Income Inequalities and Well-being in Rural Pakistan

Khadija Shams

submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

College of Social Sciences
Business School
Economics
University of Glasgow

January 2012

Thesis Advisers:	Dr. Luis Angeles
	Dr. Alexander Kovalenkov
Examination Committee:	Professor Richard Harris
	Dr. Sai Ding
	Dr. Philippe LeMay-Boucher

In the honour and memory of my father Malik Shams-ul-Haq Khattak

Abstract

Income inequalities and subjective well-being have been increasingly identified in the literature as important measures of socio-economic cohesion. This is particularly relevant for developing economies that are typically characterised by strong population growth and relatively low incomes per head. Although in those economies a considerable share of resources is derived from rural areas, data availability for these regions is often an issue which precludes important insights into the overall socio-economic tissue of the developing world. This dissertation seeks to advance our knowledge on various aspects of inequalities and well-being with particular emphasis on rural Pakistan. At the core of the present monograph lie three chapters that deal with income inequality, subjective well-being as well as physical well-being (i.e. health). The empirical analysis is based on a unique survey dataset that covers the four provinces of rural Pakistan. The dissertation seeks to contribute to the existing literature in several dimensions. We decompose overall income inequality by its different types to disentangle which sources of income are inequality-increasing and which ones reduce socio-economic divergence. The empirical measurement and assessment of both subjective and physical well-being in rural Pakistan is a rather novel aspect. We introduce and examine different well-being measures as indicators of (subjective) poverty and find that well-being in rural areas is largely driven by financial factors. When it comes to health, however, overall results are less clear-cut. The thesis is therefore able to offer several policy recommendations for important socio-economic factors in rural Pakistan. On a more general note, some of the results discussed might also illuminate the policy debate in other geographic areas with similar characteristics.

Contents

Acknowledgements

A common joke in my home country suggests that *PhD* stands for *Permanent head Damage*. I am not sure I can agree entirely. It goes without saying that such a long-term endeavour brings many challenges and, yes, at times sufferings. But I feel very blessed and thankful to have met so many helpful people along the way. It would be impossible to mention all of them explicitly. With all their support, I truly feel my studies have not caused me any (mental) damage at all but have rather been a *Pleasant and highly-rewarding Discovery*.

First and foremost, I am extremely thankful to my supervisors, Dr. Luis Angeles and Dr. Alexander Kovalenkov, for their valuable suggestions and continuous support in all aspects of my research. I am deeply indebted to Luis for his advice and guidance. I also would like to thank Alexander for his support in setting up the study and his excellent comments on the first draft of this dissertation.

My work has moreover benefited from fruitful discussions with other colleagues as well as constructive comments made by several participants at various workshops and research seminars. A special word of thanks to my *liebling* and officemate Alexander for being, above all, a dear friend.

Working on a research project of this nature is an iterative process that requires several stages of revision. I am therefore truly grateful for the suggestions and remarks made by my examiners. I believe their comments have improved my work. I am particularly pleased about the superb advice by Professor Richard Harris. Richard has introduced me to interesting techniques that allowed me to strengthen the analysis.

I would like to express my gratitude to the Department of Economics for providing me with the necessary software and for offering a pleasant research environment. I highly appreciate the training courses and research events organised by the Scottish Institute for Research in Economics (SIRE), the Scottish Graduate Programme in Economics (SGPE) and the Applied Quantitative Methods Network (AQMeN). I am particularly thankful for the personal insights offered by Professor Bill Greene and Professor Elhanan Helpman on various econometric methods for the empirical data analysis.

The biggest word of appreciation has to go to the members of the households participating in the survey across the four districts of rural Pakistan. Without their assistance and cooperation, this study would not have been possible.

I am pleased to acknowledge financial support by the Higher Education

Commission (HEC) in Pakistan and am grateful to the Frontier Women University, Peshawar, for giving me the opportunity to study abroad. Last but not least, I would like to thank my mum and my family for their understanding, patience and moral support, which has enabled me to complete my studies.

Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation 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.

The copyright of this thesis rests with the author. Due acknowledgement must always be made of the use of any materials contained in, or derived from, this thesis.

List of Tables

Chapter 1

Introduction and Motivation

This thesis investigates the determinants of inequalities and well-being in rural Pakistan using unique survey data for the year 2008; henceforth referred to as (Survey 2008). The data has been collected by means of household surveys for the special purpose of this research (see **Chapter 2** for details) and focuses on rural areas that, despite being truly representative for the Pakistani economy, have received only little attention in the literature thus far. The notion of inequalities and well-being in this dissertation is coined in terms of three topics that are at the core of the following chapters: income inequalities, (subjective) well-being poverty and health inequalities. An important aim of this work is to highlight to what extent these aspects overlap (say, are relatively rich people also relatively more healthy?), whilst at the same time investigating somewhat novel linkages in the development economics literature (say, how does poverty in subjective terms relate to household's well-being?).

The analysis is motivated by the case of rural Pakistan but we seek to benchmark our findings in the light of results for other (both developing and industrialised) nations throughout. The dissertation comprises three core chapters which we succinctly want to motivate in this introductory chapter in turn.

The first core chapter, **Chapter 3**, is concerned with income inequalities – arguably the most common type of inequality considered in the literature. Income inequality has always been a major issue for emerging economies like Pakistan.[1] According to the World Bank (2005), Pakistan's Gini coefficient (G) as natural measure of income inequality amounts to $G = 0.33$. A more

[1]It should be noted, however, that economic inequalities have increasingly been iden-

recent study by United Nations Development Programme (UNDP, 2007), on the other hand, provides a Gini coefficient of $G = 0.31$, which seems to suggest an (albeit slight) downward trend in income inequality for Pakistan over the past few years.

In the present monograph, we do not just seek to measure overall inequalities but also aim to disentangle their sources with special reference to the rural parts of Pakistan, one of the world's most populous countries. More specifically, we decompose total income inequality in rural Pakistan by its various components. This permits analysing both the contribution and the effect of each source of income on overall income inequality. All income figures are based on our field survey conducted in 2008 which allows for unique insights into geographic areas that have been hitherto largely neglected.

There exists a number of empirical studies for developing countries that have measured the contribution of different factors contributing to total income inequality, using various statistical techniques. On Pakistan see, amongst others, Kruijk (1987), Mohammad and Badar (1985) and Ercelawn (1984). On other developing countries such as Sri Lanka and India, see Glewwe (1986) and Nugent and Walther (1982), respectively. Also, see for example Pyatt et al. (1980) for a methodological overview.

Studies like these can yield important policy implications. For instance, they may assist policymakers in identifying structure and character of income inequalities but also reveal how those potentially change over time.[2] Policymakers with such information at their disposal should (at least in theory) be able to make better informed decisions when it comes to designing welfare programmes or introducing redistribution measures that seek to alter the economy-wide income distribution. This thesis intends to make several contributions in that dimension. First, it advances our knowledge about the sources of income inequality in the developing world by examining the sources of such inequalities in rural Pakistan. Second, it employs various statistical techniques to identify the contribution of different sources of income – agricultural, livestock, rental, non-farm and transfer – to overall income inequality in rural areas. Third, it provides a correlation analysis

tified as important measure of overall coherence in developed economies as well. Indeed, there is currently a rather active literature on aspects pertaining to social mobility, subprime lending, wage dumping due to globalisation, etc. – all of which are believed to lead to or are causal for rising divergence within societies.

[2]Against the backdrop of non-negligible costs in obtaining and validating the data used in this study, we can unfortunately only provide limited indications as to that latter aspect.

between income inequalities and the shares of the different income sources across the districts in order to distinguish between inequality-increasing and inequality-decreasing sources of income. Fourth, the analysis in Chapter 3 provides several policy implications and suggests ways to narrow disparities in the income distribution.

In **Chapters 4 and 5**, we turn to the rather novel aspect of measuring and assessing well-being. Indeed, there is an ongoing political debate to what extent somewhat crude activity measures such as real GDP are able to accurately reflect the state of an economy. The growing desire in the economic literature to look beyond the optimising rationale of economic agents might also explain the increasing importance of behavioural economics; a field that essentially incorporates psychological aspects into the analysis. We investigate well-being as economic concept in the present study in different contexts.

In **Chapter 4**, we analyse subjective well-being in relation to poverty from two angles (i) overall well-being and (ii) financial well-being. Thus, on a methodological level, the dissertation makes an attempt to link the emerging field of *happiness economics* with *development studies*. In particular, we integrate subjective poverty indicators with the so-called happiness function. Much of the work in that area has been done on subjective well-being for developed countries. Recently, there have been first applications to emerging economies such as Kingdon and Knight (2006) for South Africa and Knight et al. (2007) for rural China. The thesis adds to this strand of the literature on subjective well-being for the developing world by investigating subjective well-being poverty in rural Pakistan. Indeed, to the best of our knowledge this is the first study on that matter for rural Pakistan. We introduce and compare different measures of subjective well-being as means of capturing poverty. Unlike Kingdon and Knight (2006), we do not only resort to *overall well-being* and thus an indirect approach to subjective well-being measurement in relation to poverty. Instead, we additionally consider *financial well-being* that arguably corresponds more closely to the "usual" connotation of poverty as development problem and thus provide a more direct and general approach to subjective well-being measurement in this context.

We explicitly test for physical well-being (i.e. health) in relation to *absolute* as well as *relative* income standards in **Chapter 5**. The underlying motivation is to investigate to what extent health inequalities are driven by income differentials. There is increasing (both anecdotal and empirical) evidence to suggest that all over the world individual health is more and

more linked to personal wealth and the individual income status. We test in particular two key hypotheses in that regard: the *Absolute Income Hypothesis* (AIH) and the *Relative Income Hypothesis* (RIH). The AIH states that the household's health status improves with higher income. According to the RIH, the household's health is determined by its income relative to society. This implies that a relatively higher social status should ensure a better health status. For the RIH, also the distribution of income matters such that living in a society with relatively high income inequality may have a detrimental impact on the health status. The health literature captures this in a further hypothesis that is obviously closely related to the RIH, the so-called Income Inequality Hypothesis (IIH). The IIH seeks to answer the question whether inequality affects everybody in society equally or whether it is only harmful to the health of the least well-off. The former idea describes the *strong version* of the IIH, whereas the latter aspect would imply that the IIH only holds in the *weak form*.

Studies using large international datasets strongly support the AIH but provide no support for the RIH and little or no support for the IIH (see, amongst others, Wagstaff and Doorslaer (2000); Lindley and Lorgelly (2005); Gerdtham and Johannesson (2004); Lobmayer and Wilkinson (2000)). Marmot et al. (1991), in contrast, analysing health inequalities among British civil servants found that it is relative rather than absolute income that is important for health – such that a lower relative income increases the chances of ending up in a lower health category. Li and Zhu (2006) tested the two versions of the IIH for China. Their results provide evidence in favour of the strong version of the IIH, however they failed to confirm the weak version of the hypothesis.

Previewing the main results of our study for rural Pakistan, we find stronger empirical support in favour of the RIH compared to the AIH but no evidence for the validity of the IIH. We also find evidence for income inequality being beneficial for reported health in a developing economy like rural Pakistan. We furthermore observe that high income inequality particularly benefits the richer segment of society in this context.

Every researcher is confronted with a whole variety of **methodological approaches** to seek answers to her research questions. We believe that in many cases a "simple hammer is sufficient to hit the nail on its head". This general approach is reflected in the present monograph. Our analysis therefore contains descriptive as well as more formal methods throughout. We believe it is important to have a solid understanding of the overall picture to begin with and thus analyse the survey results first for different

levels of aggregation (i.e. both on the province and the district level). Our analysis proceeds in steps. Most of our results are based on regression models following a particular weighting scheme to ensure representativeness. We typically consider ordered probit models based on Maximum Likelihood (ML) estimators. In some instances, we also resort to Poisson regressions and more conventional least squares estimators.

In a nutshell, the remainder of the dissertation is **structured** as follows. Chapter 2 provides background information on Pakistan in general and particularly its rural parts and puts the study in context with the prevailing institutional environment. The chapter also contains a detailed description of the survey design, gives further details on data collection and outlines the weighting scheme that has been used for the econometric analysis. Chapter 3 deals with income inequalities. This chapter contains a review of the literature on the decomposition of income inequality by income sources. The data suggests that in particular non-farm income and transfer income are inequality increasing and inequality decreasing sources of income, respectively. Chapter 4 focuses on different approaches to well-being measurement as subjective poverty indicators. We apply several approaches and propose an appealing and original interpretation of poverty in a socio-economic context. Chapter 5 focuses on the socio-economic determinants of health. We find that two factors play an important role in determining a household's health: income and the family size of a household. However, we argue that in fact it is relative income rather than absolute income that influences a household's health status. Chapter 6 summarises the main findings and highlights important contributions of the study. Finally, we suggest various policy implications and hint at potential avenues for further research.

Chapter 2

Country context and research design

This Chapter serves two purposes. First, it provides general background information (Section 2.1) on rural Pakistan to put the study in the appropriate context. Second, it outlines in detail the research design (Section 2.2) which should facilitate understanding the econometric analysis in the subsequent core chapters.

2.1 Country context

2.1.1 Geographic and macroeconomic background

Our study of income inequalities and well-being in rural Pakistan obviously needs to take the country's geographic and demographic features into account.[1] The Islamic Republic of Pakistan is located in southern Asia and borders the Arabian Sea, India in the east, Iran and Afghanistan in the west and China to the north. According to the most recent official data provided by the Population Census Organization, Pakistan has a land area of 796,096 km^2 and a population of approximately 132 million (Government of Pakistan, Statistics Division, 1998). Excluding the national capital Islamabad, there are in total 105 districts within the four provinces of Baluchistan, the North Western Frontier Province (NWFP), Sind and Punjab. Table 2.1

[1]The following country context and anthropologic background draws upon Mailk (2005).

gives an overview of population and area figures in Pakistan.[2] We observe that Punjab is the major province in Pakistan with an overall population of more than 50%, followed by Sind and NWFP accordingly. Baluchistan, on the other hand, covers the largest area of Pakistan but has the lowest population density.

Province	**Area (km^2)**	**Population total**	**Population (%)**	**Population rural**	**Population density (per km^2)**
Punjab	206,251	74,426,525	57.6	50,854,022	361
Sind	140,914	30,439,893	23.6	15,585,225	216
NWFP	47,521	17,743,645	13.8	14,744,969	238
Baluchistan	374,190	6,565,885	5.1	4,996,638	19
Pakistan	768,876	129,175,948	100.0	86,180,855	168

Table 2.1: Area and population of Pakistan.
Source: Government of Pakistan, Statistics Division (1998).

Tables A.1 and A.2 in the Appendix report several key development indicators for Pakistan and its neighbours over the course of previous three years prior to our survey in 2008. Both Tables reveal stability in terms of the evolution of the macroeconomy in Pakistan as well as its relative position compared to neighbouring countries. We see that at the time of the survey, the Pakistani economy was on a relatively stable path. Unemployment, inflation and GPD per capita were hardly fluctuating. In comparison to India and Bangladesh, poverty in Pakistan was about three times lower. The degree of income inequality was of comparable magnitude. Thus, data obtained at time of the survey reflects a situation in the absence of major macroeconomic shocks. It can be conjectured, however, that global economic shocks after the survey may have also changed the picture even on a micro level.

2.1.2 Cultural and socio-economic background

2.1.2.1 Religion

Pakistan was formed as an Islamic nation and Islam continues to be the religion of approximately 95 percent of the population. The vast majority of the households lead a life in line with their religious and cultural values. There are also small groups of Buddhists, Christians, Parsis, and Hindus.

[2]We exclude the Federally Administrated Tribal Areas (FATA).

2.1.2.2 Family setup

In a typical Pakistani family setup, the head of the household is usually considered to be the father who is the major decision-maker for the household in various economic and non-economic aspects. Furthermore, the representative Pakistani family is set up in a joint way, in which all the family members pool their resources for the mutual economic support of the entity as a whole. Therefore, in general, Pakistani households have a preference for a family structure with many children. This ensures a sufficient amount of working and earning hands for their families. Indeed, we may say that in developing countries like Pakistan children are used as insurance mechanism. According to (CIA, 2008), the average family size in Pakistan is 4-5 children per household and it is the 6^{th} most populous country in the world.

A representative family in Pakistan is characterised by the following gender roles and statuses. The majority of Pakistani women are homemakers, and men are generally referred to as the breadwinners. The largest percentage of working women in Pakistan are nurses or teachers. There are growing numbers of violent crimes against or involving women and the government has introduced the concept of women police stations, which have been opened in major cities like Rawalpindi in Punjab, Karachi in Sind and Abbottabad in NWFP.

2.1.2.3 Social stratification

There is no caste system in Pakistan, unlike for instance India. There are high-income, middle-income and a large number of low-income persons throughout the country. Locale makes an important difference in the quality of life; a low-income person in an urban area typically is confronted with higher levels of inequalities and thus faces more social problems compared to rural, tribal or mountainous areas.

2.1.2.4 Typical lifestyle in a Pakistani village

Pakistan is an agricultural country where the two thirds of the population living in rural areas depend mostly on the agriculture sector. Village life in Pakistan depicts a true picture of our culture. Villagers are very traditional people who are hard workers. They wake up early in the morning with the Fajar prayers and start working in the fields. They work all day long in the field under the sun without caring about the harsh weather. This is the only way for them to earn their livelihood.

Most villages in Pakistan are situated away from the noise of the city life. They are peaceful and silent places. A typical Pakistani village consists of unpaved paths and streets. Its houses are made of mud. However, with lot of young members from rural families which moved to the urban sector (as part of internal migration) and to the Gulf (external migration) as part of the *Dubai Chalo* notion have benefited from the petro dollars. As a result, the villagers now build their houses from bricks and concrete though most of the village people have simple habits and limited needs.

There are green trees, vast meadows and flowery bushes in every village. In the summer they rest under shady trees, and take bath in cool water. Women also help their men in their work along with their household. They moreover take care of their domestic animals such as cows, goats, hens etc. Many small villages are still void of the facilities like safe drinking water and electricity, even hospitals and schools are typically at long distances, such that life in the village entails more struggle than the relatively modern lifestyles in the cities.

2.1.2.5 Healthcare

At a seminar at Aga Khan Medical University in 1998, medical experts reported that perinatal mortality rates in Pakistan were alarmingly high with an estimated 54 deaths per thousand births. A 1990-1994 national health survey[3] reported that 89 children per thousand under age five died in Pakistan from pneumonia, diarrhea, vaccine prevention diseases, or a combination of them, with most of these deaths occurring in the first week after birth.

A number of programs have been undertaken to attack polio; the World Health Organisation and Japan, amongst others, have participated. At the end of the twentieth century, there were one hundred thousand deaths from and at least twenty thousand new cases of paralytic polio each year.

A survey by the Federal Bureau of Statistics in Pakistan indicated that about 50 percent of the basic health units were without doctors and that about 70 percent of government health facilities are without any female staff. Only about 56 percent of the country's people have safe drinking water and just 24 percent have good sanitation.

Programs are underway to expand basic health services for women, trying to develop a women-friendly district health system, and both to

[3]More recent survey data of this scope unfortunately was not available.

strengthen and to improve human resource capacity to sustain women's health development.

2.1.2.6 Land Tenure and Property

An estimated 54.69 million acres (22.14 million hectares) of land are used for agriculture. The land is usually held by the private sector. The major crops are cotton, wheat, rice and sugarcane. A large amount of land in Pakistan has archaeological sites such as Moenjo Daro, Harappa, Taxila, Kot Dijji and Mehr Garh.

2.1.3 Commercial activities and taxation

2.1.3.1 Trade and labour

A large percentage of the commercial activities include the sale of handicraft items such as the carpets for which Pakistan is well known. Major industries of Pakistan are textiles, cement, fertilizer, steel, sugar, electric goods and shipbuilding. Pakistan's major exports include cotton, textile goods, rice, leather items, carpets, sports goods, fruit and handicrafts. Major imports are industrial equipment, vehicles, iron ore, petroleum and edible oil. Main trade partners include the United States, Hong Kong, Japan, Germany, the United Kingdom and the United Arab Emirates. With regards to the division of labour, 48 percent of workers are in the service sector, 27 percent are in industry, and 25 percent are in agriculture.

2.1.3.2 Types of taxes in Pakistan

Federal taxes in Pakistan like most of the taxation systems in the world are classified into two broad categories, i.e. direct and indirect taxes.[4] *Direct taxes* primarily comprise income tax along with a supplementary role for wealth tax. For the purpose of the charge of tax and the computation of total income, all income is classified under the following headings: Salaries, interest on securities, income from property, income from business or professions, capital gains and income from other sources. *Indirect taxes* relate to custom duties and general sales tax. A broad description regarding the nature of the administration of these taxes is explained in the following:

Personal tax applies to all individuals, unregistered firms, associations of persons, etc., that are liable to tax, at the rates ranging from 10 to 35

[4]The subsequent discussion of Pakistan's tax system is based upon Kausar (2001).

per cent. *Tax on companies* applies to all public companies (other than banking companies) incorporated in Pakistan that are assessed for tax at a corporate rate of 39%. However, the effective rate is likely to differ on account of allowances and exemptions related to industry, location, exports, etc.

Unilateral Relief: A person resident in Pakistan is entitled to a relief in tax on any income earned abroad, if such income has already been subjected to tax outside Pakistan. Proportionate relief is allowed on such income at an average rate of tax in Pakistan or abroad, whichever is lower. The Government of Pakistan has so far signed agreements to avoid double taxation with several countries including almost all of the developed countries of the world. These agreements lay down the ceilings on tax rates applicable to different types of income arising in Pakistan. They also lay down some basic principles of taxation which cannot be modified unilaterally.

Goods imported and exported from Pakistan are liable to rates of *customs duties* as prescribed in the Pakistan Customs Tariff. Customs duties in the form of import duties and export duties constitute about 37% of the total tax receipts. The rate structure of customs duty is determined by a large number of socio-economic factors. However, the general scheme envisages higher rates on luxury items as well as on less essential goods. The import tariff has been given an industrial bias by keeping the duties on industrial plants and machinery and raw material lower than those on consumer goods.

Central Excise duties are leviable on a limited number of goods produced or manufactured and services provided or rendered in Pakistan. On most of the items Central Excise duty is charged on the basis of value or the retail price. Some items are, however, chargeable to duty on the basis of weight or quantity. Classification of goods is done in accordance with the Harmonized Commodity Description and Coding system which is being used all over the world. All exports are exempted from Central Excise Duty.

Sales Tax is levied at various stages of economic activity at the rate of 15 per cent on all goods imported into Pakistan, payable by the importers as well as all supplies made in Pakistan by a registered person in the course of furtherance of any business carried on by him. There is an in-built system of input tax adjustment and a registered person can make adjustment of tax paid at earlier stages against the tax payable by him on his supplies. Thus, the tax paid at any stage does not exceed 15% of the total sales price of the supplies.

2.1.3.3 Tax evasion in Pakistan

For income tax purposes, the population of the country may be categorised as: (1) cases with no incomes or whose incomes are below the income tax threshold, (2) cases which fall within the taxable bracket but are exempt from payment of income tax under Schedule II of ITO, 1979, (3) cases which fall within the taxable bracket but have successfully avoided entering the tax net, (4) cases which are within the tax net but under-report their incomes, (5) cases which are within the tax net and correctly report their incomes but where there is the possibility of differences with the tax department on the extent of their taxable income. A measure of the extent of tax evasion is provided by the value of assets declared under the current tax amnesty scheme. Under this scheme, assets of PKR 120 billion were declared, which could not be explained through known income sources. Property and commercial surveys, if conducted frequently can identify individuals and businesses that are outside the tax net and can also limit the scope of tax evasion. Thus, further efforts might be needed to improve the existing tax administrative structure that can put tax policies into practice and simplify the process.

2.1.3.4 Social transfers

With regards to tax credits, or social transfers, the social security legislation in Pakistan includes:

- The Workmen's compensation Act,1923
- The Sindh Maternity Benefit Act, 1929
- The Punjab Maternity benefit Act, 1943
- The West Pakistan Maternity Benefit Ordinance, 1958
- The Provincial Employees Social Security Ordinance, 1965
- The West Pakistan Industrial and Commercial Employment Ordinance, 1968
- Workers Shares in Companies profits, 1968
- Workers Welfare Fund, 1969
- Workers Children Education Scheme, 1972
- The Employees Old-Age Benefits Act, (EOBI) 1976

The employees old-age benefits scheme includes old age pension, survivor pension and old age grants. Apart from that, there are also private pension funds and social welfare schemes e.g. *zakat* or similar private social charity schemes.

2.2 Research design

2.2.1 Data and sampling procedure

The data have been collected by means of a survey. Households living in rural Pakistan were interviewed in the year 2008. Our dataset comprises all four provinces of Pakistan. To ensure representativeness, we decided to sample households in 10 districts (i.e. roughly one tenth of the total number of districts) across the country (stratified sampling). Based on the figures given in Table 2.1, we could come up with the following allocation, taking population proportions into account: four districts from Punjab, three from Sind, two from NWFP and the remaining district could be assigned to Baluchistan. The selected districts in Punjab are Attock, Layyah, Rahimyarkhan and Sahiwal; Badin, Mirpurkhas and Thatta in Sind; Dir and Malakand in NWFP and Kalat from Baluchistan. The selected districts were chosen for various reasons. First, these districts are geographically in a range that offered easy access without raising security concerns for the interviewers compared to further remote areas. Second, these districts provide a representative socio-economic picture of rural Pakistan. Due to the geographic scope of the districts, great care has been taken, where necessary, in sampling households from villages which are reasonably far away from major cities such as Lahore in Punjab, Karachi in Sind, Peshawar in NWFP and Quetta in Baluchistan. Two villages were chosen from each district.[5] Within these predefined strata, households have been selected randomly. Our target was to achieve a total of 30 responses per village, that is 60 households per district, yielding an overall sample size of $N = 600$.

Summing up, we have sampled a total of 240 households from Punjab, 180 households from Sind, 120 households from NWFP and 60 households from Baluchistan. However, to ensure a good representation for rural Pakistan, we assign weights to each household with respect to the district it belongs to as shown below in Table 2.2. Our results from the econometric analysis that follows are based on that weighting scheme.

[5]Table A.3 contains a list of the selected villages.

	Census 1998			Rural Survey 2008		
District	No. of households	Total Population	Rural Population	Sample Population	pweights $[(RP)'_j/(SP)_j]$	pweights-normalized $[(pw)_j/\Sigma(pw)_j]$
(j)	$(hh)'_j$	$(TP)'_j$	$(RP)'_j$	$(SP)_j$	$(pw)_j$	$(pw)*_j$
Attock	206678	1274935	1003843	266	3773.8459	0.10
Layyah	152050	1120951	976748	289	3379.7509	0.09
RahimyarKhan	3141053	17,743,645	2524471	246	10262.0772	0.27
Sahiwal	n.a	1843194	1541204	269	5729.3829	0.15
Badin	211354	1136044	949556	267	3556.3895	0.09
Mirpurkhas	148470	905935	605760	251	2413.3865	0.06
Thatta	220068	1113194	988455	259	3816.4286	0.10
Lower Dir	76531	717649	673314	241	2793.8340	0.07
Malakand	49330	452291	409112	234	1748.3419	0.05
Kalat	34410	237834	204040	215	949.0233	0.02
Total	-	11943080	9876503	2537	38422.4606	1.00

Table 2.2: Weighting scheme (using *pweights*) of sample households. The relevant weights are reported in the last column.

Given Pakistan's cultural and social background, it goes without saying that obtaining data from such a conservative society was a rather difficult and challenging task. Without the active support of my friends and family who helped me in data collection, conducting the survey would not have been possible.

In each of the selected villages we were guided by the locals to visit different households as we were quite new to some of the places and the villages. In addition, it was also very important for us to have some locals around during the interview process for two reasons: first, to gain the respondents' support and trust, and second, to help us in translating the questionnaire to the interviewees in their local languages.

The questions were asked from the head of the household in order to collect *first-hand* data on key variables such as the household's income and the sources of income, where all income figures have been measured in Pakistani Rupees (PKR).[6] Additional household characteristics have also been obtained such as gender, age, education, marital status, employment status, total family size (children and adults) as well as insights into the household's well-being including information on overall satisfaction with life and health in terms of relative and absolute poverty measures (see Appendix B for the sample questionnaire). As mentioned before, the questions were asked directly from the head (or the father) of a household, however, in some cases when the head was away from home or in the case of a widow/separated/divorced the questions were asked from the wife/mother inside her house. After each interview, sweets were given to the household as a thank-you gesture.

We have come across various constraints throughout the process of data collection. For example, interviewees were shy and felt insecure about revealing their personal information, especially regarding their family income and wealth position for tax evasion reasons. At times, the head of the interviewed household found it difficult to disclose any sort of information about their children and especially their daughters because of cultural and religious reasons. About 10% of the households refused to be surveyed. Keeping these constraints in mind, we have deliberately sought to keep the interviews brief as well as to stick to the necessary research requirements as closely as possible.

