from https://doi.org/10.17226/25110

van der Loo, M. (2021). *Stringdist: Approximate string matching, fuzzy text search, and string distance functions.* Retrieved from https://github.com/markvanderloo/stringdist

van der Loo, M. P. J. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*, 111–122. Retrieved from https://CRAN.R-project.org/package=stringdist

van Duijn, M., & Rouwendal, J. (2013). Cultural heritage and the location choice of Dutch households in a residential sorting model. *Journal of Economic Geography*, *13*(3), 473–500. Retrieved from https://www.jstor.org/stable/26158646

van Wee, B., & Witlox, F. (2021). COVID-19 and its long-term effects on activity participation and travel behaviour: A multiperspective view. *Journal of Transport Geography*, *95*, 103144. http://doi.org/10.1016/j.jtrangeo.2021.103144

Vejchodská, E., Barreira, A. P., Auziņš, A., Jürgenson, E., Fowles, S., & Maliene, V. (2022). Bridging land value capture with land rent narratives. *Land Use Policy*, *114*, 105956. http://doi.org/10.1016/j.landusepol.2021.105956

Velandia Naranjo, D. J. (2013). *The Impact of Bus Rapid Transit System on Land Prices in Mexico City* (Working paper No. WP13DV1). Cambridge, MA: Lincoln Institute of Land Policy. Retrieved from https://www.lincolninst.edu/publications/working-papers/impact-bus-rapid-transit-system-land-prices-mexico-city

Verma, A., & Ramanayya, T. V. (2015). Different Modes and Their Characteristics. In *Public Transport Planning and Management in Developing Countries* (pp. 33–60). Boca Raton, FL: CRC Press.

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428–436. http://doi.org/10.1016/j.jbusres.2017.12.043

Von Thünen, J. H. (1966). *Isolated state: An English edition of Der isolierte Staat.* (P. Hall, Ed.). Oxford et al.: Pergamon Press.

Vuchic, V. R. (2007). *Urban Transit: Operations, Planning, and Economics.* Hoboken, NJ: John Wiley & Sons Inc.

Walters, L. C. (2012). Are Property-Related Taxes Effective Value Capture Instruments? In G. K. Ingram & Y.-H. Hong (Eds.), *Value capture and land policies.* Cambridge, MA: Lincoln Institute of Land Policy.

Webster, C. (2010). Pricing accessibility: Urban morphology, design and missing markets. *Progress in Planning*, *73*(2, 2), 77–111. http://doi.org/10.1016/j.progress.2010.01.001

WHO. (2017). *Urban green spaces: A brief for action.* Copenhagen: World Health Organization. Retrieved from https://www.euro.who.int/en/health-topics/environment-and-health/urban-health/publications/2017/urban-green-spaces-a-brief-for-action-2017

Williams, A. W. (1991). A guide to valuing transport externalities by hedonic means. *Transport Reviews*, *11*(4), 311–324. http://doi.org/10.1080/01441649108716793

Wilson, A. G. (1971). A Family of Spatial Interaction Models, and Associated Developments. *Environment and Planning A: Economy and Space*, *3*(1, 1), 1–32. http://doi.org/10.1068/a030001

Wilson, A. G. (2010). Entropy in Urban and Regional Modelling: Retrospect and Prospect. *Geographical Analysis*, *42*(4, 4), 364–394. http://doi.org/10.1111/j.1538-4632.2010.00799.x

Wirasinghe, S. C., Kattan, L., Rahman, M. M., Hubbell, J., & Thilakaratne, R. (2013). Bus rapid transit – a review, *5934*. http://doi.org/10.1080/12265934.2013.777514

Wong, D. W. S. (2009a). Modifiable Areal Unit Problem. In R. Kitchin & N. Thrift (Eds.), *International Encyclopedia of Human Geography* (pp. 169–174). Oxford: Elsevier. http://doi.org/10.1016/B978-008044910-4.00475-2

Wong, D. W. S. (2009b). The Modifiable Areal Unit Problem (MAUP). In *The SAGE Handbook of Spatial Analysis*. London: SAGE Publications, Ltd. http://doi.org/10.4135/9780857020130

