Three Essays in Macroeconomic Forecasting Using Dimensionality Reduction Methods

Yu Guo

A thesis presented in fulfilment of the requirements
for the degree of Doctor of Philosophy

Department of Economics
Adam Smith Business School
College of Social Science
University of Glasgow
June 2022

This page is intentionally left blank.

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.

Signed:

Date: June 29, 2022

Acknowledgements

Undertaking this PhD has been a truly life-changing experience for me.

I would like to express my gratitude to my supervisor Dimitris Korobilis. for all the enormous academic support and guidance from him through my PhD years. His encouragement, critique, insight, and support have benefited my work in several ways. Being a student of him has been an honour and a privilege.

This paper was funded by the School of Social Sciences of Glasgow University, for which I am deeply grateful. I would also like to thank all the employees of the Adam Smith Business School who care about our physical and mental health, as well as friendly and enthusiastic colleagues, for creating such a favorable research atmosphere.

Finally, the writing of this thesis would not have been possible without the support of my friends and family whose encouragement, patience and understanding have kept me going throughout the past four years.

Abstract

This thesis consists of three studies that concentrate on the dimensionality reduction methods used in macroeconomic forecasting.

Chapter 2 (the first study) aims to investigates the predictive ability of several indicators of consumer sentiment and perceptions about the economy. Based on seven key qualitative questions in the University of Michigan survey of consumers, I employ various quantification approaches to construct six indexes namely sentiment, disagreement, pessimism, uncertainty, price pressure, and interest rate pressure. I establish that these six indexes convey predictability for key macroeconomic indicators beyond and above the information found in existing, popular macroeconomic and financial indicators. I also provide a deep explanation of consumer indexes by monitoring their response to supply, demand, monetary policy and financial shocks using a VAR model with sign restrictions. The results indicate that price pressure and interest rate pressure are mainly correlated with financial and uncertainty shocks, while the other indicators reflect the formation of opinions that are sensitive to shocks related to supply, demand, and monetary policy.

Chapter 3 (the second study) explores the dimensionality reduction algorithm

by extracting factors from a large number of predictors that take into account correlation with the predicted (target) variable, using a novel time-varying parameter three pass-regression-filter algorithm (TVP-3PRF). The benchmark 3PRF algorithm (Kelly and Pruitt, 2015) assumes that a predictor is relevant for forecasting over the whole sample and can be represented using a series of OLS regressions. I extend this approach using time-varying parameter regressions that are conveniently represented as a series of high-dimensional time-invariant regressions which can be solved using penalized likelihood estimators. TVP-3PRF algorithm allows for a subset of variables to be relevant for extracting factors at each point in time, accounting for recent evidence that economic predictors are short-lived. An empirical exercise confirms that this novel feature of TVP-3PRF algorithm is highly relevant for forecasting macroeconomic time series.

Chapter 4 (the third study) determines which of the two main types of algorithms in the field of dimensionality reduction truely reflect the true way variables enter the model. It is know that in the area of modelling and forecasting high-dimensional macroeconomic and financial time series, two main methods, sparse modelling and dense modelling, are both popular. However, instead of simply viewing each a method for avoiding overfitting, a question that is worth exploring is which of these models can represent the real structure of the data. Another question that arises is whether the uncertainty of variable selection will affect the prediction. In line with Giannone et al. (2021), I used their spike and slab prior to explore the scenarios for six economies when forecasting production growth. The results indicate that the way macroeconomic data are employed in the model of all the economies have an obvious sparse structure albeit with different degrees. However, the pervasiveness of uncertainty causes the sparse model to fail and the model averaging technique to become the preferred method. Moreover, what is surprising is that the dense model(ridge regression) dominated after the pandemic began.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General background

> *"[...]I was using one label for a range of issues, and I wanted the simplest, shortest phrase to convey that the boundaries of computing keep advancing."* - Steve, Lohr (February 2013).

As noted by Steve, Lohr (February 2013) in The New York Times, dimensionality reduction techniques have gained popularity over the past few decades to address predictive modeling challenges. This is also the focus of this study. In fact, dimensionality reduction has always been a critical component in the fields of economics and finance. This area can be broadly classified into two categories: the creation of indicators based on economic interpretation and the dimensionality reduction of multidimensional data based on statistical algorithms. However, these two classifications are not entirely independent. For example, some dimensionality reduction algorithms such as principle component analysis are utilized in the construction of indexes. Currently, the design of dimensionality reduction algorithms also focuses on whether they have economic implications. I will elaborate on these two perspectives and delve deeper into each of them.

Economic and financial indicators play a crucial role in providing valuable insights into the state of the economy, and are essential tools for policymakers, investors, and researchers. These indicators help to monitor the overall economic performance, identify trends, and inform decision-making. In this paragraph, we will discuss the construction and application of various economic and financial indicators. Two of the most important indicators in this regard are the stock market index and the economic conditions index. For example, the Dow Jones Industrial Average (DJIA) is the world's most renowned and widely used stock market index providing a view of the US stock market and economy. Named after Charles Dow, who, along with his business partner Edward Jones, created the index in 1896, the Dow Jones Industrial Average (DJIA) comprises the 30 most prominent publicly traded companies in the US. Its purpose was to reflect the broader US economy and assist in managing investment portfolios used by a substantial number of investors. Moreover, it has been utilized as a market price proxy to examine the predictive capacity of various models concerning market returns and volatility. Similarly, The global real economic activity, which is a renowned economic indicator proposed by Kilian (2009), is based on shipping costs. This economic index has been utilized to monitor global economic activity and forecast the global crude oil price. Other examples include numerous business cycle indicators (BCI) which are constructed in each country for forecasting, dating, and confirming changes in the direction of the overall economy of a country, notable examples of which have been developed by Gehringer and Mayer (2021), Boshoff and Binge (2019), Wong et al. (2014). Regarding the state of the finances, Koop and Korobilis (2014) constructed a financial condition index(FCI) by factor augmented vector autoregressive models with time-varying coefficients and stochastic volatility. Similar to this are the St. Louis Financial Stress Index, the Kansas City Fed Financial Stress Index, and the Cleveland Fed Financial Stress Index released by St Louis Fed, Kansas Fed, Cleveland Fed, and Chicago Fed, respectively. On top of the traditional index of economic and financial conditions, the prevalence of behavioral economics has resulted in the development and popularity of various sentiment or confidence indexes and uncertainty indexes in recent years. For example, the Consumer Confidence Index (CCI) is based on a survey administered by The Conference Board, and assesses whether consumers are optimistic or pessimistic about their future financial status. In addition to this, the Consumer sentiment index is the similar one that released by University of Michigan based on their consumer surveys. As well as these, Jurado et al. (2015) introduced two uncertainty indexes : total uncertainty and economic uncertainty. The former refers to the assessment of uncertainty from all sources, while the latter refers assessment of uncertainty owing to (non-health-related) economic factors. In summary, these indexes reduce the dimensionality of the data, while providing robust and efficient information for researchers, politicians, and investors alike. This also provided inspiration for this research in how to use existing data to intelligently construct indexes that can contribute to macroeconomic forecasting.

The second part of dimensionality reduction focuses on the development of dimensionality reduction algorithms for high-dimensional data. In recent years, the proliferation of data collection technologies and the increasing complexity of data have resulted in datasets with a large number of features, commonly known as high-dimensional data, and this type of data is increasingly gaining traction both in academic research and industrial applications. High-dimensional data has emerged in tandem with advances in computer technology, which has led to data growing and updating at a faster rate, and the datasets becoming more multidimensional and unstructured. However, a wealth of helpful and effective information is buried beneath a sea of massive and complex data, making it difficult to extract its important qualities and make use of the data. The "curse of dimensionality", for example, occurs as a result of the rapid and large-scale expansion of dimensions. (see Hughes, 1968; Sammut and Webb, 2011; Ye and Sugihara, 2016). In addition, the high-dimensional data processing consumes a significant amount of computing time, storage space, and labour, which also has a negative impact on the accuracy of estimation . In mathematical terms, the condition that the number of data points, written as N, should be greater than the number of variables, denoted as q, is a basic minimum criterion for accurate estimation of linear regression coefficients, but does not guarantee high-quality parameter estimation. In traditional statistical settings, N is assumed to be substantially bigger than q in order to prove asymptotics—a common way of proving the validity of a statistical model. In practise, the linear regression model experiences difficulty in cases where the N/q ratio is small. This is where dimensionality reduction algorithms are essential for evaluating a large amount of data and extracting meaningful information needed for analysis and prediction from data with a high dimensionality, as well as removing the influence of related or repeating factors. To put it another way, the difficulties along with the high-dimensional data must be handled by reducing the number of dimensions.

To address these challenges, researchers have developed various dimensionality reduction techniques, such as Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Autoencoder Neural Networks (ANN). These techniques aim to reduce the dimensionality of the data while retaining the essential information needed for analysis and prediction. In this paragraph, we will explore the strengths and limitations of these techniques. The fundamental idea behind dimensionality reduction is to translate a data sample from a high-dimensional space into a low-dimensional space. In this domain, principle component analysis (PCA) is the basic and most wildly used method. Because some loss of original information is bound to occur in the process of mapping high-dimensional data to a low-dimensional space by projection, the goal of the projection algorithm is to obtain valuable reduction data from a high-dimensional dataset to meet accuracy and storage requirements while retaining the fundamental qualities of the original data. PCA was proposed based on the above principle. The first principle component can be defined as a direction that maximizes the variance of the projected data. The i-th principal component is a direction that maximizes the variance of the projected data and is orthogonal to the first $i-1$ principal component. However, in machine learning terms, PCA is actually a typical example of 'unsupervised learning', which means it does not consider the relationship between the information contained in the explanatory variables and the variable the researcher is interested in forecasting. Therefore, some dimensionality reduction methods that fall under the umbrella of 'supervised learning' have been proposed to improve the estimation and prediction quality of the regression. Partial Least Squares (PLS) regression, which was introduced by Wold (1966), is one such supervised learning method that is often compared to PCA. In PLS regression, similar to principal components regression, scores are generated for regression analysis by creating linear combinations of the input variables. However, PLS regression employs both the predictor variable, denoted as X, and the response variable, denoted as y, to form the scores. PLS performs a regression on a weighted X with incomplete or partial data, and is expected to perform better than PCA in terms of prediction because it additionally employs a target variable y to determine the PLS-directions which achieve both a high variance and large correlation with y. Furthermore, there have been efforts by other researchers to enhance PCA within the realm of 'supervised learning'. One such example is the work of Bair et al. (2006), where instead of utilizing all of the variables present in a dataset for principal component analysis, they only considered the predictors with the highest estimated correlation with the target variable y. This direction of development in dimensionality reduction algorithms also provides guidance for our further research, i.e. taking into account the information of the predicted variables when extracting the common factors in order to pr ovide more accurate predictions.

In addition to the supervised and unsupervised learning classification, there is another classification perspective for dimensionality reduction algorithms based on whether they are dense shrinkage or sparse shrinkage methods. Under this classification, PCA and PLS belong to dense shrinkage methods, which retain all variables in the model and assign continuous weights, so that factors are combinations of the original variables that will be hard to interpret. Ridge regression is another dense shrinkage estimation method but retains the original meaning of variables. It shrinks the size of coefficients by imposing a constraint which is an L2-norm loss function minimizing the sum of the square of the deviations between the target value and the estimated values. Therefore, ridge coefficients are estimated by a penalised residual sum of squares, and the amount of shrinkage is controlled by a chosen tuning variable λ, where the larger its value, the more are the coefficients shrink towards zero.

Under this classification criterion, the counterpart of the dense shrinkage method is the sparse method, that is, the method of variable selection. The least absolute shrinkage selection operator (LASSO)(Tibshirani, 1996) is the most similar method to ridge regression but performs a kind of continuous subset selection as it imposes an L1-norm restriction on the mean square error(MSE). The L1-norm, which has a loss function that is also known as least absolute errors (LAE), essentially minimizing the sum of the absolute deviations between the target variable and the value estimated by regressing on explanatory variable. It contains built-in feature selection and tends to create sparse coefficients. Since when p is relatively large, high variance and overfitting are a serious concern, therefore methods like LASSO, involving simple, highly regularised procedures, are frequently preferred because it is able to select a subset of the most important predictors and set the coefficients of the remaining predictors to zero. Moreover, based on a similar idea, the Elastic Net developed by Zou and Hastie (2005) arose from criticisms of LASSO, whose variable selection can be overly reliant on data, making it unstable. By combining the ridge regression and LASSO penalties, elastic net combines the advantages of both and has been found to outperform LASSO in terms of predictive power while still conducting feature selection.

The aforementioned sparse algorithms all have a common characteristic of integrating the selection and estimation processes into a single step by introducing a penalty term to the loss function minimization process. However, there are also alternative approaches that divide the process of filtering the variables from the estimation of the model coefficients into two distinct steps. Forward/backward stepwise, best subset regression, and the Leaps and Bounds algorithm are aimed at preventing overfitting by selecting variables using metrics such as Akaike information criterion (AIC), Residual Sum of Squares (RSS), or p-value. However, these methods may not be suitable for datasets with a large number of variables (p). This is because they do not account for potential correlations between variables, and exploring all possible subsets of variables may be computationally infeasible for large p. While the Leaps and Bounds technique can handle up to 30 or 40 regressor variables, its computational efficiency is unclear when p is exceptionally large. Furthermore, the search for all possible subsets becomes impractical with datasets larger than 40 input variables, making these methods unsuitable for such datasets. ortunately, numerous researchers have proposed various rules for selecting important and relevant variables, the most notable of which is the rule developed by Bai and Ng (2008). This method employs two types of thresholding rules: hard and soft. Hard-thresholding selects indicators based on their correlation coefficient with the target, and only those with correlations exceeding a predetermined threshold are chosen as predictors. However, this approach ignores the information provided by other indicators and can result in the selection of highly collinear predictors.Soft-thresholding, on the other hand, orders and selects indicators by minimizing the Residual Sum of Squares (RSS) plus a penalty term that is a function of the regression coefficients and a Lagrange multiplier λ governing the shrinkage and a function of RSS and of the N regression coefficients $\beta_i, i = 1, ..., N$. Depending on the form of the penalty term, several soft-thresholding rules can be obtained, such as Least Angle Regression (LARS), LASSO, Elastic Net Estimator (NET), and Forward Selection Regression (FWD). Researchers then rank and select the most relevant variables. These thresholding methods have been applied in several studies, including those by Bulligan et al. (2015) and Rapach et al. (2013). The widely accepted classification of sparse and dense modeling inspired some of the research in this thesis.

To conclude, this section has discussed the necessity for dimensionality reduction and the general development of dimensionality reduction methods. It provided and overview of various new indexes proposed to summarize a certain economic or financial field, and introduced dimensionality reduction algorithms belonging to supervised learning and unsupervised learning, or the dense shrinkage method and sparse variable selection process. It is apparent that numerous dimensionality reduction methods have undergone significant development since their proposal. For example, extensions to LASSO include fused LASSO (Tibshirani et al., 2005), group LASSO (Yuan and Lin, 2006), adaptive LASSO (Zou, 2006), prior LASSO (Jiang et al., 2016), and the combination of them all known as adaptive group fused LASSO (Qian and Su, 2016). However, there are also several shortcomings that can be addressed. For example, there are still gaps in the field of index construction, and hardly any dimensionality reduction algorithms for high-dimensional data have emerged that combine time-varying factor loadings under the intersection of supervised and sparse algorithms, despite the recognition of the time-varying nature of coefficients in economic models by many scholars. This provides the fundamental research motivation for this thesis. To contribute to the field of dimensionality reduction, subsequent chapters further develop various aspects of dimensionality reduction methods which include explore and present new indexes and an algorithm that is in many ways superior, to contribute to the field of dimensionality reduction. In the meantime, the intense and outgoing debate continues between the dense shrinkage method and the sparse variable selection process motivated me to determine which algorithm is appropriate for macroeconomic data.

1.2 The contribution of this thesis

This paper contributed to macro-forecasting based on data dimensionality reduction, both methodologically and empirically, based on the background presented in the previous chapter.

In the second chapter of the study, the focus is on constructing economic and financial indexes. More specifically, several indicators are created that reflect various aspects of consumer attitudes. The purpose of these indicators is twofold: to gain a deeper understanding of the consumer profile and to forecast macroeconomic variables. It is essential and valuable to study the consumer as a key player in macroeconomics, given the human element's fundamental role in economic theory. Economic growth and recession are determined by changes in business investment spending, consumer purchases, and government spending. Behavioral economics suggests that the consumer sector has as active an influence on the macro-economy as the business community and government. While Keynes recognized the importance of business expectations and the government sector, consumer activity not only directly impacts macroeconomic trends but can also reinforce or counterbalance actions taken by firms or the government. Consumers account for approximately two-thirds of US economic activity or gross domestic product (GDP), according to the Bureau of Economic Analysis's National Income and Product Accounts. The study in chapter 2 focuses on constructing various indicators that reflect consumer attitudes to gain a deeper understanding of the consumer profile and forecast macroeconomic variables. This perspective is more critical than directly modeling consumer behavior for two reasons. Firstly, consumer reactions to changes in income no longer lead to immediate and offsetting changes in spending. Consequently, consumer attitudes and expectations have become crucial components in analyzing consumer spending and saving patterns.