Furthermore, the cross-sectional nature of our data lends itself quite well to a close, detailed analysis of inequalities and well-being over a short time

[6]Note that all results that follow refer to monthly income figures.

period. By the study's very nature, however, our approach is not suitable for tackling matters on a broader national scale that are for instance associated with urban life. Also, given the scope of the study design, considering a dynamic perspective is hardly feasible. In the remainder, all results are based on our unique dataset, henceforth referred to as *Survey 2008*.

Chapter 3

Income inequality in rural Pakistan – sources and decompositions

Chapter Summary

This Chapter analyses income inequalities in rural Pakistan. Using micro data (Survey 2008) allows us to decompose income inequality according to its different sources. We calculate Gini coefficients and Theil indices both within and across provinces and districts. A partial correlation analysis extends our descriptive investigation to reveal the different impacts of the various income sources on overall income inequalities. The unique focus on rural areas and the more disaggregated (district level) approach permits more nuanced policy implications. We find that inequality between districts is higher than within districts. Non-farm and transfer income have the strongest impact on income inequality across districts both in economic and econometric terms. While the former source of income is inequality increasing, transfer income tends to reduce inequalities. Our analysis suggests important policy implications. It emphasises in particular the need for factor mobility to facilitate transfer income.

3.1 Introduction

All over the developing world policymakers are interested in devising new strategies for rebalancing skewed income distributions and reducing poverty. The choice of such strategies crucially hinges on an improved understanding of the sources of income inequality. Why do certain types of incomes go to particular groups of people? And what roles do variables such as land-ownership, migration and education play in improving income distribution and in lifting people out of poverty?

Using primary household survey data (see Chapter 2 for details), we identify different types of income to disentangle each sources' impact on overall income inequality in Pakistan's four provinces. Given that almost two thirds of Pakistan's population live in rural areas we confine ourselves to those households (Government of Pakistan, Statistics Division, 1998). Our dataset is wide-ranging, providing rather detailed information on different districts across the country. Other authors such as Adams and Alderman (1992) and Glewwe (1986) use panel data which covers up to three years, but are less detailed on the household level. We resort to a cross-sectional, yet well-designed, framework.

In comparison to similar studies conducted for countries in the region, our paper differs in two aspects. Regarding methodology, we decompose overall income inequality both by region and income sources. Furthermore, we exploit our rather detailed dataset fully, considering additional sources of income such as rental and livestock income. Our cross-sectional analysis suggests that while livestock income does not affect overall income inequality in rural Pakistan whatsoever, other types of income such as transfer, agricultural, nonfarm and rental contribute equally. A similar study has been conducted by Glewwe (1986) who finds that non-labour income (defined as the income derived from crop production by the landowners) is largely responsible for overall income inequality in rural areas.

This Chapter is structured as follows. Section 3.2 briefly discusses the related literature. In Section 3.3, we first identify the different income sources and then decompose them accordingly. We decompose inequality by income sources in Section 3.4. Section 3.5 provides a decomposition of inequality by regions. We conduct a correlation analysis between the different income sources in Section 3.6. Section 3.7 concludes and outlines potential policy implications.

3.2 Related literature

Adams (1994) uses three year panel data to analyse the impact of non-farm income on income inequality in rural Pakistan. He describes the importance of rural non-farm income for the poor by decomposing total rural income into the following five sources: non-farm, agricultural, livestock, rental and transfer income. The decomposition shows that non-farm income represents an inequality-decreasing source of income. The study then decomposes the sources of non-farm income. This analysis reveals that while non-farm unskilled labour income has an equalising effect on the income distribution, non-farm government income has a disequalising effect.

Ahmad (2000) uses micro data based on the *Household Integrated Economic Survey (HIES)* from 1992-93 to calculate the distribution of income in rural and urban areas of Pakistan. He finds that Gini coefficients show a more favourable distribution in rural areas compared to urban areas. The analysis moreover suggests that moving from household-level to individual data leads to a further improvement in the distribution of income.

In a related study, using the same HIES from 1992-93, Ahmad (2002) examines income inequality among various occupations in Pakistan such as (i) legislators/officials, (ii) teaching/health and other professionals, (iii) skilled construction, manufacturing, craft and related trades workers, and (iv) unskilled agricultural, fishing and related workers. Ahmad (2002) finds that within the given occupational groups in the four provinces of Pakistan, the highest level of inequality is observed among skilled workers (Gini coefficient of 0.299), followed by inequalities amongst the group of legislators/officials (Gini coefficient of 0.273). In contrast to that, the Gini coefficient among unskilled workers is 0.180 and is the lowest within the professional class (Gini coefficient of only 0.136). This may have been due to the fact that many of them were government employees and the wage structure was more equal in the government sector.

Anwar (2005) provides a series of Gini coefficients based on a consistent methodology using grouped household income data over 17 HIES conducted by the Pakistan Federal Bureau of Statistics (FBS) during 1963 to 2002.[1] The calculated Gini coefficients are generally higher in the urban than in the rural areas. Anwar (2005) conjectures that this was because of the urban labour force being more diversified in terms of skills and education and

[1]The household survey years are 1963-64, 1966-67, 1968-69, 1969-70, 1970-71, 1971-72, 1979, 1984-85, 1985-86, 1986-87, 1987-88, 1990-91, 1992-93, 1993-94, 1996-97, 1998-99 and 2001-02.

therefore the wage incomes being more unevenly distributed than in rural areas. Moreover, income from self-employment was more deviated in urban areas than in rural areas as urban self-employment ranged from wealthy businessmen to poor workers, whereas the bulk of the rural self-employed were a rather homogeneous group being mostly employed in informal sector enterprises.

Nugent and Walther (1982) use panel data in ungrouped (i.e. disaggregated) form to examine the sources of income inequality in India. The study is based on only three sources (agricultural, nonagricultural and transfer). The paper examines the dramatic decline in income inequality observed in rural India between 1968-69 and 1970-71. The authors describe the effects of changes in weather, technology and other factors on the distribution of incomes, which indirectly influences the labour market. Nugent and Walther (1982) find that periods of bad weather crowd out workers from the low-income group and increase inequality. Good weather, on the other hand, opens up more job opportunities for the workers from the low-income group and hence reduces imbalances in the income distribution.

Glewwe (1986)'s decomposition analysis is based on only two income sources (labour and non-labour) in Sri Lanka. It is assumed that all profits and non-monetary income are non-labor income. Specifically, non-monetary income is mainly agricultural produce consumed in the household. Additionally, incomes of those who are self-employed are counted as profits. The results suggest that labor income inequality accounts for slightly more than half of overall inequality in urban and estate sectors but not in the rural sector, where total inequality attributes largely to non-labor income and income from the sale of agricultural products (accounted for as profits). The conclusions drawn relate non-labor income inequality to an unequal distribution of land and capital in rural areas. Labor income inequality is explained by education and high wages that are paid to government employees in urban and estate sectors.

Adams and Alderman (1992) use a decomposition analysis to estimate sources of agricultural income inequality in rural Pakistan. They find that imbalances in land ownership are not the main drivers behind inequality in agricultural income. Instead, income from returns to labor and crop profits contribute the most. According to their analysis, policy makers concerned about inequality in rural Pakistan should pay attention to find ways to reduce the disparities between abilities, for instance by teaching agriculturalists managerial and technical skills.

Anwar et al. (2004) suggest that unequal landownership is one of the

most important causes of poverty and inequality in rural Pakistan as land is the principal asset in an agrarian economy. Numbers of landless households are substantially high in Pakistan. About 67% of households do not own any land. About 18% of the households own less than 5 acres of land and about 10% of the households own 5 to 12.5 acres of land, which merely provides a subsistence level of living. A very small proportion of households holds large farms. Strikingly, barely 1% of the households owns more than 35 acres of land suggesting a highly skewed landownership pattern (Anwar et al., 2004).

Overall, the various studies appear to highlight the importance of a meaningful land reform in order to tackle the poverty and inequality issues in rural Pakistan. Policies that stress labour income and advocate education and training as the main means for addressing inequality show only half of the picture. After all, how is inequality supposed to be improved through education, if access to quality education is dependent on asset and ownership inequality? Can a child coming from a landless household access a top school whose graduates enjoy high income and other privileges? In fact, this kind of purely education-focused policy may even trigger inequality in the first place. Therefore, any resulting policy implications appear only meaningful when related to land reforms. The experience of Taiwan and South Korea clearly shows that land reforms even at the early stage of development contribute to shared economic growth which does not add to inequality.

3.3 Income sources

3.3.1 Overall income shares

As laid out in the questionnaire (see Appendix B), individual households' total income may stem from up to five sources, y_i:

- **Transfer income (y_t)** includes internal and international remittances, government pensions and *zakat* (payments to the poor). Transfer income includes income earned from migration, both within and outside of Pakistan. Income earned from the first kind of migration is treated as internal remittances; income from the latter as external remittances.
- **Agricultural income (y_a)** includes net income (cash as well as in kind) from all crop production plus wage earnings from agricultural labour.

- **Nonfarm income** (y_{nf}) includes wage earnings from nonfarm labour including self employment, government and private-sector employment.
- **Rental income** (y_r) includes rents received from ownership of assets including land, machinery (tractors, threshers), buildings, and water.
- **Livestock income** (y_l) includes net returns from traded livestock (cattle, poultry) plus imputed values of home-consumed livestock.

The rather detailed data set allows for this differentiation between the various income types. Moreover, we can identify the impact of each of the different sources of income on income inequality at different levels of aggregation.

On a purely descriptive level, we can calculate each income source's share in total monthly household income data, denoted by S_i. According to Table 3.1, transfer income is the most important source of income, accounting for nearly one third of mean monthly household income. This seems to suggest that much of transfer income may come from migration to urban areas. Migration hence seems to have an important effect. In addition, the households may also tend to rely on help both from the government but especially from within their own families, neighbours and other peers. Since we are exclusively looking at rural backgrounds, agricultural income obviously is the major source of labour income. Nonfarm, livestock and rental income are of roughly equally (low) importance.

y_i	S_i
y_{nf}	18%
y_a	23%
y_t	30%
y_l	15%
y_r	14%

Table 3.1: Monthly income shares. $N = 600$ households.
Source: Survey 2008.

3.3.2 Distribution of sample households by income quintiles

The distribution of the sample households by income quintiles is shown in Table 3.2. Two points are worthwhile noticing: First, the shares of

transfer, agricultural and livestock income are relatively higher in lower income quintiles as compared to the higher income quintiles and may be considered as major sources of income for the households belonging to lower income quintiles. In contrast to that, the shares of rental and nonfarm income tend to be higher in higher income quintiles. Second, the share of livestock income is lowest in the monthly income of the households reported in the top income quintile, whereas the share of rental income is lowest in the bottom quintile. Overall, we may say that the distribution is actually very similar for the bottom 4 quintiles and the differences really come with the richest quintile.

Sample Distribution (%)	**Income Groups (in PKR)**	S_{nf}	S_t	S_a	S_r	S_l
Bottom quintile	1596-2526	16.4	31.9	23.9	9.0	18.8
2nd quintile	2550-3116	16.4	30.6	23.2	10.2	19.6
3rd quintile	3136-3359	13.2	33.1	23.6	10.3	19.8
4rth quintile	3365-4298	16.6	30.8	22.3	13.5	16.8
Top quintile	4371-6938	22.8	27.1	21.7	21.2	7.2

Table 3.2: Distribution of sample households by income quintiles.
Source: Survey 2008.
Note: All income shares are expressed in percentages.

3.3.3 Income shares by district

Table 3.3 breaks down total monthly household income y into the different shares S_i^j by the j districts surveyed. Compared with Table 3.1 which provides a more aggregate perspective, we see that the same ranking of shares applies on the district level as well.

According to Table 3.3, the share of transfer income is the highest in all the districts. Agricultural income and nonfarm income are other significant sources of income for Pakistan's rural population. Rental and livestock income contribute the least.

District	S_t	S_a	S_{nf}	S_r	S_l
Attock	26.94	22.00	26.11	18.85	6.10
Sahiwal	27.99	22.08	23.31	11.56	15.06
Layyah	28.50	22.65	20.33	10.43	18.09
Rahimyarkhan	29.56	23.30	18.33	10.49	18.32
Thatta	33.69	25.22	14.23	07.46	19.40
Badin	32.76	22.89	16.82	10.48	17.05
MirpurKhas	27.89	21.14	17.02	25.94	08.01
Malakand	29.38	21.84	12.66	16.10	20.02
Dir	34.01	24.05	12.06	08.84	21.04
Kalat	33.77	24.08	11.99	09.04	21.12

Table 3.3: Share of different income sources; by district.
Source: Survey 2008.
Note: All income shares are expressed in percentages.

3.4 Inequality decomposition by income sources

3.4.1 Decomposing overall income inequality

While calculating the shares of the different income types provides insights in the allocation of income, it does not make any statement about distributional aspects. The most common measure of income inequality is arguably the Gini coefficient G, ranging from 0 to 1, where a higher number indicates more inequality. In the extreme case of $G = 1$, the entire income goes to a single economic unit. The Gini coefficient allows for quick and easy comparison across countries, or in our case, provinces and districts. According to the World Bank (2005), Pakistan's overall Gini coefficient amounts to 0.33. More recent data suggests a downward trend.[2] In comparison to Pakistan's geographic neighbours Bangladesh and India, we find that the entire subcontinent is characterised by a similar degree of income inequality (World Bank, 2005).[3] Yet, numbers in developed economies such as the UK ($G = 0.36$) and the US ($G = 0.41$) are higher and, in contrast to Pakistan, seem to stagnate.[4]

[2]In its latest Human Development Report, the United Nations Development Programme (UNDP) provides a Gini coefficient of $G = 0.31$ for Pakistan (UNDP, 2007).

[3]Table A.2 in Appendix, gives an overview of poverty and income inequality across the neighbouring countries.

[4]The more recent study by UNDP (2007) reports identical coefficients for relevant benchmark countries such as the UK and the US. Figures for Bangladesh ($G = 0.33$) and India ($G = 0.37$), in comparison to Pakistan, seem to be on the rise which may be

Based on our sample data, we obtain the following Gini coefficient for rural Pakistan (rPk), denoted by G_{rPk}:

$$G_{rPk} = \sum_{i=1}^{5} w_i R_i G_i = 0.189. \tag{3.1}$$

Comparing G_{rPk} with Pakistan's overall Gini coefficient, G_{Pk}= 0.306 (CIA, 2008); we may say that rural inequality is lower than overall inequality, which is expected as we don't have the inequality coming from rural-urban differences. But cities (or the urban areas) themselves may have low inequality. Ahmad (2000) finds that income inequality in rural areas tends to be lower compared with urban areas. Furthermore, it appears that there is a positive relationship between the level of skilled labour and income inequality (Ahmad, 2002). The more equitable distribution of income on the countryside may hence be explained by the lower level of skilled labour there.

We need to decompose the Gini coefficient G to measure how much each particular income source contributes to overall income inequality in rural Pakistan (and also at the district and provincial level, accordingly). For that matter we break-down each household's total monthly income in the same n sources as before:

$$y = \sum_{i=1}^{n=5} y_i,$$

where y is the total monthly household income and y_i represents the income type i.

Decomposing the Gini coefficient rests on three elements: First, the weight of income source i, w_i, defined as

$$w_i = \frac{\mu_i}{\mu},$$

where μ_i represents the mean monthly income of source i and μ is the mean of total monthly household income. Second, the correlation ratio between income source and total income R_i defined by the following ratio of covariances

$$R_i = \frac{cov(y_i, r)}{cov(y_i, r_i)},$$

where r is the ranking of total income such that higher income receives a higher rank and r_i expresses the corresponding ranking of income sources.

attributed to higher and more persistent economic growth lately.

Both r and r_i follow a dense ranking from 1, 2, and so on.[5] The third part is the Gini coefficient associated with y_i given by

$$G_i = \frac{2}{\mu_i} cov(y_i, r_i).$$

The relative concentration coefficient is given by

$$g_i = R_i \frac{G_i}{G_{rPk}}.$$

This decomposition procedure further allows us to determine the individual contribution of each income source to overall income inequality. The contribution of income source i to overall income inequality in rural areas, c_i, is easily calculated by

$$c_i = w_i R_i \frac{G_i}{G_{rPk}} = w_i g_i, \tag{3.2}$$

where obviously all individual shares sum up to 1.

y_i	w_i	R_i	G_i	c_i
y_a	0.23	0.99	0.17	0.21
y_l	0.15	0.13	0.16	0.02
y_{nf}	0.18	0.88	0.31	0.26
y_r	0.14	0.97	0.37	0.27
y_t	0.30	0.95	0.17	0.26

Table 3.4: Contribution to overall income inequality by income source. Source: Survey 2008.

Table 3.4 shows that agricultural, nonfarm, rental and transfer income each account for about one quarter of total income inequality. Livestock income accounts for hardly any inequality at all. Looking at the correlation ratios (R_i) for agricultural, rental and transfer income, we can conclude that all these sources have a strong positive relationship with the total income rank, followed by nonfarm income. In contrast, the correlation ratio between livestock income and total income is small and positive. This is the result of a low covariance between livestock income and corresponding total income rank.

[5]In dense rankings, items that compare equally receive the same ranking number, and the next item(s) receive the immediately following ranking number. Equivalently, each item's ranking number is 1 plus the number of items ranked above it that are distinct with respect to the ranking order.

According to the individual Gini coefficients G_i, rental and nonfarm incomes are the most unequally distributed income sources. Agricultural and transfer income are relatively equally distributed. Inequality in livestock income is relatively small. A similar result is obtained by Ahmad (2002).

3.4.2 Decomposition of overall income inequality by province

Disaggregating the data further, we calculate in the following the Gini coefficients for the $k = 4$ provinces using the procedure outlined above. We see that the level of income inequality in Sind represents closest the overall situation in rural Pakistan. Yet, for the other provinces there is a considerable degree of dispersion. While the plain Gini index suggest hardly any income inequality in the relatively poor provinces of NWFP and Baluchistan, Punjab's coefficient is quite high considering that we only look at the rural sector.

Given our results in Table 3.5, we observe that in Punjab and Sind; nonfarm, agricultural, transfer and rental each contribute for about one fourth of overall income inequality. Livestock income hardly contribute to the overall income inequality, in both provinces. This suggests an important policy conclusion; policy makers who are concerned with income inequality in Punjab and Sind provinces, would be well-advised to pay more attention to livestock- As it has the smallest share of total income and hardly contribute to overall income inequality in the two provinces. However in NWFP and Baluchistan the situation is quite opposite. For instance, in NWFP the rental income has the largest share (that is around 50%) in overall income inequality; followed by livestock, agricultural and nonfarm income. On the other hand, the transfer income represents an inequality decreasing source of income. The inequality in rental, livestock and agricultural income may be attributed to unequal landownership in NWFP. On the other hand, in Baluchistan rental income makes the smallest contribution to the overall income inequality- While livestock and transfer income make the largest share (that is around 40% each) and nonfarm income accounts for 20% of the overall income inequality of the province. Surprisingly, agricultural income appears as inequality decreasing source of income. Which may suggest that bringing more land under cultivation may improve the distribution of income in Baluchistan.

	Punjab	**Sind**	**NWFP**	**Baluchistan**
Gini index	0.24	0.18	0.04	0.02
Income source share (w_i)				
Nonfarm	0.232	0.164	0.123	0.120
Agricultural	0.223	0.226	0.229	0.241
Transfer	0.279	0.308	0.316	0.338
Livestock	0.123	0.136	0.205	0.211
Rental	0.143	0.166	0.127	0.090
Concentration index (g_i)				
Nonfarm	0.304	0.199	0.043	0.029
Agricultural	0.230	0.136	0.025	-0.005
Transfer	0.217	0.145	-0.002	0.022
Livestock	0.016	0.011	0.046	0.035
Rental	0.368	0.394	0.182	0.011
Relative contribution (c_i)				
Nonfarm	0.298	0.187	0.124	0.190
Agricultural	0.217	0.175	0.134	-0.070
Transfer	0.255	0.255	-0.017	0.409
Livestock	0.008	0.009	0.221	0.414
Rental	0.222	0.374	0.538	0.056

Table 3.5: Inter-provincial rural household income disparities.
Source: Survey 2008.

3.4.3 Decomposition of overall income inequality by district

We cannot dismiss a priori the possibility of considerable disparities at district/tehsil levels compared to a pure province perspective. Although evidence for spatial differences at a disaggregated level is rather scanty, this may be an issue for the data at hand. This impression is endorsed by Pasha and Hasan (1982) who observe that statements about inter-provincial levels of development tend to hide major intra-provincial disparities.[6] We therefore calculate Gini coefficients for each district to clarify the contribution of each income source to overall income inequality within districts. This also allows us to draw important conclusions about the provinces. Indeed, according to our results in Table 3.5 income inequality in Punjab is much higher as compared to NWFP, but still both provinces experience similar

[6]The mean monthly income fluctuates quite strongly across the 10 districts (standard deviation of 1238.14) which could conceivably affect any decomposition effort that is based on monthly income data.

levels of inequality in some of their districts such as Attock and Layyah in Punjab and Malakand in NWFP. Although the aforementioned districts share the same overall inequality, the income sources contribute differently to overall inequality within each district (see Table 3.6).

According to Table 3.6, Sahiwal and Thatta have relatively higher inequality in terms of overall monthly household income. In Sahiwal, nonfarm income is the most unequally distributed source of income and contributes the most to overall income inequality in rural Pakistan. This may be due to a rather low literacy rate (only 40%) of the population which is the main reason of highest income inequality in the nonfarm sector. Rental income contributes the least to overall income inequality. Similarly the table indicates that in Thatta, agriculture is the most unequal source of income and contributes to 38.4% to overall income inequality, which is highest among the other income sources. According to our sample survey, 70% of the population in Thatta are small farmers (having land between one or two acres) and their livelihood continues to revolve around farm activities like agriculture and livestock. Some of these farmers are making best use of their land, while others are still using old and primitive techniques for production (as revealed by personal communication with the respondents), which may be one of the main reasons for higher income inequality in the farm sector.

The field study also provides a socio-economic profile of each district. According to the survey, the average monthly household income of Rahimyar Khan district is the lowest among all the districts analysed. Livestock income is the most unequal source of income in Rahimyar Khan as well as in Kalat district and contributes most strongly to an increasing overall income inequality in both districts (Table 3.6).

	Attock	Sahiwal	Layyah	Rahimyar-Khan	Badin	Thatta	Mirpur-Khas	Dir (lower)	Mala-Kand	Kalat
Overall Gini coefficient of monthly household income (G)	0.049	0.081	0.048	0.053	0.016	0.075	0.033	0.024	0.043	0.020
Source income weight(w_i)										
Nonfarm	0.261	0.233	0.203	0.183	0.142	0.168	0.170	0.127	0.121	0.120
Agricultural	0.220	0.221	0.227	0.233	0.252	0.229	0.211	0.218	0.241	0.241
Transfer	0.269	0.279	0.285	0.296	0.337	0.328	0.279	0.294	0.340	0.338
Livestock	0.061	0.151	0.181	0.183	0.194	0.170	0.080	0.200	0.210	0.211
Rental	0.188	0.116	0.104	0.105	0.080	0.105	0.260	0.161	0.088	0.090
Relative concentration coefficients of income sources(g_i)										
Nonfarm	0.605	1.916	0.415	1.131	2.129	0.851	0.055	-1.623	1.246	1.595
Agricultural	1.262	1.222	1.246	0.773	-1.738	1.524	-0.051	1.586	1.040	-0.294
Transfer	0.987	0.704	0.979	0.727	3.807	0.655	1.982	1.604	-0.328	1.214
Livestock	0.349	0.423	0.660	1.546	0.530	1.266	1.999	1.371	1.553	1.964
Rental	1.469	0.179	2.242	1.113	-3.034	0.340	1.091	-0.254	2.517	0.633
Source income contribution to overall income inequality(c_i)										
Nonfarm	0.158	0.447	0.084	0.207	0.358	0.121	0.009	-0.196	0.158	0.191
Agricultural	0.278	0.270	0.283	0.180	-0.398	0.384	-0.011	0.381	0.227	-0.071
Transfer	0.265	0.197	0.280	0.215	1.249	0.221	0.553	0.545	-0.096	0.410
Livestock	0.021	0.064	0.120	0.283	0.090	0.246	0.160	0.288	0.310	0.414
Rental	0.278	0.022	0.233	0.117	-0.319	0.026	0.284	-0.023	0.405	0.057

Table 3.6: Decomposition of overall income inequality by district using the Gini coefficient.
The more unequally distributed the income source, the higher the coefficient of concentration. The concentration coefficient shows how much a given income source "pushes" up the overall income inequality. The relative concentration coefficient is given by $g_i = R_i \frac{G_i}{G_{rPk}}$.The "negative" sign indicates that the source income decreases as the total income increases and thus "pushes" down the overall income inequality . The contribution of each income source (c_i) to overall income inequality is determined by $w_i g_i$, where $w_i = \frac{\mu_i}{\mu}$ and $g_i = R_i \frac{G_i}{G_{rPk}}$.
Source: Survey 2008.

Similarly Table 3.6 shows that in Attock and Layyah, agricultural income has maximum contribution to overall income inequality due to the uneven distribution of land in favour of the rich, while livestock income contributes the least to income inequality in the two districts. Table 3.6 indicates that transfer income is the most unequally distributed income source in Badin, Mirpurkhas and Dir and is mainly responsible for pushing up overall income inequality. Furthermore inequality in transfer income is the result of internal and external remittances (according to personal communication with the respondents). Evidently, the uneven land distribution in rural Pakistan forces the poor to seek the bulk of their livelihood by migrating within Pakistan or even by emigrating away from the country.

The study reveals that 60% of the households in Malakand are landless which may explain why rental income is the most unequally distributed income source. Rental income contributes with 40.5% to overall income inequality in the district which is the highest contribution as compared to the other income sources (see Table 3.6).

3.5 Inequality decomposition by regions

3.5.1 Decomposition of income inequality within and between provinces

Although the Gini index provides a transparent and easy way of comparing income inequality across regions, it is flawed in that it is not perfectly decomposable in some situations (Shorrocks, 1982). One should therefore be careful in interpreting this index in empirical studies. Other inequality measures seem more suitable for inequality decompositions. Shorrocks (1980) derives an entire class of measures which are additively decomposable under relatively weak restrictions on the form of the index. The subclass of mean independent measures turns out to be a single parameter family which involves the square of the coefficient of variation and two entropy formulae proposed by Theil.

In the context of additive decomposability, the generalised entropy (GE) class of inequality indices is a good alternative to the Gini index. Unlike the Gini coefficient, the members of this class are perfectly decomposable without a residual term. The most common index within the GE class is

the Theil Index (I) as introduced in Equation 3.3:

$$I = \frac{1}{N} \sum^{N} \frac{y}{\bar{y}} ln \left(\frac{y}{\bar{y}} \right) = 0.055, \tag{3.3}$$

where, as above, N is the sample size, y is the individual household's monthly income (summable over N) and $\bar{y}$ the corresponding mean. Given the design of our survey, the sample is readily partitioned to differentiate between inequality within (I_w) and between (I_b) regions where it obviously holds that $I = I_b + I_w$.

Using this methodology we can first decompose the Theil Index on a more aggregate province level as follows:

$$\begin{aligned} I &= \underbrace{\sum_{k=1}^{K=4} \frac{N_k}{N} \left(\frac{\bar{y}_k}{\bar{y}} \right) ln \left(\frac{\bar{y}_k}{\bar{y}} \right)}_{Between} + \underbrace{\sum_{k=1}^{K=4} \left(\frac{N_k}{N} \frac{\bar{y}_k}{\bar{y}} \right) I_k}_{Within} \\ &= 0.0016 + 0.0534 = 0.055. \end{aligned} \tag{3.4}$$

The results of Equation 3.4 suggest that inequality on this level of aggregation is almost exclusively driven *within* rather than between the provinces.

3.5.2 Decomposition of income inequality within and between districts

We can decompose the Theil Index for the district level in a similar fashion. Decomposing income inequality within and between districts helps to further disentangle total inequality in the given sample.

On the more disaggregated level we have $J = 10$ districts. The Theil Index decomposition hence takes the following form:

$$\begin{aligned} I &= \underbrace{\sum_{j=1}^{J=10} \frac{N_j}{N} \left(\frac{\bar{y}_j}{\bar{y}} \right) ln \left(\frac{\bar{y}_j}{\bar{y}} \right)}_{Between} + \underbrace{\sum_{j=1}^{J=10} \left(\frac{N_j}{N} \frac{\bar{y}_j}{\bar{y}} \right) I_j}_{Within} \\ &= 0.0535 + 0.0013 = 0.055, \end{aligned} \tag{3.5}$$

where N_j is the number of households in district j, $\bar{y}_j$ is mean monthly household income in district j and I_j is the resulting Theil Index in district j.