Wong, D. W. S., Lasus, H., & Falk, R. F. (1999). Exploring the variability of segregation index D with scale and zonal systems: An analysis of thirty US cities. *Environment and Planning A*, *31*(3, 3), 507–522. http://doi.org/10.1068/a310507

Xie, Y. (2021a). *Bookdown: Authoring books and technical documents with r markdown.*

Xie, Y. (2021b). *Knitr: A general-purpose package for dynamic report generation in r.* Retrieved from `https://yihui.org/knitr/`

Xie, Y. (2021c). *Tinytex: Helper functions to install and maintain TeX live, and compile LaTeX documents.* Retrieved from `https://github.com/yihui/tinytex`

Ye, X. (2022). Estimating sigma squared for the Classical Linear Regression Model (CLRM) with the Presence of the Modifiable Areal Unit Problem (MAUP). *Geographical Analysis*, *54*(2, 2), 382–404. http://doi.org/10.1111/gean.12291

Ye, X., & Rogerson, P. (2022). The Impacts of the Modifiable Areal Unit Problem (MAUP) on Omission Error. *Geographical Analysis*, *54*(1, 1), 32–57. http://doi.org/10.1111/gean.12269

Yen, B. T. H., Mulley, C., & Shearer, H. (2019). Different Stories from Different Approaches in Evaluating Property Value Uplift: Evidence from the Gold Coast Light Rail System in Australia. *Transportation Research Record*, *2673*(3, 3), 11–23. http://doi.org/10.1177/0361198118823001

Yen, B. T. H., Mulley, C., Shearer, H., & Burke, M. (2018). Announcement, construction or delivery: When does value uplift occur for residential properties? Evidence from the Gold Coast Light Rail system in Australia. *Land Use Policy*, *73*(July 2017, July 2017), 412–422. http://doi.org/10.1016/j.landusepol.2018.02.007

Yin, L., Zhang, H., Patterson, K., Silverman, R., & Wu, L. (2020). Walkability, Safety, and Housing Values in Shrinking Cities: Spatial Hedonic Study in Buffalo, Pittsburgh, and Detroit. *Journal of Urban Planning and Development*, *146*(3, 3), 04020029. http://doi.org/10.1061/(ASCE)UP.1943-5444.0000595

Yoshino, N., & Stillman, G. B. (2018). Back to the Future: Instructive Features from Past Innovations in Raising Private Finance for Infrastructure. In N. Yoshino, M. Helble, & U. Abidhadjaev (Eds.), *Financing Infrastructure in Asia and the Pacific: Capturing Impacts and New Sources* (pp. 333–365). Tokyo: Asian Development Bank Institute.

Zhao, Z. J., Das, K. V., & Larson, K. (2012). Joint Development as a Value Capture Strategy in Transportation Finance. *Journal of Transport and Land Use*, *5*(1, 1), 5–17. http://doi.org/10.5198/jtlu.v5i1.142

Zhao, Z. J., Iacono, M., Lari, A., & Levinson, D. (2012). Value Capture for Transportation Finance. *Procedia - Social and Behavioral Sciences*, *48*, 435–448. http://doi.org/10.1016/j.sbspro.2012.06.1023

Ziccardi, A., & González Reynoso, A. (2015). Política de vivienda y municipios en México. In *Habitabilidad y política de vivienda en México*. México, D.F.: Universidad Nacional Autónoma de México. Retrieved from `https://www.puec.unam.mx/index.php/publicaciones/145-publicaciones-digitales/pd-instituciones/415-habitabilidad-y-politica-de-vivienda-en-mexico`

Zolnik, E. (2019). A longitudinal analysis of the effect of public rail infrastructure on proximate residential property transactions. *Urban Studies*, *00*(0, 0), 004209801983656–004209801983656. http://doi.org/10.1177/0042098019836564

Zuk, M., Bierbaum, A. H., Chapple, K., Gorska, K., & Loukaitou-Sideris, A. (2018). Gentrification, Displacement, and the Role of Public Investment. *Journal of Planning Literature*, *33*(1, 1), 31–44. http://doi.org/10.1177/0885412217716439