Secondly, high-cost physical assets like houses, vehicles, and household durables are relatively infrequent purchases that are usually planned in advance. Replacement demand is more commonly driven by subjective valuation and upgrading preferences than by the item's usability. This involves not only consumer attitudes but also the fact that the debt used to purchase these assets needs to be repaid from expected future real earnings, which is primarily a matter of consumer expectations regarding future business conditions. Earlier research on consumers has either focused on particular consumer behaviors or provided general index of consumer attitudes like sentiment and confidence without delving into the consumer profile in greater detail. This chapter bridges this gap by proposing indicators that reflect various aspects of consumer attitude, and highlights the significance of individual index in forecasting macroeconomic variables.

Therefore, six consumer indicators have been constructed which capture consumer attitudes and expectations from a diverse and novel perspective. specifically, these are: overall consumer sentiment about the future business environment, consumer pessimism about the future, the extent to which consumers' opinions are divided, consumer perceived price pressures, consumer perceived interest rate pressures, and consumer uncertainty about durable goods markets. Thus, by using these diverse indexes, researchers can identify various facets of consumer attitudes. Moreover, because of their different connotations, these six indicators respond to supply shocks, demand shocks, monetary policy shocks, and financial shocks to varying degrees, making them capable of forecasting particular macroeconomic variables.

Chapter 3 of the study focused on improving algorithms for extracting common factors. Existing methods for extracting common factors have been developed in the fields of supervised learning and time-varying coefficients respectively, this research innovatively considers both elements simultaneously. The proposed dimensionality reduction algorithm enhanced both supervised learning and time-varying factors, and was advantageous due to its linear model-based nature, resulting in faster processing times. Specifically, the time-varying parameter three-pass regression filter (TVP-3PRF) took into account the relationship between the predicted variable and the predictors, thus making more efficient use of the information in the explanatory variables. Given the variability of economic structures and the short-lived nature of economic variables, the algorithm incorporated time-varying parameters and/or factor loadings into the model. Compared to other commonly used factor extraction methods, the proposed algorithm provided more efficient and accurate estimation of common factors for macroeconomic forecasting, regardless of whether the actual factor loading was constant, regime-switching, or time-varying, and whether the common factors were strong or weak. At the same time, the results of the empirical study revealed that US macroeconomic variables exhibited a clear short-lived response to the common factor, which provided further evidence of the necessity for time-varying modeling.

Chapter 4 presented evidence supporting the debate between dense and sparse modeling in macroeconomic forecasting. Recent discussions have focused on whether sparse models outperform dense models due to statistical reasons or because variables in real models are sparse. This question is raised because several researchers have observed that sparse models perform better than dense models (Stock and Watson, 2006; Korobilis, 2013; Tian and Timmermann, 2015). This phenomenon was also observed in the study described in Chapter 3, where factor loading exhibited strong sparsity, indicating that many variables were not useful for predicting the target variable at a certain period. This leads to the study in Chapter 4, and contributes to this debate. Taking advantage of the spike-and-slab prior nature that allows the model to be sparse or dense, I experimented with sparsity or density situations for six large economies when forecasting production growth. These are the economies of the UK, the EU as a whole, and the four largest countries in the EU (Germany, France, Italy, and Spain). The results show that all economies prefer sparse models, but there is uncertainty regarding the selection of important variables. Sparse models are indeed appropriate for countries with low uncertainty. However, model averaging techniques with different sets of regressors provided better forecasts for countries with high uncertainty. Additionally, during periods of chaos and disorder, combining all available data may be the best option. This result provides guidance for macroeconomic modeling in each country studied and provides general guidelines for other countries to explore their own modeling techniques.

I provide a clear motivation for each method used in each chapter for the self-contained nature of each chapter. Furthermore, each chapter offers the information necessary for the reader to grasp the rationale underpinning the approaches described in this thesis. All of the relevant technical information is provided in the appendices.

Chapter 2

Consumer opinions and the business cycle

2.1 Introduction

As direct participants in numerous economic decisions, consumers' perspectives regarding the current and future trajectory of the economy matter a great deal. According to Armantier et al. (2017), consumers' opinions drive the various economic choices they make, including decisions related to savings, investment, durable goods purchases, wage negotiations, and more. The set of choices made further affect macroeconomic outcomes, including real gross domestic product and industrial production. Related simulation schemes not only accelerate consumption, but also save tens of thousands of jobs thus boosting the economy. Building on these observations, the aim of this research is to construct consumer indicators from different angles that reflect various aspects of consumers' opinion, explore their performance they have in forecasting economic variables, especially consumption and GDP. The existing literature relating to consumer indexes mostly focuses on consumer sentiment and uncertainty. For example, the official consumer sentiment index proposed by the university of Michigan has been employed in numerous research studies such as those by Gillitzer and Prasad (2018) and Shapiro et al. (2020). Notably, another index related to consumer perception, consumer uncertainty, became popular after the financial crisis of 2007–2008. A vast array of studies use the interquartile range or the standard deviation for a point forecasting survey of consumers as the proxy of consumer uncertainty, and map the evolution between the proxy and real economies, including studies by Moessner et al. (2011) and Dovern et al. (2012). However, targeting only one perception of consumers is not sufficient to explore the information consumers provide to the real economy, which is one of the contributions of the current study.

In this paper, given the qualitative nature of consumers' responses, survey data needs to be quantified in order to be used in a statistic model. To achieve this goal, our objective is to use different survey questions as well as different summary statistics for categorical data in order to obtain comprehensive, quantitative measures of consumers' attitudes, and then examine the role these indexes play in the economic and financial field. The first group of questions to focus on the consumer's opinions about business conditions. Compared to alternative questions regarding household finance or income, which are relatively more objective markers, attitudes towards the business condition are highly related to the consumers' answer to the question asking whether now is the right time to make major purchases, to save, or to make investments. This kind of attitude can directly influence how business managers manage product development, distribution strategy, messaging, and promotions, thus impacting the economy. Another question category is the reasons behind buying or not buying high-value household assets, i.e. durable goods, vehicles and houses. Many pieces of literature such as Alvarez-Parra et al. (2013), Lanteri (2018) and Kydland et al. (2016) have proven that these high-value consumptions play important roles in the business cycle, thus providing a sufficient foundation in studying them. Based on these survey questions, we build six consumer survey indexes to delve deeper into the information from consumers on various measures. They are consumer sentiment, consumer pessimism, consumer disagreement, price pressure on consumers, interest rate pressure on consumers, and consumer uncertainty. Details of the questions and the quantification approaches adopted are presented in section two.

Our results provide empirical evidence for the predictive ability of the consumer survey indexes. Targeted on forecasting GDP, investments, and consumption, by comparing the results using the constant parameter model, constant parameter model with stochastic volatility(SV), and time-varying parameter model with stochastic volatility(TVP-SV), we find that regardless of whether the choice of benchmark model is AR(2) or AR(2) with 5 macroeconomic common factors, a constant parameter model with SV always exhibits the strongest predictive abilities. It is also important to note that the inclusion of the consumer index can keep improving the forecasting performance in this estimation framework. Moreover, consumer survey indexes have a better ability to enhance predictions than other indexes used for comparison in the medium and long run under the AR(2) framework and short and medium horizon under the AR(2) plus macroeconomic common factors framework.

The impulse response function and variance decomposition analysis of a sign-restricted VAR model provides a deeper insight into the nature of the six consumer survey indexes. Consumer sentiment, consumer pessimism, and consumer disagreement are three indicators that mainly react to aggregate supply shocks and monetary shocks. Unlike sentiment and pessimism, which contain intuitive meaning, the way consumer disagreement works is that people update and acquire information in different manners. Therefore, the emergence of disagreement and the change in its degree indicates the occurrence of new uncertain information in the economic environment, reflecting the upcoming changes in the economy. According to the current study, consumer disagreement only falls when the recession appears, therefore the innovation shock that causes economic recession will decrease the consumer disagreement index. Conversely, the pressure of price and interest rate are more demand- and financial-oriented because they exhibit strong positive and negative contemporaneous reactions to aggregate demand shocks and financial shocks, respectively. The last index, consumer uncertainty, barely reacts to financial shocks but responds at an average level to supply, demand, and monetary policy shocks, indicating that it is a completely different uncertainty index from the financial uncertainty index proposed by Jurado et al.

(2015). Therefore, each index has distinct intrinsic meanings.

The remainder of the paper is organized as follows. The next section explains the details of constructing the consumer index. The forecasting analysis, including the data, estimation methods, and results, are discussed in the third section. The fourth section then provides the opportunity to look inside these consumer survey indexes by using the VAR model with sign restriction. Finally, concluding comments are presented in the fifth section.

2.2 Data description and data reliability

The Survey Research Center at the University of Michigan conducts the Surveys of Consumers, which were founded in 1946 by George Katona. These surveys have emphasized the significant role of consumer spending and saving decisions in shaping the national economy. The Surveys of Consumers have been proven to be a reliable predictor of the future direction of the economy. The Index of Consumer Expectations, which is produced by the Surveys of Consumers, is included in the Leading Indicator Composite Index published by the U.S. Department of Commerce, Bureau of Economic Analysis, demonstrating the Surveys' ability to understand and forecast changes in the economy. The composite Index of Leading Indicators selects data based on its economic significance, statistical adequacy, consistency in timing, conformity to business cycles, smoothness, and prompt availability, making it a rigorous and reliable measure.

The Expectations Index of the Surveys of Consumers focuses on how consumers view prospects for their own financial situation, the general economy in the near and long term. The survey contains approximately 50 core questions, with a minimum of 600 interviews conducted by telephone each month from the Ann Arbor facility. The surveys cover three broad areas of consumer sentiment: personal finances, business conditions, and buying conditions. The core questions track different aspects of consumer attitudes and expectations, including overall assessments of past and expected changes in personal finances, expected changes in nominal and real family income, attitudes towards business conditions in the economy as a whole, and specific questionnaire items concerning expected changes in inflation, unemployment, and interest rates, as well as confidence in government economic policies. Finally, the survey includes several questions to gauge respondents' views on current market conditions for large household durables, vehicles, and houses.

In various regions, consumers are asked to not only provide their general opinions but also explain their reasons for those opinions in their own words. These additional inquiries indicate a growing interest in not only predicting consumer behavior but also comprehending why they make certain spending and saving decisions. By understanding the rationale behind consumer actions, we can grasp why they react differently to the same economic events at different times.

While purchases of homes, vehicles, and household goods and the incurrence of debt and acquisition of financial assets are significant economic decisions for individual families, their timing influences the course of the entire economy. These infrequent and large spending and saving decisions are often associated with planning and thoughtfulness on the part of consumers, rather than impulsive or habitual behavior. Additionally, these decisions are not solely based on consumers' current economic circumstances but also rely on their expectations of household income, employment, prices, and interest rates.

The University of Michigan also studied consumers' accuracy in predicting certain economic variables based on survey data. For example, as for interest rate, they used a balance score equal to the percentage of consumers who expected interest rates to decrease minus the percentage that expected interest rates to increase, plus 100. To be consistent with the survey question that asks consumers about the expected direction of change, the annual percentage point change in the prime rate was used for the objective measure. And as for unemployment rate, they use a balance score equal to the percentage of consumers who thought the unemployment rate would increase minus the percentage who thought it would decline, plus 100. Since consumers are asked about the direction of expected change in the unemployment rate, not its level, the annual percentage point change in the unemployment rate is used as the objective measure.The result found that consumer expectations had a strong correlation with actual changes in economic variables such as interest rates, unemployment rate, and home and vehicle sales. On average, consumers anticipated changes in these variables several months in advance, with a lead time of two to three quarters. For instance, changes in interest rate expectations were on average anticipated six months ahead of the actual change, while changes in unemployment rate expectations were generally anticipated nine months ahead. Consumer attitudes towards home and vehicle buying also preceded changes in sales by two quarters, with a time series correlation of 0.77 and 0.73 respectively. Overall, these findings suggest that consumer expectations are an important factor in predicting and understanding changes in the economy.

The researchers in univerisity of Michigan also made efforts to guarantee the accuracy of the survey according to the survey information released by the University of Michigan. Taking the example of whether the inflation rate is expected to increase or decrease in the future (the same applies to price level). The most common misunderstanding is about the answer "same". To investigate the potential bias resulting from respondents misinterpreting the price expectations question, a probe was added to the survey starting from March 1982. The probe specifically asked respondents who answered "same" whether they meant the level of prices or the rate of inflation would remain unchanged. After doing that, they use some statistical tools to compensate for the downward bias caused by the misinterpretation of "same" responses prior to the introduction of the probe. In addition, the researchers at the University of Michigan used a range of methods to handle various types of missing values and outliers which only arises in survey questions where consumers are required to provide specific numerical values, though those survey questions will not be used in this study.

In conclusion, whether in the means of avoiding misunderstandings during the data collection process or in the processing of the collected data, the University of Michigan has fully ensured the accuracy of the survey data and demonstrated its predictive power.

2.3 Indexes of consumer opinion

We used data from the University of Michigan surveys of consumers[1] for the period 1960Q1 to 2021Q2. Our focus is questions pertaining to business conditions and major household purchases (durables, vehicles, and houses). All the questions we use are defined in Table 2.1. The first four questions relate to the business expectations[2], the answer to which falls into one of three categories depending on whether respondents believe economic conditions will improve, remain the same, or deteriorate. Although the label of each of these categories might differ, they are always comparable. For example, in the first category, depending on the question, the response might be labeled as better, favorable, or good times, but in all instances, the response expresses a positive opinion. We label these three categories as *pos*, *neu*, and *neg*, indicating positive, neutral, and negative opinions, respectively.

For each of these four questions, we extract the following three measures

$$
\begin{aligned}
Sent_i &= pos_i - neg_i + 100, && (2.1)\\
Pess_i &= p_{neg,i}, && (2.2)\\
Disag_i &= 1 - p^2_{pos,i} - p^2_{neg,i}, && (2.3)
\end{aligned}
$$

for each question $i = NEWS, BAGO, BUS12, BUS5$, where *pos* and *neg* indi-

[1]https://data.sca.isr.umich.edu/data-archive/mine.php.

[2]The variable BAGO asks participants to evaluate current economic conditions compared to a year ago. This is not necessarily a backward-looking variable, as it also entails expectations about the future as well as the present. It is not clear what is the discounting that each respondent uses when asked such general questions, and for that reason, we find indices based on BAGO to be highly correlated with NEWS and BUS12 and BUS5.

cate the number of responses in each category while p_{pos} and p_{neg} are the same numbers expressed as a proportion, for instance, $p_{pos} = \frac{pos}{pos+neg}$. The first indicator is the measure of consumer sentiment, and this is the exact formula that the University of Michigan uses to compute the Index of Consumer Sentiment as a factor from several subquestions. The difference between the percentage of favorable replies and the percentage of unfavorable replies provides a measure of the overall sentiment towards the product, service, or experience. If the percentage of favorable replies is higher than the percentage of unfavorable replies, then the sentiment is generally positive. Conversely, if the percentage of unfavorable replies is higher than the percentage of favorable replies, then the sentiment is generally negative. This index exhibits traits of being straightforward and comprehensible, quantifiable, standardized, impartial, and easily interpretable. Moreover, it furnishes a set of principles that can be utilized for the development of other indexes. The second indicator is a measure of consumer pessimism, which simply counts the percentage of respondents that thought economic conditions are or will be worse. Finally, the third indicator is a measure of dispersion for ordinal data, commonly known as the Gini-Simpson Index. In ecology, this formulation is employed to measure the diversity of species, and in economics, it is commonly used to measure income inequality. This index is named disagreement, as higher values of this indicator capture higher dispersion between the two extreme categories, while lowed values indicate higher consensus.[3]

[3]When $p_{pos} = p_{neg}$ the indicator takes its maximum value 1 and disagreement is at its highest. This is because respondents are equally split between the categories. Similarly, when either $p_{pos} = 0$ or $p_{neg} = 0$ all respondents give the same answer (whether it is optimistic or not about business conditions) and consensus is at its highest. Of course because in practice there are three categories of answers, it could be the case that all respondents agree that business

The second group of questions numbered as questions 36, 38, and 42 in the University of Michigan surveys (see Table 2.1), refer to reasons for buying or not buying durable goods, vehicles and houses. These questions allow us to pin down in detail the exact reasons for these significant purchases, with responses focusing on prices, interest rates, as well as a general concept of uncertainty. Regarding prices, optimistic consumers respond that times are good either due to prices being low or due to the expectation that prices will increase. Pessimistic consumers respond that they consider times to be bad because prices are high. For the interest rate category consumers either respond that times are good (interest rate is low) or times are bad (interest rate is high). Regarding the more general concept of uncertainty, the reason respondents given for not purchasing include explicitly stating that the future is uncertain or that they can't afford to buy these goods. We use the relative balance of 'good times' minus the 'bad times' for interest rate and price category, and the sum of the 'future is uncertain' and 'can't afford' to extract the following indicators

$$Unc_j^{price} = (GT/PricesLow + GT/PricesIncrease) - BT/PricesHigh, \quad (2.4)$$

$$Unc_j^{int} = GT/InterestLow - BT/InterestHigh, \quad (2.5)$$

$$Unc_j^{gen} = BT/Can'tAfford + BT/UncertainFuture, \quad (2.6)$$

for each question $j = 36, 38, 42$ (see Table 2.1), where GT/X indicates the number of consumers giving reason X for selecting the category "Good Times", and similarly BT/Y is the number of consumers giving reason Y for selecting

conditions will remain the same (such that $p_{neu} = 1$), in which case there is no disagreement but it holds that $p_{pos} = p_{neg} = 0$ and the $Disag$ index attains its highest value. However, this is never the case in these consumer survey data, as the vast majority of respondents in these surveys belong to the extreme two categories.