Results are summarised in Table 3.7. We find that between-district inequality is greater than within-district inequality. This may seem at odds with the results for the province level, where exactly the opposite effect

	I_b	I_w	I
Overall	–	–	0.055
Provinces	0.002	0.053	0.055
Districts	0.054	0.001	0.055

Table 3.7: Regional comparison of Theil indices

occurs. Yet, as Table 3.7 illustrates, if the entire degree of inequality is determined within the provinces on a more macro level, this necessarily implies that the "whole action" occurs between the districts which are *nested within* the provinces. Therefore, while overall measures are obviously the same (up to some rounding margin) regardless of the level of aggregation, it appears much more meaningful to look at the district level rather than comparing between provinces as commonly done. We may also suggest that on the whole, all the rural Pakistan gives the same picture irrespective of which province is chosen as its sample representation.

To substantiate this point further, we can easily determine how much total inequality is explained by between-district inequality by looking at the following ratio R_b:

$$\begin{aligned} R_b &= \frac{I_b}{I} \\ &= \frac{0.0535}{0.0550} = 0.97. \end{aligned} \tag{3.6}$$

According to Equation 3.6, 97% of total inequality is explained by inequality between the districts. The remaining 3% is explained by inequality within the districts. In the remainder we therefore exclusively consider inequality on a district level.

3.6 Partial correlation analysis

In this Section, we consider partial correlation coefficients to measure the statistical relationship between income inequality (G_j) and the shares of different income sources ($S_{i,j}$) across the districts, where all variables are expressed in terms of *percentages* and G_j represents the Gini coefficient as measure of overall income inequality in the j^{th} district and the share $S_{i,j}$ is defined as the ratio of average income of type i in district j over total

average income in district j:

$$\begin{aligned} S_{a,j} &= \frac{\bar{y}_{a,j}}{\bar{y}_j} \times 100 \\ S_{t,j} &= \frac{\bar{y}_{t,j}}{\bar{y}_j} \times 100 \\ S_{nf,j} &= \frac{\bar{y}_{nf,j}}{\bar{y}_j} \times 100 \\ S_{r,j} &= \frac{\bar{y}_{r,j}}{\bar{y}_j} \times 100 \\ S_{l,j} &= \frac{\bar{y}_{l,j}}{\bar{y}_j} \times 100 \end{aligned}$$

Similar to a regression analysis, partial correlation seeks to measure a relationship between dependent and independent variable, whilst eliminating potential effects of a third variable. The partial correlation coefficients measure in this case the degree of statistical association between income source and district-wide Gini coefficient, where the latter one is considered as the dependent variable. Results are shown in Tables 3.8. Columns two and three report the partial correlation coefficient and the corresponding significance level, respectively. We find a strong positive and statistically significant correlation of 0.74 between the share of nonfarm income and income inequality across the districts. Thus, nonfarm income appears as inequality increasing source of income. In contrast to this, we detect a similarly strong (and also statistically significant) but negative correlation (-0.74) between the share of transfer income and income inequality. Unlike nonfarm income, transfer income is a source of income that is capable of reducing income inequalities in rural Pakistan.

Partial Correlation of G_j with		
Variable	Corr.	Sig.
$S_{a,j}$	0.6494	0.114
$S_{t,j}$	-0.7354*	0.060
$S_{nf,j}$	0.7354*	0.060
$S_{r,j}$	0.6291	0.130
$S_{l,j}$	0.6517	0.113

Table 3.8: Correlation Analysis

3.7 Conclusion and Policy implications

This Chapter analysed the impact of various sources of income on income inequality in Pakistan. Unlike other studies, we exclusively focus on the rural sector. Our unique survey dataset allows for detailed comparisons and decompositions at different levels of aggregation. The descriptive analysis suggests that it is much more meaningful to analyse income inequalities in rural Pakistan at a district rather than province level. We find that transfer income is the most important source of total monthly household income, accounting for almost 30%. Shares of rental and livestock income are the lowest in the sample. Shares of transfer, agricultural and livestock income are relatively higher in lower income quintiles as compared to higher income quintiles. Shares of rental and nonfarm income tend to be higher in higher income quintiles.

We consider both Gini coefficients and Theil indices to measure income inequality and the corresponding contributions of the various sources of income. Agricultural, nonfarm, rental and transfer income each contribute to income inequality in the sample area to a similar degree. Livestock income hardly affects the results. Decomposing overall income inequality within and between districts implies that inequality between the districts is greater than within the districts. The Theil index decomposition suggests that almost the entire degree of inequality (97%) can be explained by inequality between the districts.

The correlation analysis confirms the conjectured importance of transfer income in reducing income inequalities. While the derived coefficient for transfer income is negative, nonfarm income is found to be an inequality-increasing source of income. Considering both effects together appears to neutralise each other's role in driving income inequalities.

With regards to policy implications, given that agricultural income and rental income each contribute one fourth to overall income inequality in rural Pakistan, the policies should aim first to correct both of these sources of inequality. It is well known that addressing these inequality sources requires addressing asset inequality. Land is the main source of agricultural and rental income, and Pakistan is characterised by high land ownership inequality as identified by Anwar et al. (2004) and one of the main causes of poverty and inequality in rural Pakistan. Therefore, the policy makers should seriously consider land reforms in rural Pakistan in order to enhance pro-poor growth.

Secondly, as an inequality-decreasing source of income, policy makers

should also put more emphasis on transfer income in fighting inequalities across the districts. They should try to foster its flows both within Pakistan and from abroad but particularly try to increase factor mobility within the country's urban sector.

In order to reduce spatial imbalances, policy makers should try to provide stronger technical education for the rural (and often poor) unskilled labour force so as to increase mobility and to improve job opportunities within Pakistan and abroad. To raise the overall income level, policy makers should take steps to help poorer households send migrants abroad. One supporting measure that may be considered in that regard is the establishment of "rural migration centers" to process visas, work contracts and loan arrangements for prospective external migrants.

Chapter 4

Well-being and poverty in rural Pakistan: a subjective approach

Chapter Summary

This Chapter uses our dataset (Survey 2008) to estimate well-being functions in rural Pakistan using regressions with categorical variables. We investigate the impact of socio-demographic factors on life satisfaction with particular emphasis on subjective well-being measurement to evaluate poverty and its different components. We present a happiness model which is general enough to use different potential measures, highlighting their similarities and differences. Contrary to much of the literature, we find effects that are positive and of considerable economic magnitude. In particular the number of children seems to matter for the household's life satisfaction. We show, however, that all effects crucially hinge on individual household-specific characteristics. This Chapter's main contributions are as follows. First, we link the emerging field of happiness economics with development studies. Second, we intend to challenge the view that poverty is best understood from a more macro-level without properly accounting for individuals' own valuation of their well-being.

4.1 Introduction

The analysis of life satisfaction is a relatively new but rapidly emerging topic in economics. While much of the literature surveys evidence for developed

countries, little economic research has been carried out thus far for the developing world – notable exceptions are Kingdon and Knight (2006) on South Africa and Knight et al. (2007) on rural China. Using our household survey data (2008) for rural Pakistan, we investigate *subjective well-being poverty*. By that, we refer to poverty evaluated using happiness functions to measure the household's life satisfaction in subjective terms. Subjective well-being refers in this context to the life satisfaction as declared by the household. Poverty measurement is stated here exclusively in terms of households' views without demarcating a standard poverty threshold.

This methodological twist obviously raises the question whether subjective well-being can be used as measure of poverty. Conventional poverty measures are typically defined by minimum income or consumption levels. We instead intend to explore the subjective well-being approach to poverty. We do so because in a liberal and democratic spirit, we place a value on the individuals' own valuation of their welfare such that subjective well-being is based on value judgment, which underlies much of the personal judgment of the respondents about what we call *well-being poverty*.

Elaborating on Kingdon and Knight (2006), we introduce and compare different measures of well-being poverty based on self-assessments. Unlike them, however, we do not only resort to the holistic (and thus indirect) concept of overall happiness to measure well-being poverty but also introduce more direct measures for poverty that focus on monetary terms; for instance, income satisfaction and satisfaction with expenditures.[1] The economic analysis of psychological aspects such as happiness has increasingly received attention in the recent literature.[2] There are several studies which try to establish a relationship between happiness and different socioeconomic variables such as age, gender, income, employment, marital status, etc. to investigate the degree of households' satisfaction with their status quo.[3]

Several studies on the economics of happiness add the number of children to the list of explanatory variables in a so-called happiness equation as done for instance by Angeles (2009), Blanchflower (2008) and Clark et al. (2008). However, the empirical evidence is inconclusive. While some authors (Tella et al. (2003); Alesina et al. (2004)) find a negative or, respectively, no effect

[1]The convention in the happiness literature is to define happiness as the overall satisfaction with the socio-economic status quo. We elaborate on this idea in Section 4.2.

[2]We shall use the terms *happiness* and *life satisfaction* interchangeably in the remainder.

[3]See for example Blanchflower (2008) and the references therein.

(Clark (2006)), others (Stutzer and Frey (2006)) detect a positive effect of having children at home on overall household's happiness.

None of these studies, however, tries to relate the impact of children on happiness to the specific characteristics of the individuals such as gender, age, marital status, income and education. Only few papers account for the individual characteristics of the households. For instance, Frey and Stutzer (2000) using Swiss household survey data of 1992, find that having children has no effect on the happiness of married couples but a sizable (and negative) impact on single parents.

The purpose of this Chapter is to link both overall satisfaction (capturing both monetary and non-monetary factors) and financial satisfaction with demographic as well as socio-economic variables (which we may summarise as social well-being).[4] In particular, we want to investigate whether concepts of the economics of happiness can indeed be used in development economics to employ the subjective well-being approach to poverty as suggested by Kingdon and Knight (2006). We shall do so by estimating a comprehensive model for rural Pakistan, using highly correlated but alternative poverty measures. Comparing results, we shall argue that while subjective approaches indeed offer a viable alternative to conventional measurement techniques in development economics, distinguishing between direct and indirect measures of poverty may matter and should always be considered to ensure robustness, given the self-assessment nature of such a metric.

To our knowledge there is no comparable adaption of the economics of happiness to exclusively rural areas. We therefore believe that this Chapter fills an important gap in the literature and may well serve as blueprint for the analysis of other developing countries with similar demographic features both from a time series and cross-sectional perspective.

This Chapter is structured as follows. Section 4.2 lays the necessary ground for the further analysis. We develop the notion of life satisfaction in terms of overall satisfaction and financial satisfaction and also present some stylised facts. Section 4.3 investigates the correlation between the alternative measures of subjective well-being poverty. It furthermore provides additional descriptive statistics for the poverty measures both from an indirect and a direct angle. Section 4.4 analyses the different measurement approaches applied to data on rural Pakistan, accounting particularly for its demographic features. In so doing, we investigate the link between children and household's special characteristics such as family type, age, income and

[4]Financial satisfaction is measured by reported satisfaction with income and expenditure.

education using our baseline model. We moreover investigate alternative poverty measurement approaches in Section 4.5. Section 4.6 concludes.

4.2 Some stylised facts and basic concepts

There are numerous conventional measures of poverty in monetary terms. One may construct a so-called poverty line and then measure the distance of a household's income from a certain reference threshold, typically defined in a particular social context. The World Bank, on the other hand, defines poverty in absolute terms suggesting that any income in terms of Purchasing Power Parity (PPP) below a certain subsistence level classifies households as being "poor". Rather than considering income, one may also categorise poverty by the uses of income, particularly focusing on consumption. Either way, the conventional poverty metric in the development studies literature suggests an easily quantifiable way of assessing a households' economic situation. We believe, however, that the issue of poverty has more dimensions to it: what really matters is not how households may be classified in monetary or purchasing power terms but rather the household's self-reported degree of well-being.

The measurement of subjective well-being poverty is by its very nature closely related to the assessment of happiness as also brought forward by Kingdon and Knight (2006). Well-being is typically measured by means of an ordinal scale, where a higher value indicates a higher level of individual satisfaction (Blanchflower and Oswald, 2004). Studies on happiness are usually based on micro data and findings by and large seem to be materially robust regardless of whether estimation is done in an ordered logit model or by employing Ordinary Least Squares (OLS) regressions with categorical dependent variables.

Subjective well-being poverty can be measured along two dimensions. First, one may simply resort to the more holistic notion of happiness which encompasses both monetary and non-monetary, that is socio-economic, factors. Ceteris paribus (i.e. keeping everything else constant), higher income (and hence less poverty in the conventional sense) should induce households to feel happier; although effects are likely to be disproportional and at a diminishing rate (Frey and Stutzer, 2002b). Analysing happiness may thus be considered an indirect method of assessing poverty. This approach is flexible enough to also measure poverty in more broader terms for instance by considering *capabilities poverty* – see Sen (1983) and Sen (1993) for fur-

ther details.[5] Second, one may take a more direct route and assess the household's subjective well-being poverty directly in financial terms – such that it simply asks how satisfied households feel about their incomes and expenditures.

To appreciate the subsequent estimation results in a broader context, it may be helpful to first discuss some stylised facts from the happiness economics literature. Based on US and European panel data, those may be summarised as follows (Blanchflower, 2008):

- well-being depends positively on these controls:
  - being female
  - married couples
  - age (U-shaped behaviour!)
  - level of education
  - active religious involvement
  - level of health
  - level of income
  - regular sexual engagement
  - monogamy
  - being childless
- well-being is decreasing among people with the following characteristics:
  - newly divorced (or separated)
  - adults in their mid to late 40s
  - unemployed
  - immigrants and minorities
  - commuters
  - people with poor health (e.g. high blood pressure)

[5]Sen introduced the capabilities approach to well-being and poverty. He suggested that the capabilities approach to a person's advantage is concerned with evaluating it in terms of his or her ability to achieve various valuable functions as part of living. Sen claimed that absolute deprivation in terms of a person's capabilities relates to relative deprivation in terms of commodities, income and resources. Thus, happiness is more related to psychological rather than material deprivation.

- less educated
- poor
- sexually inactive
- parents (having children)

As this list reveals, both macroeconomic factors but also subjective assessments play a role in the perception of happiness. To take an example, the general unemployment rate has a depressing effect, suggesting that a higher risk of becoming laid off and the associated economic uncertainty reduce happiness. At the same time, subjective well-being is also influenced by several factors that are non-economic such as age, sex, marital status, health status, education, social capital, religion, as well as social and political institutions (Helliwell, 2002).

Psychologists and sociologists thus seem to rightly focus on the possible influence of personality-related factors (such as optimism, self-esteem and perceived personal control) in conjunction with socio-demographic factors (e.g. gender, age, education, health, family size, income, marital status, employment status etc.) when studying why people are happy or unhappy.

As Frey and Stutzer (2002a) argue, happiness from an economist's perspective is best measured in terms of health and wealth. Age is an important third variable in that it is likely to determine both these factors' impacts on happiness. However, the role of age is not so straightforward to assess for several reasons. The notion of well-being, after all, may change its connotation with varying age depending on the level of financial and physical well-being.

Life satisfaction, as the list of stylised facts reveals, may be best thought of as some "umbrella concept" capturing various aspects of a person's life, including both social and financial satisfaction. Given that in the developing world poverty generally encompasses all these aspects, we feel it is reasonable to use those insights from this strand of the literature to construct an alternative, more subjective poverty metric which more closely reflects the specific socio-economic context.

4.3 Descriptive statistics

The relevant endogenous variables are all constructed by means of an ordinal scale. The measure of overall satisfaction is based on the following question: *"How satisfied are you with your current socio-economic status?"*. Answers

were recorded on a numerical scale ranging from 1 to 4, where 1 is coded as *"Not at all satisfied"*, 2 as *"Less than satisfied"*, 3 as *"Rather satisfied"* and 4 as *"Fully satisfied"* to capture overall subjective well-being poverty.

The more direct approach to poverty (in terms of income and expenditure exclusively) was constructed using the same scale to ensure comparability. We distinguish between two alternatives: satisfaction with income and satisfaction with expenditure; responses for each of the two alternatives were described according to the the 1-4 scale, such that 1 is recorded as *"Not at all satisfied"*, 2 as *"Less than satisfied"*, 3 as *"Rather satisfied"* and 4 as *"Fully satisfied"*. These measures refer to subjective well-being poverty in its monetary terms.

Table 4.1 reports some summary statistics for each of the subjective well-being poverty measures mentioned.[6]

	Subjective well-being poverty measure in terms of		
	overall satisfaction with *socio-economic status (1-4)*	**satisfaction with *income (1-4)***	**satisfaction with *expenditure (1-4)***
Mean	2.11	2.40	2.37
Standard deviation	1.41	0.61	0.58
Frequency of value:			
4	33.67%	4.50%	2.33%
3	5.50%	30.50%	32.67%
2	4.00%	63.33%	62.50%
1	56.83%	1.67%	2.50%

Table 4.1: Summary statistics of subjective well-being poverty measures. Source: Survey 2008.

The distribution of the satisfaction variables in terms of income and expenditure are rather similar with the first two moments of both measures being nearly identical. We see that only few people report very high or very low values of the satisfaction index. With regards to satisfaction with

[6]Table C.1 reports additional background data on the given poverty measures. The Table consistently ranks Punjab as being the most prosperous province in subjective terms independent of the measure used. We may consider overall satisfaction an encompassing concept which not only covers monetary aspects but also considers other socio-demographic factors and may thus be treated as the *most comprehensive* approach to estimate subjective well-being poverty.

current socio-economic status (i.e. overall satisfaction), we notice that the answers are more dispersed. Most of the responses lie between the two extremes: 56.83% of the respondents are *"not at all satisfied"* and one third of the respondents are on the other extreme of the scale.

	Overall satisfaction with socio-economic status (1-4)	**Satisfaction with income** (1-4)	**Satisfaction with expenditure** (1-4)
Overall satisfaction with socio-economic status	1		
Satisfaction with income	0.81***	1	
Satisfaction with expenditure	0.81***	0.94***	1

Table 4.2: Correlation matrix.
$N = 600$ households. *, **, *** indicates significance level of 10%, 5% and 1%, respectively.

Given that the subjective poverty measures in terms of income and expenditure are not only based on the same scale but also seem to point in a similar direction, it may be worthwhile to further investigate their potential interrelationships. Bruni and Porta (2007), after all, suggest the presence of such linkages in that they argue that certain approaches in the happiness literature could not be dealt with adequately without employing insights from economics, psychology, sociology and philosophy. Table 4.2 reports correlation coefficients between the subjective well-being poverty variables. Indeed, the three measures we consider are strongly and positively correlated. We see that a high degree of income satisfaction goes hand in hand with high expenditure satisfaction (correlation of 0.94). Both measures correlate with the overall satisfaction level essentially in the same way (correlations of 0.81, each).[7] We therefore consider both income and expenditure approaches as alternatives for measuring subjective poverty directly in monetary terms.

[7]The correlation matrix clearly supports Easterlin's micro approach towards happi-

4.4 Determinants of overall satisfaction

4.4.1 The model

The model we use to evaluate subjective well-being poverty is a straightforward application from the happiness literature. As the findings summarised in Section 4.2 suggest, happiness is best thought of as a function of various factors. We follow the common approach and estimate a happiness function for rural Pakistan of the following form:

$$\begin{aligned} happiness = \beta_0 + \boldsymbol{\beta_1}(sex) + \beta_2(age) + \beta_3(age)^2 + \beta_4(educ) + \\ \beta_5(numberofchildren) + \boldsymbol{\beta_6}(unemployment) + \\ \beta_7 ln(income) + \beta_8 ln(relativeincome) + \\ \boldsymbol{\beta_9}(maritalstatus) + \boldsymbol{\beta_{10}}(health) + \boldsymbol{\beta_{11}}(region) + \varepsilon. \end{aligned} \tag{4.1}$$

The regression model (4.1) implies that happiness is not simply a binary case but is measured in terms of the ordered categories (1-4) introduced above. We employ a standard ordered probit (oprobit) model that is widely used to analyse discrete data of this variety. Our framework is based on an underlying latent model with single index function and constant thresholds. The control variables included are sex, age, education, number of children, employment status, the household's monthly nominal income (both in absolute and relative terms) expressed in natural logs, marital status, overall family's health position and regional allocation of household i, where the corresponding $\boldsymbol{\beta}$ represents the vector of coefficients of the dummy variables.

We moreover restructure model (4.1) by imposing dummies on the respective number of children per household instead of using the actual number of children per household:

$$\begin{aligned} happiness = \beta_0 + \boldsymbol{\beta_1}(sex) + \beta_2(age) + \beta_3(age)^2 + \beta_4(educ) + \\ \boldsymbol{\beta_5}(numberofchildren_{dummy}) + \boldsymbol{\beta_6}(unemployment) + \\ \beta_7 ln(income) + \beta_8 ln(relativeincome) + \\ \boldsymbol{\beta_9}(maritalstatus) + \boldsymbol{\beta_{10}}(health) + \boldsymbol{\beta_{11}}(region) + \varepsilon. \end{aligned} \tag{4.2}$$

ness, which confirms the positive correlation between individual income and individual measures of subjective well-being. The so-called Easterlin Paradox suggests that within a society rich people tend to be happier than poor people, whereas on a more macro level, relatively rich societies tend not to be happier (or not by much) than relatively poor societies (Easterlin, 2001).

Potential gender differences are captured by means of a dummy, where *sex* is 1 if the respondent is male and 0 otherwise. Similarly, employment and marital status take the value 1 when the head of the household is unemployed or living as a couple and 0 otherwise. Age effects, as usual, are allowed to be non-linear. All the remaining variables, except for income, are dummy variables. We therefore use matrix notation to remind ourselves that these coefficients are really expressed in terms of vectors relating to different household groupings. More specifically, the impact of the number of children on happiness in (4.2) is captured by $\boldsymbol{\beta_5}$.[8] We created six dummies in total to separate the cases when one, two, three, four, five, six, seven or more children are present in a household.[9] We chose households with seven (or more) number of children as the reference group. Note that children are here defined as individuals aged less than 16 years who live with their parents.

We moreover constructed a health index to evaluate the general health status, where a higher value refers to a higher level of health. The index is based on the following response from the household's head: *"During the last 12 months, how many times has someone in your household visited a doctor?"* The answer options: *none*, *once*, *twice*, *three times* and *four times or more* were then mapped correspondingly onto dummy variables which may be translated into the health index as follows: 4, 3, 2, 1 and 0 for excellent, good, fair, poor and very poor health status, respectively. The excluded category is being in excellent health condition indicated by an index value of 4 – our baseline case. In a similar fashion, the region refers to three mutually exclusive dummies for respondents living in Punjab, NWFP and those who live in Sind. The baseline category here corresponds to households living in Baluchistan.

We may furthermore classify households by their specific characteristics such as their family structure (dual/single parent family), age, educational or financial background to analyse if and how happiness functions differ across various socio-economic groups.[10] Therefore, the third latent regression model used to develop a happiness function for the households belonging to different socio-economic backgrounds is given as:

[8]It might be interesting to explore how results differ across the children's gender; an issue we do not tackle here for data reasons.

[9]There is no childless household in the sample.

[10]This approach is line with Kahneman and Krueger (2006) who, using data from different countries, investigate how individual responses to subjective well-being questions vary with their circumstances as described by marital status or income.

$$happiness = \beta_0 + \beta_1(educ) + \boldsymbol{\beta_2}(unemployment) + \beta_3 ln(relativeincome) + \boldsymbol{\beta_4}(numberofchildren_{dummy}) + \boldsymbol{\beta_5}(region) + \varepsilon. \quad (4.3)$$

Regression model (4.3) is general enough to address all the various sub-groupings among the sample population based on the respective household's characteristics.

Results for all three models are discussed in the subsequent Sections. Seeking to ensure that all models are properly specified, we moreover conducted the linktest (or specification error test) for each of the ordered probit regressions.[11]

Variable	Obs	Mean	Std. Dev.	Min	Max
Male*	558	0.93	0.26	0	1
Age	600	52.97	8.88	30.00	70.00
Education	600	7.71	3.39	2.00	16.00
Number of Children	600	4.24	1.04	1.00	8.00
Unemployed*	187	0.31	0.46	0	1
Total monthly income (in PKR)	600	3515.58	1207.88	1596	6938
Relative income	600	1.24	0.39	0.79	1.68
Couple*	471	0.79	0.41	0	1
Health Satisfaction index	600	2.26	1.01	0.00	4.00
Province:					
Punjab*	240	0.4	0.49	0	1
NWFP*	180	0.2	0.40	0	1
Sind*	120	0.3	0.46	0	1
Baluchistan*	60	0.1	0.30	0	1

Table 4.3: Summary statistics of socio-economic determinants of happiness.
Note: (*) stands for a dummy variable.
Source: Survey 2008.

Table 4.3 provides summary statistics of the given controls, where all the variables with asterisks (*) refer to dummies. For instance, as mentioned above, *sex* is 1 if the respondent is male and 0 otherwise. According to

[11]The logic of these type of model misspecification tests is discussed in more detail in Chapter 5.

Table 4.3, the mean for male is 0.93, which implies that 93% of the sample observations comprise men. With regards to unemployment, the mean amounts to 0.31 or we may say that about 31% of the sample households are unemployed. The average for the variable *couple* is approximately 0.79; in other words, 79% of the interviewed households were living as a couple. By the same token, 40% of the sample observations live in Punjab (mean of 0.4), 30% in Sind, 20% in NWFP and 10% in Baluchistan. As far as the other controls are concerned, the average age is 53 years, the average education level in rural Pakistan is about 7 to 8 years of schooling, the average number of children is about 4 per household, and the average relative income amounts to 1.24 with an average monthly income of PKR 3515.58.

4.4.2 Baseline results

Ordered probit regression		
Number of obs	=	600
Wald $\chi^2(18)$	=	4962.88
Prob $> \chi^2$	=	0.0000
Pseudo R^2	=	0.1471
Log pseudolikelihood = -537.14467		
Dependent variable: Overall satisfaction		
Independent Variable	coef.	Robust Std. Err.
Male	-0.6588***	0.17902
Age	0.0131	0.0702
AgeSquared	-0.0001	0.0007
Years of Education	0.0402**	0.0190
No. of children	0.1290**	0.0628
Unemployed	-0.2958**	0.1535
Log of household's income	0.0373	0.1649
Log of relative income	-0.0271	0.7456
Couple	.1591	0.1727
Health Satisfaction index:		
4	*Reference Group*	
3	-0.3250	0.2076
2	-0.2625	0.2077
1	-0.5792**	0.2279
0	-9.9443***	0.2724
Region:		
Punjab	1.5061***	0.1805
NWFP	0.1397	0.2080
Sind	0.5689***	0.1953
Baluchistan	*Reference Group*	
/cut1	1.7275	2.1887
/cut2	1.8541	2.1865
/cut3	2.1074	2.1861

Table 4.4: Baseline results. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

Specification error test		
Number of obs	=	600
Wald $\chi^2(2)$	=	127.59
Prob $> \chi^2$	=	0.0000
Pseudo R^2	=	0.1474
Log pseudolikelihood = -536.96266		
Dependent variable: Overall satisfaction		
Independent Variable	coef.	Robust Std. Err.
-hat	0.7766***[0.000]	0.1397
-hatsq	0.0646[0.126]	0.0423
/cut1	1.5637	0.1620
/cut2	1.6903	0.1653
/cut3	1.9438	0.1735

Table 4.5: Specification error test: Baseline Model (4.1).
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets[ ] are p values.

The convention in the happiness economics literature is to define happiness as the overall satisfaction with the current socio-economic status. Therefore, we first estimate the model with the described happiness index (that is overall satisfaction) as dependent variable and use those results as benchmark. Table 4.4 gives an overview of the results which may be considered the relevant baseline. Our findings largely confirm the literature on the determinants of happiness. As expected, happiness depends on gender, education, family size, unemployment, health and region. For instance, the probability of being happy increases with an increasing family size or educational achievements. On the other hand, being a male, unemployed or having a low health profile lowers one's chance of being satisfied. We furthermore observe that happiness is region-dependent. Relatively speaking, living in Punjab indicates an elevated chance of being happy compared to Sind. The linktest (_hatsq) for the baseline model (4.1) is statistically insignificant as shown in Table 4.5. This means that we have no omitted relevant variable(s) and our link function is correctly specified.

Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.50403465
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= .05032489
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= .09738449
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= .34825596

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
Male*	0.2492***	0.0619	0.0064	0.0042	0.0011	0.0054	-0.2567***	0.0690
Age	-0.0052	0.0280	0.0000	0.0003	0.0003	0.0018	0.0048	0.0260
AgeSquared	0.00003	0.00028	-2.44E-07	0.0000	-1.69E-06	0.00002	-2.5E-05	0.00026
Education	-0.0161**	0.0076	0.0001	0.0001	0.0010	0.0006	0.0149**	0.0071
No. of children	-0.0515**	0.0251	0.0005	0.0005	0.0033*	0.0019	0.0477**	0.0232
Unemployed*	0.1173**	0.0602	-0.0019	0.0017	-0.0089*	0.0058	-0.1065**	0.0538
Log of household's income	-0.0149	0.0658	0.0001	0.0006	0.0010	0.0042	0.0138	0.0610
Log of relative income	0.0108	0.2974	-0.0001	0.0028	-0.0007	0.0190	-0.0100	0.2757
Couple*	-0.0633	0.0684	0.0010	0.0016	0.0047	0.0060	0.0576	0.0610
Health Satisfaction Index:								
3*	0.1288	0.0812	-0.0021	0.0023	-0.0099	0.0078	-0.1168	0.0720
2*	0.1042	0.0818	-0.0016	0.0020	-0.0077	0.0073	-0.0950	0.0732
1*	0.2249***	0.0839	-0.0062	0.0045	-0.0213*	0.0119	-0.1974***	0.0696
0*	0.5517***	0.0256	-0.0503***	0.0134	-0.1000***	0.0188	-0.4013***	0.0271
Region:								
Punjab*	-0.5416***	0.0541	0.0134***	0.0050	0.0454***	0.0110	0.4829***	0.0522
NWFP*	-0.0556	0.0825	0.0002	0.0006	0.0029	0.0034	0.0525	0.0796
Sind*	-0.2215**	0.0720	-0.0024	0.0030	0.0057	0.0040	0.2182***	0.0763

Table 4.6: Marginal effects after oprobit regression of baseline model (4.1). Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

The marginal effects for model (4.1) are given in Table 4.6. According to Table 4.6, being a male increases the probability of being observed in a lower happiness category, say (1), and decreases its chance of being in higher outcomes such as (4). Similarly, unemployed people and people with low health status, poor or very poor (1 and 0) are usually observed in a lower happiness category (1) compared to the higher category (4). With increasing education and family size, on the other hand, the probability of being observed in the higher happiness category (4) increases and at the same time it decreases the chance to be observed in the lower happiness category (1). Our results also support regional effects. For example, people living in Punjab and Sind are most likely to be observed in higher happiness categories compared to the lower happiness outcomes.

Our results with respect to gender, marital status and health are supported by several other studies of comparable scope. Knight et al. (2007), for instance, analyse national household survey data for subjective well-being in rural China and also find that men report lower happiness than women. Taking singles as reference category, married couples are relatively more blissful according to their study, whereas divorced couples or widowhood dampen happiness. They furthermore suggest that happiness increases with a higher level of education, income or health respectively. They suggest that happiness is U-shaped in age which is somewhat surprising, given that for developing countries one may suspect an inverted shape due to poor social security systems in place especially for older citizens.

It is generally believed in the developing world that old people become increasingly less happy "by nature" not only because their physical and cognitive capacities deteriorate, but also because of psychological factors such as the increasing likelihood of suffering from depression. From a socio-economic point of view, older people tend to be in poorer health and have lower income which makes them less happier indeed (Frey and Stutzer, 2002a). The situation may be fundamentally different in developed countries. Easterlin (2006), for instance, based on data from General Social surveys from 1973-1994, finds that both health and financial satisfaction in the US follow a U-shaped pattern in age. This would imply that well-being falls with rising age, reaching a particular turning point and then starts rising again. Contrary to the results for rural China, the US or many EU countries, our results, however, seem to support the idea of an inverted U-shaped relatioship between age and happiness with a theoretical turning point of 65.5 years of age. On purely statistical grounds, however, it is not obvious per se whether any relationship exists at all.

While one might expect happiness to be of inverted-U shape in age, a general relationship between age and happiness seems to be difficult to establish. A U-shaped age effect on happiness has been challenged on empirical grounds.[12] The age-happiness pattern is found to differ across countries and time periods. The overall empirical evidence appears blurred at best and it is thus difficult to draw any robust conclusion.

We further extend the baseline results by using dummies for the different number of children per household rather than using the actual number of children per household, using the regression model (4.2). The results are given in Table 4.7 and suggest that the probability of being happy is low for the households with a small number of children and a low health status. Similarly, being male and being unemployed reduces the chances of reporting happiness. On the other hand, the likelihood of happiness increases in the level of education. One possible explanation for this positive association may be that higher education increases the chances of employability. As far as regional effects are concerned, in Punjab the probability of being happy is about thrice as high compared to the likelihood feeling satisfied in Sind.

The linktest for the model (4.2) appears to be statistically insignificant as shown in Table 4.8. The variable _hatsq is statistically insignificant which indicates that the model is properly specified. There are no omitted relevant variables and one should not be able to find any additional predictors that are statistically significant except by chance. Apart from assessing the model specification, we also test for the joint significance of the dummies used on the number of children (one, two, three, four, five and six). We use the *test* command in STATA as shown in Table 4.9 for this purpose, where 1, 2, 3, 4, 5 and 6 corresponds to the coefficients when one, two, three, four, five and six children are present in the household, respectively. Amid a highly significant test statistic ($\chi^2(6) = 917.84$) with six degrees of freedom, we reject the null hypothesis of all the coefficients being equal to zero, and accept the alternative hypothesis that at lease one of the coefficients is significantly different from zero instead. In other words, we may say that the given coefficients for the number of children per household are *not* jointly significant.

[12] See Horley and Lavery (1995) and the references therein for details.

Ordered probit regression		
Number of obs	=	600
Wald $\chi^2(18)$	=	5420.19
Prob > χ^2	=	0.0000
Pseudo R^2	=	0.1482
Log pseudolikelihood = -536.45646		
Dependent variable: Overall satisfaction		
Independent Variable	coef.	Robust Std. Err.
Male	-0.6865***	0.1903
Age	0.0095	0.0718
AgeSquared	-3.2E-05	0.0007
Years of Education	0.0418**	0.0192
No. of children:		
1	-7.6685***	0.4249
2	-0.9727**	0.4857
3	-0.5966	0.3836
4	-0.5308	0.3603
5	-0.3789	0.3704
6	-0.2798	0.3896
7 or more	*Reference Group*	
Unemployed	-0.2897*	0.1538
Log of household's income	0.0434	0.1659
Log of relative income	-0.0689	0.7494
Couple	0.1716	0.1731
Health Satisfaction index:		
4	*Reference Group*	
3	-0.3362	0.2082
2	-0.2554	0.2097
1	-0.5889***	0.2287
0	-8.8264***	0.2889
Region:		
Punjab	1.5068***	0.1822
NWFP	0.1442	0.2107
Sind	0.5766***	0.1990
Baluchistan	*Reference Group*	
/cut1	0.6481	2.2762
/cut2	0.7749	2.2734
/cut3	1.0287	2.2730

Table 4.7: Baseline results; using dummies for number of children. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

Specification error test		
Number of obs	=	600
Wald $\chi^2(2)$	=	125.50
Prob > χ^2	=	0.0000
Pseudo R^2	=	0.1483
Log pseudolikelihood = -536.3563		
Dependent variable: Overall satisfaction		
Independent Variable	coef.	Robust Std. Err.
-hat	0.9316***[0.000]	0.1763
-hatsq	0.0520[0.693]	0.1318
/cut1	0.6490	0.0865
/cut2	0.7759	0.0902
/cut3	1.0298	0.0983

Table 4.8: Specification error test: Model (4.2) using dummies for number of children.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets [] are p-values.

. test 1 2 3 4 5 6	
(1)	[Overall satisfaction]1 = 0
(2)	[Overall satisfaction]2 = 0
(3)	[Overall satisfaction]3 = 0
(4)	[Overall satisfaction]4 = 0
(5)	[Overall satisfaction]5 = 0
(6)	[Overall satisfaction]6 = 0
$\chi^2(6) = 917.84$***	
Prob > $\chi^2 = 0.0000$	

Table 4.9: Joint hypothesis testing; using the dummies for the number of children.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

Table 4.10: Marginal effects after oprobit regression of model (4.2).
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.50649323
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= 0.05041426
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= 0.09737584
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= 0.34571666

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
Male*	0.2589***	0.0650	0.0068	0.0045	0.0015	0.0059	-0.2672***	0.0731
Age	-0.0038	0.0287	0.0000	0.0003	0.0002	0.0019	0.0035	0.0265
AgeSquared	0.00001	0.00028	-1.28E-07	0.00000	-8.36E-07	0.00002	-0.00001	0.00026
Education	-0.0167**	0.0077	0.0002	0.0002	0.0011*	0.0006	0.0154**	0.0071
No. of children:								
1*	0.4966***	0.0257	-0.0504***	0.0134	-0.0976***	0.0183	-0.3486***	0.0259

Table 4.10: (continued)

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
2*	0.3346***	0.1220	-0.0209	0.0161	-0.0500*	0.0296	-0.2637***	0.0802
3*	0.2298*	0.1379	-0.0075	0.0082	-0.0238	0.0200	-0.1985*	0.1112
4*	0.2089	0.1381	-0.0032	0.0034	-0.0154	0.0120	-0.1903	0.1242
5*	0.1491	0.1420	-0.0035	0.0055	-0.0132	0.0161	-0.1324	0.1210
6*	0.1105	0.1511	-0.0025	0.0054	-0.0096	0.0165	-0.0984	0.1295
Unemployed*	0.1149*	0.0604	-0.0019	0.0017	-0.0089	0.0058	-0.1041*	0.0538
Log of household's income	-0.0173	0.0662	0.0002	0.0007	0.0011	0.0043	0.0160	0.0612
Log of relative income	0.0275	0.2989	-0.0003	0.0030	-0.0018	0.0196	-0.0254	0.2763
Couple*	-0.0682	0.0684	0.0011	0.0017	0.0052	0.0063	0.0618	0.0607
Health Satisfaction Index:								
3*	0.1331	0.0812	-0.0023	0.0024	-0.0105	0.0080	-0.1203*	0.0717
2*	0.1014	0.0826	-0.0016	0.0020	-0.0076	0.0074	-0.0922	0.0738
1*	0.2283***	0.0838	-0.0065	0.0047	-0.0221*	0.0121	-0.1996***	0.0693
0*	0.5494***	0.0255	-0.0505***	0.0134	-0.1001***	0.0188	-0.3988***	0.0271
Region:								
Punjab*	-0.5413***	0.0546	0.0137***	0.0051	0.0461***	0.0112	0.4816***	0.0526
NWFP*	-0.0574	0.0836	0.0002	0.0006	0.0031	0.0035	0.0541	0.0806
Sind*	-0.2246***	0.0734	-0.0023	0.0031	0.0060	0.0041	0.2208***	0.0777

Table 4.10 shows the marginal effects resulting from the ordered probit regression analysis of model (4.2); using dummies for the number of children per household. According to Table 4.10, if a household has a limited number of children (i.e. no more than two), the likelihood of being in the higher categories of happiness levels (i.e. 2, 3 and 4) decreases and at the same time it is more likely to be observed in the lowest category of happiness. The same applies to health. Households reflecting a low health satisfaction index (i.e. 0 or 1) are most likely to fall in the lowest category of happiness (i.e. 1) against the higher categories. Similarly, if a respondent is unemployed or in case of a male-headed family, the risk of being in the lowest happiness category increases, while it diminishes the highest possible outcome. Furthermore, happiness favours the more educated households as shown in Table 4.10. With the increase in educational attainment, the probability of belonging to higher happiness outcomes increases, while it decreases for lower happiness outcomes. However, we may argue that the marginal effects of education are considerably small in either case. Lastly, happiness is sensitive to the regional background. Households for instance belonging to Punjab are more likely to be observed in the higher level categories of happiness, relative to those who belong to Sind.

While our baseline results are in keeping with economic intuition, only gender, education, number of children, unemployment, health and region turn out to be of some statistical meaning. In particular, a large family size seems to be vital. On the other hand, a higher nominal income (both in absolute and relative terms) does not seem to significantly improve households' perceived well-being. A result which seems at odds with economic reasoning, as one would expect households in developing countries to feel strongly about monetary factors.[13] This seems to suggest that in order to capture poverty in monetary terms, we should resort to direct poverty measures, for instance, satisfaction with income or expenditure. We therefore consider in the remainder two issues in turn. First, we analyse the interplay between having children and household's specific features in improving the family's overall subjective well-being, using model (4.3). Second, we employ the baseline model (4.1) for other response variables (i.e. direct poverty measures) for the sake of comparison with the baseline results.

[13]Our results suggest that happiness is the broader measure which seems to capture reasonably well other important aspects of poverty such as health and level of education matter most in economic and econometric terms using this well-being metric. With income being insignificant, happiness here encompasses the idea of capabilities poverty, but it is more general than this approach. It moreover incorporates important and often neglected socio-economic factors.

4.4.3 Children and household's specific characteristics

The regression analysis so far (see Tables 4.4 and 4.7) has treated the entire sample as homogeneous unit. This implicitly assumes that having children at home affects respondents in a similar vein, irrespective of their age, marital status, gender, income level or educational background. We shall relax this and assess in this Section whether household-specific characteristics significantly impact on the experience of having children, using regression model (4.3).[14]

Table C.2 distinguishes different experiences of parenthood according to marital status. The first column produces the aggregate findings for all 600 families which is then broken down according to different family structures. Four out of five families are of the "traditional" dual type. The remaining 20% are single-parent families, i.e. spearated/divorced or widowed parents. It appears the basic picture is robust across dual-parent families. The more children dual-parent families have, the greater is the probability of being happy. Single-parent families suffer fom the task of rearing children, and are unhappy with the increasing number of children; yet they are unhappiest in absolute terms with just one child which reflects the idea of children being supportive for general family matters. To take a comparison from a developed country, a similar study by Frey and Stutzer (2000) based on Swiss survey data for the year 1992 also suggests that single-parent families are unhappier than dual-parent families with children.

In order to test for model specification, we use the *linktest* command in STATA. Results are reported in Table C.2. The linktest based on the variable *_hatsq* is statistically insignificant which shows that we do not have an omitted variable bias in the model and the link function of the model is correctly specified. We use the *test* command provided by STATA to test for the joint significance of the dummies used for the number of children; as shown in Table C.2. The test statistic suggests that the dummies used for the number of children is statistically significant in each set-up, i.e. for the case of modelling all the families together or for modeling/grouping them separately as dual and single parent family, as we have to reject the null hypothesis of all the coefficients being equal to zero in each case.

Table C.3 shows the marginal effects of the number of children on the probability of happiness outcomes based on family types. In case of all (pooled) families or dual parent families, having a single child decreases the

[14]Given the nature of this question, it seems most natural to only consider happiness as broad subjective poverty measure.

probability for the higher happiness outcomes (2, 3 and 4), while it increases the likelihood of being in the lowest outcome (1). On the other hand, for single parent families, having one child decreases the probability for the highest happiness outcome (4) and reveals an increased chance to end up in the lowest outcome (1).

Overall, a particular family type in itself may not sufficiently capture feelings of parenthood and their perceived effects on subjective well-being poverty. The age structure of the parents may matter. Parents may feel differently about having children depending on their age. This question is addressed by Table C.4.

Table C.4 shows differences in experiences about having children at home between individuals of different age groups. Columns two and three separate individuals into those who are aged 40 or less and those who are aged above 40, respectively. We shall call the former group "younger group" and identify the latter as "older group". Well-being of older group households overall increases by having more children. However, looking more closely, clear differences emerge. The results show that as people get older and older, they feel relatively happier than before about having even the same number of children at home. In contrast, in the younger age group, any additional child may imply an extra burden for the young parents who still struggle for achieving their life goals and most probably the gap between the goals and actual achievements is relatively larger for younger individuals compared to older persons.

In the younger group which comprises individuals aged 40 years or less, people are moreover more likely to be unhappy by having two children as compared to the older individuals. However, on pure statistical grounds, we found no evidence for children influencing happiness of the younger parents.[15]

As for the older groups, the probability of happiness is clearly increasing with the number of children up till seven (or more) children. Considering all individuals together, this high number of children yields the largest increase in the probability of household's happiness. One may conjecture that young individuals may also want to establish a similarly large family but having so many children in such a young age is quite rare. Moreover, peer pressure may be lower as there is still a fairly large time span to increase the size of the family.

[15]No family of the 40 or less years of age group reported having seven or more children at home; hence the excluded reference category corresponds to households having one child in total.

We also test for an appropriate model specification using the *linktest*. As Table C.4 shows, the variable *_hatsq* is statistically insignificant which implies that based on the *linktest* there is no evidence for an omitted variable bias inherent in the model and that the model is well-specified properly as the link function is correctly specified. Furthermore, in order to test for the joint significance of the dummies used for the number of children, we use the *test* command in STATA as shown in Table C.4. The reported chi-square for the dummies used for the number of children is statistically insignificant in case of individuals aged 40 or below; and we may accept the null hypothesis of all the coefficients being equal to zero. On the other hand, for all individuals or for individuals aged above 40, the resulting test statistics are highly significant which gives us reason to reject the null hypothesis and accept the alternative that at least one of the coefficients is significantly different from zero.

Table C.5 shows the marginal effects of the number of children on the probability of achieving happiness outcomes based on age-groups of the individuals. Compared to the reference category (7 or more children), having one child decreases the probability for the higher happiness outcomes (2, 3 and 4), while it increases the likelihood of being in the lowest outcome (1) in case of all individuals or individuals aged above 40. On the other hand, for households aged 40 or less, no significant effect of the number of children on household's happiness has been detected.

The happiness model (4.3) is comprehensive enough to allow for further clusterings. While the poverty-reducing effect of children seems to hinge on family type and the parents' age, one may also expect that the "meaning" of a child differs across households depending on their level of income and education. With respect to the former dimension we separate families into two groups: those whose income exceeds average income and those whose income falls short of this benchmark. Results indeed differ among these groups as shown in Table C.6.

Table C.6 implies that households in the upper income group are less likely to be happy with increasing number of children as compared to the lower income group. Not surprisingly, the lower income group is the larger of the two and estimates for that group most closely resemble those of the population at large: well-being is increasing in the number of children living in the household. Results for column (2) suggest that as income rises, parents preferences shift more towards the attainment of personal goals and career enhancement. In other words, we may speculate that as income goes up, the "demand" (or desire) for children drops as the opportunity cost

of raising a child rises.[16] In relatively poor families, on the other hand, the demand for children is higher because the higher the number of children around, the more helping hands are available. This supportive role is crucial to the functioning of the household given that in such lower income classes, marked by extreme poverty, children start working (inside and outside the home) at the age of 12 or even below and also need to look after the elders in their family. In that sense, children are more productive and "beneficial" for the relatively poor group. Indeed, column (3) reports a statistically significant and negative estimate for one child against a large number of children (7 or more).

As before, in an attempt to assess the appropriateness of the specified model, we use the *linktest* command. Results reported in Table C.6 provide evidence for the absence of an omitted variable bias in the model suggesting that the link function of the model is correctly specified. To test for the joint significance of the dummies used for the number of children, we use the *test* command in STATA as shown in Table C.6. Test statistics for the dummies used for the number of children are statistically significant in all the three cases, i.e. regardless of modeling all the households together or grouping them separately as households having above average income and households having below average income. We therefore have to reject the null hypothesis of all the coefficients being equal to zero and accept the alternative that at least one of the coefficients is significantly different from zero for each of the given cases.

Table C.7 summarises the marginal effects of the number of children on the likelihood of happiness outcome(s) based on the income of the households. Results suggest that, compared to the reference category (7 or more children), having one child decreases the probability for the higher happiness outcomes (2, 3 and 4), while it increases the likelihood of being in the lowest outcome (1) in case of all individuals or for individuals whose income is less than the average income. On the other hand, for households whose income exceeds average, an additional child reduces the likelihood for the higher happiness outcomes (2, 3 and 4) and at the same time it increases the chance of being in the lowest happiness outcome (1).

The level of education is a further important socio-economic feature for the role of children in reported poverty assessments. Similar to the income variable, we split the respondents in two major groups, choosing 10 years

[16]Considering the upper income group, no family reported having seven or more children at home; hence the excluded/reference category corresponds to households having one child in total.

of schooling as the relevant benchmark. Individuals with such a complete level of high school education are summarised in column (2). Individuals who did not complete the 10 years of schooling are grouped together in the last column of Table C.8.

This classification allows us to analyse the effects of children on households' happiness with different levels of education. We find that education does matter in this regard. Less educated individuals are relatively less likely to be happy with increasing number of children compared to the more educated ones.[17]

The results of the model specification testing are reported in Table C.8 and show that the *linktest* or at least the variable *_hatsq* are statistically insignificant. Thus, there is no evidence for an omitted variable bias problem and the link function of the model is correctly specified. In order to test for the joint significance of the dummies used for the number of children, we employ the *test* STATA command as shown in Table C.8. The test results imply that the dummies used for the number of children are statistically insignificant in case of individuals having less than full school education. We therefore fail to reject the null hypothesis of all the coefficients being equal to zero. On the other hand, for all individuals or for individuals having full school education, the chi-square test statistics are highly significant and we have to reject the null hypothesis and accept the alternative that at least one of the coefficients is significantly different from zero in that case.

Research in this area has shown that parental education plays a positive role in the child's care and outcome. Guryan et al. (2008), for instance, observes that higher-educated parents spend more time with their children. This relationship is striking, given that higher-educated parents also spend more time working outside the home. This relationship is robust and holds across all subgroups examined, including both non-working and working parents. It also holds across all four subcategories of child care: basic, educational, recreational, and travel related to child care.[18]

[17]Though, we didn't find any statistical evidence for less educated households being affected by having children at home.

[18]"Total child care" is defined as the sum of those four primary time use components. "Basic" child care is time spent on the basic needs of children, including breast feeding, rocking a child to sleep, general feeding, changing diapers, providing medical care (either directly or indirectly), grooming, and so on. "Educational" child care is time spent reading to children, teaching children, helping children with homework, attending meetings at a child's school, and similar activities. "Recreational" child care involves playing games with children, playing outdoors with children, attending a child's sporting event, going to the zoo and taking walks with children. "Travel" child care is any travel related to any of the three categories of child care. For example, driving a child to school, to a

From an economic perspective, this positive education gradient in child care can be viewed as surprising, given that the opportunity cost of time is higher for higher-educated parents. Yamauchi (2009) departs from previous research by providing evidence on the relationship between parental education and children's outcomes. His results suggest that educated and mentally healthier parents are likely to have children with better outcomes. Educated parents are more frequently engaged in education-oriented activities with their children and mentally healthier parents exhibit more favourable parenting practices relating to children's behavioural and socio-emotional outcomes.

Table C.9 provides an overview of the marginal effects of the number of children on the likelihood of the different happiness outcomes based on educational background of the households. The presence of only one child decreases the probability for the higher happiness outcomes (2, 3 and 4), while it increases the likelihood of being in the lowest outcome (1) in case of all individuals or for individuals having full school education. These findings are statistically significant. Considering individuals with less than full school education, five to six may be the ideal number of children, but on pure statistical grounds we fail to provide conclusive evidence.

Overall, the results obtained are in line with the cited literature. We may conclude based on Table C.8 that parental education is important to the children's outcomes. For instance, the higher-educated parents may put a lot of efforts in bettering their children in terms of education, food, health etc. and thus reap higher both monetary and non-monetary rewards in return. Less-educated parents, in contrast, may lack parental skills for taking good care of their children in those dimensions which might adversely affect their children's outcomes. In addition, the chances of infants' mortality are also higher in less educated families as a consequence of poor health care. Parental characteristics may therefore well affect children's outcomes which in turn will have a bearing on the household's subjective poverty.

4.5 Alternative measures

To test whether income remains insignificant for alternative poverty measurement approaches we moreover employ model (4.2) for the other response variables. Since happiness is a rather "soft" concept, it seems more straightforward to use subjective categories which focus more on financial terms.

doctor, or to sports practise are all included in "travel" child care.

Table 4.11: Results by financial satisfaction.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets[ ] are p values.

ordered probit regression	Dependent Variable: Satisfaction with			
	income		expenditure	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Male	-0.5428***	0.2116	-0.5736***	0.2307
Age	-0.0554	0.0665	-0.0890	0.0683
AgeSquared	0.0005	0.0006	0.0008	0.0007
Years of Education	0.0327*	0.0197	0.0255	0.0195
No. Of children:				
1	-1.3544***	0.3910	-1.2625***	0.4075
2	-0.9664**	0.4285	-1.1644**	0.5467
3	-0.7074*	0.3808	-0.7840**	0.4063
4	-0.4815	0.3763	-0.5477	0.4036
5	-0.4180	0.3809	-0.4836	0.4000
6	-0.5291	0.3891	-0.5070	0.4207
7 or more	*Reference Group*			
Unemployed	-0.1556	0.1665	-0.0840	0.1717
Log of household's income	0.3272*	0.1718	0.4261***	0.1687

Table 4.11: (continued)

ordered probit regression				
	Dependent Variable: Satisfaction with			
	income		expenditure	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Log of relative income	0.4816	0.6905	0.5014	0.6951
Couple	0.0978	0.1741	0.0600	0.1816
Health Satisfaction index:				
4	*Reference Group*			
3	-0.2383	0.1902	-0.2678	0.1824
2	-0.1207	0.1967	-0.0231	0.1953
1	-0.1651	0.2213	-0.0987	0.2178
0	-1.5871***	0.3416	-1.4108***	0.3177
Region:				
Punjab	1.3073***	0.1444	1.3593***	0.1484
NWFP	-0.2809*	0.1604	-0.1118	0.1501
Sind	*(dropped)*			
Baluchistan	*Reference Group*			
/cut1	-2.4251	2.2049	-2.1088	2.2152
/cut2	1.0449	2.1940	0.8925	2.2007
/cut3	2.8372	2.2031	2.9559	2.2169

Table 4.11: (continued)

ordered probit regression	Dependent Variable: Satisfaction with income		Dependent Variable: Satisfaction with expenditure	
Log pseudolikelihood	-446.98337		-432.94983	
Obs	600		600	
Wald $\chi^2(12)$	382.10		340.14	
Prob $> \chi^2$	0.0000		0.0000	
Pseudo R^2	0.1866		0.1907	
.linktest	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
-hat	0.8943***	0.1595	0.9116***	0.1454
	[0.000]		[0.000]	
-hatsq	0.0597	0.0866	0.0603	0.0928
	[0.491]		[0.516]	
/cut1	-2.3990	0.1329	-2.0829	0.1260
/cut2	1.0312	0.0834	0.8931	0.0803
/cut3	2.8298	0.1484	2.9661	0.1684

Table 4.11: (continued)

	Dependent Variable: Satisfaction with income		Dependent Variable: Satisfaction with expenditure	
.linktest				
Log pseudolikelihood	-446.80168		-432.76776	
Obs	600		600	
Wald $\chi^2(12)$	170.43		164.82	
Prob $> \chi^2$	0.0000		0.0000	
Pseudo R^2	0.1869		0.1911	
. test	$\chi^2(6)$	27.92	$\chi^2(6)$	25.13
(using dummies for No. of children)	prob>χ^2	0.0001	prob>chi2	0.0003

Table 4.12: Marginal effects after oprobit; results by financial satisfaction.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

Outcome: Satisfaction with income

. mfx, predict(p outcome())
Marginal effects after oprobit
y = Pr(Satisfaction with income==1) (predict, p outcome(1))
= 0.00032308
y = Pr(Satisfaction with income==2) (predict, p outcome(2))
= 0.52300195
y = Pr(Satisfaction with income==3) (predict, p outcome(3))
= 0.444576
y = Pr(Satisfaction with income==4) (predict, p outcome(4))
= 0.03209897

Outcome	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
Male*	0.0003*	0.0002	0.2105***	0.0773	-0.1513***	0.0467	-0.0595*	0.0338
Age	0.0001	0.0001	0.0220	0.0264	-0.0181	0.0217	-0.0040	0.0049
AgeSquared	0.0000	0.0000	-0.0002	0.0003	0.0002	0.0002	0.0000	0.0001
Education	0.0000	0.0000	-0.0130*	0.0078	0.0107*	0.0065	0.0024	0.0015

Table 4.12: (continued)

Outcome: Satisfaction with income

Outcome	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.0195	0.0194	0.3787***	0.0450	-0.3667***	0.0601	-0.0315***	0.0082
2*	0.0066	0.0085	0.3199***	0.0969	-0.2958***	0.1007	-0.0307***	0.0088
3*	0.0021	0.0025	0.2639**	0.1267	-0.2310**	0.1166	-0.0350**	0.0150
4*	0.0007	0.0008	0.1886	0.1438	-0.1564	0.1200	-0.0329	0.0257
5*	0.0008	0.0012	0.1622	0.1421	-0.1386	0.1255	-0.0244	0.0187
6*	0.0013	0.0018	0.2012	0.1372	-0.1752	0.1249	-0.0273*	0.0155
Unemployed*	0.0002	0.0003	0.0616	0.0656	-0.0512	0.0554	-0.0106	0.0107
Log of household's income	-0.0004	0.0003	-0.1299**	0.0683	0.1068*	0.0563	0.0235*	0.0134
Log of relative income	-0.0006	0.0009	-0.1912	0.2742	0.1572	0.2250	0.0347	0.0506
Couple*	-0.0001	0.0003	-0.0387	0.0687	0.0322	0.0576	0.0067	0.0115
Health Satisfaction Index:								
3*	0.0003	0.0004	0.0940	0.0743	-0.0785	0.0630	-0.0159	0.0121
2*	0.0002	0.0003	0.0478	0.0776	-0.0396	0.0651	-0.0084	0.0129
1*	0.0002	0.0004	0.0652	0.0867	-0.0545	0.0740	-0.0110	0.0134
0*	0.0319	0.0247	0.4021***	0.0294	-0.4004***	0.0442	-0.0336***	0.0086
Region:								
Punjab*	-0.0044**	0.0021	-0.4737***	0.0461	0.3922***	0.0414	0.0859***	0.0194
NWFP*	0.0005	0.0004	0.1099*	0.0613	-0.0934*	0.0539	-0.0169**	0.0088

Table 4.12: (continued)

Outcome: Satisfaction with expenditure

. mfx, predict(p outcome())

Marginal effects after oprobit

y = Pr(Satisfaction with expenditure==1) (predict, p outcome(1))

= 0.00161063

y = Pr(Satisfaction with expenditure==2) (predict, p outcome(2))

= 0.52052488

y = Pr(Satisfaction with expenditure==3) (predict, p outcome(3))

= 0.46081717

y = Pr(Satisfaction with expenditure==4) (predict, p outcome(4))

= 0.01704732

Outcome	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std.Err.	dy/dx	Std.Err.	dy/dx	Std.Err.	dy/dx	Std.Err.
Male*	0.0016**	0.0007	0.2202***	0.0829	-0.1810***	0.0595	-0.0409	0.0268
Age	0.0005	0.0004	0.0350	0.0269	-0.0317	0.0243	-0.0038	0.0032
AgeSquared	0.0000	0.0000	-0.0003	0.0003	0.0003	0.0002	0.0000	0.0000
Education	-0.0001	0.0001	-0.0100	0.0077	0.0091	0.0070	0.0011	0.0009
No. of children:								
1*	0.0444	0.0405	0.3400***	0.0377	-0.3677***	0.0704	-0.0167***	0.0058
2*	0.0344	0.0440	0.3362***	0.0679	-0.3533***	0.1075	-0.0173***	0.0061
3*	0.0096	0.0101	0.2822**	0.1241	-0.2706**	0.1265	-0.0212**	0.0098
4*	0.0036	0.0037	0.2112	0.1502	-0.1928	0.1380	-0.0219	0.0171

Table 4.12: (continued)

Outcome: Satisfaction with expenditure

Outcome	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std.Err.	dy/dx	Std.Err.	dy/dx	Std.Err.	dy/dx	Std.Err.
5*	0.0040	0.0053	0.1835	0.1430	-0.1719	0.1382	-0.0156	0.0111
6*	0.0048	0.0067	0.1899	0.1452	-0.1797	0.1429	-0.0150	0.0099
Unemployed*	0.0005	0.0010	0.0330	0.0672	-0.0300	0.0617	-0.0034	0.0066
Log of household's income	-0.0022*	0.0013	-0.1675***	0.0664	0.1517***	0.0604	0.0180**	0.0089
Log of relative income	-0.0026	0.0038	-0.1971	0.2733	0.1785	0.2475	0.0212	0.0302
Couple*	-0.0003 0.0011		-0.0235	0.0711	0.0214	0.0649	0.0024	0.0073
Health Satisfaction Index:								
3*	0.0016	0.0015	0.1043	0.0701	-0.0956	0.0651	-0.0103	0.0070
2*	0.0001	0.0011	0.0091	0.0767	-0.0082	0.0697	-0.0010	0.0081
1*	0.0006	0.0014	0.0387	0.0850	-0.0353	0.0784	-0.0040	0.0080
0*	0.0582	0.0375	0.3541***	0.0285	-0.3946***	0.0505	-0.0178***	0.0061
Region:								
Punjab*	-0.0169***	0.0061	-0.4774***	0.0450	0.4399***	0.0439	0.0545***	0.0160
NWFP*	0.0007	0.0010	0.0437	0.0584	-0.0400	0.0539	-0.0044	0.0056

Table 4.11 shows the results of the oprobit regressions that can be compared with the baseline results. According to Table 4.11, male are less likely to be satisfied with their finances (income/expenditure) compared to female households. Education increases the probability of financial satisfaction, particularly in terms of satisfaction with income. We find that financial satisfaction is a positive function of education; however, the estimate is significantly positively associated with the income measure. Furthermore, households having more children are more likely to be financially satisfied compared to households with less or fewer children. As expected, a higher income increases the likelihood of financial satisfaction and vice versa. Similarly, a higher level of health satisfaction increases the probability of the household's financial satisfaction. Regarding regional effects, people in Punjab are most likely to be satisfied with their income/expenditure compared to the other provinces.

As far as the model specification is concerned, our model has been properly specified as shown in Table 4.11. The *linktest* is statistically insignificant. The variable _hat is highly significant in either case, which indicates that the model is properly specified. The variable _hatsq appears to be statistically insignificant, which suggests that all the relevant variables have been included in the model and the link function is correctly specified. We also test the joint significance of the dummies used for the number of children (per household) by applying the *test* command in STATA as reported in Table 4.11. The higher chi-square values and the correspondingly lower p-values indicate that the estimates are not jointly significant; and thus we reject the null hypothesis of all the coefficients being equal to zero and accept the alternative hypothesis that at least one of the coefficients is significantly different from zero.

Table 4.12 shows marginal effects after applying an ordered probit regression model to (4.2), using financial satisfaction (satisfaction with income/expenditure) as response variable. According to Table 4.12, being a male increases the probability to be observed in the lower categories (1 and 2) of financial satisfaction; and decreases the probability to be in the higher financial satisfaction categories (3 and 4). A large family size is beneficial in terms of financial satisfaction of a household. Compared to the reference category of seven or more children, a small family of one to three children reduces the chances of a household of being in the higher financial satisfaction categories (3 and 4), whilst increasing the likelihood of being in the lower outcome category (2). Similarly, households with lower health outcomes have an increased chance to be observed in the lower category (2) of finan-

cial satisfaction while it lowers the chances of being in the higher response categories (3 and 4). Higher income leads to higher financial outcomes even on pure subjective grounds. For instance, higher income increases the likelihood of being in the higher categories (3 and 4) of financial satisfaction and at the same time decreases the probability of being in the lower response categories (i.e. 1 or 2). Similarly, higher education ensures higher income satisfaction; for instance, education increases the likelihood of being in the higher category (3) of income satisfaction and decreases the probability to be observed in the lower category (2) of income satisfaction. However, we cannot confirm such a pattern in case of satisfaction with expenditure on pure statistical grounds. As far as regional effects are concerned, we find that compared to other provinces, being in Punjab increases the chances to be observed in the categories of higher financial satisfaction (i.e. 3 and 4), whereas it reduces the chances of being in the lower categories of financial satisfaction (i.e. 1 and 2). In other words, we may say that people in Punjab are most likely to be satisfied with their income or expenditure compared to the people belonging to the rest of the given provinces.

Summing up, we see that using this more direct approach, income does matter also on subjective grounds. Estimates are significantly positively associated with the corresponding measure. Closely related to higher income levels is the degree of schooling in the developing world. Indeed, the variable capturing years of education is significantly positively associated with the subjective poverty measures, all other things held constant. Similarly, a higher health index and an increased number of children also seem to be positively associated with the household's reported financial well-being (that is satisfaction with income/expenditure).[19] Unlike our baseline results, financial satisfaction appears to be U-shaped in age. The estimated turning point corresponds to the age of 55 years. This is similar to Easterlin (2006) who also reports evidence of a U-shaped relationship between financial satisfaction and age in the US. Such a pattern, however, cannot be confirmed for the case at hand in purely statistical terms. Given the limited public provision of social security (for instance, pensions, old age benefits, etc.) in Pakistan, one would expect an inverted U-shaped relationship between these two variables.

Comparing all three measures of subjective well-being poverty (both direct and indirect measures), a somewhat consistent picture seems to emerge: We may conclude that the level of education, health and the number of chil-

[19]The expenditure or income approach can be considered as two sides of the same coin here. We elaborate on this idea in section 4.2.

dren matter the most both in econometric and economic terms. In particular, a large family size of at least 7 children seems to be vital. Our findings thus support the notion of children as insurance mechanism in developing countries. In areas such as rural Pakistan children are integrated in the family life early on. Many of them contribute considerably to the overall household income already at a young age. Our results suggest that the number of children is one of the major determinants of subjective well-being poverty in rural Pakistan.

4.6 Conclusion

We use in this Chapter the dataset for rural Pakistan to shed light on issues associated with subjective well-being. This is to our knowledge the first study of life satisfaction in this part of the developing world. We not only provide a good deal of descriptive information but also test various measures in relation to overall and financial well-being using a regression model with categorical variables. Our main contributions are the following. First, we link the emerging field of the economics of happiness with development studies. In particular, we construct subjective well-being measures to evaluate poverty, highlighting their differences but also similarities. Second, we intend to challenge the view of poverty being a purely macro-level phenomenon which is based on a conventional nominal (either absolute or relative) metric. We demonstrate that analysing the issue on a more micro-level allows for a much richer analysis and more differentiated insights.

Based on data for rural Pakistan, we employ different ways of measuring subjective well-being: overall satisfaction with socio-economic status (happiness), satisfaction with income and satisfaction with expenditure. Our results suggest that happiness is the broader measure which seems to capture reasonably well other important components of poverty. In particular health and the level of education matter most in economic and econometric terms using this well-being metric. With income being insignificant, the notion of happiness in this context encompasses the idea of capabilities poverty. However, the approach here is more general than this as it also incorporates important and often neglected socio-economic factors such as the number of children in a household.

On the other hand, alternative approaches to well-being are found to be more appropriate in capturing the conventional notion of income poverty; with income being highly significant in either case. We find that both overall

and financial satisfaction are positive functions of income.

Our model is general enough to be mapped onto different categorical variables capturing subjective poverty: more indirect but comprehensive measures such as happiness but also reported satisfaction with income and expenditure.

The baseline results moreover confirm findings common in the well-being literature: Happiness increases among females, married couples, educated and healthy individuals. Unlike other studies, particularly for industrialised nations, an inverted (as opposed to a conventional) U-shaped pattern characterises the age-happiness profile. While this finding matches theoretical considerations, the jury is still out to provide convincing empirical evidence for either shape.

Pakistan shares distinct demographic features with many other developing countries: It is characterised by a large population as well as high population growth and fertility rates. As expected, the socio-economic environment is crucial for explaining perceived poverty. Our analysis suggests a positive effect of the number of children on individual household's life satisfaction – a result which has not yet been established in such a framework.

Elaborating on the children effect, our analysis furthermore suggests that children increase well-being among elderly, highly educated and those with low income. This may be attributed to the rural environment in which children typically are integrated in domestic activities early on; yet given the generally low level of education, children become increasingly less productive and rather create an additional financial burden for the family. A large family size only seems to be associated with happiness in educated families.

The economics of happiness is still in its infancy, yet it seems to offer promising approaches for development studies. This Chapter is a further contribution to linking these two fields. Ultimately, further ground is to be established from which also other development issues may be analysed from a more psychological perspective in conjunction with solid economic underpinnings.

Chapter 5

Socio-economic determinants of health in rural Pakistan: relative or absolute standards?

Chapter Summary

This Chapter addresses key aspects of health inequality. We analyse in particular to what extent income determines household-specific health outcomes in rural Pakistan using our survey data. Controlling for various socio-economic characteristics, we investigate the validity of three income-health hypotheses: the Absolute Income Hypothesis (AIH), the Relative Income Hypothesis (RIH) and the Income Inequality Hypothesis (IIH). Whilst these hypotheses crucially differ in their exact substance, broadly speaking, those refer to the idea that a household's health status might be linked to the existing socio-economic environment. Households with a more favourable income position (either in absolute, relative or distributional terms) might enjoy a better health status. We employ a general empirical specification that nests different health functions as special cases. This permits testing the income-health hypotheses separately and jointly. To ensure robustness of our results, we moreover employ different estimation techniques which allow for alternative health measures. We find that in rural Pakistan especially relative income (with respect to the relevant community) is a major determinant of health. This is in contrast to results typically reported for developed countries, where in particular the household's absolute income position appears to

matter. The study provides two important insights into the causes of health inequalities. Firstly, higher relative income improves health directly because of higher social support and other psychosocial reasons. Secondly, in reference to the IIH, higher income inequality improves health outcomes, specifically, of the households with higher relative income in a community.

5.1 Introduction

It has become increasingly acknowledged in both social and health sciences that a range of socio-economic factors contribute to inequalities in health. If this was the case, individual health outcomes would more and more hinge on the household's wealth and income status rather than purely immaterial determinants such as genetics or luck. In this Chapter we aim to investigate the link between income and health in rural Pakistan. Against the backdrop of a relatively underdeveloped public health system there, we seek to analyse in particular several key income-health hypotheses that are commonly considered in the health economics literature. We do so using different econometric modelling strategies and by comparing our results with other both developed and emerging economies.

The relationship between health and income tends to be framed in terms of three related hypotheses. The *Absolute Income Hypothesis* (AIH) states that a household's health depends on its own level of income, independent of the financial situation of its peers. In other words, the AIH suggests that the higher an individual's income, the lower the risk of being unhealthy. The *Relative Income Hypothesis* (RIH), on the other hand, claims that at any given level of income, the household's health depends on its income in relation to society. This implies that a higher social status provides increasing psychosocial satisfaction and thus ensures better health. For the RIH the *distribution of income* matters such that living in an environment characterised by a relatively unequal income distribution is hypothesised to cause psycho-social stress, leading to a worsening of health outcomes. This is captured by the related *Income Inequlity Hypothesis* (IIH). One may argue that living in a place with an unequal income distribution leads to social or ethnic inequalities such that relatively deprived people do not have access to a proper and efficient health care system (which might explain social and ethnic inequalities in health), and are prone to get sick and thus visit doctors more frequently. Relative income as well as the income distribution are often labelled as "psychosocial determinants" of health, even though

they are triggered by material standards such as income.

From a policy maker's perspective, it is therefore essential to disentangle the empirical importance of each of these hypotheses. Does relative income dominate over absolute one? In which form does the distribution of income matter at all? Different policy recommendations would obviously arise depending on the estimation outcomes. We therefore employ different estimators to robustify our findings. In particular, we resort to OLS estimations, ordered probit as well as Poisson regression models.

The majority of the studies for developed countries thus far tends to provide evidence in favour of the AIH in comparison to the RIH with only a few exceptions. Results on the validity of the IIH are mixed, however. Studies based on US population data, overall, tend to favour the AIH, rejecting the RIH and providing little or no support for the IIH (Wagstaff and Doorslaer, 2000). Similarly, Lindley and Lorgelly (2005) used data for the UK provided by the British Household Panel Survey *(BHPS)* to test for and to distinguish between the AIH vis-à-vis the RIH. The longitudinal nature of their data moreover allowed them to investigate the dynamic evolution of the given hypotheses. They provide strong evidence to support the AIH, and found that the RIH does not seem to hold over time within the UK. Gerdtham and Johannesson (2004) analysed Swedish panel data comprising more that 40,000 adults who were followed over 10-17 years. Whilst their results are consistent with the AIH, they fail to confirm either the RIH or IIH. Marmot et al. (1991), on the other hand, analysing health inequalities among British Civil Servants (Whitehall II study) found that in lower administrative ranks, it is indeed relative rather than absolute income that is important for health such that a low relative income has a detrimental effect on the health status. Lobmayer and Wilkinson (2000) conducted a multi-country study based on the Luxembourg Income Study (LIS).[1] Overall, the results cannot confirm the idea of more egalitarian societies reporting better health suggesting that the IIH does not hold.

Amid the inconclusive evidence on the IIH, Mellor and Milyo (2002) suggest investigating this postulated health-income relationship from two angles (i.e. in the strong or weak form). The strong version of the IIH implies that inequality adversely affects all members in a society equally, regardless of their financial status. The weak version states that income in-

[1]Their analysis covered the following 14 OECD countries in alphabetical order: Australia, Belgium, Canada, Denmark, Finland, France, (former West) Germany, Italy, the Netherlands, Norway, Spain, Sweden, the UK, the US; excluding Luxembourg given its relatively small population size.

equality is harmful to the health of only the least well off in a society. Mellor and Milyo (2002)'s analysis is based on the US Current Population Survey (CPS) data from 1995-1999 and examines the effect of income inequality on individual health status for both the general individuals and those individuals living in poverty. However, they fail to establish consistent evidence of the IIH in either form. On the other hand, Li and Zhu (2006) tested the two versions of the IIH for China. They used China Health and Nutrition Survey (CHNS) data for five years (1989, 1991, 1993, 1997 and 2000). Overall, their results provide evidence supporting the strong version but failed to confirm the IIH in its weak form.

Research has shown that micro data rather than macro data appears to be more appropriate to discriminate between the competing income-health hypotheses. For instance, Smith (1999) using aggregate data for international comparisons found that a distinction between the effects of (absolute and relative) income and income inequality on health can hardly be drawn. This is due to a potential concave association between income and individual health. This form of non-linearity in the income-health relationship causes income inequality and health to move in opposite directions on an aggregate level. Wagstaff and Doorslaer (2000), therefore, favour the use of micro data over macro data as well.

Our study is based on micro survey data and comprises 600 representative households of rural Pakistan. We test the three income-health hypotheses (AIH, RIH, IIH), while controlling for individual socio-economic characteristics of the households like gender, age, education, marital status and family size. There is a considerable literature investigating these hypotheses using individual-level data. Those studies are typically based on self-assessed health, infant health or mortality rates as proxies of health outcomes. In particular, Meara (1999) focused on infant health (i.e. low birth weight), Mellor and Milyo (2002) used the self-reported health status and Gerdtham and Johannesson (2004) considered mortality as relevant health measure. In contrast to those studies, we use as health variable the number of *visits to a doctor* ($DocVisits$) made at the household level during the given year. Our study is particularly in line with the paper by Gerdtham and Johannesson (2004), who were the first to explicitly discriminate between the three health-income hypotheses in a coherent setting. Further studies to test the hypotheses jointly are Fiscella and Franks (1997); Daly et al. (1998); Meara (1999) and Mellor and Milyo (2002). However, most of them tend to focus on the IIH without in some cases even reporting results for RIH. As an important methodological contribution to the analysis,

we consider it is important to realise the links and associations among the different health-income hypothesis. We therefore strive to set up a framework that is general enough to permit testing all three key hypotheses in a coherent setting.

Despite these unique strengths (micro data, simultaneous testing set-up, robustness checks), our analysis and dataset also carry some limitations which need to be mentioned. First, it should be noted that Pakistan is an agrarian economy with 70% of its population living in rural areas that are engaged in subsistence agriculture. Income is as a result more equally distributed because most workers have very low levels of income. In other words, incomes are concentrated at low levels and that concentration dominates the overall distribution of income. This may make the rural side of Pakistan not an ideal laboratory to test the IIH. However, it may be the case that apart from the level of income inequality, the variations in income inequality across geographical regions also matter. In such a case, there is sufficient variations in income inequality across the sample districts. That is, 97% of the total income inequality in rural Pakistan is explained by inequality between the districts. The remaining 3% are explained by inequality within the districts.[2] A further limitation is that we have measured relative income and income inequality only at the district level rather than considering the sample as a whole. It might be the case, however, that it is important to test the given hypotheses at the country level as well – which is not supported by the available data.[3] Furthermore, we assume that the association between health and socio-economic status (SES), whether measured by education, gender, or income is largely due to the effects of SES on health, not vice versa following (Doornbos and Kromhout, 1990; Fox et al., 1985; Power et al., 1990; Wilkinson, 1986). This one-way causality is in line with the most part of the health economics literature.

The remainder of the Chapter is structured as follows. We first provide in Section 5.2 general background information on the relationship between healthcare systems and their impact on health inequalities with special reference to Pakistan to motivate our health measure. Section 5.3 summarises the data and describes the empirical specification. Section 5.4 presents and discusses our empirical results. The Chapter ends with concluding remarks and policy recommendations in Section 5.5.

[2]See Chapter 3 for details.

[3]Recall that our dataset considers cross-sections and focuses on rural Pakistan only.

5.2 Inequalities in health and the healthcare system

Health depends on a number of factors, including biological factors, environmental factors, nutrition, and the standard of living. In other words, health can be seen as a function of welfare. Few of the issues which cause ill health are dealt with directly by 'health services'; they are, rather, issues in the 'welfare state' as a whole. When in the 19th century, Chadwick identified poor health as a major cause of pauperism, his response was to improve sanitation, not to introduce more extensive medical care. Most of the world's diseases are attributable to poor water supply or nutrition. 'Health services' are thus better described as medical services.

There are clear differences in the incidence of ill health by social class. All studies cited above show a close link between health and social inequality. People in lower social classes, including children, are more likely to suffer from infective and parasitic diseases, pneumonia, poisonings or violence. Adults in lower social classes are more likely, in addition, to suffer from cancer, heart disease and respiratory disease. Lower class people have more time off work, pay more visits to the doctor and are likely to be chronically ill. As part of 1999 General Household Survey, ONS statisticians looked at 1,200 workless households, containing at least one person of working age. They found that 32% of the members of workless households reported chronic illnesses, compared with 12.5% of those in working household.[4]

There are several possible explanations for these inequalities. For instance, poverty leads to ill health through nutrition, housing and environment. Another explanation may lie in cultural and behavioural aspects. There are, for example, differences in the diet and fitness of different social classes and in certain habits like smoking. There are moreover often major inequalities in access to health care according to social class. The problem becomes what Tudor Hart once called an 'inverse care law' implying that those individuals in the worst health condition receive the least services. The inverse care law proposed by Julian Tudor Hart in 1971 states that 'the availability of good medical care tends to vary inversely with the need

[4]The General Household Survey (GHS) is a survey conducted on an annual basis by the Office for National Statistics (ONS) and collects data about private households in Great Britain. The aim of this survey is to provide government departments and organisation with information on a range of topics concerning private households for monitoring and policy purposes.

for it in the population served'. The law explains the fact that poor people with chronic illness and diseases actually need and deserve good medical care but they cannot afford it and vice-versa. Apart from income, location, race/ethnicity and gender may also explain health inequalities in general.

Before proceeding with testing all the three income-health hypotheses in rural Pakistan, it may be helpful to obtain a better understanding of the existing healthcare system in Pakistan. In Pakistan, the public health services and hospitals are relatively cheap compared to the private ones, but the standard they provide is not satisfactory. It involves a lot of administrative delays and poor health facilities. The public hospitals are very inefficient compared to the private clinics/hospitals in providing proper treatments to their patients and on proper time. This leads to prolonged and chronic sickness and diseases and the patients have to suffer rather strongly and for long and need to pay as a result more visits to doctors. However, the majority of the people still go to these hospitals because they cannot afford the private doctors. Those who are relatively better-off financially, on the other hand, can visit the private doctors and receive the proper treatment within good time, and are thus able to avoid frequent visits to a doctor. Based on those features inherent in the healthcare system in Pakistan, it thus appears plausible to use *the number of times* a household visits to a doctor as a variable in order to evaluate the household's health as well as its financial status within society.

5.3 Data and methodology

Our survey data provides household-level data on health and different socio-economic variables like income, family size, education, marital status, age and sex of the household's head. The household's health status is determined by the number of visits to a doctor. The following question has been asked from the household head: "During the last 12 months, how many times has someone in your household visited a doctor?". The answers included: none, once, twice, three times, four times (or more). In the following, we define a health satisfaction index (h) that is considered here as *continuous variable*. The index is inferred from the number of visits to a doctor (DocVisits). More specifically, households are ranked according to the number of visits. A higher rank (4, 3, 2, 1, 0) corresponds to a lower frequency of visits. This implies that a lower health index index reflects a lower health status (captured by a relatively higher number of doctor visits).

5.3.1 Descriptive statistics

	No. of doctor visits (DocVisits)	**Health index** (h)
Mean	2	-
Standard deviation	1.25	-
Frequency of visits:		
0	11.33%	4
1	31.50%	3
2	30.33%	2
3	25.17%	1
4 or more times	1.67%	0

Table 5.1: Descriptive statistics for household's health variables. Source: Survey 2008.

Table 5.1 illustrates the concept and provides some descriptive statistics. The health distribution is skewed towards the lower health scales with the mass of the distribution being concentrated on that side. For instance, the given distribution shows that 11.33% and 31.50% of the observations are reporting higher health indices i.e. 4 and 3, respectively. While more than half of the population (nearly 60%) finds itselt at the lower health scales. This suggests a overall fairly unequal distribution of health.

Inequalities in health can arise for various reasons. Most countries identify differences in health status by social grouping and economic status. To single out the effects of income on health inequalities, we investigate in particular the AIH, RIH and IIH. We first want to ensure that such a proposed link between income and health exists for our data. We therefore calculate the correlation coefficient between health and the rank of a household within the income distribution (used as proxy of relative income).[5] We use *relative* income to capture the socio-economic status which is said to affect the health status (see e.g. Wagstaff et al. (1989); Humphries and van Doorslaer (2000); Gerdtham and Johanneson (2000); Doorslaer et al. (1997); Doorslaer and Koolman (2000) and Bommier and Stecklov (2002)).

[5]See Li and Zhu (2006) for further methodological details.

This can be computed as follows:

$$C_{h,R} = \frac{cov(h_i, R_i)}{\sqrt{var(h)var(R)}} = 0.224, \quad (i = 1, 2, 3.....600) \tag{5.1}$$

where h_i is the health variable of a household, R_i is the ith household's fractional rank in the income distribution, *cov* is the covariance between the two and *var* measures the variance of the given variables.[6] Equation (5.1) suggests that health is positively correlated with the income ranking of a household in the given income distribution such that higher income groups are relatively better-off in terms of health and vice versa.

The household's relative income in social comparisons is assessed with respect to the average income of the overall society or community.[7] The average income, however, may vary across a certain group, community or region (Wagstaff and Doorslaer, 2000). Economists therefore prefer to measure relative household income at the community level for cross-sectional data. In our case, the community might be considered to correspond to a district.[8] Thus, the relative income at community/district level is given by

$$(y_r)_{ij} = \frac{y_i}{\overline{y}_j},$$

where $(y_r)_{ij}$ is the relative income of i^{th} household in district j, y_i denotes absolute income of the i^{th} household and $\overline{y}_j$ represents the average income of the j^{th} district.

In analysing the relationship between income and health, we consider in particular two key hypotheses: the AIH and the RIH. The AIH seeks to investigate a relationship between the household's income and health status, whereas according to the RIH, health is rather affected by relative income differentials. The RIH builds upon the claim that low relative income increases psychosocial stress which may lead to physical illness (Cohen et al. (1991); Cohen et al. (1997)).[9] Similarly, several studies suggest that it is in

[6] In fractional ranking, items that compare equal receive the same ranking number, which is the mean of what they would have under ordinal rankings. Furthermore, in ordinal ranking, all items receive distinct ordinal numbers (1, 2, 3, and so on..), including items that compare equal.

[7] This implies that relative incomes with positive or negative values indicate the household's income to be greater than or less than the average income of society, respectively.

[8] We are using the words "community" and "district" interchangeably in the remainder.

[9] Low social status/prestige and lack of control and awareness are often labelled as psychosocial determinants of health, even though they may be triggered by material factors such as lack of income or bad housing (Kawachi et al., 2002).

fact an individual's relative income instead of absolute income that matters (Marmot et al. (1991); Wilkinson (1997); Wilkinson (1998)). If this were the case, a doubling of everyone's income would have no effect on health. Such arguments explain income effects psycho-socially, rather than in materialistic terms and have lead to a model of health in which social coherence plays an important role. Additionally, the individual's health is also attributed to the distribution of income within a society. For instance, living in a region with an unequal distribution of income by itself afflicts health (Wilkinson, 1996), which is related to psychosocial mechanisms rather than material deprivation.

Table 5.2 provides summary statistics of material and psychosocial determinants of health for our sample.[10] Income inequality is measured here in terms of the Gini coefficient. This is arguably the most commonly used measure of income inequality in testing the IIH in an attempt to establish a relationship between income inequality and health (see, amongst others, Kennedy et al. (1998); Mellor and Milyo (2001); Soobader and LeClere (1999)).

Variable	Obs	Mean	Std.Dev	Min	Max
Absolute income y_i	600	3515.58	1207.88	1596	6938
Relative income$(y_r)_{ij}$	600	1.00	0.08	0.79	1.24
Income inequality of district G_j%	10	4.42	2.19	1.6	8.1

Table 5.2: Material and psychosocial determinants of health.
Source: Survey 2008.
Note: All income figures are related to the households' monthly incomes in Pakistani currency (PKR).

In our survey, individuals are grouped into 10 districts. We estimated the inequality index G_{ij} at the household level such that households living within the same district have been assigned the same index. Inequality hence differs across districts but not across households within any given district.[11]

[10] See Kawachi et al. (2002) for further methodological details.

[11] G_{ij} is a contextual variable that varies across districts but has the same value for all the households within a district. A similar idea has been presented by (Blalock, 1984) and (Lindley and Lorgelly, 2005) in order to explain individual-level variables by using group-level variables.

	h_i	y_i	$(y_r)_{ij}$
h_i	1		
y_i	0.1865***	1	
$(y_r)_{ij}$	0.7045***	0.2148***	1

Table 5.3: Correlation matrix between health and income variables. $N = 600$ households. *, **, *** indicates significance level of 10%, 5% and 1%, respectively.

Table 5.3 reports the correlations between health and the income variables. According to Table 5.3, health is positively correlated to income in general. But the correlation is much stronger in case of relative income ($(y_r)_{ij}$) compared to the absolute income (y_i) of the household.