"Bad Times". Therefore, we use the different categories of answers to create three indicators that measure the pressure of price($Press^{price}$), the pressure of interest rate($Press^{int}$) and the general uncertainty(Unc_j^{gen}). This last uncertainty indicator can be thought of as a proxy to the concept of Knightian uncertainty, that is, the lack of any quantifiable knowledge about some possible occurrence.[4]

We extract these six indexes for each relevant question as described above, such that we have four from each of the Sentiment, Pessimism and Disagreement indexes, and three from each of the Price, Interest Rate, and General uncertainty indexes. These indices fulfill the requirements of being comprehensible, measurable, and standardized, as does the index proposed by the University of Michigan. The final aggregated indexes are obtained as the first principal component from all the indexes in each individual question. For example, we denote, hereafter, as *Sent* the Sentiment index that is the first principal component of $Sent_{NEWS}, Sent_{BAGO}, Sent_{BUS12}, Sent_{BUS5}$. Similarly, Unc^{price} is the Price Pressure index that is the first principal component of $Press^{price}_{DURRN}$, $Press^{price}_{VEHRN}$, $Press^{price}_{HOMRN}$. These factors are plotted in Figure 2.1

As shown in Figure 2.1, the disagreement index and sentiment index present a similar pattern (although their correlation is less than 0.6), which means that these two indicators may respond in a similar way to economic and/or financial shocks. It can be clearly seen in this figure that during several economic recessions defined by NBER (the early 1980s, early 1990s, early 2000s, the great

[4]Most consumers do classify bad times as being driven by specific types of uncertainty e.g. higher prices or interest rates, or some other related reason (e.g. bad credit). Therefore, those consumers that end up giving the generic response "uncertain future" or "can't afford" are those that are not able to quantify uncertainty as being in any of the other categories.

Table 2.1: Survey questions

No†	Mnemonic	Description
23	NEWS	News Heard of Recent Changes in Business Conditions
25	BAGO	Current Business Conditions Compared with a Year Ago
28	BUS12	Business Conditions Expected During the Next Year
29	BUS5	Business Conditions Expected During the Next 5 Years
36	DURRN	Reasons for Opinions About Large Household Durables
38	VEHRN	Reasons for Opinions for Buying Conditions for Vehicles
42	HOMRN	Reasons for Opinions About House Buying Conditions

† Series number refers to the respective question number in the Michigan survey.

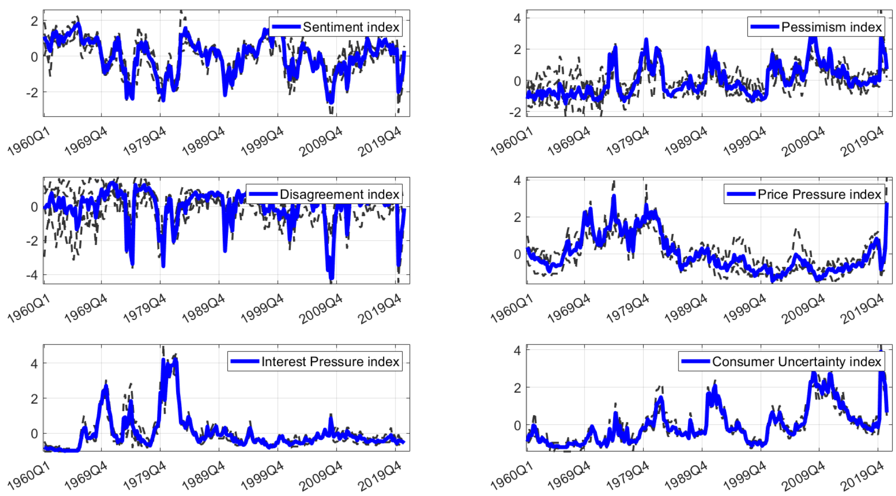

Figure 2.1: *Plots of aggregate consumer survey indexes. Thick blue lines represent final, aggregate indexes of six different categories, and thin black dotted lines represent individual components in this category.*

recession in Dec 2007-June 2009, and the recent economic recession caused by the global pandemic), the sentiment and the disagreement indexes all declined, albeit at different magnitudes and speeds. In contrast, consumer pessimism increased during these periods. The decrease in consumer sentiment and reduction of consumer pessimism during the recession both occur because consumers are generally pessimistic about the economy at this stage. Consumer disagreement also decreased, because almost all consumers heard negative news and felt that current living conditions were not as good as before, and their views also interact. Therefore, consumers' attitudes and views on the future were relatively consistent throughout the recession. In addition, other indicators are also related to each other to varying degrees.

2.4 Forecast comparison of indexes of consumer opinions

2.4.1 Data and Model setting

For the consumer survey indexes, we use 1960Q1-1990Q1 as in-sample data and 1990Q2-2021Q2 as out-of-sample data to predict the target variable recursively.

Four other indexes are selected for comparison with our consumer survey indexes, which are National Financial Conditions Index (NFCI), adjusted NFCI (ANFCI), Financial Uncertainty Index, and Economic Uncertainty Index. Specifically, NFCI is an index released by the Federal Reserve Bank of Chicago every week, and is used to summarize the U.S. financial conditions in money markets, debt and equity markets, as well as the traditional and "shadow" banking systems. It is a weighted average of 105 indicators of the national financial activity. ANFCI is an index that excludes the influence of economic conditions and only describes the financial conditions as the two aspects tend to be highly correlated. The economic and financial uncertainty indexes proposed by Jurado et al. (2015) and Ludvigson et al. (2021) are for the broader macroeconomy and the financial sector, respectively. Given the availability of data, the start date for other indexes is 1970Q1. We also include an indicator called Consumer Index as the comparison index because this indicator can be regarded as a more comprehensive indicator. Unlike a single consumer survey index which focuses on a specific aspect, the Consumer Index is the first common factor extracted from all index series constructed by all single questions.

Two benchmark models are selected to test whether each consumer indicator can improve their forecasting performance when forecasting GDP and other measurements of the real economy, i.e. investment and consumption. The first model is the AR(2) model(labeled AR), and the other is the model AR(2) plus the first five common factors extracted from FRED's 100[5] quarterly series(labeled ARF). Based on these two basic frameworks, this paper compares the prediction ability when adding each index into a constant parameter model, a constant parameter model with stochastic volatility(SV), and a time-varying parameter model with stochastic volatility(TVP-SV). The specific model setting and estimation method

[5]The FRED quarterly large macroeconomic database has 215 variables in total, of which 102 variables can be used in extracting factors. As well as targeted variables, the variable sentiment index should also be excluded in the factor extracting process because it is exactly the same as one of our consumer survey indexes.

are as follows.

We use the basic linear model to do forecasting.

$$\boldsymbol{y_{t+h}} = \boldsymbol{\beta x_t} + \boldsymbol{u_t} \tag{2.7}$$

In the benchmark model AR(2), $\boldsymbol{x_t}$ in Equation 2.7 equals $[\boldsymbol{y_{t-1}, y_{t-2}}]'$, while in the other benchmark model AR(2) plus five macro factors, $\boldsymbol{x_t}$ equals $[\boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}}, \text{factor1}, \boldsymbol{...}, \text{factor5}]'$. We then add one of our indicators one time in the benchmark model, and employed an evaluation measurement i.e. relative mean square prediction error(RMSPE) to compare the prediction accuracy.

For the constant parameter model, OLS is used to estimate them, while for the SV and TVP-SV models, the Bayesian inference with a horseshoe prior is used, as in Korobilis et al. (2021). The prior is

$$\begin{aligned}
\boldsymbol{\beta} \mid \lambda^2, \{\psi_i^2\}_{i=1}^{Tp} &\sim N(0, V), \\
V_{i,i} &= \lambda^2 \psi_i^2, \quad i = 1, \ldots, Tp, \\
\lambda^2 \mid \xi &\sim \text{inv} - \text{Gamma}(1/2, 1/\xi), \\
\xi &\sim \text{inv} - \text{Gamma}(1/2, 1) \\
\psi_i^2 \mid \zeta_i(\tau) &\sim \text{inv} - \text{Gamma}\left(1/2, 1/\zeta_i\right), \\
\zeta_i &\sim \text{inv} - \text{Gamma}(1/2, 1)
\end{aligned} \tag{2.8}$$

Because Equation 2.7 is a direct multi-step ahead forecasting model, the way to set and deal with stochastic volatility is to refer to Chan (2013) for the sake of acknowledging the effect of autocorrelated errors in the estimator variances.

$$
\begin{aligned}
y_t &= \tau_t + u_t, \qquad u_t \sim \mathcal{N}\left(0, e^{h_t}\right), \\
h_t &= h_{t-1} + \varepsilon_t^h, \quad \varepsilon_t^h \sim \mathcal{N}\left(0, \sigma_h^2\right).
\end{aligned}
\tag{2.9}
$$

where τ_t is any mean process.

As mentioned above, we also test the forecasting ability of the consumer survey indexes in the time-varying parameter model setting. The time-varying parameter model is normally written as in 2.10.

$$
\begin{aligned}
y_t &= x_t\beta_t + u_t \\
\beta_t &= \beta_{t-1} + v_t
\end{aligned}
\tag{2.10}
$$

where u_t and v_t are $p \times 1$ vectors of normally distributed error terms. By writing the time-varying parameter regression model Equation 2.10 in a incremental form,

$$y_t = x_t\beta_t + u_t \tag{2.11}$$

$$= x_t\Delta\beta_t + x_t\beta_{t-1} + u_t \tag{2.12}$$

$$= x_t\Delta\beta_t + x_t\Delta\beta_{t-1} + x_t\beta_{t-2} + u_t \tag{2.13}$$

$$\cdots \tag{2.14}$$

$$= x_t\Delta\beta_t + x_t\Delta\beta_{t-1} + \ldots + x_t\Delta\beta_2 + x_t\beta_1 + u_t \tag{2.15}$$

the time-varying parameter model can be written in a static form:

$$\begin{aligned} y &= \mathcal{Z}\beta^{\Delta} + u, \quad & u \sim N(0, \Sigma) \\ \beta^{\Delta} &= v, & v \sim N(0, S) \end{aligned} \tag{2.16}$$

where

$$\mathcal{Z} = \begin{bmatrix} x_1 & 0 & \dots & 0 & 0 \\ x_2 & x_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ x_{T-1} & x_{T-1} & \dots & x_{T-1} & 0 \\ x_T & x_T & \dots & x_{T-1} & x_T \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\beta}^{\Delta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \Delta\boldsymbol{\beta}_2 \\ \dots \\ \Delta\boldsymbol{\beta}_{T-1} \\ \Delta\boldsymbol{\beta}_T \end{bmatrix} \tag{2.17}$$

Here the matrix $\mathcal{Z}$ has Tp covariates but only T observations, so it is a model with more predictors than observations. The horseshoe prior introduced above is perfectly suitable for this form because it is capable of accommodating high dimensional models. Moreover, the stochastic volatility displayed in Equation 2.9 is also trivial to incorporate in this form. The models and methods described above are used to forecast GDP, consumption and investment for h=1,2,4, and 8 quarters ahead.

2.4.2 Forecasting Analysis

As mentioned in the previous section, the evaluation measurement used to compare the predictive accuracy for different models and indexes is the out-of-sample mean square forcasting error(MSFE). Here is a brief introduction.

In practice, forecast accuracy is often evaluated by measuring the discrepancy between the actual and predicted values. One widely used measure of forecast accuracy is the mean square forecasting error (MSFE). The mean square forecasting error (MSFE) is a widely used measure of forecast accuracy in econometrics and time series analysis. It is calculated as the average of the squared differences between the actual values and the predicted values over a specified time period. MSFE is a useful tool for comparing the accuracy of different forecasting models or evaluating the performance of a single model for a specific time period. A lower MSFE indicates better forecasting accuracy and can be used to guide model selection and parameter tuning.

The mean square forecasting error is defined as:

$$\text{MSFE} = \frac{1}{h} \sum_{t=os+1}^{T} (y_t - \hat{y}_t)^2,$$

where y_t is the actual value at time t, $\hat{y}_t$ is the predicted value at time t, and os is the number of in-sample observations, while h is the length of out-of-sample data, which equals to $T - os$. The MSFE measures the average prediction error over the entire time series, and it is often used to compare the performance of different forecasting models.

In normal cases, it is more useful to compare the accuracy of different models based on the relative mean square forecasting error (RMSFE). RMSFE is used to select the best forecasting model among several alternatives, with the numerator representing the model being tested and the denominator representing the benchmark model. By comparing the size of RMSFE and 1, we can determine which model has better predictive ability compared to the benchmark model. Additionally, when using the same benchmark, this value can also be used to compare the predictive abilities of different models, with smaller RMSFE values indicating better predictive performance for the model being tested. The RMSPE is defined as:

$$\text{RMSFE} = \frac{\text{MSFE}_{\text{the model being tested}}}{\text{MSFE}_{\text{benchmark model}}}$$

Therefore, in this section, RMSFEs are provided for easier comparison.

Table 2.2, 2.4, and 2.5 present the evaluation of short-term, medium-term, and long-term predictions of various indicators on GDP, consumption, and investment, respectively. The bold numbers indicate that the prediction performance of the corresponding index is better than the benchmark model, that is, the model without any index. The number that is underlined indicates that the corresponding index has the smallest RMSFE among all indicators. From the three tables, it can be seen that Model 2(M2), a constant parameter model with stochastic volatility, is the most suitable model and provides better forecasting performance than the constant parameter model(M1) and time-varying parameter model(M3) whose accuracy is even poorer than the benchmark model in general. Although the TVP-SV model itself worsens the forecasts, the addition of the consumer survey indexes alleviate this to some extent. However, a detailed analysis of the TVP-SV model is not given in later section because of the overall poor performance of this model. Furthermore, by comparing the predictions for three targeted variables, it is clear that the improvement in GDP prediction by each index is not as substantial as that of consumption and investment. Never-

Table 2.2: Evaluation for GDP Forecasting

		AR				ARF			
		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
M1	No Index	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Sentiment	**0.991**	**0.978**	**0.997**	1.010	1.005	**0.998**	1.008	1.034
	Pessimism	**0.988**	**<u>0.975</u>**	**<u>0.967</u>**	1.001	**0.983**	1.002	**0.997**	1.012
	Disagreement	1.014	**0.997**	**0.993**	**0.999**	**<u>0.892</u>**	1.008	1.000	1.002
	Price Pressure	1.017	1.023	1.033	1.002	**0.977**	1.010	1.035	1.025
	Interest Rate Pressure	**0.994**	1.007	1.012	**<u>0.991</u>**	**0.992**	1.007	1.018	1.004
	General Uncertainty	1.008	1.015	1.002	1.020	1.015	1.005	1.002	1.005
	NFCI	1.003	1.005	1.002	**0.998**	**0.992**	**0.994**	**0.999**	1.004
	ANFCI	**<u>0.974</u>**	**0.999**	1.008	1.006	**0.999**	**<u>0.992</u>**	1.008	1.004
	Financial Uncertainty	1.001	1.004	1.002	**1.000**	1.070	1.004	**0.992**	1.015
	Economic Uncertainty	1.015	1.029	1.018	1.004	1.176	1.003	**<u>0.986</u>**	**<u>0.988</u>**
	Consumer Index	**0.997**	1.007	**1.000**	**0.998**	**0.997**	1.007	**1.000**	**0.998**
M2	No Index	**0.837**	**0.998**	**0.982**	1.009	**0.983**	**0.986**	1.024	**0.998**
	Sentiment	**0.801**	**<u>0.951</u>**	**0.971**	1.012	1.015	**0.984**	1.036	1.004
	Pessimism	**0.811**	**0.957**	**<u>0.969</u>**	1.015	1.027	**0.982**	1.029	1.002
	Disagreement	**0.845**	**0.998**	**0.979**	1.014	**<u>0.929</u>**	**0.991**	1.029	1.001
	Price Pressure	**0.849**	1.007	**0.984**	**<u>0.998</u>**	**0.993**	**0.993**	1.034	**<u>0.993</u>**
	Interest Rate Pressure	**0.825**	**0.988**	**0.979**	1.002	**0.981**	**0.990**	1.025	**<u>0.993</u>**
	General Uncertainty	**0.831**	**0.989**	1.000	1.017	1.028	**0.994**	1.042	1.004
	NFCI	**0.748**	**0.979**	1.005	1.027	**0.964**	**0.927**	1.005	1.017
	ANFCI	**0.761**	**0.995**	1.013	1.027	**0.968**	**0.927**	1.014	1.018
	Financial Uncertainty	**0.775**	1.010	1.015	1.032	**0.965**	**0.938**	1.016	1.015
	Economic Uncertainty	**0.757**	1.012	1.008	1.038	**0.958**	**0.944**	1.012	1.022
	Consumer Index	**<u>0.742</u>**	**0.971**	1.006	1.029	**0.982**	**<u>0.926</u>**	1.003	1.019
M3	No Index	1.564	1.669	1.544	1.017	5.821	1.432	1.453	**0.984**
	Sentiment	1.670	1.656	6.088	1.035	3.198	1.578	6.455	1.001
	Pessimism	1.439	1.258	1.083	1.119	5.567	1.623	1.699	1.065
	Disagreement	1.466	1.052	1.614	1.033	3.226	1.593	1.678	1.170
	Price Pressure	1.485	1.482	1.496	1.045	3.175	1.050	1.267	1.007
	Interest Rate Pressure	1.278	1.445	1.548	1.020	6.096	2.532	1.579	**<u>0.982</u>**
	General Uncertainty	**<u>0.729</u>**	1.031	1.681	1.008	3.218	1.462	1.447	**0.989**
	NFCI	1.431	1.658	1.128	1.023	2.784	1.308	1.056	1.012
	ANFCI	1.555	1.056	1.138	1.018	2.831	1.531	1.116	**0.998**
	Financial Uncertainty	1.403	1.629	1.838	1.032	2.053	1.730	1.815	1.011
	Economic Uncertainty	1.553	1.789	2.757	1.044	2.831	1.154	1.030	1.020
	Consumer Index	1.543	1.654	1.033	1.035	3.959	1.425	1.023	1.035

Notes: M1 represents the constant parameter model estimated using OLS; M2 denotes the constant parameter model with stochastic volatility; M3 is the time-varying parameter model with stochastic volatility. Under each type of model, no index means that an AR(2) or AR(2) plus macroeconomic common factor model is used. RMSFE means the mean square error relative to the benchmark model. For all the models, the benchmark is the constant model without any indexes. A bold number means the corresponding model exhibits better forecasting performance than the benchmark model. A number underlined indicates that it is the the best result in each model category.

theless, consumer survey indexes generally exhibit better prediction ability than other indexes.