5.3.2 Empirical specification

The measure of health that we use as endogenous variable throughout the Chapter is the health h of i^{th} household, h_i. Health is measured on a numerical scale ranging from 0 to 4, derived from the answer to the question: "During the last 12 months, how many times has someone in your household visited a doctor?". As discussed above, answers include: 0, 1, 2, 3, 4 or greater.

Our regressions include a list of control variables relating to the households whose effects on health have been shown to be important in the literature. These control variables are in particular: age, sex, income (both absolute and relative), district-specific income inequality, marital status, kids, family size and education. More specifically, to capture *age* we use the household's head age. We create two dummies for *sex* of the household's head such that if $female = 1$ otherwise 0. The household's absolute income and relative income (with respect to the district) are expressed in logarithmic terms. Similarly, the variable measuring the degree of income inequality in the districts will be considered in percentage terms in our regressions. For marital status and kids we create dummies. If it is a *couple* = 1 (dual parent family) or otherwise 0. Similarly, the case of the household having *kids* (aged less than 16 years and living with their parents) is assigned 1 or 0 otherwise. The total number of household members determines the $familysize$. The *education* variable is measured by the number of years

of education of the household's head. Table D.1 presents the correlations between the given controls that we use as socio-economic determinants of health in our regressions. Our discussion focuses on the two relationships that are found to be statistically significant. According to Table D.1, age is negatively correlated with being a female (r = -0.1264), while the significance of the estimated coefficient confirms that life-expectancy of females is lower compared to males in developing countries like Pakistan. In line with Pakistan's demographic features (Pakistan is the sixth most populous country in the world with an average birth rate is 27.52 births/1,000 population (CIA, 2008)), this negative relationship is in line with intuition: women give birth to a relatively high number of children which adversely affects their health and life expectancy. The correlation between family size and kids is positive and statistically significant (r = 0.1592). This positive relationship confirms our initial conjecture because having kids at home obviously adds to the family size.

Following much of the literature, the baseline empirical specification that we use for studying the determinants of health is as follows

$$h_i = \alpha + \beta_1 log(y_i) + \beta_2 log(y_r)_{ij} + \gamma(G)_{ij} + B(X)_i + \epsilon_i, \qquad (5.2)$$

where the subscripts i and j in (5.2) refer to households and districts, respectively. The household's absolute income is denoted by y_i; $(y_r)_{ij}$ and G_{ij} represent district-specific measures of relative incomes and income inequality. X_i denotes the remaining controls mentioned.

Model (5.2) can be considered as the empirical counterpart of a health function of the general form h(y,y_r,G,X) which we first estimate using OLS. This specification will serve in the remainder as our benchmark model to investigate the three income hypotheses. As additional robustness check, we shall estimate the model with alternative procedures; more specifically ordered probit and using Poisson regressions.

5.4 Empirical results

5.4.1 OLS estimation

5.4.1.1 Absolute and Relative Income Hypotheses: separate and joint tests

Assuming for now that it is absolute rather than relative levels of income that matter, we would consider a health function of the form $h(y, X)$, where

y measures absolute income and X represents the given set of control variables. We estimate this health function with the following empirical specification in order to test the AIH:

$$h_i = \alpha + \beta log(y_i) + B(X)_i + \epsilon_i. \tag{5.3}$$

Alternatively, we may also consider that it is not absolute but rather relative levels of income that affect health. This would suggest a health function of the form $h(y_r, X)$, where y_r represents the relative income of the household with respect to the given district and X represents the set of control variables. For testing the RIH, we estimate the health function as follows:

$$h_i = \alpha + \beta log(y_r)_{ij} + B(X)_i + \epsilon_i. \tag{5.4}$$

We may moreover assume that people care about both absolute and relative levels of income. This implies a combined health function of the form $h(y, y_r, X)$, with y, y_r and X defined as above. This can be translated in econometric terms as follows to test the income hypotheses jointly in absolute and relative terms:

$$h_i = \alpha + \beta_1 log(y_i) + \beta_2 log(y_r)_{ij} + B(X)_i + \epsilon_i. \tag{5.5}$$

According to the RIH, however, the household's health is additionally affected by the distribution of income (G_{ij}) within a society so that living in a place with an unequal income distribution is anticipated to lead to a worsening of the health experience. This would suggest a health function of the form $h(y, y_r, G, X)$, which could be expressed in econometric terms as follows:

$$h_i = \alpha + \beta_1 log(y_i) + \beta_2 log(y_r)_{ij} + \gamma_1 (G)_{ij} + B(X)_i + \epsilon_i. \tag{5.6}$$

Since health is usually assumed to be curvilinear in income inequality, we introduce a squared inequality term (G^2) to allow for such potential non-linearities in health outcomes. This would imply accordingly a health function of the form $h(y, y_r, G, G^2, X)$, which can be specified as:

$$h_i = \alpha + \beta_1 log(y_i) + \beta_2 log(y_r)_{ij} + \gamma_1 (G)_{ij} + \gamma_2 (G)^2_{ij} + B(X)_i + \epsilon_i. \tag{5.7}$$

The estimation strategy is as follows. First, we use (5.3) and (5.4) to investigate the AIH as well as the RIH in turn. Having estimated both models separately, we use in a next step (5.5) to test both hypotheses jointly. Model (5.6) will then be used to test all three hypotheses (i.e. including the IIH) simultaneously. Finally, (5.7) is estimated to capture any potential non-linearity in health with respect to the distribution of income within a district.

Table 5.4: AIH and RIH: separate and joint tests using OLS.
Note: regression coefficients are in bold and standard errors appear below them. *,**,*** denote statistical significance at 10%, 5% and 1% levels.

OLS regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
Constant	**1.3229**	**5.5201**	**1.5811**	**1.0614**	**0.4843**
	1.7395	1.4566	1.7568	1.8542	1.9429
Age	**-0.0820***	**-0.0866***	**-0.0853***	**-0.0844***	**-0.0766**
	0.0465	0.0490	0.0471	0.0469	0.0474
Age Squared	**0.0007**	**0.0008**	**0.0008***	**0.0008***	**0.0007**
	0.0005	0.0005	0.0005	0.0005	0.0005
Female	**-0.2081**	**-0.2490**	**-0.2130**	**-0.2203**	**-0.2144**
	0.1777	0.1767	0.1759	0.1753	0.1759

Table 5.4: (continued)

OLS regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
Absolute income	**0.4948*** **		**0.4853*** **	**0.5382*** **	**0.5604*** **
	0.1308		0.1289	0.1431	0.1455
Relative income		**1.2525** **	**1.1818** **	**1.1713** **	**1.1728** **
		0.5631	0.5431	0.5392	0.5399
Income-inequality				**0.0314**	**0.1762**
				0.0266	0.1228
(Income-inequality)^2					**-0.0134**
					0.0114
Couple	**0.1343**	**0.1437**	**0.1302**	**0.1203**	**0.1157**
	0.1032	0.1043	0.1034	0.1037	0.1038
Kids	**-0.6438**	**-0.7824**	**-0.7212**	**-0.7019**	**-0.7481**
	0.5553	0.5977	0.5699	0.5647	0.5881

Table 5.4: (continued)

OLS regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
Familysize	**-0.0723**	**-0.0729**	**-0.0774***	**-0.0773***	**-0.0742***
	0.0448	0.0449	0.0447	0.0449	0.0447
Education	**-0.0070**	**0.0002**	**-0.0063**	**-0.0068**	**-0.0077**
	0.0133	0.0136	0.0132	0.0132	0.0132
Region:					
Punjab	**0.1831**	**0.1982**	**0.1916**	**0.0718**	**-0.0553**
	0.1514	0.1495	0.1477	0.1831	0.2020
NWFP	**-0.0311**	**-0.0324**	**-0.0296**	**-0.0723**	**-0.1566**
	0.1594	0.1594	0.1563	0.1601	0.1743
Sind	**-0.0111**	**0.0190**	**-0.0043**	**-0.0758**	**-0.1271**
	0.1525	0.1519	0.1492	0.1607	0.1652
Baluchistan	*Reference Group*				

Table 5.4: (continued)

OLS regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
Number of obs	600	600	600	600	600
F(k, n)	F(11, 588) = 3.06	F(11, 588) = 2.01	F(12, 587) = 3.30	F(13, 586) = 3.10	F(14, 585) = 2.94
Prob > F	0.0005	0.0258	0.0001	0.0002	0.0002
R^2	0.0597	0.0394	0.0695	0.0718	0.0739
Root MSE	0.99376	1.0044	0.98937	0.98902	0.98874
.ovtest	F(3, 585)	F(3, 585)	F(3, 584)	F(3, 583)	F(3, 582)
(Ho: Model has no omitted variables)	= 2.4 Prob > F = 0.0668	= 2.79 Prob > F = 0.0398	= 1.25 Prob > F = 0.2908	= 1.74 Prob > F = 0.1573	= 1.95 Prob > F = 0.1200

Table 5.4 summarises the benchmark results of estimating the various models using OLS. In addition, we test for a potential omitted-variable bias in each model resorting to the *ovtest* in STATA, the equivalent of the well-known Ramsey RESET test. The first column of Table 5.4 provides the test results for the AIH. We find a positive and strongly significant effect of absolute income. The second column reports in a similar vein results for the separate estimation of the relative income measure. Relative income has a sizable positive and statistically significant effect on health.

In column three of Table 5.4, we present a joint picture of the two income hypotheses. Results suggest evidence in favour of both the AIH as well as the RIH. Unlike Li and Zhu (2006) who failed to establish a significant impact of relative income on health, we find that both measures are statistically significant and of positive signs. However, as expected, compared to the effect of absolute income, the effect of relative income on household's health is more than twice as high. Ferrer-i-Carbonell (2005) use German data to empirically analyse the importance relative income for individual well-being. He finds that absolute income has a very small and not significant coefficient when included alongside relative income. It is also interesting to note that once we control for both absolute and relative income, the effects of the relative and absolute incomes remain unaffected and statistically significant as well.

We test additionally for the effects of income inequality in linear (column 4) and quadratic (column 5) terms. The positive coefficient of "income-inequality linear" and the negative coefficient of "income-inequality squared" suggest an inverted U-shaped pattern between health and income inequality. The maximum corresponds to an income inequality of 6.57%. This suggests that any increase in income inequality is beneficial for health till reaching the 6.57% threshold after which higher income inequality poses threats to health. We may therefore conclude that the IIH is only supported by districts with high income inequalities – with "high" being pinned down to about 7% in this context. However, on pure statistical grounds this relationship does not exist at all. A similar story applies to China as for instance suggested by Li and Zhu (2006) who found an inverted U-shaped (and statistically siginificant) association between individual health and community level income inequality.

As far as the model specification is concerned, the ovtest is statistically significant for the two separate tests for the AIH and the RIH and we reject the null hypothesis that the models do not have an omitted-variable bias. However, in the joint tests for the income hypotheses the ovtest remains

insignificant with higher p-values than the conventional threshold of 0.05 (95% significance). Thus, we fail to reject the null and conclude that we do not need to consider additional variables in our model specification. The test results reported in Table 5.4 suggest that both the (absolute and relative) income variables significantly influence household's health and so should remain in the model.

The Table 5.4 reveals that, apart from the income variables, all the other controls are found to be statistically insignificant for all the models considered. Age and the family size of the interviewed households, however, are negatively related to the household's health in some cases. The relationship between age and the household's health is found to be U-shaped (and statistically significant) with a minimum age of around 53 years. This suggests that the household's health is lowest around an age of 53. In other words, young and the old individuals are capable of having healthier families than the middle-aged group. Several reasons may be adduced to explain the observed U-shaped relationship between age and the household's health. For instance, many young people tend to be very successful in their lives, but with increasing age and family size their financial burden potentially increases which may adversely affect the health status of a household.

Older people, on the other hand, have developed skills to adjust their work situations to their needs. Moreover, they learn how to cope with negative life events and how to absorb negative shocks more effectively. Similarly, family size is negatively related to the household's health and does appear to describe a statistically meaningful relationship.

5.4.2 Robustness tests

We evaluate in this Section the robustness of our results employing alternative model specifications. In particular, we amend the definition of the dependent variable which thus far has been framed in rather crude terms (see Table 5.1). Rather than considering an index number, we define health in terms of a categorical variable using an ordered probit model (OPM) in Section 5.4.2.1. As further robustness check, we use a Poisson regression model in Section 5.4.2.2. This permits using the raw data (i.e. the number of doctor visits) directly without the need for mapping this information into the health variable, h.

5.4.2.1 Ordered probit modelling

Given that we want to re-estimate (5.2) in terms of health categories, we resort to an OPM. The OPM is a regression model that incorporates ordinal dependent variables. It can be thought of as an extension of the probit regression model for dichotomous dependent variables, allowing for more than two (ordered) response categories. The model is usually estimated using maximum likelihood.

In our case, the response variable health is treated as a latent variable ranging from $-\infty$ to $+\infty$ given that the exact level of health is unobserved and is denoted by $\tilde{h}$. The variable has been distributed into the following five ordinal health categories (c) conditional on the reported number of doctor visits ($DocVisits$):

$$h = \begin{cases} 0 \Rightarrow VeryPoor \quad if \quad \tau_0 = -\infty \leq \tilde{h} < \tau_1 \\ 1 \Rightarrow Poor \quad if \quad \tau_1 \leq \tilde{h} < \tau_2 \\ 2 \Rightarrow Fair \quad if \quad \tau_2 \leq \tilde{h} < \tau_3 \\ 3 \Rightarrow Good \quad if \quad \tau_3 \leq \tilde{h} < \tau_4 \\ 4 \Rightarrow Excellent \quad if \quad \tau_4 \leq \tilde{h} < \tau_5 = +\infty \end{cases}$$

Thus the latent $\tilde{h}$ is divided into J=5 ordinal categories,

$$h = c \quad if \quad \tau_{c-1} \leq \tilde{h} < \tau_c \quad for \quad c = 0 \quad to \quad J$$

When the latent variable $\tilde{h}$ crosses a cut point, the observed category changes, where the cut points τ_1 through τ_{J-1} are estimated. Some authors refer to these as thresholds. Ordered probit model results from modeling the *probit* of the cumulative response probabilities as a linear function of the covariates.

The resulting OPM in structural form is as follows:

$$\tilde{h}_i = \alpha + \beta_1 log(y_i) + \beta_2 log(y_r)_{ij} + \gamma(G)_{ij} + B(X)_i + \epsilon_i. \tag{5.8}$$

Table D.2 catalogues the coefficient estimates of five different specifications of the ordered probit regression in order test for the three income-health hypotheses, separately as well as jointly. Ensuring well-specified models throughout, we also conducted the *linktest*. The idea of this test is that if the model is properly specified, one should not be able to find any additional predictors that are statistically significant except by chance. The variable _hat should thus be a statistically significant predictor since it is the predicted value from the model. This will be the case unless the model is misspecified. On the other hand, if our model is properly specified, variable _hatsq should not have much predictive power except by chance.

Therefore, the linktest is significant for a significant _hatsq. Such a test outcome would usually suggest that either we have an omitted-variable bias or it might be the case that the link function is not correctly specified. According to Table D.2, the variable _hatsq appears to be statistically insignificant for all the given specifications of the ordered probit model for health. The main aspect to consider in this context is the significance of _hat. This basically checks whether we need more variables in our model by running a new regression with the observed outcome variable. In our case, the _hat is significant only for the joint test of the AIH and the RIH as shown in Table D.2, column (3), which means that our ordered probit model for health is correctly specified and we therefore do not require any additional variables that significantly determine the health outcome. The corresponding marginal effects for the joint test of the AIH and the RIH are reported in Table D.3.

The results of the ordered probit regression in Table D.2 are consistent with the results obtained by OLS in Table 5.4. In the ordered probit regression, a positive coefficient indicates an increased chance that a subject with a higher score on the independent variable will be observed in a higher health category. Similarly, a negative coefficient implies a relatively greater chance that a subject with a higher score on the independent variable will be observed in a lower health category. For instance, as shown in Table D.2, higher income increases the chances of being in a higher health category and the result obtained is statistically significant. However, since the protective effect of absolute income on health is relatively uncontested (compared with the effect of relative income and income inequality), we do not place very much emphasis on it.

Similarly, the results in Table D.2 show a positive association between relative income and the household's health, which means higher relative income brightens the chances of being in a higher health category. Furthermore, this association is found to be stronger and statistically significant in case of relative income compared to absolute income. In addition, we obtain a significant and negative relationship between family size and health, which indicates that households with larger families are most likely to be observed in a lower health category and vice versa. Health appears to be U-shaped in age – the estimated turning point corresponds to the age of 53 years. Such a pattern might explain the mid-age crises as young parents with children struggle relatively strongly for their career and financial management. As they turn old and their children grow up, however, their lives normally become more stable which positively affects the overall health

status of a household. The remaining controls included in the model are found to be statistically insignificant.

As far as the ancillary parameters (or cut points) are concerned, Cut1 is the estimated cut point of the latent variable $\tilde{h}$ (which is a continuous and unobservable response variable) used to differentiate very poor health from higher health categories (i.e. poor, fair, good and excellent). For example, households that had a value of -1.7035 or less on the underlying latent variable that gave rise to our health category variable would be classified as very poor. Cut2 is the estimated cut point used to differentiate very poor and poor health categories from higher health categories (fair, good and excellent). This means that households that had a value of -0.2750 or greater on the latent variable would be classified in the higher health categories. Cut3 differentiates very-poor, poor and fair health categories versus higher health categories like good and excellent such that households that had a value of 0.5775 or greater on the given latent variable would be classified in those higher health categories. Cut4 distinguishes very poor, poor, fair and good health categories from the highest category; excellent. Cut4 which is equal to 6.262 indicates that households that had a value 1.6451 or greater on health variable would fall in the excellent category for health.

The corresponding marginal effects are presented in Table D.3. According to these marginal calculations, health appears to be inverted U-shaped in age for the lower health category (1), the estimated turning point is around 56 years. In contrast, health is U-shaped in age for the higher health categories (3) and (4) with estimated tipping points of 54 to 56 years, respectively. We may conclude that before 54-56 years (middle-age) with each year increase in the age of the household, the probability to be in the lower health category (1) increases by 2.59% and at the same time the likelihood to be observed in the higher health categories (3) and (4) decreases by 1.96% and 1.80%, respectively. Conversely, after crossing the middle-age, the probability of being in the lower health category (1) decreases by 0.023% and at the same time the chance to be in the higher health categories (3) and (4) increases by 0.018% and 0.016%, respectively. Overall, we may conclude that age has a non-linear (U-shaped) effect on household's health status. Before the age of around mid-fifties households are more likely to be observed in lower health categories but after crossing that age they are most probably to lie in the higher health categories. Furthermore, the likelihood of being in the lower health categories (0) and (1) increases by 1.6% and 17.87% respectively, if the households are having

kids in their homes. These findings are in line with initial conjectures as households with kids potentially have more health problems related with kids compared to the households without children. Similarly, a larger family size is inversely related to the health of a household. For instance, a unit increase in the family size of a household increases its chance to lie in the lower health categories (0), (1) and (2) at the rate of 0.35%, 2.27% and 0.68%, respectively. At the same time, it lowers its chance to be observed in the higher health categories (3) and (4) by the amount 1.72% and 1.58%, respectively. As far as household's absolute and relative incomes are concerned, those have a positive influence on household's health. If, for example, the household income increases by 1%, its probabilty to be observed in the lower health categories (0), (1) and (2) decreases, while at the same time it is more likely to be observed in the higher health categories such as (3) and (4). However, as expected, compared to the marginal effects of absolute income, the marginal effects of relative income on health are more than twice as high.

Overall, we find evidence supporting both the AIH and the RIH. However, the effect of the hypothesis in relative terms appears more pronounced than in absolute form. We may conclude that being in better health involves more psychosocial factors rather than absolute material standards. This means that wealthier households' health is positively influenced by their higher incomes relative to their reference group. On the other hand, lower relative income weakens one's power in the allocation of efficient local health-related resources and thus leads to a poor health status, stress and potential depression.[12]

5.4.2.2 Poisson regression modelling

Thus far, we have been measuring the health of a household based on the number of visits to a doctor. Given that our collected data refers to a count variable ($DocVisits$), we additionally resort to a Poisson model that uses this information directly. We therefore fit a Poisson model to the count data that expresses the log outcome rate as a linear function of a set of predictors as follows:[13]

[12]See the discussion by Deaton (2003).

[13]The observed counts follow a Poisson distribution with probability P, where the Poisson distribution is a discrete probability distribution. The probability distribution of a Poisson random variable X (in our case the number of doctor visits per household $DocVisits$), represents the number of successes occurring in a given time interval or a specified region of space and is given by the formula:

$$log(DocVisits) = \alpha + \beta_1 log(y_i) + \beta_2 log(y_r)_{ij} + \gamma_1 (G)_{ij} + B(X)_i \quad (5.9)$$

In interpreting the results of our model (5.9), we present both the regression coefficients as well as the corresponding marginal effects in Table D.4 and Table D.5, respectively. The marginal effects are conventionally calculated at the mean of the independent variables. The number of doctor visits serves here as a proxy for household health such that higher number of visits reflects poorer health of a household and vice-versa. In addition, we tested for the appropriateness of the Poisson regression model (5.9) using the linktest. The variable _hat is statistically significant which implies that the model is properly specified; the variables included are appropriate and the functional form of the model or the variables included is correct. Furthermore, the variable _hatsq is statistically insignificant which usually means that either we have no omitted relevant variable(s) or our link function is correctly specified.[14]

Table D.4 and Table D.5, respectively, report the coefficient estimates and corresponding marginal effects of the Poisson model (5.9) of doctor visits. The age of the interviewed household head, which is specified as a quadratic function, has a nonlinear effect on number of doctor visits, peaking at an age of around 55 years and falling thereafter. This is in line with intuition as younger parents of childbearing age have young children and more health issues compared to old parents with grown up children. While, as expected, the household size is positively associated with the number of doctor visits by a household, household incomes in absolute as well as in relative terms have negative effects on the number of doctor visits. This indicates that higher income has a positive influence on household's health. However, the effect of relative income is about four times larger compared to absolute income. This result appears plausible for a developing country

$$P(X) = \frac{e^{-\mu}\mu^x}{x!},$$

where x= 0,1,2,3,4...., e= 2.71828, μ= mean number of successes in the given time interval or region of space. If μ is the average number of successes occurring in a given time interval or region in the Poisson distribution, then the mean and the variance of the Poisson distribution are both equal to μ. E(X) = μ and V(X) = $\sigma^2 = \mu$. Note that in a Poisson distribution only one parameter, μ, is needed to determine the probability of an event. In our case, μ, or the average number of doctor visits per household is equal to 1.8.

[14]In a Poisson regression model, we consider a logarithmic link function of the outcome variable on the left hand side of the equation.

like Pakistan, where the majority of the people are poor and it is in fact the relative terms that matter. Income inequality is negatively associated with the number of doctor visits. Similarly, the interaction term between income inequality and relative income (II*RI) is also negative (Table D.4, column(6)) which suggests that income inequality favors the richer people in a community in terms of their health. The other controls are found to be statistically insignificant and are thus not discussed further.

Summing up, the poisson regression model (5.9) confirms both the AIH and RIH as shown in Table D.4 and Table D.5. However, the implications surrounding the RIH are much stronger than those for the AIH. The RIH suggests that people in less advantaged circumstances are prone to get sick and visit a doctor more frequently amid their relatively unhealthy life style. However, we failed to confirm the hypothesis that more egalitarian societies are characterised by better health. Hence, we cannot provide evidence for the IIH. This finding is in line with Lobmayer and Wilkinson (2000) who, using data on 14 OECD countries, failed to accept the IIH either suggesting that income inequality is rather beneficial for health.

5.5 Concluding remarks

In this Chapter we have sought to investigate the validity of three key income-health hypotheses: the AIH, the RIH and the IIH using OLS, ordered probit and Poisson regression modelling. The analysis is based on household survey data (Survey 2008) for rural Pakistan. We specify a health function that is general enough to permit a separate as well as simultaneous investigation. We find that all estimation techniques, controlling for various socio-economic characteristics, provide a stronger evidence in favour of the RIH compared to AIH and no support for the IIH. Relative income appears to have a significantly positive effect on health outcomes in rural Pakistan. This finding is in contrast to the general view established for developed countries according to which incomes in absolute terms seem to be the main driver of a household's health status. As robustness check, we employ alternative health measures. Given that the ordered probit model captures the exact (but unobserved) level of health with respect to the number of doctor visits, we prefer this model over OLS estimations or Poisson specifications.

Conventional economic models tend to assume that absolute income levels are the primary determinant of individual well-being. This assumption is

not confirmed by our analysis. Rather, we find that if everyone's income is roughly the same (thus narrowing income inequalities), well-being remains nearly unaffected. Relative income seems to be a better predictor of well-being in rural Pakistan. Contrary to IIH, we find that rising inequality improves health and favors the richer segment of the society. Understanding the relationship between income and health is of obvious relevance to policymakers. In a "first-best" world, health differentials should not hinge on material factors after all. The relationship between the socio-economic environment and health outcomes may be particularly relevant for regions with weak public health provision such as rural Pakistan. The Pakistani government should therefore take note of this issue in areas, where relatively poor health care system exists. Thus, policymakers concerned about health would be well advised to improve the quality as well as the number of public health care units and hospitals, especially in areas which are considerably far away from the major cities like Islambad in Punjab, Karachi in Sind, Peshawar in NWFP and Quetta in Baluchistan.

The scope of our analysis is confined to the rural parts of Pakistan. Our results apply to those districts and should not simply be generalised to Pakistan as a whole. Further research on the urban and more modern sectors would greatly inform the debate and indeed our analysis proposed here may serve as blueprint in that regard.

Chapter 6

General conclusions and directions for future research

Income inequality and poverty have always been at the core of policy debates in the developing world. The common way of trying to tackle these issues lies in redistributing income from the rich to the poor either within society or even across countries. It is hoped that the mere flow of funds would serve as panacea to all the problems faced by households even in remote areas. However, as we seek to argue in this monograph, more nuanced policy responses might be available, once main sources and drivers of inequalities have been identified and understood. This dissertation intends to advance that knowledge in particular for rural Pakistan using our unique survey data (Survey 2008) that has been especially collected to seek answers to these issues.

We investigate different forms of material and immaterial inequalities and their relationship to well-being. We decompose in Chapter 3 income inequality by its root sources. This is crucial for policymakers as it allows them to better understand the potential repercussions of negative shocks such as natural disasters or household members' illness on income inequality and poverty. In contrast to previous studies, we do not only confine poverty to the monetary dimension but also consider immaterial factors such as subjective well-being and health. Indeed, the crossover between happiness economics and development studies introduces a subjective approach to poverty measurement which we investigate in more detail in Chapter 4. The relationship between health (i.e. physical well-being) and income (i.e. material well-being) is analysed in absolute, relative as well distributional

terms in Chapter 5.

Following is an overall conclusion of the monograph, summarising the thesis' main findings, highlighting important contributions to the literature and providing further policy implications as well as pointing out avenues of future research.

Main findings

In relation to the issue of income inequality, we find that transfer income in the rural sector contributes the largest share of total household income and tends to decrease overall inequality in that sector. A decrease in the share of transfer income would imply an increase in inequality within the rural sector, whereas there appears to be a positive relationship between the share of non-farm income and income inequality. Our results contradict most of the other studies on rural Pakistan that in particular highlight agricultural income as important driver of inequalities.

The dissertation moreover contributes to the novel field of happiness studies, particularly in relationship with development economics. We distinguish among various well-being measures to assess poverty in rural Pakistan. The estimated subjective well-being functions contain variables pertaining to both the income as well as the capabilities approach to poverty measurement. For instance, our well-being metric for satisfaction with income or expenditure captures important explanatory variables like sex, education, children, income and health. It is interesting to note that income matters to poverty also on subjective grounds. However, we have failed to establish statistically significant evidence for income using alternative well-being measures such as overall satisfaction of a household. This finding seems to challenge the conventional approach according to which happiness (or overall satisfaction) is often used for poverty measurement. Our data for rural Pakistan suggest that financial well-being (i.e. satisfaction with income or expenditure) contains more socio-economic information of a household than the conventional approach. We conclude that it is possible to view financial well-being as an encompassing concept that corresponds not only to the income approach but also to the physical functioning approach as well as the social functioning approach to poverty.

We establish further key findings regarding the relationship between income and health. This is a topical issue because there is compelling empirical evidence to suggest that there is indeed an association between these two variables. Unlike much of the literature for developed countries, the study for rural Pakistan based on our survey data (Survey 2008) strongly

supports the RIH but provides little support for the AIH and no support for the IIH. Our study moreover reveals that income inequality has positive effects on health in rural Pakistan generally and in particular favours those households with higher relative income.

Contributions to the literature

Our analysis tackles novel aspects in the literature, seeking to strengthen links between (sub-)disciplines of economics. In doing so, we fill an important gap in the literature, particularly by considering the rural parts of Pakistan that have typically been neglected in the literature. Our results are largely in line with other studies on developing economies, whilst at the same making further distinct contributions.

A number of empirical studies in developing countries have identified the contribution of different sources of income to total income inequality. Glewwe (1986)'s decomposition analysis, for example, simply differentiates between labour and non-labour income, while the work by Nugent and Walther (1982) is based on only three sources (agricultural, nonagricultural, and transfer). Chinn (1979) includes rental income in nonfarm income, while Matlon (1979) includes livestock income in nonfarm income. The present dissertation adds to this as it uses data in more disaggregated form to show the contribution of five different sources of income (i.e. transfer, agricultural, nonfarm, livestock and rental) to total income inequality.

While income inequality may effect a society and its economic development in many ways, we focus in this thesis on a particular aspect of the socio-economic effects of inequality, namely its impact on health. The relationship between income inequality and health is an issue which has attracted the attention of a variety of social science disciplines such as economics, sociology and public health. We are amongst the first to provide evidence that rising income inequality leads to increasing inequalities in health in developing countries as diverging incomes favour the relatively richer segments of society and improves their provision of health services.

Little research has been done on the relationship between well-being and the notion of poverty in general and in particular for poor countries. Our analysis also makes a contribution to the literature in that regard by providing three different well-being measures: (i) overall satisfaction (life satisfaction), (ii) satisfaction with income and (iii) satisfaction with expenditure in order to capture so-called subjective well-being poverty. We investigate to what extent these concepts are competing and to what extent they are complementary to encompass poverty. We conclude that it is possible to

view subjective well-being as an encompassing concept, which permits us to quantify the relevance and importance of the other approaches and of their various components. Thus, our methodology effectively provides weights of the relative importance of these various measures of poverty. Our definition of poverty involves the household's own perception of well-being as to what constitutes a good quality of life or a bad one. We moreover develop a methodology for using subjective well-being as the criterion for evaluating poverty, and illustrate its use by reference to the survey data for rural Pakistan containing detailed socio-economic information on the individual and the household, as well as information on reported subjective well-being. Finally, we inquire how the insights gained from the study of households' happiness or well-being in economics helps policymakers to alleviate poverty. For instance, our results seem to suggest that higher income satisfaction is closely related to higher educational achievement. All of these constitute important contributions to the literature.

Policy implications and directions for further research

Overall, the findings derived from this thesis may provide important policy insights on reducing income inequalities and poverty in rural Pakistan. We find that in the rural areas of Pakistan transfer income appears to be an effective way to help poor households out of poverty, hereby lowering income inequalities at the same time. The local government should consider encouraging policies that provide access to credit for the poor, education and job training opportunities for formal sector employment in order to increase income opportunities for low income groups in the urban sector. Rural migrants to urban centres (especially to industrial zones) are usually members of low income households that lack employment and/or agricultural land. Therefore, if these migrants were employed in the urban sector, their remittances (i.e. transfer income) to their rural households would help increase income of the poorer rural households. Such a practice will help to reduce income disparities in the rural areas of Pakistan. However, further research on this proposed mechanism is required to identify the sources of income inequality within the urban sector. This would help in analysing and comparing the urban and rural dimensions of income inequality in Pakistan.

Furthermore, the thesis indicates that income inequality favours the richer households in terms of their health status. This justifies a stronger spatial and ethnic dimension of redistribution practices to counter the impacts of this imbalance. Rural and more disadvantaged areas need to experience a greater level of investment in infrastructure and human capital

(for instance, hospitals and schools etc.) in order to change their resources, which may improve their living standard in general.

In addition to income inequality, poverty deserves more attention as it could cease future economic development. Our results suggest that apart from income, education and health are other important tools for poverty alleviation programmes in rural Pakistan. Policymakers interested in poverty reduction should broaden income, education and health opportunities in rural Pakistan in general and particularly for the low income households.

Appendix A

Appendix Chapter 2

	2006	2007
GDP per capita (average annual growth)	4.7	4.2
Inflation (CPI)	7.9	7.7
Unemployment rate (% of total labour force)	6.0	5.3

Table A.1: Overview of selected main economic indicators in Pakistan. Source: World Bank

Country	Poverty measure	Income inequality measure
Pakistan	13.4	33.0
India	34.7	32.5
Bangladesh	36.0	31.8

Table A.2: Overview of poverty and income inequality measures of neighbouring countries.
Note: Poverty measure is the percentage of population living below $1 per day. Inequality is measured by the Gini coefficient (%) for each country.
Source: World Bank (2005)

Province	**District**	**Village/Chak**
Punjab	Attock	Village Behboodi
		Village Ghorghushti
	Layyah	Village Kotsultan
		Village Jakhar
	Rahimyar Khan	Village Trinda Ali Murad Khan
		Village Bapraula
	Sahiwal	Ayub's village
		Arifabad village
Sind	Badin	Village Allah Dino soomro
		Village Muhammad Suleman Thebo
	MirpurKhas	Village Khudabad
		Village FakerGhulam Ali Lighari
	Thatta	Village Jati
		Village Chato Chand
NWFP	Dir	Village Karo
		Village Toormang
	Malakand	Village Landakay
		Village Brikot
Baluchistan	Kalat	Village Takht
		Village Mastang

Table A.3: List of selected villages

Appendix B

Sample Questionnaire

INCOME INEQUALITIES AND WELL-BEING IN RURAL PAKISTAN: A HOUSEHOLD SURVEY

This questionnaire is especially designed for the comparative analysis of socio-economic conditions of different districts of rural Pakistan. It is meant for academic purposes only. The information provided here will not be disclosed to any public authority and will be treated strictly confidential. We therefore ask the interviewee to give actual data about their income and other socio-economic profiles.

Cover Page

Province:————
Total rural population of the province: ————
District:————
Village:————-
Surveyor:————
Supervisor:————-
Period of Survey:————-

Interview Schedule

SECTION 1: RESPONDENTS CHARACTERISTICS

1. Name of the Household Head (If applicable):————
2. Sex: i. Male———— ii. Female————-

3. Age:——————————

4.(a) Family size (with complete information about the age and sex of each household, starting with the household head and end at young children):——————————————————————————
————————————————

——————————————————————

(b)Earners in the Family————————————

(c) Number of working adults in the family——————————

(d) Family Structure——————————————

5. What is the highest grade you have completed in school? In which level?

(a) None——————

(b) "5 Years" School (Primary)————

(c) "8 Years" School (Middle)————-

(d) Secondary General——————

(e) Intermediate——————————

(f) Graduate——————————

(g) Post Graduate——————————-

(h) Higher Education——————-

(i) Other Please Specify——————

6. Marital Status:

(a) Couple————

(b) Widow/Widower————

(c) Divorced————

(d) Seperated————

7. Current Work Status of the Household Head?

(a) Employed——————

(b) Unemployed——————

SECTION 2: INCOME

8. Daily family income (in Pak .Rs) ——————————

9. Total Monthly Income (in Pak. Rs) ——————

10. Total annual Income(in Pak. Rs)——————

11. Distribution of income:

Monthly distribution of income

(a) Non Farm income——————

(b) Farm income——————-

(c) Transfer income——————

(d) Livestock income——————

(e) Rental income——————

(f)Total monthly income————

Annual distribution of income

(a) Non Farm income————

(b) Farm income————-

(c) Transfer income————

(d) Livestock income————

(e) Rental income————

(f)Total annual income————

SECTION 3: SUBJECTIVE WELL-BEING POVERTY

12. How satisfied are you with your current socio-economic status?

(a) Not at all satisfied————————

(b) Less than satisfied————————

(c) Rather satisfied————————

(d) Fully satisfied————————-

13. How satisfied are you with your current income?

(a) Not at all satisfied————————

(b) Less than satisfied————————

(c) Rather satisfied————————

(d) Fully satisfied————————-

14. How satisfied are you with your current expenditure?

(a) Not at all satisfied————————

(b) Less than satisfied————————

(c) Rather satisfied————————

(d) Fully satisfied————————-

SECTION 4: HEALTH

15. During the last 12 months, how many times has someone in your household visited a doctor?

(a) None————————-

(b) Once————————-

(c) Twice————————

(d) Three Times————————-

(e) Four Times or More————-

Appendix C

Appendix Chapter 4

Overall satisfaction	Punjab	Sind	NWFP	Baluchistan
4	57.08	25	15	3.33
3	10.83	1.11	3.33	1.67
2	5.83	0	0.83	15
1	26.25	73.89	80.83	80
Total(%)	100	100	100	100
Average overall satisfaction	2.99	1.77	1.53	1.28
Satisfaction with income	Punjab	Sind	NWFP	Baluchistan
4	11.67	0	0	0
3	57.5	22.22	12.5	0
2	30.83	77.78	87.5	83.33
1	0	0	0	16.67
Total(%)	100	100	100	100
Average satisfaction with income	2.81	2.22	2.13	1.83
Satisfaction with expenditure	Punjab	Sind	NWFP	Baluchistan
4	6.25	0	0	0
3	62.92	22.22	12.5	0
2	30.83	72.78	87.5	90
1	0	5	0	10
total(%)	100	100	100	100
Average satisfaction with expenditure	2.75	2.17	2.13	1.90

Table C.1: Alternative measures of subjective poverty across provinces.
Note: $N = 600$. Source: Survey 2008.

Table C.2: Results by family type.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets [ ] are p-values.

ordered probit regression	Dependent Variable: Overall satisfaction					
	All families		Dual parent family		Single parent family	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Education	0.0362**	0.0184	0.0365*	0.0209	0.0196	0.0373
Unemployed	-0.3857***	0.1365	-0.3658**	0.1803	0.0344	0.3332
Log of relative income	0.3841	0.7359	0.3971	0.8025	-0.5452	1.9942
No. Of children:						
1	-7.2185***	0.4046	-7.6688***	0.4539	*(dropped)*	
2	-0.2984	0.4255	-0.0130	0.4991	-1.1416*	0.6605
3	-0.4199	0.3692	-0.2742	0.4311	-1.2334*	0.7332
4	-0.3868	0.3562	-0.2606	0.4059	-1.2904*	0.7390
5	-0.2603	0.3658	-0.2218	0.4157	-0.6127	0.7482
6	-0.1653	0.3869	-0.1821	0.4330	0.1286	0.7676
7 or more	*Reference Group*				*(NaN)*	

Table C.2: (continued)

ordered probit regression	Dependent Variable: Overall satisfaction					
	All families		Dual parent family		Single parent family	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Region:						
Punjab	1.4575***	0.1843	1.5522***	0.2068	1.3293***	0.4420
NWFP	0.1050	0.2121	0.2028	0.2313	-0.3103	0.5509
Sind	0.4920**	0.2014	0.5332**	0.2272	0.3217	0.4710
Baluchistan	*Reference Group*					
/cut1	0.7568	0.4138	0.9303	0.4723	0.1285	0.8037
/cut2	0.8750	0.4134	1.0395	0.4721	0.2905	0.8019
/cut3	1.1170	0.4173	1.2703	0.4764	0.6040	0.8030
Log pseudolikelihood	-555.8310		-433.8928		-115.8851	
Obs	600		471		129	
Wald $\chi^2(12)$	2597.23		3181.27		47.47	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1130		0.1498	

Table C.2: (continued)

	Dependent Variable: Overall satisfaction					
	All families		Dual parent family		Single parent family	
.linktest	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
-hat	0.9450***	0.2383	1.1921***	0.4286	1.0198***	0.1878
	[0.000]		[0.005]		[0.000]	
-hatsq	0.0371	0.1684	-0.1003	0.2337	0.2491	0.1576
	[0.826]		[0.668]		[0.114]	
/cut1	0.7507	0.0856	0.9844	0.1471	0.2355	0.1738
/cut2	0.8689	0.0890	1.0937	0.1506	0.3972	0.1775
/cut3	1.1110	0.0960	1.3246	0.1539	0.7143	0.1737
Log pseudolikelihood	-555.8045		-433.7863		-114.9549	
Obs	600		471		129	
Wald $\chi^2(12)$	121.07		96.66		29.49	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1132		0.1567	
. test	$\chi^2(6)$	1041.04	$\chi^2(6)$	1163.64	$\chi^2(5)$	13.82
(using dummies for No. of children)	prob>χ^2	0.0000	prob>χ^2	0.0000	prob>χ^2	0.0168

Table C.3: Marginal effects after oprobit; results by family type. Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

All families

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.46258935

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.04708689

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= 0.09532151

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.39500225

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5403***	0.0254	-0.0471***	0.0126	-0.0954***	0.0180	-0.3978***	0.0260
2*	0.1185	0.1665	-0.0015	0.0049	-0.0079	0.0161	-0.1091	0.1457
3*	0.1662	0.1432	-0.0019	0.0041	-0.0106	0.0133	-0.1537	0.1265
4*	0.1531	0.1395	0.0000	0.0012	-0.0064	0.0072	-0.1468	0.1323
5*	0.1035	0.1449	-0.0005	0.0021	-0.0053	0.0100	-0.0977	0.1331
6*	0.0658	0.1541	-0.0002	0.0016	-0.0032	0.0095	-0.0625	0.1431

Table C.3: (continued)

Dual parent family

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.43675021

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.0432888

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= .09169836

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.42826263

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5670***	0.0282	-0.0432***	0.0145	-.0918***	0.0193	-0.4320***	0.0297
2*	0.0051	0.1961	0.0001	0.0020	-0.0001	0.0036	-0.0051	0.1954
3*	0.1088	0.1710	0.0001	0.0018	-0.0039	0.0097	-0.1051	0.1601
4*	0.1027	0.1594	0.0001	0.0014	-0.0020	0.0042	-0.1016	0.1569
5*	0.0879	0.1651	0.0004	0.0008	-0.0026	0.0074	-0.0857	0.1576
6*	0.0722	0.1722	0.0003	0.0007	-0.0021	0.0075	-0.0704	0.1645

Table C.3: (continued)

Single parent family

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.56578813

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.06260246

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= 0.11088915

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.26072027

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	*(dropped)*							
2*	0.3427***	0.1268	-0.0376	0.0272	-0.0773	0.0510	-0.2278***	0.0702
3*	0.4241**	0.2016	-0.0306	0.0232	-0.0748	0.0505	-0.3187**	0.1469
4*	0.4606**	0.2230	-0.0251	0.0181	-0.0694	0.0440	-0.3660**	0.1812
5*	0.2274	0.2549	-0.0146	0.0234	-0.0387	0.0532	-0.1741	0.1829
6*	-0.0509	0.3055	0.0016	0.0070	0.0062	0.0336	0.0431	0.2652

Table C.4: Results by age.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets[ ] are p values.

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals aged 40 or below		Individuals aged above 40	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Education	0.0362**	0.0184	0.1234***	0.0285	0.0050	0.0233
Unemployed	-0.3857***	0.1365	-0.7060***	0.2465	-0.2805*	0.1651
Log of relative income	0.3841	0.7359	0.1400	1.2600	0.4267	0.9438
No. Of children:						
1	-7.2185***	0.4046	*(dropped)*		-7.0798***	0.4574
2	-0.2984	0.4255	-0.6267	0.7467	-0.1175	0.5480
3	-0.4199	0.3692	-0.2096	0.7460	-0.4671	0.4248
4	-0.3868	0.3562	-0.4317	0.7293	-0.3146	0.3970
5	-0.2603	0.3658	-0.4446	0.7363	-0.1002	0.4106
6	-0.1653	0.3869	-0.5668	0.7458	-0.0730	0.4448
7 or more	*Reference Group*					

Table C.4: (continued)

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals aged 40 or below		Individuals aged above 40	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Region:						
Punjab	1.4575***	0.1843	1.8123***	0.4491	1.3218***	0.2283
NWFP	0.1050	0.2121	0.1265	0.5105	0.0825	0.2502
Sind	0.4920**	0.2014	0.8080*	0.4790	0.3228	0.2381
Baluchistan	*Reference Group*					
/cut1	0.7568	0.4138	1.5162	0.8684	0.5024	0.4682
/cut2	0.8750	0.4134	1.6815	0.8720	0.6042	0.4672
/cut3	1.1170	0.4173	2.0964	0.8773	0.7763	0.4737
Log pseudolikelihood	-555.8310		-191.6650		-349.9552	
Obs	600		206		394	
Wald $\chi^2(12)$	2597.23		75.88		2336.42	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1762		0.1099	

Table C.4: (continued)

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals aged 40 or below		Individuals aged above 40	
.linktest	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
-hat	0.9450***	0.2383	0.5221*	0.3167	1.1026***	0.2697
	[0.000]		[0.090]		[0.000]	
-hatsq	0.0371	0.1684	0.1511	0.1038	-0.1102	0.2882
	[0.826]		[0.145]		[0.702]	
/cut1	0.7507	0.0856	1.2248	0.2418	0.4874	0.1051
/cut2	0.8689	0.0890	1.3900	0.2435	0.5893	0.1075
/cut3	1.1110	0.0960	1.8074	0.2638	0.7615	0.1124
Log pseudolikelihood	-555.8045		-190.9327		-349.8696	
Obs	600		206		394	
Wald $\chi^2(12)$	121.07		66.90		69.90	
Prob > χ^2	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1793		0.1101	
. test	$\chi^2(6)$	1041.04	$\chi^2(5)$	2.32	$\chi^2(6)$	790.39
(using dummies for No. of children)	prob>χ^2	0.0000	prob>χ^2	0.8034	prob>χ^2	0.0000

Table C.5: Marginal effects after oprobit; results by age.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

All individuals

Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.46258935
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= 0.04708689
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= 0.09532151
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= 0.39500225

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5403***	0.0254	-0.0471***	0.0126	-0.0954***	0.0180	-0.3978***	0.0260
2*	0.1185	0.1665	-0.0015	0.0049	-0.0079	0.0161	-0.1091	0.1457
3*	0.1662	0.1432	-0.0019	0.0041	-0.0106	0.0133	-0.1537	0.1265
4*	0.1531	0.1395	0.0000	0.0012	-0.0064	0.0072	-0.1468	0.1323
5*	0.1035	0.1449	-0.0005	0.0021	-0.0053	0.0100	-0.0977	0.1331
6*	0.0658	0.1541	-0.0002	0.0016	-0.0032	0.0095	-0.0625	0.1431

Table C.5: (continued)

Individuals aged 40 or below

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.43232246

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.06563723

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= 0.16104829

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.34099202

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	*(dropped)*							
2*	0.2443	0.2725	-0.0078	0.0211	-0.0424	0.0701	-0.1942	0.1839
3*	0.0830	0.2962	0.0004	0.0022	-0.0086	0.0372	-0.0747	0.2575
4*	0.1696	0.2838	0.0011	0.0029	-0.0166	0.0328	-0.1542	0.2518
5*	0.1757	0.2879	-0.0011	0.0083	-0.0220	0.0482	-0.1526	0.2330
6*	0.2230	0.2832	-0.0044	0.0153	-0.0336	0.0611	-0.1850	0.2090

Table C.5: (continued)

Individuals aged above 40

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.47398748

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.040592

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= 0.0680626

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.41735791

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5303***	0.0317	-0.0406***	0.0137	-0.0681***	0.0191	-0.4216***	0.0322
2*	0.0468	0.2183	-0.0002	0.0023	-0.0014	0.0087	-0.0452	0.2075
3*	0.1840	0.1620	-0.0025	0.0051	-0.0083	0.0113	-0.1732	0.1464
4*	0.1249	0.1565	-0.0001	0.0011	-0.0030	0.0047	-0.1217	0.1516
5*	0.0399	0.1637	-0.0001	0.0008	-0.0010	0.0050	-0.0388	0.1579
6*	0.0291	0.1774	0.0000	0.0007	-0.0007	0.0052	-0.0283	0.1715

Table C.6: Results by income.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets [ ] are p-values.

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals above average income		Individuals below average income	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Education	0.0362**	0.0184	0.0678***	0.0270	0.0173	0.0242
Unemployed	-0.3857***	0.1365	-0.1771	0.2153	-0.4611***	0.1760
Log of relative income	0.3841	0.7359	0.5866	1.0349	0.0105	1.0360
No. Of children:						
1	-7.2185***	0.4046	*(dropped)*		-6.9450***	0.4687
2	-0.2984	0.4255	-0.8636	0.7245	-0.2712	0.4852
3	-0.4199	0.3692	-1.5465**	0.6552	0.0948	0.4341
4	-0.3868	0.3562	-1.1201*	0.6410	-0.1488	0.4114
5	-0.2603	0.3658	-0.7487	0.6364	-0.3398	0.4544
6	-0.1653	0.3869	-0.7518	0.6699	0.0539	0.4547
7 or more	*Reference Group*					

Table C.6: (continued)

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals above average income		Individuals below average income	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Region:						
Punjab	1.4575***	0.1843	0.8901***	0.1904	1.5897***	0.2081
NWFP	0.1050	0.2121	0.1254	0.3814	0.1081	0.2232
Sind	0.4920**	0.2014	*(dropped)*		0.2016	0.2768
Baluchisatn	*Reference Group*					
/cut1	0.7568	0.4138	-0.1976	0.6584	0.9242	0.4818
/cut2	0.8750	0.4134	-0.0810	0.6550	1.0509	0.4822
/cut3	1.1170	0.4173	0.1553	0.6548	1.3125	0.4871
Log pseudolikelihood	-555.8310		-209.12822		-332.59519	
Obs	600		228		372	
Wald $\chi^2(12)$	2597.23		51.12		2377.05	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1217		0.1493	

Table C.6: (continued)

	Dependent Variable: Overall satisfaction					
	All individuals		Individuals above average income		Individuals below average income	
.linktest	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
-hat	0.9450***	0.2383	0.9990***	0.1479	1.0590***	0.3118
	[0.000]		[0.000]		[0.001]	
-hatsq	0.0371	0.1684	0.2914	0.2220	-0.0369	0.1994
	[0.826]		[0.189]		[0.853]	
/cut1	0.7507	0.0856	-0.0928439	0.1319251	0.9283	0.1104
/cut2	0.8689	0.0890	0.0241273	0.1265235	1.0551	0.1147
/cut3	1.1110	0.0960	0.2617294	0.1268558	1.3166	0.1280
Log pseudolikelihood	-555.8045		-208.16165		-332.57497	
Obs	600		228		372	
Wald $\chi^2(12)$	121.07		47.06		91.33	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1258		0.1493	
. test	$\chi^2(6)$	1041.04	$\chi^2(5)$	12.25	$\chi^2(6)$	800.62
(using dummies for No. of children)	prob>χ^2	0.0000	prob>χ^2	0.0315	prob>χ^2	0.0000

Table C.7: Marginal effects after oprobit; results by income. Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

All individuals

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.46258935

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.04708689

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= 0.09532151

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.39500225

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5403***	0.0254	-0.0471***	0.0126	-0.0954***	0.0180	-0.3978***	0.0260
2*	0.1185	0.1665	-0.0015	0.0049	-0.0079	0.0161	-0.1091	0.1457
3*	0.1662	0.1432	-0.0019	0.0041	-0.0106	0.0133	-0.1537	0.1265
4*	0.1531	0.1395	0.0000	0.0012	-0.0064	0.0072	-0.1468	0.1323
5*	0.1035	0.1449	-0.0005	0.0021	-0.0053	0.0100	-0.0977	0.1331
6*	0.0658	0.1541	-0.0002	0.0016	-0.0032	0.0095	-0.0625	0.1431

Table C.7: (continued)

Individuals
above
average income
Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.41757285
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= 0.04596276
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= 0.09400195
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= 0.44246244

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	*(dropped)*							
2*	0.3276	0.2378	-0.0093	0.0179	-0.0296	0.0406	-0.2886	0.1819
3*	0.5377***	0.1598	-0.0190	0.0162	-0.0527*	0.0310	-0.4660***	0.1227
4*	0.4239**	0.2201	0.0000	0.0054	-0.0156	0.0162	-0.4083**	0.2038
5*	0.2916	0.2380	-0.0007	0.0058	-0.0126	0.0193	-0.2784	0.2153
6*	0.2926	0.2452	-0.0036	0.0108	-0.0182	0.0287	-0.2708	0.2077

Table C.7: (continued)

Individuals
below
average income
Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.49286505
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= 0.05045866
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= 0.10111774
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= 0.35555855

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5119***	0.0350	-0.0505***	0.0166	-0.1014***	0.0261	-0.3601***	0.0340
2*	0.1072	0.1878	-0.0024	0.0073	-0.0096	0.0225	-0.0951	0.1584
3*	-0.0378	0.1726	0.0001	0.0005	0.0020	0.0081	0.0357	0.1648
4*	0.0593	0.1637	-0.0004	0.0014	-0.0037	0.0106	-0.0553	0.1520
5*	0.1343	0.1758	-0.0027	0.0063	-0.0115	0.0200	-0.1201	0.1502
6*	-0.0215	0.1811	0.0001	0.0003	0.0012	0.0092	0.0202	0.1717

Table C.8: Results by education.
Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. Figure in brackets [ ] are p-values.

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals with full school education		Individuals with less than full school education	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Education	0.0362**	0.0184	-0.0431	0.0536	0.1122***	0.0434
Unemployed	-0.3857***	0.1365	-0.5083**	0.2406	-0.2611	0.1698
Log of relative income	0.3841	0.7359	-0.5756	1.2257	0.7207	0.9136
No. Of children:						
1	-7.2185***	0.4046	-6.7719***	0.3401	*(dropped)*	
2	-0.2984	0.4255	0.4699	0.5970	-0.3720	0.4713
3	-0.4199	0.3692	-0.2031	0.3275	-0.0616	0.4510
4	-0.3868	0.3562	0.2368	0.2852	-0.2291	0.4279
5	-0.2603	0.3658	*(dropped)*		0.2092	0.4441
6	-0.1653	0.3869	-0.0091	0.3182	0.2772	0.4818
7 or more	*Reference Group*					

Table C.8: (continued)

ordered probit regression	Dependent Variable: Overall satisfaction					
	All individuals		Individuals with full school education		Individuals with less than full school education	
Independent variable	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Region:						
Punjab	1.4575***	0.1843	1.1989***	0.3623	1.5735***	0.2215
NWFP	0.1050	0.2121	-0.3330	0.4136	0.1829	0.2534
Sind	0.4920**	0.2014	0.0777	0.3964	0.5135	0.2355
Baluchistan	*Reference Group*					
/cut1	0.7568	0.4138	-0.1140	0.7530	1.5643	0.5767
/cut2	0.8750	0.4134	-0.0391	0.7524	1.7101	0.5779
/cut3	1.1170	0.4173	0.1698	0.7590	1.9822	0.5861
Log pseudolikelihood	-555.8310		-168.69566		-377.4454	
Obs	600		198		402	
Wald $\chi^2(12)$	2597.23		2114.75		101.23	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1320		0.1252	

Table C.8: (continued)

	Dependent Variable: Overall satisfaction					
	All individuals		Individuals with full school education		Individuals with less than full school education	
.linktest	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
-hat	0.9450***	0.2383	0.9974***	0.1614	1.3836***	0.4558
	[0.000]		[0.000]		[0.002]	
-hatsq	0.0371	0.1684	0.0719	0.1613	-0.1248	0.1540
	[0.826]		[0.656]		[0.418]	
/cut1	0.7507	0.0856	-0.0845	0.1304	1.8093	0.2982
/cut2	0.8689	0.0890	-0.0096	0.1343	1.9556	0.3009
/cut3	1.1110	0.0960	0.1994	0.1357	2.2277	0.3071
Log pseudolikelihood	-555.8045		-168.64825		-377.12979	
Obs	600		198		402	
Wald $\chi^2(12)$	121.07		40.18		89.10	
Prob $> \chi^2$	0.0000		0.0000		0.0000	
Pseudo R^2	0.1174		0.1322		0.1259	
. test	$\chi^2(6)$	1041.04	$\chi^2(5)$	595.35	$\chi^2(5)$	8.63
(using dummies for No. of children)	prob>chi2	0.0000	prob>chi2	0.0000	prob>chi2	0.1250

Table C.9: Marginal effects after oprobit; results by education. Note: 1. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 2. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

All individuals

Marginal effects after oprobit

y = Pr(Overall satisfaction==1) (predict, p outcome(1))

= 0.46258935

y = Pr(Overall satisfaction==2) (predict, p outcome(2))

= 0.04708689

y = Pr(Overall satisfaction==3) (predict, p outcome(3))

= 0.09532151

y = Pr(Overall satisfaction==4) (predict, p outcome(4))

= 0.39500225

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.5403***	0.0254	-0.0471***	0.0126	-0.0954***	0.0180	-0.3978***	0.0260
2*	0.1185	0.1665	-0.0015	0.0049	-0.0079	0.0161	-0.1091	0.1457
3*	0.1662	0.1432	-0.0019	0.0041	-0.0106	0.0133	-0.1537	0.1265
4*	0.1531	0.1395	0.0000	0.0012	-0.0064	0.0072	-0.1468	0.1323
5*	0.1035	0.1449	-0.0005	0.0021	-0.0053	0.0100	-0.0977	0.1331
6*	0.0658	0.1541	-0.0002	0.0016	-0.0032	0.0095	-0.0625	0.1431

Table C.9: (continued)

Individuals
with
full school education
Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.39305323
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= 0.02907138
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= 0.082819
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= 0.49505639

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	0.6141***	0.0430	-0.0289*	0.0172	-0.0827***	0.0260	-0.5025***	0.0447
2*	-0.1646	0.1846	-0.0057	0.0104	-0.0115	0.0227	0.1817	0.2164
3*	0.0793	0.1292	0.0010	0.0012	0.0004	0.0021	-0.0806	0.1289
4*	-0.0901	0.1069	-0.0018	0.0028	-0.0023	0.0042	0.0942	0.1129
5*				*(dropped)*				
6*	0.0035	0.1226	0.0001	0.0021	0.0001	0.0023	-0.0036	0.1269

Table C.9: (continued)

Individuals
with less than
full school education
Marginal effects after oprobit
y = Pr(Overall satisfaction==1) (predict, p outcome(1))
= 0.49947521
y = Pr(Overall satisfaction==2) (predict, p outcome(2))
= 0.05799063
y = Pr(Overall satisfaction==3) (predict, p outcome(3))
= 0.10406146
y = Pr(Overall satisfaction==4) (predict, p outcome(4))
= 0.3384727

Outcome: Overall satisfaction	(1)		(2)		(3)		(4)	
Variable	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
No. of children:								
1*	*(dropped)*							
2*	0.1452	0.1763	-0.0052	0.0109	-0.0163	0.0264	-0.1237	0.1398
3*	0.0246	0.1798	-0.0003	0.0029	-0.0019	0.0146	-0.0223	0.1623
4*	0.0912	0.1696	-0.0011	0.0027	-0.0069	0.0134	-0.0831	0.1539
5*	-0.0831	0.1750	0.0001	0.0019	0.0047	0.0073	0.0784	0.1696
6*	-0.1097	0.1873	-0.0005	0.0040	0.0052	0.0046	0.1050	0.1873

Appendix D

Appendix Chapter 5

	Age	Female	Absolute income	Relative income	Income inequality	Couple	Kids	Family size	Education
Age	1								
Female	-0.1264***	1							
Absolute income	0.0105	-0.0577	1						
Relative income	0.0191	0.0001	0.0428	1					
Income inequality	-0.0535	0.0344	-0.2529	0.0002	1				
Couple	0.0033	-0.0313	0.0171	0.0066	0.0734	1			
Kids	-0.0003	0.0225	-0.0148	0.0527	0.0165	-0.0429	1		
Family size	0.0153	-0.2476	0.0555	0.0378	0.1356	0.0536	0.1592***	1	
Education	-0.0058	0.0119	0.1161	-0.0157	0.0503	0.0199	-0.0615	0.0560	1

Table D.1: Correlation matrix between health determinants.
$N = 600$ households. *, **, *** indicates significance level of 10%, 5% and 1%, respectively.