For GDP, as indicated in Table 2.2, under the AR framework without macroeconomic common factors, more than half of the consumer survey indexes show their forecasting ability under the M1, and other than 1-step ahead forecasting, their forecasting improvement is unquestionably superior to other comparable indicators. This trend is more obvious when stochastic volatility is taken into account. Under the condition of a forecasting horizon that equals 2 and 4, almost all consumer survey indexes display additional prediction ability to some extent, while other indexes basically only perform well in short-term prediction. Moreover, in the medium and long-term forecast, consumer survey indexes are superior to other indexes. Among these, Pessimism is the best for the the medium term forecast, while Price Pressure produces the most accurate results for the long-term. Conversely, in the short-term 1-step ahead forecast, other indicators clearly outperform consumer survey indexes in terms of GDP prediction, but the super factor, the Consumer Index, gives the best result, achieving 25.8% improvement in GDP forecasting compared to the benchmark model fortunately. This scenario also appears in other forecasting circumstances, namely, when the performance of each consumer index is not as good as that of other indexes in general, the performance of the Consumer Index is often similar to or even better than that of other indexes. The pattern of the result changes after five macroeconomic factors are included in the model. This indicates to some extent that there is an overlap in the information contained in certain indexes and macroeconomic common factors. With the model including macroeconomic common factors, the strongest predictive power lies in the different consumer indexes among the various forecast horizons. Specifically, disagreement exhibits better predictive ability than other indexes when h=1 in the short term, but is not as good as other indexes when h=2 generally, while the Consumer Index is the best among all indexes of SV. Moreover, the long-term forecasting shows the best results for the two pressure metrics.

The Granger causality test result also confirms the above findings.

In finance and economics, the Granger causality test is commonly used to investigate causal relationships between two time series variables. However, in some cases, the standard Granger causality test may not be suitable, as it assumes that the volatility of the series is constant over time. To address this issue, researchers have developed the Granger causality test that considers stochastic volatility in the time series. This approach allows for a more accurate evaluation of the causal relationship between two variables by taking into account the volatility of the series. When applying Granger causality test to stochastic volatility models, we are interested in testing whether one series's volatility is Granger-causal for series stock's volatility, i.e., whether past volatility of one stock can help predict future volatility of another stock, taking into account the potential stochastic nature of the volatility process.

In this study, GARCH(1,1) was used to fit the stochastic volatility of two time series.The results of the Granger causality test overall support the aforementioned prediction results. Taking Model2 of the AR model as an example, the p-value of the test is shown in Table 2.3.

Table 2.3: Granger Causality Test

Index	h=1	h=2	h=4	h=8
Sentiment	0.0139**	0.0057***	0.0002***	0.1331
Pessimism	0.0014***	0.0018***	0.0169**	0.7598
Disagreement	0.0042***	0.0011***	0.0000***	0.9932
Price Pressure	0.3155	0.4498	0.0015***	0.0262**
Interest Rate Pressure	0.7066	0.4534	0.0021***	0.8755
General Uncertainty	0.0000***	0.0000***	0.0000***	0.7450

Notes: Notes: the null hypothesis H_0 represents that the proposed index is not a Granger cause of GDP. A p-value less than 0.1 (or 0.05, or 0.01) indicates the rejection of the null hypothesis at the 0.1 (or 0.05, or 0.01) significance level, suggesting that the proposed index is a Granger cause of GDP, i.e., that the proposed index has predictive power for GDP beyond its own past values.

The results presented in Table 2.3 corroborate those shown in Table 2.2. Specifically, among the eight periods of forward forecasting, only the Price Pressure Index was identified as a Granger cause of GDP. However, while the Price Pressure Index was not found to be a Granger cause of GDP in short-term forecasts, its impact was observed to be significant in longer-term predictions.

Regarding the consumption prediction presented in Table 2.4, in the constant AR(2) model, the result exhibits a similar trend to that in GDP forecasting. Sentiment and Pessimism perform best in the medium term and long term forecast, and ANFCI hits a peak in one-quarter ahead prediction, and the Consumer Index has similar accuracy to ANFCI. When adding macroeconomic factors to the constant model, although the improvement is not large for any of the indexes, consumer survey indexes show the best in each forecasting horizon. The SV model still has the best performance in general. Almost all the consumer survey indexes

Table 2.4: Evaluation for Consumption Forecasting

		AR				ARF			
		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
M1	No Index	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Sentiment	1.021	**<u>0.974</u>**	**0.995**	1.008	1.008	**<u>0.991</u>**	1.007	1.025
	Pessimism	1.011	**0.975**	**<u>0.980</u>**	**<u>0.994</u>**	**0.996**	**0.999**	**1.000**	1.004
	Disagreement	1.012	1.016	1.015	1.004	**<u>0.933</u>**	1.005	1.004	1.001
	Price Pressure	**1.000**	1.004	1.013	**0.996**	**0.999**	1.007	1.018	1.023
	Interest Rate Pressure	1.014	1.009	1.010	**0.996**	1.003	1.002	1.002	**0.996**
	General Uncertainty	1.010	1.004	1.003	**1.000**	1.004	1.012	1.000	**<u>0.993</u>**
	NFCI	**0.997**	**0.998**	**0.997**	**0.998**	1.015	1.000	**1.000**	**0.997**
	ANFCI	**<u>0.995</u>**	1.007	1.026	**0.999**	1.021	1.004	1.002	1.001
	Financial Uncertainty	1.003	**0.999**	**0.997**	1.002	1.038	1.030	1.015	1.004
	Economic Uncertainty	1.017	**0.996**	1.003	1.013	1.173	1.039	1.041	1.001
	Consumer Index	**0.996**	**0.995**	**0.998**	**0.997**	1.001	**0.999**	**<u>0.993</u>**	1.002
M2	No Index	**0.746**	**0.967**	**0.974**	**0.995**	**0.675**	**0.901**	1.038	**0.988**
	Sentiment	**0.683**	**0.949**	**<u>0.965</u>**	**0.998**	**0.690**	**0.890**	1.038	**0.991**
	Pessimism	**0.691**	**0.955**	**0.967**	**0.999**	**0.683**	**0.893**	1.038	**0.990**
	Disagreement	**0.752**	**0.972**	**0.977**	**0.998**	**0.684**	**0.911**	1.040	**0.991**
	Price Pressure	**0.760**	**0.969**	**0.968**	**<u>0.985</u>**	**0.671**	**0.914**	1.042	**0.981**
	Interest Rate Pressure	**0.741**	**0.970**	**0.966**	**0.992**	**0.667**	**0.915**	1.036	**0.986**
	General Uncertainty	**0.715**	**0.963**	**0.982**	1.001	**0.689**	**0.896**	1.048	**0.992**
	NFCI	**0.633**	**<u>0.920</u>**	**0.977**	1.003	**0.529**	**0.823**	1.022	**0.980**
	ANFCI	**0.636**	**0.926**	**0.975**	1.004	**0.534**	**0.826**	1.024	**0.981**
	Financial Uncertainty	**0.665**	**0.939**	**0.980**	1.006	**<u>0.522</u>**	**0.835**	1.026	**<u>0.979</u>**
	Economic Uncertainty	**0.665**	**0.940**	**0.976**	1.013	**0.543**	**0.844**	1.030	**0.982**
	Consumer Index	**<u>0.615</u>**	**<u>0.920</u>**	**0.979**	1.009	**0.556**	**<u>0.801</u>**	1.024	**0.981**
M3	No Index	1.358	1.533	1.617	1.015	1.441	1.493	1.071	**0.978**
	Sentiment	**<u>0.502</u>**	1.308	5.640	1.009	3.133	**0.898**	6.048	**0.988**
	Pessimism	**0.663**	1.488	1.087	1.026	3.575	**0.906**	1.102	1.060
	Disagreement	1.193	1.361	1.034	**<u>1.000</u>**	3.418	**0.912**	1.104	**0.985**
	Price Pressure	1.307	1.150	1.291	1.035	1.577	**0.943**	1.071	**0.976**
	Interest Rate Pressure	1.269	1.453	1.629	1.006	2.307	1.073	1.086	**0.976**
	General Uncertainty	**0.747**	1.519	1.230	1.004	**<u>0.704</u>**	**0.918**	1.126	**0.981**
	NFCI	1.209	1.169	1.559	1.007	1.458	**<u>0.829</u>**	1.076	**0.971**
	ANFCI	1.028	1.012	1.035	1.008	1.516	**0.889**	1.087	**0.975**
	Financial Uncertainty	1.288	1.530	1.798	1.015	1.477	**0.903**	2.026	**0.980**
	Economic Uncertainty	1.376	2.172	1.829	1.027	1.673	**0.968**	1.058	**<u>0.970</u>**
	Consumer Index	1.260	**<u>0.892</u>**	1.203	1.010	1.518	**0.836**	1.043	**0.975**

Notes: M1 represents the constant parameter model estimated using OLS; M2 denotes the constant parameter model with stochastic volatility; M3 is the time-varying parameter model with stochastic volatility. Under each type of model, no index means that an AR(2) or AR(2) plus macroeconomic common factor model is used. RMSPE means the mean square error relative to the benchmark model. For all models, the benchmark is the constant model without any indexes. A bold number means the corresponding model exhibits better forecasting performance than the benchmark model. A number underlined indicates it is the best result in each model category.

provide predictive capabilities in all the forecasting horizons, while the indexes used for comparison fail in long-term prediction. The Consumer Index, NFCI(and the Consumer Index), Sentiment, and Price Pressure give the best consumption forecast on each horizon and achieve 38.5%, 8%, 3.5%, and 1.5% improvements, respectively. Financial Uncertainty and the Consumer Index perform well when the five macroeconomic factors are included in the SV model.

Table 2.5 displays the evaluation results for forecasting investment. The pattern in the constant AR(2) model framework is exactly the same as in GDP and consumption. Except for the 1-step ahead prediction that the ANFCI tops out all the indexes, Disagreement is the one that performs best in the rest of the forecasting horizons. The inclusion of macroeconomic common factors underscores the forecasting ability of the consumer survey indexes in short and medium term prediction but not for long-term forecasting. The consideration of stochastic volatility is continues to be essential and the Sentiment, Pessimism, and the Consumer Index lead in 1-,2-,4- and 8- step ahead forecasting of investment among all the other indexes, achieving 31.5%, 5.5%, 4.2% and 5% improvement, respectively. Almost identical trends are evident when macroeconomic factors are taken into account. This further suggests that the consumer indexes proposed in this study are beneficial for macroeconomic forecasting, but possibly due to the overlap of information with macroeconomic common factors, the performance of consumer indexes is weakened relative to the other comparable indicators when the model incorporates the macroeconomic common factor.

Table 2.5: Evaluation for Investment Forecasting

		AR				ARF			
		h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
M1	No Index	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Sentiment	1.037	1.003	1.021	**0.986**	1.006	**0.987**	**0.992**	1.026
	Pessimism	1.059	1.025	1.018	**0.992**	1.029	**0.989**	**0.992**	1.008
	Disagreement	1.045	<u>**0.984**</u>	<u>**0.970**</u>	<u>**0.984**</u>	1.011	<u>**0.986**</u>	<u>**0.976**</u>	1.004
	Price Pressure	1.017	1.015	1.002	1.021	1.024	**0.989**	1.017	1.009
	Interest Rate Pressure	**0.998**	**0.997**	1.005	1.001	<u>**0.990**</u>	1.009	1.016	1.013
	General Uncertainty	1.028	1.078	1.041	**0.997**	1.035	1.020	1.021	1.026
	NFCI	1.001	1.003	**0.997**	**0.995**	**0.998**	1.016	1.014	1.017
	ANFCI	<u>**0.936**</u>	**0.990**	**0.997**	1.004	**0.991**	1.005	1.016	1.000
	Financial Uncertainty	**0.994**	1.001	1.001	**1.000**	1.033	1.013	1.007	1.044
	Economic Uncertainty	1.060	1.070	1.009	1.009	1.088	1.080	1.008	<u>**0.992**</u>
	Consumer Index	**0.993**	1.000	**0.998**	**0.999**	1.028	1.006	**0.982**	1.024
M2	No Index	**0.743**	**0.968**	**0.967**	**0.995**	1.010	**0.929**	**0.972**	**0.989**
	Sentiment	<u>**0.685**</u>	**0.951**	**0.960**	**0.998**	1.014	<u>**0.923**</u>	**0.975**	1.001
	Pessimism	**0.705**	<u>**0.945**</u>	<u>**0.958**</u>	**0.996**	1.013	**0.928**	**0.973**	**0.992**
	Disagreement	**0.753**	**0.975**	**0.974**	**0.996**	**0.992**	**0.931**	**0.971**	**0.991**
	Price Pressure	**0.755**	**0.969**	**0.964**	**0.981**	1.020	**0.937**	**0.973**	**0.992**
	Interest Rate Pressure	**0.738**	**0.975**	**0.964**	**0.989**	**0.998**	**0.928**	<u>**0.971**</u>	**0.989**
	General Uncertainty	**0.712**	**0.953**	**0.978**	**0.999**	1.033	**0.935**	**0.984**	**0.988**
	NFCI	**0.872**	**0.950**	**0.988**	**0.985**	**0.986**	**0.949**	**0.991**	**0.928**
	ANFCI	**0.866**	**0.947**	**0.996**	**0.986**	<u>**0.980**</u>	**0.959**	**0.994**	<u>**0.925**</u>
	Financial Uncertainty	**0.932**	**0.996**	**0.987**	**0.975**	**1.000**	**0.951**	**0.984**	**0.934**
	Economic Uncertainty	1.070	1.091	1.002	**0.978**	1.017	**0.966**	**0.987**	**0.937**
	Consumer Index	**0.984**	**0.972**	**0.990**	<u>**0.950**</u>	**0.996**	**0.940**	**0.992**	**0.934**
M3	No Index	1.319	1.567	1.591	1.011	2.130	4.090	1.272	1.043
	Sentiment	1.415	1.404	5.638	1.009	1.910	1.290	3.310	1.044
	Pessimism	1.374	1.389	1.479	1.027	2.044	2.484	1.434	1.053
	Disagreement	1.444	1.483	1.677	<u>**0.999**</u>	1.919	1.315	1.315	3.989
	Price Pressure	1.227	1.367	1.385	1.025	1.874	1.273	1.467	1.432
	Interest Rate Pressure	1.104	1.382	1.617	1.005	1.987	1.267	1.567	1.104
	General Uncertainty	1.361	1.421	1.641	1.005	1.945	1.711	1.570	1.031
	NFCI	1.213	1.584	1.606	1.367	1.846	1.413	1.411	**0.974**
	ANFCI	1.399	1.518	1.878	1.352	2.816	2.094	1.789	1.112
	Financial Uncertainty	1.399	1.791	1.715	1.375	1.947	1.591	1.472	1.044
	Economic Uncertainty	1.559	2.288	1.660	1.301	2.406	4.861	1.350	1.045
	Consumer Index	1.415	1.520	1.950	1.123	2.099	1.422	1.455	<u>**0.931**</u>

Notes: M1 represents the constant parameter model estimated using OLS; M2 denotes the constant parameter model with stochastic volatility; M3 is the time-varying parameter model with stochastic volatility. Under each type of model, no index means that an AR(2) or AR(2) plus macroeconomic common factor model is used. RMSPE means the mean square error relative to the benchmark model. For all models, the benchmark is the constant model without any indexes. A bold number means the corresponding model has better forecasting performance than the benchmark model. A number underlined indicates it is the best result in each model category.

2.5 Deep Understanding of the Consumer Index

Korobilis (2020) proposed a new sign restriction for structural identification associated with the MCMC algorithm for estimating parameters of reduced-form vector autoregression(VARs). We use this algorithm to measure the effect of four shocks (aggregate supply shock, aggregate demand shock, monetary policy shock, and financial shock) on our consumer survey indexes.

In our VAR, except for the six consumer survey indexes, we add another 10 endogenous variables that are helpful for identifying four shocks needed. The restrictions used for each variable are presented in Table 2.6.

Table 2.6: Restrictions to identified shocks

Variables	Proxies	Shocks			
		Supply	Demand	Monetary	Financial
GDP	GDP	+	+	+	0
consumption	Consumption	+	+	+	0
stock price	SP500Ind	N A	N A	N A	+
	SP500	N A	N A	N A	+
financial uncertainty	JLNF12	N A	N A	N A	-
stock market volatility	VIX index	N A	N A	N A	-
prices	CPI	-	+	+	N A
	GDPDEF	-	+	+	N A
interest rate	TB3MS	N A	+	-	N A
	Federal funds rate	N A	+	-	N A

Note: The Table describes the restrictions used identifying shocks (in columns) for each variable (in rows) in our VAR. + and − means positive and negative restriction, respectively. 0 is zero restriction, and NA indicates that the response of the variable is left unrestricted.

Figure 2.2 to Figure 2.5 present the impulse responses of 16 endogenous variables to four shocks in our large VAR. The green lines and the shaded areas are posterior medians and their 90% probability bands, respectively.

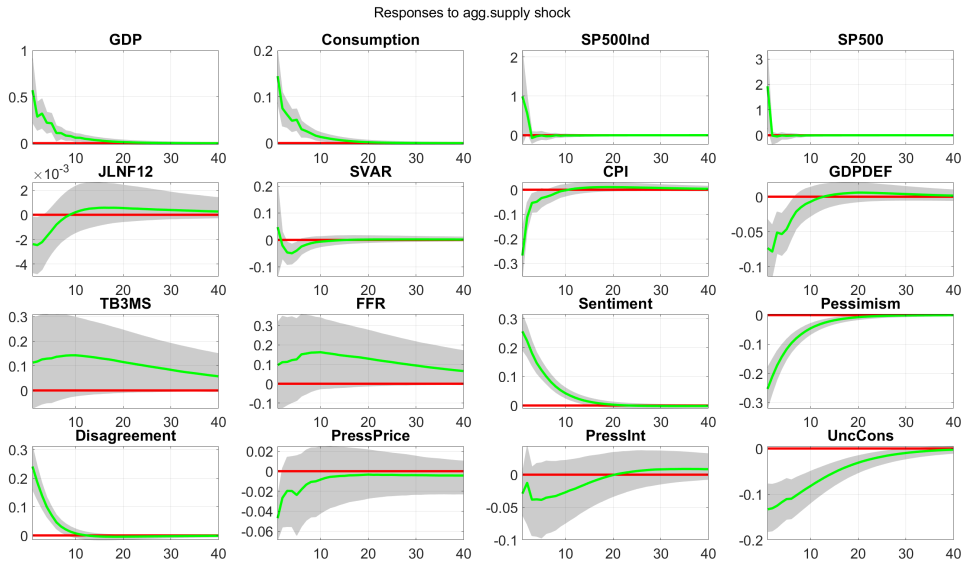

Figure 2.2: *Impulses response functions to an aggregate supply shock*

As for the aggregate supply shock, it is identified as a shock that causes GDP, consumption to react positively and two price proxies, CPI and GDP deflator, to react negatively, contemporaneously, which is also displayed in Figure 2.2 clearly. In this figure, stock price and interest rate do not make any significant response in all periods. However, what is most notable is that the consumer sentiment index and disagreement index exhibit strong positive contemporaneous responses before subsequently disappearing, while consumer pessimism and uncertainty index fall contemporaneously. The mechanism is that when there is a supply boom, consumers' positive attitudes towards general economic and financial conditions will be obviously encouraged, leading to high consumer sentiment and low pessimism. At the same time, consumption intention for a large appliance, vehicle, and house is stimulated, which is reflected in the reduction of consumer uncertainty. The reason for the positive reaction of consumer disagreement is the connotation behind this measurement. During recession periods, people are generally becoming more pessimistic and consumers are less divided because of the bleak market, which is represented in the small consumer disagreement index. By contrast, when people experience an economic explosion, they have a hazy sense of future conditions, thereby exhibiting less disagreement. Therefore, the disagreement index reacts positively to aggregate supply shock contemporaneously.

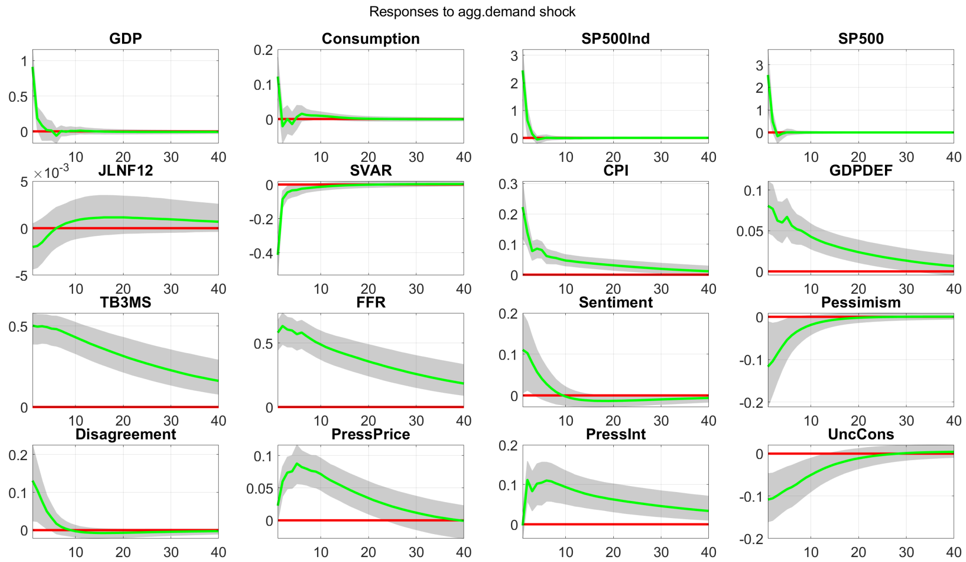

Figure 2.3: *Impulses response functions to a aggregate demand shock*

Figure 2.3 illustrates the impact of an aggregate demand shock, which is defined as a shock that prompts simultaneous positive reactions in GDP, consumption, prices, and interest rates. It is seen that the two indexes, price pressure and interest rate pressure, which do not respond strongly to an aggregate supply shock, react positively contemporaneously and even proceed up before thereafter progressively declining until they lose significance. Intuitively, the increase in purchase pressure of consumers for large durable goods, cars and houses lies behind the interest rate and price innovation induced by demand shock. Consumer pressure stems from fears that prices and interest rates will continue to rise, which is why these two indicators are named price pressure and interest rate pressure. A conspicuous feature of these two indexes is that they are lagging indicators, as the peak appears several periods later than the time at which the shock happens. Unsurprisingly, consumer uncertainty drops as demand shock is always a reflection of a sudden increase in purchasing power, which means people do not have uncertain opinions about buying conditions. As expected, the disagreement index increase slightly in response to the demand shock.

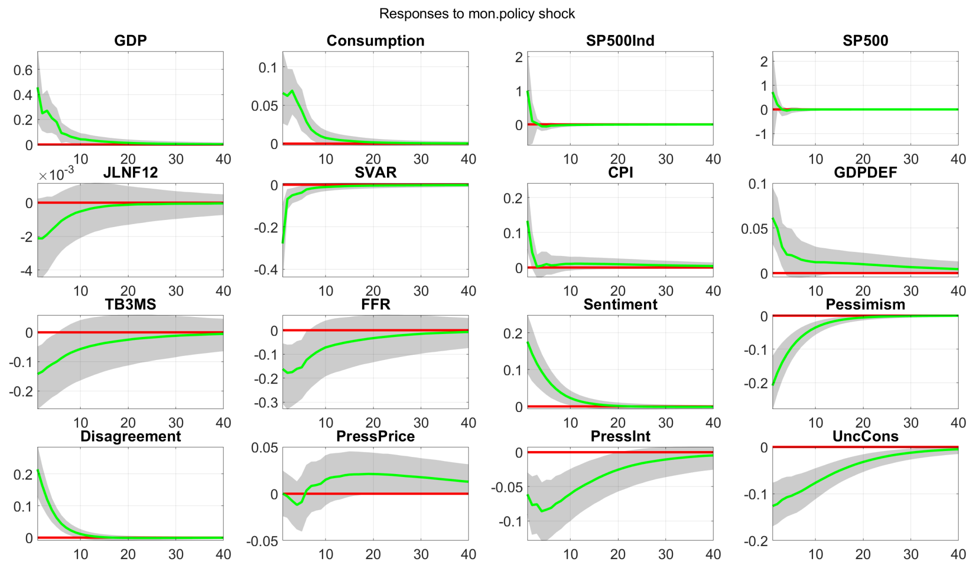

Figure 2.4: *Impulses response functions to a monetary shock*

The impulse response function to a monetary policy shock, which is identified by causing a positive contemporaneous reaction in GDP, consumption and price, and a negative contemporaneous reaction in interest rate is plotted in Figure 2.4. The consumer sentiment gains an increase as before because of the simulation of lower interest rates and/or loose money supply control. The response of the consumer disagreement index exhibits a similar pattern to the previous two shocks, because it only reacts negatively in recession. Also, as is the case in the first two shocks, the effect of monetary policy shock on the stock market-related index and financial uncertainty index is still not statistically significantly different from zero. Moreover, consumer pessimism, interest pressure and uncertainty give negative contemporaneous reactions, which means that the monetary policy shock eases the purchase stress caused by interest rate and other uncertainty factors, reducing the negative emotions of consumers. This validates the implication of the consumer uncertainty index, which can reflect the general uncertainty of consumers.

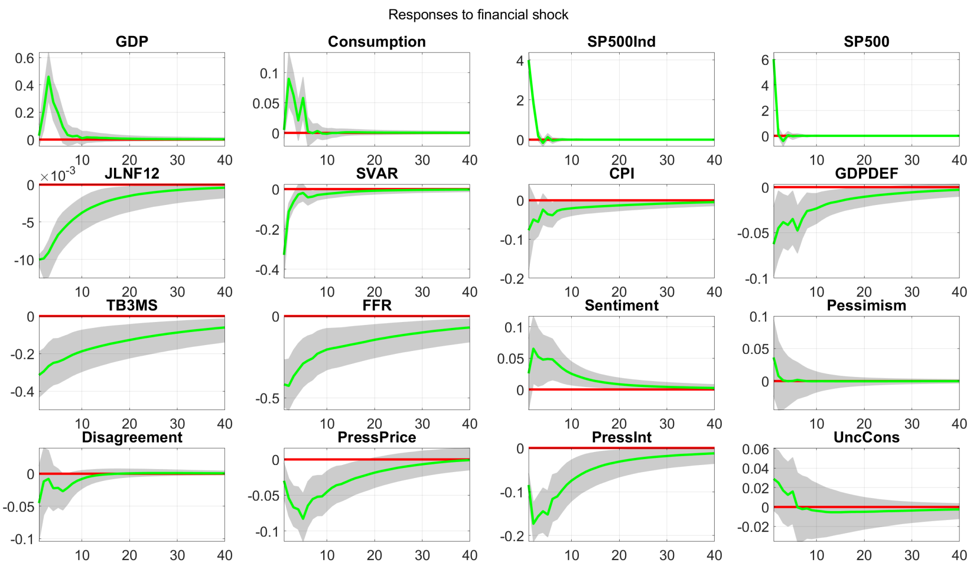

Figure 2.5: *Impulses response functions to a financial shock*

Figure 2.5 represents the impulses response function of the financial shock which moves the stock price index and volatility in a different direction, causes financial uncertainty to react negatively, and imposes a zero contemporaneous reaction of GDP and consumption. It is unsurprising that two interest rate proxies, the three-month treasury bill, which is seen as a safer investment relative to stocks, and the federal funds rate, which is seen as the target interest rate, show strong negative contemporaneous responses since stock price and interest rate are always negatively correlated. Among all consumer survey indexes, only price pressure and interest rate pressure index have strong responses to a financial shock, and the negative reaction means the increase in stock price and decrease of interest rate that comes with financial shock relieve the consumption pressure because of the potential increase of interest rate and price. Last but not least, financial shocks are not a driver of the consumer uncertainty index, and this result is consistent with the variance decomposition shown in Figure 2.6.

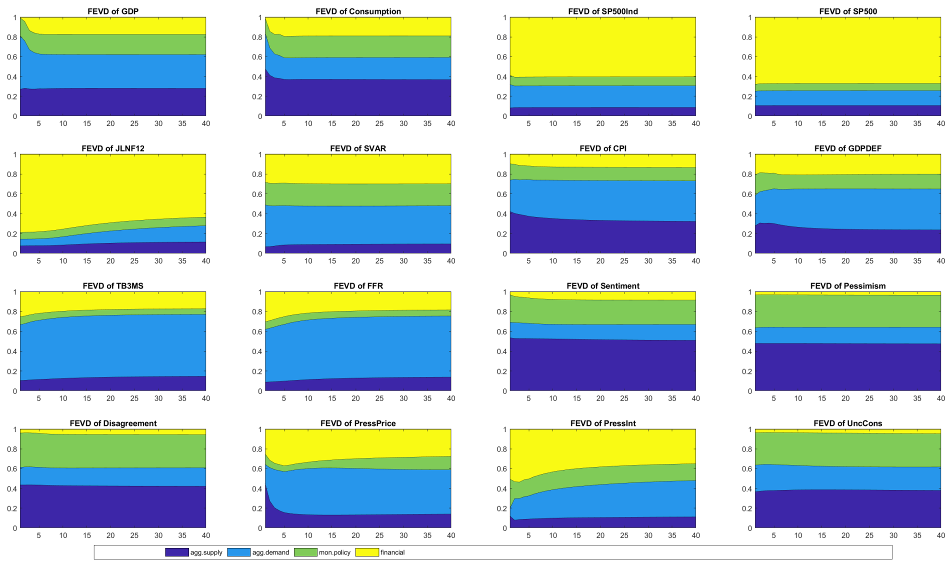

Figure 2.6: *Median Forecast Error Variance Decomposition*

In order to understand the essence of these indicators, we use the variance decomposition to illustrate the percentage that a variable can be explained by supply, demand, monetary policy and financial shocks, which is shown in Figure 2.6. I conjunction with the analysis above, forecast error variance decomposition confirms that consumer uncertainty is not explained by financial shocks at all. It captures all three shocks equally but excludes financial shock, which means that this uncertainty is only relates to macroeconomic conditions. Yet the financial uncertainty(the JLNF 12) proposed by Jurado et al. (2015) is mainly concerned with financial uncertainty. The forecast error variance decomposition of JLNF12 indicates that 80% of its forecast error can be explained by financial shocks in the economy. Moreover, the price and interest pressure are also more financially located variables as they are substantially affected by financial shocks. Regarding the other variables, consumer disagreement and pessimism captured more monetary policy shocks and aggregate supply shocks, and supply shock is the major driver for consumer sentiment. Furthermore, demand shocks dominate the three-month treasury bill and federal funds rate.