Table D.2: AIH and RIH: separate and joint tests using ordered probit regression.
Note: regression coefficients are in bold and standard errors appear below them. *,**,*** denote statistical significance at 10%, 5% and 1% levels. Figure in brackets[] are p values.

Oprobit regression	Dependent Variable: Health (h)				
	Separate tests			Joint tests	
Specification:	1	2	3	4	5
Age	**-0.0913***	**-0.0943***	**-0.0953***	**-0.0943***	**-0.0863***
	0.0502	0.0522	0.0511	0.0510	0.0514
Age Squared	**0.0008***	**0.0009***	**0.0009***	**0.0008***	**0.0008**
	0.0005	0.0005	0.0005	0.0005	0.0005
Female	**-0.2338**	**-0.2736**	**-0.2401**	**-0.2485**	**-0.2432**
	0.1881	0.1841	0.1871	0.1866	0.1873

Table D.2: (continued)

Oprobit regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
Absolute income	**0.5202*** **		**0.5134*** **	**0.5715*** **	**0.5956*** **
	0.1480		0.1467	0.1624	0.1660
Relative income		**1.3936** **	**1.3458** **	**1.3362** **	**1.3398** **
		0.5830	0.5713	0.5679	0.5683
Income-inequality				**0.0340**	**0.1839**
				0.0281	0.1303
(Income-inequality)^2					**-0.0139**
					0.0120
Couple	**0.1475**	**0.1549**	**0.1441**	**0.1334**	**0.1289**
	0.1092	0.1091	0.1100	0.1105	0.1105
Kids	**-0.8694**	**-0.9962**	**-0.9507**	**-0.9320**	**-0.9717**
	0.7676	0.7983	0.7809	0.7757	0.7943

Table D.2: (continued)

Oprobit regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
Familysize	**-0.0779**	**-0.0776***	**-0.0836***	**-0.0836***	**-0.0806***
	0.0474	0.0468	0.0476	0.0477	0.0476
Years of Education	**-0.0079**	**-0.0002**	**-0.0072**	**-0.0078**	**-0.0088**
	0.0144	0.0145	0.0143	0.0143	0.0143
Region:					
Punjab	**0.1861**	**0.1988**	**0.1965**	**0.0667**	**-0.0647**
	0.1589	0.1553	0.1557	0.1928	0.2126
NWFP	**-0.0378**	**-0.0387**	**-0.0361**	**-0.0823**	**-0.1695**
	0.1662	0.1646	0.1638	0.1678	0.1834
Sind	**-0.0133**	**0.0180**	**-0.0065**	**-0.0841**	**-0.1375**
	0.1591	0.1570	0.1564	0.1685	0.1737
Baluchistan	*Reference Group*				

Table D.2: (continued)

Oprobit regression	Dependent Variable: Health (h)				
	Separate tests		Joint tests		
Specification:	1	2	3	4	5
\cut1	**-1.4137**	**-5.7580**	**-1.6970**	**-1.1352**	**-0.5236**
	1.9256	1.6332	1.9473	2.0520	2.1524
\cut2	**-0.0074**	**-4.3379**	**-0.2677**	**0.2976**	**0.9074**
	1.9381	1.6281	1.9605	2.0645	2.1695
\cut3	**0.8404**	**-3.5019**	**0.5849**	**1.1518**	**1.7625**
	1.9418	1.6272	1.9642	2.0685	2.1744
\cut4	**1.9055**	**-2.4557**	**1.6522**	**2.2194**	**2.8327**
	1.9483	1.6273	1.9709	2.0763	2.1827
Log pseudolikelihood	-821.229	-826.863	-817.727	-816.992	-816.367
Number of obs	600	600	600	600	600
Wald χ^2(k)	Wald $\chi^2(11)$	Wald $\chi^2(11)$	Wald $\chi^2(12)$	Wald $\chi^2(13)$	Wald $\chi^2(14)$
	= 28.49	= 22.02	= 34.67	= 35.30	= 35.86
Prob > χ^2	0.0027	0.0242	0.0005	0.0008	0.0011
Pseudo R^2	0.0217	0.0150	0.0258	0.0267	0.0275

Table D.2: (continued)

Oprobit regression	Dependent Variable: Health (h)				
	Separate tests			Joint tests	
Specification:	1	2	3	4	5
.linktest					
_hat	**0.6224**[0.329]	**0.7526**[0.842]	**0.9568****[0.029]	**0.9298**[0.319]	**1.2930**[0.378]
	0.6262	3.7695	0.4370	0.9329	1.4672
_hatsq	**0.2457**[0.529]	**-0.0346**[0.947]	**0.0425**[0.912]	**0.0324**[0.937]	**-0.0859**[0.839]
	0.3904	0.5245	0.3826	0.4130	0.4214
\cut1	**-1.5376**	**-5.3180**	**-1.7035**	**-1.1696**	**-0.2836**
	0.2460	6.7612	0.1517	0.4918	1.2288
\cut2	**-0.1342**	**-3.8977**	**-0.2750**	**0.2627**	**1.1490**
	0.2397	6.7715	0.1119	0.4992	1.2463
\cut3	**0.7132**	**-3.0616**	**0.5775**	**1.1168**	**2.0044**
	0.2435	6.7719	0.1165	0.5049	1.2518
\cut4	**1.7800**	**-2.0155**	**1.6451**	**2.1846**	**3.0740**
	0.2505	6.7662	0.1294	0.5095	1.2547

Table D.2: (continued)

Oprobit regression	Dependent Variable: Health (h)				
	Separate tests			Joint tests	
Specification:	1	2	3	4	5
.linktest					
Log pseudolikelihood	-821.028	-826.8599	-817.7186	-816.9880	-816.3364
Number of obs	600	600	600	600	600
Wald $\chi^2(2)$	27.18	21.46	33.43	34.26	33.85
Prob $> \chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000
Pseudo R^2	0.0219	0.0150	0.0259	0.0267	0.0275

Table D.3: Marginal effects after oprobit using AIH and RIH joint test.
Note: 1. marginal effects are in bold and standard errors appear below them. 2. *,**,*** denote statistical significance at 10%, 5% and 1% levels. 3. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

Marginal effects after oprobit
y = Pr(Health category==0) (predict, p outcome(0))
= 0.01680492
y = Pr(Health category==1) (predict, p outcome(1))
= 0.22656655
y = Pr(Health category==2) (predict, p outcome(2))
= 0.31904463
y = Pr(Health category==3) (predict, p outcome(3))
= 0.32718251
y = Pr(Health category==4) (predict, p outcome(4))
= 0.1104014

Outcome	(0)	(1)	(2)	(3)	(4)
Variable	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx
Age	**0.0040***	**0.0259***	**0.0077***	**-0.0196***	**-0.0180***
	0.0024	0.0139	0.0045	0.0106	0.0098
Age Squared	**-0.00004**	**-0.00023***	**-0.00007**	**0.00018***	**0.00016***
	0.00002	0.00013	0.00004	0.00010	0.00010

Table D.3: (continued)

Outcome	(0)	(1)	(2)	(3)	(4)
Variable	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx
Female*	**0.0125**	**0.0677**	**0.0122****	**-0.0526**	**-0.0398**
	0.0124	0.0545	0.0051	0.0427	0.0275
Absolute income	**-0.0214*****	**-0.1394*****	**-0.0415*****	**0.1055*****	**0.0968*****
	0.0075	0.0414	0.0149	0.0324	0.0277
Relative income	**-0.0562***	**-0.3654****	**-0.1088****	**0.2766****	**0.2537****
	0.0315	0.1548	0.0486	0.1205	0.1084
Couple*	**-0.0066**	**-0.0398**	**-0.0099**	**0.0305**	**0.0258**
	0.0058	0.0310	0.0066	0.0240	0.0188
Kids*	**0.0160*****	**0.1787****	**0.1546**	**-0.0686**	**-0.2807**
	0.0062	0.0794	0.1475	0.0721	0.2990
Family size	**0.0035***	**0.0227***	**0.0068***	**-0.0172***	**-0.0158***
	0.0021	0.0131	0.0041	0.0098	0.0091
Years of Education	**0.0003**	**0.0020**	**0.0006**	**-0.0015**	**-0.0014**
	0.0006	0.0039	0.0012	0.0029	0.0027

Table D.3: (continued)

Outcome	(0)	(1)	(2)	(3)	(4)
Variable	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx
Region:					
Punjab*	**-0.0084**	**-0.0535**	**-0.0152**	**0.0406**	**0.0366**
	0.0068	0.0426	0.0124	0.0321	0.0292
NWFP*	**0.0016**	**0.0099**	**0.0028**	**-0.0075**	**-0.0067**
	0.0073	0.0450	0.0120	0.0344	0.0298
Sind*	**0.0003**	**0.0018**	**0.0005**	**-0.0013**	**-0.0012**
	0.0066	0.0425	0.0125	0.0322	0.0294

Table D.4: AIH and RIH: separate and joint tests using Poisson regression.
Note: regression coefficients are in bold and standard errors appear below them. *,**,*** denote statistical significance at 10%, 5% and 1% levels. Figure in brackets[] are p values. II and RI stands for Income-inequality and RI relative income at district level, respectively.

Poisson regression	Dependent Variable: Docvisits					
	Separate tests			Joint tests		
Specification:	1	2	3	4	5	6
Constant	**0.0269**	**-2.2334***	**-0.2490**	**0.2311**	**0.4670**	**-0.1816**
	1.4562	1.1596	1.5004	1.5479	1.6980	1.5031
Age	**0.0723****	**0.0772****	**0.0773****	**0.0764****	**0.0734****	**0.0751****
	0.0325	0.0347	0.0338	0.0335	0.0341	0.0335
Age Squared	**-0.0007****	**-0.0007****	**-0.0007****	**-0.0007****	**-0.0007****	**-0.0007****
	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
Female	**0.2181**	**0.2441**	**0.2240**	**0.2316**	**0.2287**	**0.2203**
	0.1851	0.1754	0.1839	0.1842	0.1864	0.1840

Table D.4: (continued)

Poisson regression	Dependent Variable: Docvisits					
	Separate tests			Joint tests		
Specification:	1	2	3	4	5	6
Absolute income	**-0.2541****		**-0.2461****	**-0.2950****	**-0.3055****	**-0.2472****
	0.1186		0.1171	0.1263	0.1341	0.1173
Relative income		**-1.1626*****	**-1.1320*****	**-1.1189*****	**-1.1259*****	
		0.4188	0.4168	0.4124	0.4114	
Income-inequality				**-0.0279***	**-0.0826**	
				0.0167	0.0883	
(Income-inequality)2					**0.0051**	
					0.0082	
(II * RI)						**-0.1768****
						0.0762
Couple	**-0.0879**	**-0.0899**	**-0.0843**	**-0.0752**	**-0.0732**	**-0.0826**
	0.0694	0.0701	0.0700	0.0699	0.0694	0.0699

Table D.4: (continued)

Poisson regression	Dependent Variable: Docvisits					
	Separate tests			Joint tests		
Specification:	1	2	3	4	5	6
Kids	**0.5630**	**0.6660**	**0.6319**	**0.6156**	**0.6328**	**0.6314**
	0.6458	0.6761	0.6596	0.6520	0.6607	0.6780
Familysize	**0.0471**	**0.0488**	**0.0516***	**0.0511***	**0.0501***	**0.0520***
	0.0293	0.0300	0.0298	0.0299	0.0298	0.0298
Education	**0.0012**	**-0.0027**	**0.0005**	**0.0008**	**0.0012**	**0.0008**
	0.0100	0.0098	0.0098	0.0098	0.0099	0.0098
Region:						
Punjab	**-0.0507**	**-0.0605**	**-0.0634**	**0.0411**	**0.0877**	**-0.0637**
	0.0956	0.0907	0.0904	0.1119	0.1195	0.0926
NWFP	**0.0498**	**0.0494**	**0.0475**	**0.0842**	**0.1142**	**0.0468**
	0.0948	0.1594	0.0918	0.0946	0.1051	0.0938
Sind	**0.0289**	**0.0131**	**0.0203**	**0.0868**	**0.1047**	**0.0210**
	0.0873	0.0857	0.0840	0.0933	0.0963	0.0862
Baluchistan	*Reference Group*					

Table D.4: (continued)

Poisson regression	Dependent Variable: Docvisits					
	Separate tests			Joint tests		
Specification:	1	2	3	4	5	6
Log pseudolikelihood	-263.42706	-263.2062	-262.03188	-261.75132	-261.70438	-262.45038
Number of obs	600	600	600	600	600	600
Wald χ^2(k)	Wald χ^2(11) =24.47	Wald χ^2(11) =22.38	Wald χ^2(12) =32.96	Wald χ^2(13) =33.22	Wald χ^2(14) =33.41	Wald χ^2(12) =30.71
Prob > χ^2	0.0109	0.0216	0.0010	0.0016	0.0025	0.0022
.linktest						
_hat	**2.1468****	**0.7588**	**1.3199***	**1.2717***	**1.2283***	**1.4500***
	[0.024]	[0.398]	[0.072]	[0.082]	[0.095]	[0.068]
	0.9544	0.8978	0.7336	0.7304	0.7355	0.7942
_hatsq	**-1.0339**	**0.2175**	**-0.2873**	**-0.2451**	**-0.2075**	**-0.4026**
	[0.196]	[0.794]	[0.652]	[0.710]	[0.756]	[0.555]
	0.8001	0.8341	0.6361	0.6600	0.6692	0.6812
-cons	**-0.2964**	**0.0618**	**-0.0803**	**-0.0673**	**-0.0559**	**-0.1146**
	[0.305]	[0.805]	[0.706]	[0.740]	[0.782]	[0.622]
	0.2892	0.2500	0.2127	0.2029	0.2024	0.2324

Table D.4: (continued)

Poisson regression	Dependent Variable: Docvisits					
	Separate tests			Joint tests		
Specification:	1	2	3	4	5	6
.linktest						
Log pseudolikelihood	-263.21061	-263.19398	-262.0028	-261.72914	-261.68847	-262.40046
Number of obs	600	600	600	600	600	600
Wald $\chi^2(2)$	19.36	17.23	28.36	29.64	29.89	26.16
Prob $> \chi^2$	0.0001	0.0002	0.0000	0.0000	0.0000	0.0000

Marginal effects after poisson
Y= predicted number of DocVisits (predict)
= 1.7523524

Variable	dy/dx	X
Age	**0.1338****	52.8567
	0.0591	
Age Squared	**-0.0012****	2874.4
	0.0006	
Female*	**0.4492**	0.071868
	0.3979	
Absolute income	**-0.5169****	8.11056
	0.2165	
Relative income	**-1.9606*****	-0.00366
	0.7426	
Income-inequality	**-0.0489***	4.80887
	0.0294	
Couple*	**-0.1347**	0.791253
	0.1276	
Kids*	**0.8084**	0.994149
	0.6216	
Familysize	**0.0896***	4.30496
	0.0523	
Years of Education	**0.0014**	7.79433
	0.0172	
Region:		
Punjab*	**0.0718**	0.553191
	0.1961	
NWFP*	**0.1525**	0.113475
	0.1773	
Sind*	**0.1552**	0.265957
	0.1712	

Table D.5: Marginal effects for poisson regression: a joint test of the Income-hypotheses. Note: marginal effects are in bold and standard errors appear below them. *,**,*** denote statistical significance at 10%, 5% and 1% levels. X stands for the mean value of each regressor. (*) dy/dx is for discrete change of dummy variable from 0 to 1.

Bibliography

Adams, R. H. (1994). Non-farm income and inequality in rural Pakistan: a decomposition analysis, *Journal of Development Studies* **31**: 110–133.

Adams, R. H. and Alderman, H. (1992). Sources of income inequality in rural Pakistan: a decomposition analysis, *Oxford Bulletin of Economics and Statistics* **54**: 591–608.

Ahmad, M. (2000). Estimation of distribution of income in Pakistan, using micro data, *The Pakistan Development Review* **39**(4): 807–824.

Ahmad, M. (2002). Income inequality among various occupations/professions in Pakistan – estimates based on household income per capita, *The Lahore Journal of Economics* **7**(1).

Alesina, A., Tella, R. D. and MacCulloch, R. (2004). Inequality and happiness: are Europeans and Americans different?, *Journal of Public Economics* **88**: 2009–2042.

Angeles, L. (2009). Do children make us happier?, Discussion Paper Series University of Glasgow.

Anwar, T. (2005). Long-term changes in income distribution in Pakistan: Evidence based on consistent series of estimates, *Discussion Paper Series No.3, Centre for Research on Poverty and Income Distribution (CRPRID)* .

Anwar, T., Qureshi, S. K. and Ali, H. (2004). Landlessness and rural poverty in Pakistan, *The Pakistan Development Review* **43**(4): 855–874.

Blalock, H. M. (1984). Contextual-effects models: theoretical and methodological issues, Department of Sociology, University of Washington, Seattle, Washington 98195.

Blanchflower, D. (2008). International evidence on well-being, NBER Working Paper.

Blanchflower, D. and Oswald, A. (2004). Wellbeing over time in Britain and the USA, *Journal of Public Economics* **88**: 1359–1386.

Bommier, A. and Stecklov, G. (2002). Defining health inequality: why Rawls succeeds where social welfare theory fails, *Journal of Health Economics* **21**: 497–513.

Bruni, L. and Porta, P. L. (2007). *Handbook on the Economics of Happiness*, Edward Elgar Publishers.

Chinn, D. (1979). Rural poverty and the structure of farm household income in developing countries: Evidence from Taiwan., *Economic Development and Cultural Change* **27**: 283–301.

CIA (2008). The World Factbook.

Clark, A. E. (2006). Born to be mild? Cohort effects don't explain why well-being is U-shaped in age, Working Paper 35 Paris-Jourdan Sciences Economiques.

Clark, A. E., Frijters, P. and Schields, M. A. (2008). Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles, *Journal of Economic Literature* **46**(1): 95–144.

Cohen, S., David, T. and Andrew, S. (1991). Psychological stress and susceptibility to the common cold, *New England Journal of Medicine* **325**(9): 606–12.

Cohen, S., Scott, L., Stephen, M., Bruce, R., Eugene, H. and Jay, K. (1997). Chronic social stress, social status and susceptibility to upper respiratory infections in nonhuman primates, *Psychosomatic Medicine* **59**(3): 213–21.

Daly, M., Duncan, G., Kaplan, G. and Lynch, J. (1998). Macro-to-micro links in the relation between income inequality and mortality, Milbank Quarterly.

Deaton, A. (2003). Health inequality and economic development, *Journal of Economic Literature* **41**: 113–158.

Doornbos, G. and Kromhout, D. (1990). Education level and mortality in a 32-year follow-up study of 18-year-old men in the netherlands, *International Journal of Epidemiology* **19**: 374–79.

Doorslaer, E. V. and Koolman, X. (2000). Income-related inequalities in health in Europe: evidence from the European Community Household Panel., Ecuity II Project, Working Paper 1, Erasmus University, Rotterdam.

Doorslaer, E. V., Wagstaff, A. and Bleichrodt, H. (1997). Income-related inequalities in health: some international comparisons, *Journal of Health Economics* **16**: 93–112.

Easterlin, R. (2001). Income and happiness: Towards a unified theory, *Economic Journal* **111**: 465–84.

Easterlin, R. (2006). Life cycle happiness and its sources: Intersections of psychology, economics, and demography, *Journal of Economic Psychology* **27**: 463–482.

Ercelawn, A. (1984). Income inequality in rural Pakistan: A study of sample villages, *Pakistan Journal of Applied Economics* **3**: 1–28.

Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect, *Journal of Public Economics* **89**: 997–1019.

Fiscella, K. and Franks, P. (1997). Poverty or income inequality as predictor of mortality: longitudinal cohort study, *British Medical Journal* **314**(7096): 1724–27.

Fox, A., Goldblatt, P. and Jones, D. (1985). Social class mortality differentials: artifact, selection or life circumstances?, *Journal of Epidemiology and Community Health* **39**: 1–8.

Frey, B. and Stutzer, A. (2000). Happiness, economy and institutions, *The Economic Journal* **110**: 918–938.

Frey, B. and Stutzer, A. (2002a). *Happiness and Economics*, Princeton University Press.

Frey, B. and Stutzer, A. (2002b). What can economists learn from happiness research?, *Journal of Economic Literature* **40**: 402–435.

Gerdtham, U.-G. and Johanneson, M. (2000). Income-related inequality in life-years and quality-adjusted life-years., *Journal of Health Economics* **19**(6): 1007–1026.

Gerdtham, U.-G. and Johannesson, M. (2004). Absolute income, relative income, income inequality and mortality, *The Journal of Human Resources (JHR)* **XXXIX**(1): 228–247.

Glewwe, P. (1986). The distribution of income in Sri Lanka in 1969-70 and 1980-81: a decomposition analysis, *Journal of Development Economics* **24**: 255–274.

Government of Pakistan, Statistics Division (1998). Population Census.

Guryan, J., Hurst, E. and Kearney, M. (2008). Parental education and parental time with children, *Journal of Economic Perspectives* **22**: 23–46.

Helliwell, J. F. (2002). How's life? Combining individual and national variables to explain subjective well-being, *Economic Modelling* **20**: 331–60.

Horley, J. and Lavery, J. J. (1995). Subjective well-being and age, *Social Indicators Research* **34**: 283–286.

Humphries, K. and van Doorslaer, E. (2000). Income-related inequalities in health in Canada, *Social Science and Medicine* **50**: 663–671.

Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being, *Journal of Economic Perspectives* **20**: 3–24.

Kausar, F. (2001). Tax structure and tax evasion in Pakistan, Islamabad AIOU.

Kawachi, I., Subramanian, S. V. and Almeida-Filho, N. (2002). A glossary for health inequalities, *J Epidemiol Community Health* **56**: 647–652.

Kennedy, B., Kawachi, I., Glass, R. and Prothrow-Stith, D. (1998). Income distribution, socioeconomic status and self-rated health in the United States: Muti-level analysis, *British Medical Journal* **317**(7163): 917–921.

Kingdon, G. G. and Knight, J. (2006). Subjective well-being poverty vs. income poverty and capabilities poverty?, *Journal of Development Studies* **42**(7): 1199–1224.

Knight, J., Song, L. and Gunatilaka, R. (2007). Subjective well-being and its determinants in rural China, Discussion Paper Series University of Oxford No. 334.

Kruijk, H. (1987). Sources of income inequality in Pakistan, *Pakistan Development Review* **26**: 659–672.

Li, H. and Zhu, Y. (2006). Income, income inequality and health: evidence from China, *Journal of Comparative Economics* **34**: 668–693.

Lindley, J. and Lorgelly, P. (2005). The relative income hypothesis: does it exist over time?, Sheffield Economic Research Paper Series.

Lobmayer, P. and Wilkinson, R. (2000). Income, inequality and mortality in 14 developed countries, *Sociology of Health and Illness* **22**(4): 401–414.

Mailk, I. H. (2005). *Culture and Customs of Pakistan (Culture & Customs of Asia)*, Greenwood Press.

Marmot, M., George, S., Fiona, N. and Amanda, F. (1991). Health inequalities among British civil servants: the Whitehall II study, *Lancet* **337**(8754): 1387–93.

Matlon, P. (1979). Income distribution among farmers in Northern Nigeria: empirical results and policy implications., African Rural Economy Paper No. 18. East Lansing, Mich., U.S.A.: Michigan State University.

Meara, E. (1999). Inequality and infant health, Harvard Medical School. Manuscript. Boston, Massachusetts:.

Mellor, J. and Milyo, J. (2001). Reexamining the evidence of an ecological association between income inequality and health, *Journal of Health Politics, Policy and Law* **26**(3): 487–522.

Mellor, J. and Milyo, J. (2002). Income inequality and health status in the United states: evidence from the current population survey, *Journal of Human Resources* .

Mohammad, F. and Badar, G. (1985). Structure of rural income in Pakistan: some preliminary estimates, *Pakistan Development Review* **24**: 385–403.

Nugent, J. and Walther, R. (1982). Short-run changes in rural income inequality: A decomposition analysis, *Journal of Development Studies* **18**: 239–269.

Pasha, N. and Hasan, T. (1982). Development ranking of districts in Pakistan, *Pakistan Journal of Applied Economics* **1**(2).

Power, C., Manor, A. F. and Fogelman, K. (1990). Health in childhood and social inequalities in health in young adults, *Journal of the Royal Statistical Society* **153**: 17–28.

Pyatt, G., Chen, C. and Fei, J. (1980). The distribution of income by factor compnents, *Quarterly Journal of Economics* **95**: 451–473.

Sen, A. (1983). Poor, relatively speaking, *Oxford Economic Papers* **35**: 153–69.

Sen, A. K. (1993). *The Quality of life*, Oxford University Press.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures, *Econometrica* **48**: 613–625.

Shorrocks, A. F. (1982). Inequality decomposition by factor components, *Econometrica* **50**: 193–211.

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status, *The Journal of Economic Perspectives* **13**: 145–166.

Soobader, M. and LeClere, F. (1999). Aggregation and the measurement of income inequality: effects of morbidity, *Social Science and Medicine* **48**(6): 733–744.

Stutzer, A. and Frey, B. S. (2006). Does marriage make people happy or do happy people get married?, *The journal of Socio-Economics* **35**: 326–347.

Tella, R. D., MacCulloch, R. J. and Oswald, A. J. (2003). The macroeconomics of happiness, *Review of Economics and Statistics* **85**: 809–827.

UNDP (2007). Human Development Report.

Wagstaff, A. and Doorslaer, E. V. (2000). Income inequality and health: what does the literature tell us?, *Public Health* **21**: 543–67.

Wagstaff, A., van Doorslaer, E. and Paci, P. (1989). Equity in the finance and delivery of health care: some tentative cross-country comparisons, *Oxford Review of Economic Policy* **5**(1): 89–112.

Wilkinson, R. (1996). Unhealthy societies: the afflictions of inequality, London Routledge.

Wilkinson, R. (1997). Health inequalities: relative or absolute material standards?, *British Medical Journal* **314**(7080): 591–5.

Wilkinson, R. (1998). Low relative income affects mortality, *British Medical Journal* **316**(7144): 1611.

Wilkinson, R. G. (1986). Class and health: research and longitudinal data, London, England: Tavistock.

World Bank (2005). World development indicators.

Yamauchi, C. (2009). Parental investment in children: different pathways of parental education and mental health, Working Paper, The Australian National University, Centre for Economic Policy Research.