2.6 Conclusion

By constructing indicators using six quantification approaches for seven questions in the data set of the Michigan consumer survey, we have six consumer survey indexes with different meanings. They are consumer sentiment, consumer pessimism, consumer disagreement, price pressure, interest pressure, and general uncertainty for consumers. We focus on the performance of these six indicators in forecasting real economic variables, GDP, consumption and investment. The result indicates that, in general, the consumer survey indexes display more predictive accuracy than the other indexes in the literature. Simultaneously, the inclusion of stochastic volatility in the constant parameter model is necessary when analyzing the performance of various indexes. In detail, consumer survey indexes outperform other indexes in the medium and long-run when the model does not include macroeconomic common factors, but perform better in the short and medium forecasting horizons when macroeconomic common factors are added. The final result that needs to be emphasized is that when single consumer survey indexes fall behind other indexes in general, the Consumer Index as the super factor always exhibits a similar or even better predictive capability than other indexes. The follow-up VAR model with sign restriction provides us with a better understanding of the forecasting results given by consumer survey indexes. It illustrates that, compared with other variables, interest rate pressure and price pressure are the only two indexes that will obviously react to financial shocks, while others are more macroeconomic-orientated. Besides, financial uncertainty and consumer uncertainty are both uncertainty indexes, but their emphases are completely different. The first emphasizes finance while the second emphasizes the general economy. Finally, regarding the disagreement index, the degree of disagreement among consumers will rise in an economic boom such as positive supply-side shock and monetary policy shock.

Chapter 3

Time-varying Three-Pass Regression Filter

3.1 Introduction

Factor models have long been established as the default approach for modeling high-dimensional economic data. Numerous algorithms and estimators for extracting unobserved factors exist in the literature, and they can be classified into two main strands. The first strand focuses on factors being a lower dimensional representation of a large vector of macroeconomic, financial or other variables that might (or not) provide useful information for some variable(s) of interest that we want to model as economists. In economics such factors are notoriously difficult to interpret, but they typically provide good explanatory power in different settings (e.g. macroeconomic forecasting, yield curve modeling, exchange rate comovements, etc). However, a second and more recent strand of the literature argues in favor of extracting factors that are supervised, that is, they are estimated with reference to the variable(s) of interest. This paper builds on this second strand of the literature and introduces a novel methodology for extracting "targeted" factors that account for instabilities (structural breaks) in parameters and the information set.

There are numerous ways in which high-dimensional information sets can be incorporated into empirical settings. These range from principal components (Stock and Watson, 2002a,b), to Bayesian shrinkage (De Mol et al., 2008), to reduced rank regression (Velu et al., 1986), and random projection methods (Koop et al., 2019). However, factor analysis by means of principal components has been by far the most popular methodology for forecasting when high-dimensional data sets are available. The main idea underlying factor analysis and related techniques is to reduce the dimension of the subspace spanned by the large data set and select the most important features that capture maximum information about this data set. Principal components (PC) constitute the standard and most widely used method among all the possible ways to estimate common factors.[1] However, forecasting the target variable using a function of unobserved common factors extracted from PC imply a far-fetched assumption that the target variable depends directly on the whole set of predictor variables constantly in every moment. While the stylized fact is that predicted economic variables vary in predictors set, and not only the composition of this set may different in each time period, but also they constitute common factors with varying weight. This leads

[1] For example, Stock and Watson (2002a) and others such as Forni et al. (2000, 2005), Bai and Ng (2008) and Kim and Swanson (2014) work with principal components.

to two possible improvements relative to PCA. The first direction is that of targeting when extracting common factors; The second is considering time-varying factor loadings. We will discuss how our proposed improvements address both of these aspects later in this chapter.

The first improvement builds on the observation that more variables do not necessarily lead to estimated factors with lower uncertainty or better forecasting results, due to the fact that some predictors might not have any predictive power over the target variable. For example, Boivin and Ng (2006) argue that factors estimated from large macroeconomic panels are less useful for forecasting when the idiosyncratic errors are cross-correlated, and also if forecasting power is provided by a factor that is dominant in a small data set but a larger data set. Therefore, the process of reducing the dimensions among the predictors should relate to the forecast goal in order to improve the forecasting ability of the factors. Much of the literature focuses on improvement of this issue. For example, Bai and Ng (2008), who apply the method of principal components to "targeted predictors" selected using hard and soft thresholding rules; and Fuentes et al. (2015), who introduce the sparse partial least squares method, taking into account the response variable for the component estimation, by constructing a factor-forming subset before estimate common factors. The second strand of papers centers on the relationship between all predictors and the predicted variables, and mainly focuses on adjusting the factor loading of the explanatory variables to emphasize those variables that are more important to the target variables when extracting the common factors. The first literature solving this problem in this logic is Wold

(1966) who proposed partial least squares (PLS) regression, in which the common factors tailored to the target variable are constructed such that the covariance between a target variable and these common components is maximized. Recently, the literature considered this including Kelly and Pruitt (2015), who developed a new estimator for factor models—the three-pass regression filter (3PRF)—that relies on a series of ordinary least squares (OLS) regressions. Their work forms the basis of the algorithm proposed in this paper.

The second improvement over traditional PCA is based on the fact that the predictor set might change over time. We implemented this by means of time-varying factor loadings when extracting common factors. Several researchers have challenged the assumption of constant parameters in model estimation and macroeconomic forecasting, see for instance Aastveit et al. (2017). Nevertheless, the number of papers incorporating time instabilities in factor models remains relatively small. Stock and Watson (2009) allow for the possibility of instabilities in the factor loadings when using a factor model to forecast macroeconomic variables, but only consider structural breaks at a single, known point in time. Another paper that considered structural breaks is that by Cheng et al. (2016), who allow the number of pre- and post-break factors to be differ from each other. However, the driving forces of structural changes such as preference changes, technological progress and policy changes, play a role gradually over a long period of time, or some abrupt policy changes also take a period of time to take effect. Therefore, it is more reasonable to consider the time-varying loading instead of setting a break point. Moreover, structural change, the reallocation of economic activity, also causes the feature of short-lived effect between economic variables, which can be captured by time-varying factor loading.

Given facts that the set of original variables required to predict different variables varies, and the contribution of the original variable set to common factors changes over time in extracting common factors for high-dimensional data, this paper proposes a new approach for estimating time-varying factor loadings that incorporates information from predicted variables. We call this method ”time-varying parameter three-pass regression(TVP-3PRF),” which extends the linear three-pass regression filter (3PRF) introduced by Kelly and Pruitt (2015). The 3PRF involves a series of univariate OLS regressions in three passes, as the name suggests. To extend the linear 3PRF into one that involves time-varying factor loading, we propose an efficient algorithm that greatly reduces the computational burden compared to traditional state-space methods. We follow Chan et al. (2014) and Korobilis (2021) and write the TVP regression as a linear regression in stacked form (using banded matrices). This formulation allows the TVP regression to be viewed as a high-dimensional linear regression shrinkage problem, instead of a period-by-period problem (as is the case with the state-space formulation of TVPs). When comparing the computational time needed for estimating time-varying coefficients models using our proposed TVP algorithm versus the traditional Kalman filter algorithm under T=100 and N=500, we found that the Kalman filter algorithm requires 45.1632 seconds to run, while our proposed TVP algorithm takes only 25.9348 seconds. These timings were obtained on a computer with an Intel(R) Core(TM) i5-8250U CPU, which has 4 cores and 8 logical processors, 8GB of RAM, and was running MATLAB version 2022a. In this respect, we then use a simple transformation in conjunction with lasso shrinkage to estimate time-varying loadings within the 3PRF setting.

Given facts that the set of original variables required to predict different variables varies, and the contribution of the original variable set to common factors changes over time in extracting common factors for high-dimensional data, this paper proposes a new approach for estimating time-varying factor loadings that incorporates information from predicted variables. We call this method "time-varying parameter three-pass regression(TVP-3PRF)," which extends the linear three-pass regression filter (3PRF) introduced by Kelly and Pruitt (2015). The 3PRF involves a series of univariate OLS regressions in three passes, as the name suggests. To extend the linear 3PRF into one that involves time-varying factor loading, we propose an efficient algorithm that greatly reduces the computational burden compared to traditional state-space methods. We follow Chan et al. (2014) and Korobilis (2021) and write the TVP regression as a linear regression in stacked form (using banded matrices). This formulation allows the TVP regression to be viewed as a high-dimensional linear regression shrinkage problem, instead of a period-by-period problem (as is the case with the state-space formulation of TVPs). In this respect, we then use a simple transformation in conjunction with lasso shrinkage to estimate time-varying loadings within the 3PRF setting. When comparing the computational time needed for estimating time-varying coefficients models using our proposed TVP algorithm versus the traditional Kalman filter algorithm under T=100 and N=500, we found that the

Kalman filter algorithm requires 45.1632 seconds to run, while our proposed TVP algorithm takes only 25.9348 seconds. These timings were obtained on a computer with an Intel(R) Core(TM) i5-8250U CPU, which has 4 cores and 8 logical processors, 8GB of RAM, and was running MATLAB version 2022a.

The method proposed in this study fills the gap in the field of dimensionality reduction algorithms for high-dimensional data. Specifically, it tackles the challenge of extracting time-varying, sparse factor loadings that are targeted on the predicted variables. Notably, this approach has demonstrated strong performance in both simulated and empirical settings. By means of a Monte Carlo simulation of synthetic data, we study the finite sample accuracy of the TVP-3PRF, both in in-sample factor estimation and out-of-sample forecasting, and compare this to a number of alternative methodologies. We find that the TVP-3PRF performs well regardless of whether the data-generating process (DGP) contains assumptions about regime-switching factor loadings or sparse factor loadings. The TVP-3PRF performs especially well when there are weak factors present in the DGP. When it comes to forecasting performance, factors estimated with the TVP-3PRF algorithm are, in several cases superior to factors coming from alternative algorithms. Finally, we evaluate the performance of the TVP-3PRF in a real-data scenario. In particular, we provide strong empirical evidence that the TVP-3PRF performs well when forecasting major U.S. macroeconomic variables based on the FRED-MD macroeconomic data set. Compared to other related methods – including hard-thresholding method, soft-thresholding methods, PLS, and the original linear-3PRF – our TVP-3PRF performs the best in most cases.

We attribute this excellent forecasting performance of the TVP-3PRF to evidence that the effect of economic variables is short-lived. Based on in-sample evidence, the estimated factor loadings from TVP-3PRF are sparse and characterized by abrupt structural break, revealing that constant parameter and non-sparse factor model methodologies are missing important features of macroeconomic and financial data.

In summary, our proposed algorithm for extracting common factors from high dimensional data constitutes a significant methodological contribution to the field of high dimensional data analysis. Our Monte Carlo simulations demonstrate that the algorithm performs well, not only in accurately extracting common factors from historical data but also in out-of-sample forecasting for data with various features. This highlights the versatility and robustness of our algorithm and its potential to generalize to a wide range of datasets. In terms of empirical evidence, the method proposed in this study also proved to have good results in forecasting macroeconomic variables because it captures the characteristic that the effect of economic variables is short-lived. Our algorithm has practical applications in making accurate predictions, which is crucial in various fields such as finance, economics, and marketing. Overall, our contribution represents a valuable addition to the field of high dimensional data analysis and has the potential to advance research in numerous domains.

This paper is organized as follows: Section 2 introduces the new time-varying parameter approach and the TVP-3PRF algorithm. Section 3 presents a Monte Carlo experiment to study the finite sample accuracy of the TVP-3PRF. Section

4 gathers empirical applications devoted to macroeconomic variables forecasting. Section 5 concludes.

3.2 Time-varying Three-Pass Regression

3.2.1 Time-varying parameter regression estimation

For the scalar observation y_t, $t = 1, \ldots, T$, our model is given by:

$$y_t = x_t \beta_t + \varepsilon_t \tag{3.1}$$

$$\beta_t = \beta_{t-1} + u_t \tag{3.2}$$

where y_t is a 1×1 vector of observable variables; x_t is a $1 \times p$ matrix of observable variables; β_t is a $p \times 1$ vector of time-varying parameters; and ε_t and u_t are $p \times 1$ vectors of normally distributed error terms.

In this form, β_t is a p-dimensional state vector for each t and its estimation can become computationally cumbersome for large T or p. According to Chan and Jeliazkov (2009), more efficient computation can be achieved by writing the system in static(time-invariant) regression form.

Define the matrices:

$$H = \begin{bmatrix} I_p & 0 & \dots & 0 & 0 \\ -I_p & I_p & \ddots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & -I_p & I_p & 0 \\ 0 & 0 & 0 & -I_p & I_p \end{bmatrix}, \quad \mathcal{X} = \begin{bmatrix} x_1 & 0 & \dots & 0 & 0 \\ 0 & x_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & x_{T-1} & 0 \\ 0 & \dots & 0 & 0 & x_T \end{bmatrix} \tag{3.3}$$

are $Tp \times Tp$ and $T \times Tp$ matrices, respectively.

Then using the definitions of the matrices $\mathcal{X}$ and H in Equation 3.3, the time-varying parameter model in Equation 3.1 and 3.2 can be written as

$$y = \mathcal{X}\beta + \varepsilon, \varepsilon \sim N(0, \Sigma) \tag{3.4}$$

$$\beta = H^{-1}u, u \sim N(0, S) \tag{3.5}$$

where $y = (y_1 \dots y_T)'$, and $\varepsilon = (\varepsilon_1 \dots \varepsilon_T)'$ are $T \times 1$ vectors that are formed by stacking all T observations in y_t and ε_t, $\beta = (\beta_1 \dots \beta_T)'$, and $u = (u_1 \dots u_T)'$ are $Tp \times 1$ vectors that are formed by stacking all T observations in β_t and u_t, respectively. The first equation in 3.4 can be treated as a constant parameter regression with parameters β, despite the fact that the $Tp \times 1$ vector of parameters β comprises of T "time copies" of p predictor coefficients.

From a different point of view, we can rewrite the time-varying parameter regression model (Equation 3.1) in a incremental form:

$$
\begin{aligned}
y_t &= x_t\beta_t + \varepsilon_t && (3.6)\\
&= x_t\Delta\beta_t + x_t\beta_{t-1} + \varepsilon_t && (3.7)\\
&= x_t\Delta\beta_t + x_t\Delta\beta_{t-1} + x_t\beta_{t-2} + \varepsilon_t && (3.8)\\
&\dots && (3.9)\\
&= x_t\Delta\beta_t + x_t\Delta\beta_{t-1} + \ldots + x_t\Delta\beta_2 + x_t\beta_1 + \varepsilon_t && (3.10)
\end{aligned}
$$

hence we can rewrite Equation 3.2 as:

$$\Delta\beta_t = u_t \tag{3.11}$$

To compare these two form more intuitively, the matrix form is written:

$$
\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{T-1} \\ y_T \end{bmatrix}
=
\begin{bmatrix}
x_1 & 0 & \dots & 0 & 0 \\
x_2 & x_2 & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots \\
x_{T-1} & x_{T-1} & \dots & x_{T-1} & 0 \\
x_T & x_T & \dots & x_{T-1} & x_T
\end{bmatrix}
\begin{bmatrix} \beta_1 \\ \Delta\beta_2 \\ \dots \\ \Delta\beta_{T-1} \\ \Delta\beta_T \end{bmatrix}
+
\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_{T-1} \\ \varepsilon_T \end{bmatrix}
\tag{3.12}
$$

Then we have

$$
\begin{aligned}
y &= \mathcal{Z}\beta^{\Delta} + \varepsilon && (3.13)\\
\beta^{\Delta} &= u && (3.14)
\end{aligned}
$$

where $\beta^{\Delta} = [\beta_1', \Delta\beta_2', \ldots, \Delta\beta_T]'$. To identify the difference between Equation 3.14 and 3.5, we find that $\mathcal{Z} = \mathcal{X}H^{-1}$, and $\beta^{\Delta} = H\beta$, which means Equation 3.13 and 3.14 are a trivial rotation of Equation 3.4-3.5 by the matrix H. Once we get the result for β^{Δ}, it is easy to obtain the estimation for β.

After writing the time-varying parameter model in this incremental form and treating Equation 3.13 as a constant parameter regression with parameters β^{Δ}, this becomes a high-dimensional problem because the independent variable matrix Z is a $T \times Tp$ matrix in which the number of covariates, Tp, is larger than the number of observations, T.

To estimate such a regression model, we need to use penalty regression to apply shrinkage to some of the parameters to zero. Lasso (Least Absolute Shrinkage and Selection Operator) estimator (Tibshirani, 1996) is one of the most popular methods in high-dimensional data analysis. The lasso estimate is defined by

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \tag{3.15}$$

$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \leq c \tag{3.16}$$

where c is a prespecified free parameter that determines the degree of regularization. The dual to this problem is

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{3.17}$$

The parameter c or λ controls the amount of shrinkage, and this tunning paramter can be chosen by minimizing an estimate of mean square error based on k-fold cross-validation. An important feature of the $L1$ penalty (constrain is in an absolute form) is that some coefficient estimates can be exactly zero.

The estimation of β^{Δ} gotten by using Lasso reveals the dynamic structure of the time-varying parameter β. Specifically, if for a certain increment $\Delta\beta_s = 0$, then the parameter β_t has not varied from period $s-1$ to s, in a other word, $\beta_s = \beta_{s-1}$. Moreover, $\beta_t = (\Delta\beta_t + \Delta\beta_{t-1} + \ldots + \Delta\beta_2 + \beta_1)$. In this way, we can easily get the estimation for time-varying parameter β by performing a static regression.

3.2.2 TVPs in a three-pass regression filter setting

To understand the idea of the whole TVP-3PRF algorithm more intuitively, we first provide an informal introduction: There is a relatively large number N of predictors $\mathbf{x}$, from which we extract factors so as to forecast a target variable y. While $\mathbf{x}$ depends on two sets of common factors, say $\mathbf{F}$ and $\mathbf{G}$ (plus idiosyncratic components), y depends only on $\mathbf{F}$, therefore we wish to extract only $\mathbf{F}$ from $\mathbf{x}$. In addition, there also exist proxy variables, $\mathbf{z}$, whose common components are also driven only by $\mathbf{F}$ as well. This setting is the same as that in Kelly and Pruitt (2015), that introduced the linear 3PRF for estimation of $\mathbf{F}$ and forecasting of y, but the key and novel difference is that we include time variation in factor loadings estimation.

More formally, let us consider the following model:

$$y_{t+1} = \beta_0 + \boldsymbol{\beta_t}'\mathbf{F}_t + \eta_{t+1} \tag{3.18}$$

$$\mathbf{z}_t = \lambda_0 + \boldsymbol{\Lambda_t}\mathbf{F}_t + \omega_t \tag{3.19}$$

$$\mathbf{x}_t = \phi_0 + \Phi_t\mathbf{F}_t + \varepsilon_t \tag{3.20}$$

where y is the target variable of interest; $\mathbf{F}_t = (\mathbf{f}_t', \mathbf{g}_t')'$ are the $K = K_f + K_g$ common driving forces of all variables, the unobservable factors. $\boldsymbol{\beta_t} = \left(\boldsymbol{\beta_t}_f', \mathbf{0}'\right)'$ so that y depends only on f. z is a small set of L proxies that are driven by the same underlying forces as y, such that $\boldsymbol{\Lambda_t} = \left(\boldsymbol{\Lambda_t}_f, \mathbf{0}\right)$; $\mathbf{x}_t$ is a large set of N variables, driven by both $\mathbf{f}_t$ and $\mathbf{g}_t$; and $t = 1, \ldots, T$. In addition,

$$\boldsymbol{\beta_t} = \boldsymbol{\beta_{t-1}} + u_{1t} \tag{3.21}$$

$$\boldsymbol{\Lambda_t} = \boldsymbol{\Lambda_{t-1}} + u_{2t} \tag{3.22}$$

$$\Phi_t = \Phi_{t-1} + u_{3t} \tag{3.23}$$

For brevity, we refer to Kelly and Pruitt (2015) for precise conditions on the factors, the permitted temporal and cross-sectional dependence of the residuals, and the existence of proper central limit theorems.

Based on the model in Equation 3.18 to 3.23, our algorithm for the TVP-3PRF model consists of the following three steps:

Step 1: Time-series regressions of each element of $\mathbf{x}$, x_i, on z; that is, run N lasso regressions

$$x_{i,t} = \phi_{0,i} + \mathbf{z}_t'\phi_{i,t} + \epsilon_{i,t} \tag{3.24}$$

Step 2: Cross-section regressions of x_i, $\hat{\phi}_{i,t}$; that is, run T linear regressions

$$x_{i,t} = \alpha_{0,t} + \hat{\phi}'_{i,t}\mathbf{F}_t + \epsilon_{i,t} \tag{3.25}$$

Step 3: Regression of y_t on factor $\hat{\mathbf{F}}_{t-h}$; that is, run a linear regression for each forecast horizon of interest, h:

$$y_t = \beta_0 + \hat{\mathbf{F}}'_{t-h}\boldsymbol{\beta} + \eta_t \tag{3.26}$$

Note that one can also estimate a time-varying parameter model in the third step, which means it is possible to model time variation in the predictive power of the estimated factors $\hat{\mathbf{F}}_t$ for the target variable y_{t+h}. However, to highlight the importance of time variation in the process of extracting common factors (not time instability in the relationship between the predicted variable and the predictors), and to facilitate comparison with the benchmark algorithm, principle component, we stick to using linear regression in the third step.

Kelly and Pruitt (2015) developed asymptotic theory for the linear 3PRF approach, indicating that the 3PRF-based forecast converges in probability to the infeasible best forecast as cross-section N and sample size T become large. We do not need additional and special conditions to claim that their consistency results could be extended to our algorithm. The reason is quiet obvious because there are no other advanced estimator except for the basically OLS estimator. The consistency and asymptotic normality of the OLS estimator can be found in any econometric textbook. Hence, based on this, the TVP-3PRF should conserve the consistency properties of the linear 3PRF.

3.3 Monte Carlo Simulation

3.3.1 Design

Suppose that $\mathbf{x}_t$ and y_t for $t = 1, 2, ..., T$ are generated via the following factor structure considered by Kelly and Pruitt (2015):

$$\begin{aligned} \mathbf{x}_t &= \Phi_t F_t + \varepsilon_t \\ y_{t+1} &= \Lambda F_t + \eta_t \end{aligned} \tag{3.27}$$

where $F_t = (f_t, \mathbf{g}_t')'$, $\Phi_t = (\Phi_{f,t}, \Phi_{g,t})$ and $\Lambda = (1, 0)$ The relevant and irrelevant factors are generated according to the following dynamics, respectively:

$$\begin{aligned} f_t &= \rho_f f_{t-1} + u_{f,t} \\ \mathbf{g}_t &= \rho_g \mathbf{g}_{t-1} + \mathbf{u}_{g,t} \end{aligned} \tag{3.28}$$

where $u_{f,t} \sim N(0, 1)$, and $\mathbf{u}_{g,t} \sim N(0, \Sigma_g)$, with $u_{f,t}$ and $\mathbf{u}_{g,t}$ uncorrelated. We consider $K_g = 4$ irrelevant factors and $K_f = 1$ relevant factor. Parameters of the diagonal matrix $\sum_g$ are chosen so that irrelevant factors are dominant, in the sense that they have variances 1.25, 1.75, 2.25 and 2.75 times larger than the relevant factor. The parameters ρ_f and ρ_g govern serial correlation among factors and take values of 0, 0.3, or 0.9. We set $y_{t+1} = f_t + \sigma_y \eta_{t+1}$ for $\eta_{t+1} \sim \text{IIN}(0, 1)$ and adjust σ_y to ensure that the infeasible best forecast has an R^2 of 50%. The idiosyncratic terms are assumed to follow autoregressive dynamics,

$$\varepsilon_{i,t} = a\varepsilon_{i,t-1} + \tilde{\varepsilon}_{i,t} \tag{3.29}$$

where a governs their serial correlation and takes values of 0, 0.3 or 0.9. Cross-sectional correlation among idiosyncrasies is specified via $\tilde{\varepsilon}_{i,t} = (1 + d^2)\, v_{i,t} + dv_{i-1,t} + dv_{i+1,t}$, where $v_{i,t}$ is standard normally distributed and the cross-correlation parameter d takes values of 0 or 1.

There are three data-generating process for the Monte-Carlo Simulation, which are constant factor loadings, regime-switching factor loadings, and the factor loading containing the sparse feature which shows the economic predictors are short-lived. The constant factor loadings for each predictor are drawn as standard normal. Regarding the regime-switching factor loadings, without loss of generality, we assume there are two regimes. Two half part factor loadings are drawn from a $U(2, 3)$ uniform distribution and a $U(-3, -2)$ uniform distribution plus a standard normal distribution, respectively. Discuss about the sparse factor loadings, we generate the initial time-varying factor loading matrices first, then set some of them to zeros, and the time-varying factor loadings are produced by $\Phi_t = \mu + 0.99\,(\Phi_{t-1} - \mu) + T^{-3/4}\eta_t$, where $\mu \sim U(-2, 2)$ and η_t is a standard normal. We start from Φ_0 which drawn from $U(0, 4)$ and discard the first 50.

We also consider the different factor strength in accordance with Kelly and Pruitt (2015) so as to determine the circumstance in which our algorithm performs well and circumstance in which it is likely to encounter problems. The factor strength marked by the median percentage of predictor variation is explained by the following factors: 30% for normal factors, 20% for moderately weak factors and 10% for weak factors. For simulations labeled "Non-pervasive Factors", half of the predictors have a loading of zero on the relevant factor, otherwise all predictors have non-zero loadings on all factors. This is achieved by adjusting the variance of the idiosyncratic terms.

3.3.2 Models and evaluation criteria

We perform L = 1000 Monte Carlo replications for each configuration of parameters ρ_f, ρ_g, α and β and sample sizes T and N. Once $\mathbf{x}_t$ and y_t are generated, we apply the TVP-3PRF to extract the factor and predict the target variable. In particular, first, we estimate a time-series (time-varying) regression, $x_{i,t} = \mathbf{z}_t'\boldsymbol{\phi}_i + \epsilon_{i,t}$, for $t = 1, 2, \ldots, T$. For simplicity, we take the proxy variable as the target variable, $\mathbf{z}_t' = y_t$. Second, we run a cross-section OLS regression, $x_{i,t} = \hat{\boldsymbol{\phi}}_{i,t}'\mathbf{F}_t + \epsilon_{i,t}$, for $t = 1, 2, \ldots, T$, using the weighted average of the time-varying factor loadings obtained in the previous step. Third, we run a time-series OLS regression, $y_t = \beta_0 + \hat{\mathbf{F}}_{t-1}'\boldsymbol{\beta} + \eta_t$, and produce the forecast $\hat{y}_{t+1} = \beta_0 + \hat{\mathbf{F}}_t'\hat{\boldsymbol{\beta}}$, which is the final result obtained using the TVP-3PRF approach introduced in this paper.

We compare the predictive performance of our proposed method with several benchmark methodologies. First, we compute the forecast obtained with the linear version of the 3PRF proposed in Kelly and Pruitt (2015). Second, Bates et al. (2013) demonstrate that PCA methods can be applied to consistently estimate dynamic factor models under certain instabilities in the loadings. Therefore, we compute the forecast obtained with the method of principal components. Third, Bai and Ng (2008) argue that the principal components methodology, as it stands, does not take into account the predictive ability of $\mathbf{x}_t$ for y_{t+h} when the factors are estimated. Therefore, Bai and Ng (2008) propose using only predictors that are informative for y_t in the process of the factor estimation. Accordingly, we also compute the forecast obtained with Threshold PCA. Among various thresholds, we consider the lasso and elastic net as soft-thresholding rules, and two hard-thresholding rules based on t statistic associated with X larger than 1.28 and 1.65. Fourth, PLS, which calculates the factor loading maximizing covariance between a target variable and these common components, is also included in the comparison. Ultimately, our focus is on comparing the median out-of-sample MSFE over the 1000 replications associated with each of the seven methods (TVP-3PRF approaches and six competitors) to evaluate their relative predictive performance. Finally, in our simulation experiments, we assume that the number of relevant factors are known (i.e., across all procedures, we extract one factor).

3.3.3 Result

We first compare the in-sample estimation of the relevant factor $\hat{f}_t$ with the real f_t used in the data-generation process. Table 3.1 to 3.3 demonstrate the in-sample fitting results for DGP1 to DGP3.

Table 3.1: MSE for DGP1

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
						Normal					
0	0	0	0	0.36	0.35	1.96	**0.33**	3.05	3.24	2.16	1.90

continued

Table 3.1: MSE for DGP1

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
0.3	0.9	0.3	0	0.46	0.40	2.09	**0.36**	2.10	1.01	2.07	2.14
0.3	0.9	0.3	1	0.43	0.45	1.91	**0.42**	1.72	2.59	1.77	1.82
0.3	0.9	0.9	0	0.68	**0.60**	1.78	0.63	1.22	0.98	1.72	1.48
0.3	0.9	0.9	1	**0.53**	0.74	1.98	0.73	1.03	1.07	1.79	1.97
0.9	0.3	0.3	0	0.48	**0.41**	2.08	0.41	2.14	1.42	2.11	2.50
0.9	0.3	0.3	1	**0.32**	0.38	2.15	0.38	3.17	3.19	1.99	1.98
0.9	0.3	0.9	0	0.39	**0.32**	1.98	0.33	0.62	0.52	1.73	1.43
0.9	0.3	0.9	1	1.06	0.88	2.29	**0.87**	3.14	3.24	2.24	2.78
Moderately weak											
0	0	0	0	0.53	**0.48**	2.02	0.49	2.88	0.99	2.25	2.44
0.3	0.9	0.3	0	0.52	0.50	2.07	**0.50**	2.63	1.35	2.01	1.99
0.3	0.9	0.3	1	**0.53**	0.60	2.04	0.61	1.38	2.84	2.04	1.91
0.3	0.9	0.9	0	**0.65**	0.71	2.14	0.71	2.80	2.88	2.31	2.19
0.3	0.9	0.9	1	0.78	0.77	2.06	**0.76**	2.28	2.46	1.97	2.11
0.9	0.3	0.3	0	**0.52**	0.55	1.95	0.55	2.46	3.36	2.60	2.75
0.9	0.3	0.3	1	0.60	0.49	1.76	**0.46**	1.20	1.09	1.46	1.27
0.9	0.3	0.9	0	0.67	0.54	1.55	**0.53**	2.07	1.58	2.71	2.41
0.9	0.3	0.9	1	**0.34**	0.39	2.37	0.39	2.85	0.82	2.50	1.92
weak											
0	0	0	0	**0.45**	0.60	1.90	0.57	1.06	1.05	1.87	1.88
0.3	0.9	0.3	0	**0.47**	0.58	1.90	0.58	2.69	0.93	2.06	2.06
0.3	0.9	0.3	1	**0.65**	0.74	1.94	0.73	2.57	2.72	2.00	2.50
0.3	0.9	0.9	0	**0.69**	0.74	1.94	0.74	2.83	1.20	2.17	2.17
0.3	0.9	0.9	1	**0.66**	0.95	2.05	0.94	2.64	1.43	2.02	1.84
0.9	0.3	0.3	0	0.50	**0.50**	1.87	0.51	0.91	1.03	2.89	2.00
0.9	0.3	0.3	1	**0.50**	0.74	2.44	0.71	1.48	2.55	2.29	2.46
0.9	0.3	0.9	0	0.51	0.42	1.64	**0.41**	3.04	3.30	2.13	2.15
0.9	0.3	0.9	1	**0.53**	0.76	2.06	0.74	2.71	2.97	2.28	2.12
Non-pervasive Factors											
0	0	0	0	0.47	0.41	1.95	**0.41**	1.42	3.04	1.91	1.91
0.3	0.9	0.3	0	0.54	0.49	1.87	**0.48**	0.86	0.83	1.80	1.83
0.3	0.9	0.3	1	**0.46**	0.53	2.02	0.53	1.40	1.89	2.33	2.11
0.3	0.9	0.9	0	0.58	**0.49**	1.83	0.50	2.89	3.15	1.92	1.78
0.3	0.9	0.9	1	0.70	0.65	1.82	**0.65**	1.83	2.71	2.16	2.25
0.9	0.3	0.3	0	0.57	0.58	2.21	**0.56**	3.23	2.64	2.84	2.83
0.9	0.3	0.3	1	0.77	**0.76**	2.17	**0.76**	0.89	1.34	2.06	1.88

continued

Table 3.1: MSE for DGP1

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
0.9	0.3	0.9	0	**0.44**	0.45	0.93	0.45	0.79	0.67	1.27	1.17
0.9	0.3	0.9	1	**0.61**	0.72	2.68	0.70	2.70	1.06	2.84	1.70

Note: The table reports the median MSE based on 5000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation, respectively in the predictors' residuals. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

Table 3.2: MSE for DGP2

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
						Normal					
0	0	0	0	**0.88**	0.98	2.02	1.34	2.26	2.17	2.04	2.04
0.3	0.9	0.3	0	**0.67**	0.80	2.10	1.07	2.38	2.39	2.12	2.12
0.3	0.9	0.3	1	**0.68**	0.95	2.10	1.11	1.87	1.67	2.13	2.10
0.3	0.9	0.9	0	**0.83**	1.16	1.91	1.23	1.84	2.29	1.92	1.92
0.3	0.9	0.9	1	**0.92**	1.22	1.97	1.23	2.38	2.19	2.11	1.93
0.9	0.3	0.3	0	**0.58**	0.87	2.06	1.16	2.27	2.73	2.08	2.10
0.9	0.3	0.3	1	**1.00**	1.12	2.35	1.59	2.50	2.70	2.47	2.41
0.9	0.3	0.9	0	**0.52**	0.56	2.20	0.70	3.04	2.44	2.16	2.10
0.9	0.3	0.9	1	0.72	**0.65**	2.08	0.69	2.36	2.76	2.04	1.83
						Moderately weak					
0	0	0	0	**0.69**	1.05	1.99	1.28	1.95	1.88	1.99	2.00
0.3	0.9	0.3	0	**0.66**	0.89	1.90	1.07	2.28	2.52	1.97	1.96
0.3	0.9	0.3	1	**0.73**	0.98	2.09	1.29	2.19	2.52	2.05	2.06
0.3	0.9	0.9	0	**1.17**	1.26	1.85	1.50	1.53	2.24	2.18	2.17
0.3	0.9	0.9	1	**1.27**	1.38	2.00	1.34	1.47	1.42	1.80	1.75
0.9	0.3	0.3	0	**0.53**	0.69	2.21	0.86	1.44	1.44	1.77	1.77
0.9	0.3	0.3	1	**0.70**	0.82	2.43	1.08	2.47	2.54	2.46	2.42
0.9	0.3	0.9	0	**0.52**	0.53	2.14	0.70	2.92	3.02	2.76	2.69
0.9	0.3	0.9	1	0.65	**0.53**	2.04	0.70	2.83	2.87	2.10	2.09
						weak					
0	0	0	0	**0.58**	0.87	1.97	0.88	2.00	1.69	2.10	2.11
0.3	0.9	0.3	0	**0.79**	0.99	1.89	0.99	1.61	1.67	1.70	1.70
0.3	0.9	0.3	1	**0.72**	1.15	2.05	1.17	1.56	1.65	2.00	1.93

continued

Table 3.2: MSE for DGP2

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
0.3	0.9	0.9	0	**0.82**	1.19	2.07	1.20	2.36	1.85	2.15	2.13
0.3	0.9	0.9	1	**0.89**	1.19	2.03	1.25	1.60	1.60	2.06	2.20
0.9	0.3	0.3	0	**0.66**	0.80	1.99	0.87	2.49	2.64	1.88	1.95
0.9	0.3	0.3	1	**0.64**	0.95	1.83	1.04	1.78	1.74	1.99	1.91
0.9	0.3	0.9	0	**0.52**	0.72	1.81	0.76	2.85	2.41	2.25	2.64
0.9	0.3	0.9	1	1.31	**1.29**	2.14	1.33	1.64	1.62	3.27	2.69
				Non-pervasive Factors							
0	0	0	0	**0.64**	0.93	2.04	1.23	2.26	2.36	2.13	2.12
0.3	0.9	0.3	0	**0.77**	1.07	2.04	1.09	2.27	2.03	2.04	2.04
0.3	0.9	0.3	1	**0.87**	1.20	1.79	1.39	1.62	1.70	1.83	1.82
0.3	0.9	0.9	0	**1.06**	1.28	2.32	1.29	2.52	2.50	2.20	2.22
0.3	0.9	0.9	1	**0.80**	1.16	1.67	1.26	1.57	1.45	1.88	1.64
0.9	0.3	0.3	0	**0.59**	0.83	2.29	1.00	2.07	2.09	2.25	2.21
0.9	0.3	0.3	1	**0.80**	0.98	2.45	1.34	2.70	2.72	2.39	2.43
0.9	0.3	0.9	0	0.84	**0.76**	1.67	0.80	1.00	1.05	1.95	2.03
0.9	0.3	0.9	1	**0.56**	0.73	2.18	0.82	1.25	1.09	1.79	2.23

Note: The table reports the median MSE based on 5000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation, respectively in the predictors' residuals. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

Table 3.3: MSE for DGP3

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
				Normal							
0	0	0	0	**0.87**	1.09	2.12	1.41	2.38	2.45	2.18	2.17
0.3	0.9	0.3	0	**0.71**	1.10	2.02	1.37	2.22	2.34	2.13	2.12
0.3	0.9	0.3	1	**0.74**	1.15	2.08	1.28	2.25	2.42	2.10	2.08
0.3	0.9	0.9	0	**0.88**	1.16	2.08	1.22	2.29	2.48	2.07	2.06
0.3	0.9	0.9	1	**0.92**	1.22	2.04	1.23	2.31	2.48	2.09	2.10
0.9	0.3	0.3	0	**0.85**	1.17	2.17	1.35	2.46	2.59	2.17	2.17
0.9	0.3	0.3	1	**0.88**	1.25	2.36	1.46	2.62	2.74	2.43	2.46
0.9	0.3	0.9	0	**0.54**	0.64	2.18	0.79	2.30	2.35	2.25	2.34
0.9	0.3	0.9	1	**0.67**	0.76	2.21	0.87	2.26	2.64	2.21	1.98

continued

Table 3.3: MSE for DGP3

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PC	PLS	Hard1	Hard2	Soft1	Soft2
					Moderately weak						
0	0	0	0	**0.83**	1.03	2.11	1.32	2.49	2.67	2.19	2.18
0.3	0.9	0.3	0	**0.74**	1.10	2.08	1.34	2.56	2.64	2.14	2.14
0.3	0.9	0.3	1	**0.72**	1.02	2.12	1.23	2.45	2.50	2.16	1.91
0.3	0.9	0.9	0	**0.84**	0.97	2.12	1.28	2.25	2.46	2.11	2.15
0.3	0.9	0.9	1	**0.79**	1.22	2.12	1.11	1.94	2.59	1.96	2.03
0.9	0.3	0.3	0	**0.76**	1.11	2.20	1.37	2.43	2.53	2.23	2.28
0.9	0.3	0.3	1	**0.79**	1.24	2.17	1.20	2.46	2.67	2.25	2.25
0.9	0.3	0.9	0	**0.52**	0.84	2.19	1.06	2.37	2.54	2.19	2.20
0.9	0.3	0.9	1	**0.62**	0.73	2.26	0.73	2.43	2.82	2.52	1.84
					Weak						
0	0	0	0	**0.99**	1.37	2.12	1.43	2.36	2.49	2.10	2.10
0.3	0.9	0.3	0	**0.83**	1.05	2.09	1.23	2.35	2.49	2.14	2.13
0.3	0.9	0.3	1	**0.68**	0.87	2.08	1.05	2.31	2.58	2.10	2.08
0.3	0.9	0.9	0	**0.84**	1.15	2.11	1.36	2.38	2.50	2.11	2.11
0.3	0.9	0.9	1	**1.07**	1.42	2.07	1.43	2.29	2.34	2.02	2.01
0.9	0.3	0.3	0	**0.79**	1.10	2.20	1.34	2.41	2.64	2.28	2.25
0.9	0.3	0.3	1	**0.68**	1.17	2.27	1.32	2.37	2.44	2.33	2.35
0.9	0.3	0.9	0	**0.74**	0.76	2.32	1.23	2.52	2.64	2.36	2.33
0.9	0.3	0.9	1	**0.64**	0.86	2.17	0.92	2.53	2.60	2.21	1.99
					Non-pervasive Factors						
0	0	0	0	**0.80**	0.99	2.14	1.10	2.44	2.49	2.17	2.17
0.3	0.9	0.3	0	**0.72**	1.05	2.14	1.24	2.38	2.57	2.14	2.15
0.3	0.9	0.3	1	**0.77**	1.20	2.19	1.29	2.36	2.53	2.17	2.20
0.3	0.9	0.9	0	**0.92**	1.16	2.17	1.33	2.39	2.52	2.19	2.22
0.3	0.9	0.9	1	**0.98**	1.17	2.11	1.23	2.14	1.64	2.03	2.00
0.9	0.3	0.3	0	**0.89**	1.31	2.15	1.45	2.35	2.45	2.17	2.16
0.9	0.3	0.3	1	**1.05**	1.37	2.48	1.43	2.69	2.71	2.46	2.45
0.9	0.3	0.9	0	**0.66**	0.85	2.20	1.10	2.53	2.63	2.24	2.21
0.9	0.3	0.9	1	**0.62**	0.93	2.58	0.98	2.71	2.87	2.33	2.47

Note: The table reports the median MSE based on 5000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation, respectively in the predictors' residuals. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

Table 3.1 reports the Monte Carlo simulation results for the DGP in which the factor loadings are constant throughout the sample length. When assuming that co-movement among macro variables is assumed to be stable over time, TVP-3PRF, linear 3PRF and PLS all exhibit similar performance in terms of mean square error (MSE) and outperform the principal component as well as methods that perform variable selection prior to estimating the common factor. This implies that if the relationship between all large macroeconomic variables and the common factor does not change over time, the four algorithms mentioned before are all able to provide relatively equal accuracy in estimating of the underlying relevant factor. When a break point exists in the factor loading series, the TVP-3PRF algorithm exhibits greater advantages, as can be seen in Table 3.2. In most cases, the MSE for TVP-3PRF is the smallest , and the constant 3PRF performs the best in only 4 cases.

The results reported in Table 3.3 are based on a DGP where the factor loadings are not only time-varying but are also sparse. This more complex DGP is able to reflect more accurately the time-varying and short-lived relationship between economic variables in the real world. In this table, the TVP-3PRF performs the best in all cases whatever the strength of relevant factors, which means this algorithm is able to capture the relationship between the set of predictors and its relevant common factors.

Because the performance of TVP-3PRF on DGP3 is so excellent that it clearly exceeds all other algorithms, we need to explore its out-of-sample performance on DGP3. For the sake of brevity, we do not distinguish factor strength on this occasion. The relative mean square forecasting error(RMSFE) for the out-of-sample forecasting is presented in Table 3.4.

Table 3.4: *RMSFE for DGP3*

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	hard1	hard2	soft1	soft2
0.3	0.9	0.3	0	**0.988**	1.069	1.013	1.011	1.013	1.001	**0.995**
0.3	0.9	0.3	1	1.030	1.066	1.026	1.013	1.023	0.996	**0.995**
0.3	0.9	0.9	0	**0.990**	1.060	1.025	1.003	1.026	1.010	1.001
0.3	0.9	0.9	1	1.035	1.095	1.047	1.025	1.035	**0.998**	**0.995**
0.9	0.3	0.3	0	**0.820**	**0.888**	**0.943**	**0.992**	**0.977**	**0.998**	**0.997**
0.9	0.3	0.3	1	**0.725**	**0.873**	**0.900**	**0.986**	**0.974**	1.003	1.001
0.9	0.3	0.9	0	**0.708**	**0.854**	**0.857**	**0.987**	**0.956**	1.001	1.000
0.9	0.3	0.9	1	**0.728**	**0.800**	**0.790**	**0.907**	**0.878**	**0.961**	1.002

Note: The table reports the out-of-sample median relative mean square forecasting error(RMSFE) based on 1000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation, respectively in the predictors' residuals. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

Table 3.4 presentes the RMSFE of 7 algorithms, and the benchmark model is PCA. Therefore, a number less than 1 indicates that the predictive ability of this model is better than that of PCA, and vice versa. It is evident that except for two cases where soft-threshold PCA performs best, TVP-3PRF performs best for all the other settings. The next section examines the performance of TVP-3PRF when applied to real economic data.

3.4 Empirical Application

The full dataset consists of quarterly observations on 248 U.S. macroeconomic time series from 1960Q1 through 2021Q4. According to the FRED-QD Appendix provided by Michael W. McCracken, we only use variables which have the factor code 1 to extract common factors,[2] therefore the overall number of predictors is

[2] These are lower level disaggregates, while variables with factor code 0 are aggregates. In this way way, we use, for example, savings and consumption to extract factors from, but not total product (which is a linear combination of these variables).

101 in the end. We use the data from 1960Q1 to 1989Q4 to begin our in-sample estimation while the remaining observations are used for evaluating forecast accuracy. The series are transformed by taking logarithms and/or differencing according to the standard transformation codes provided by FRED-QD. Let Y_{t+h}^{h} denote the variable to be forecasted in a h-period ahead forecast at annual frequency. Before forecasting each target, we standardize the data to ensure unit variance in the full sample, and then transform the data by partialing the target and predictors with respect to a constant and four lags of the target. Specifically, autoregressive dynamics are partialed out by initially regressing Y_{t+h}^{h} and X_t on $1, Y_t^1, Y_{t-1}^1, Y_{t-2}^1$, and Y_{t-3}^1. $\tilde{Y}_{t+h}^{h,cv}$ and $\tilde{X}_t^{cv}$ are the residuals from these regressions, and also the variables used in the formal three passes. For the target-proxy TVP-3PRF (which means we set the target variables as the proxy): the first pass regressing $\tilde{X}_{i,t-h}^{cv}$ on $\tilde{y}_t^{cv}$ and a constant for $t = 1, 2, \ldots, T$, which is run separately for each $i = 1, 2, \ldots, N$, yielding $\hat{\phi}_{i,t}$ the dimension of which is $(T-h, N)$. We then complete the last h number of $\hat{\phi}_{i,t}$ as the same value as $\hat{\phi}_{i,T-h}$; the second pass regressing $\tilde{x}_{i,t}^{cv}$ on $\hat{\phi}_{i,t}$ and a constant for $i = 1, 2, \ldots, N$, separately run for each $t = 1, 2, \ldots, T$, yielding $\hat{f}_t$; the third pass regressing $\tilde{y}_t^{cv}$ on $\hat{f}_{t-h}$ and a constant for $t = 1, 2, \ldots, T$, yielding $\hat{\beta}_0, \hat{\beta}$. The out-of-sample forecast is then constructed as $\hat{\beta}_0 + \hat{f}_t\hat{\beta}$.

3.4.1 Forecasting economic activity

We use eight competing approaches: PCA from all the 101 predictors; PLS; PCA where hard-thresholding (t=1.28&t=1.65) has been performed before extracting the first principal component to forecast (HardThres1 and HardThres2); PCA where soft thresholding (shrinkage method=Lasso&Lars) has been performed before extracting the first principal component to forecast (SoftThres1 and SoftThres2); linear 3PRF; TVP-3PRF. For the 3PRF-related approaches, we use one factor and use the predicted variable as a target proxy in the first step of the 3PRF approach (called target-proxy 3PRF in Kelly and Pruitt (2015)).

Table 3.5: RMSFE for macroeconomic variables

	h=1	h=2	h=3	h=4	h=8	h=1	h=2	h=3	h=4	h=8
	GDP					Consumption				
PLS	0.94	1.00	0.89	0.98	0.94	1.22	1.06	1.06	1.12	0.95
3PRF	0.95	0.97	0.93	**0.91**	0.98	1.01	1.00	**0.94**	**0.95**	0.98
HardThres1	1.11	1.14	1.22	1.28	1.09	0.97	0.94	1.01	0.98	1.02
HardThres2	1.08	1.10	1.34	1.26	1.12	**0.89**	**0.87**	1.00	0.96	1.04
SoftThres1	1.04	1.22	0.89	1.19	1.04	0.95	1.03	1.02	1.10	1.02
SoftThres2	1.04	1.17	1.04	1.21	1.04	0.97	1.04	0.96	1.16	1.02
TVP-3PRF	**0.78**	**0.83**	**0.87**	**0.91**	**0.89**	1.00	1.05	1.05	1.09	**0.82**
	Industrial Production					Employment				
PLS	1.15	1.12	0.98	1.06	1.04	1.02	**0.88**	0.92	0.96	0.95
3PRF	0.84	0.94	0.93	0.94	1.01	0.95	1.01	0.96	0.96	0.92
HardThres1	1.01	1.17	1.25	1.32	1.05	1.09	1.06	1.15	1.14	0.92
HardThres2	0.93	1.17	1.29	1.33	1.06	1.10	1.06	1.30	1.20	**0.84**
SoftThres1	0.94	1.05	1.03	1.21	1.04	0.94	1.05	1.14	1.05	0.98
SoftThres2	0.90	1.05	1.02	1.22	1.04	0.94	1.11	1.15	1.05	1.00
TVP-3PRF	**0.76**	**0.83**	**0.86**	**0.89**	**0.94**	**0.88**	**0.88**	**0.90**	**0.94**	0.93
	Hours					CPI				
PLS	0.87	0.77	0.85	0.94	1.01	0.88	0.87	0.82	0.75	0.99
3PRF	0.96	0.97	0.96	0.99	0.96	0.91	0.86	0.83	0.81	0.85
HardThres1	1.07	1.07	1.15	1.18	1.05	1.01	0.99	0.98	0.99	0.97
HardThres2	1.06	1.13	1.17	1.24	**0.93**	1.02	0.98	0.98	0.99	0.97

continued

Table 3.5: RMSFE for macroeconomic variables

	h=1	h=2	h=3	h=4	h=8	h=1	h=2	h=3	h=4	h=8
SoftThres1	0.97	1.12	1.11	0.94	1.03	1.06	0.92	0.93	0.80	1.04
SoftThres2	0.94	1.15	1.13	1.02	1.05	1.06	0.91	0.91	0.81	1.06
TVP-3PRF	**0.84**	**0.77**	**0.81**	**0.86**	0.96	**0.77**	**0.68**	**0.65**	**0.57**	**0.59**
	GS10					Investment				
PLS	1.39	1.44	1.31	1.26	1.39	1.10	0.91	0.93	1.02	1.54
3PRF	1.30	1.51	1.56	1.39	1.26	0.99	0.86	0.82	0.84	0.98
HardThres1	1.15	1.12	1.10	1.15	**0.90**	1.02	0.98	1.00	1.02	1.00
HardThres2	1.12	1.20	1.19	1.16	1.02	1.02	1.00	1.01	1.05	1.01
SoftThres1	0.99	1.06	1.04	1.07	1.03	1.07	1.05	1.22	0.92	1.01
SoftThres2	0.99	1.07	1.03	1.06	1.03	1.06	1.04	1.10	1.07	1.01
TVP-3PRF	**0.97**	**0.97**	**0.97**	**0.94**	1.10	**0.90**	**0.71**	**0.65**	**0.69**	**0.92**
	S&P500					Export				
PLS	1.26	1.11	1.13	1.04	**0.95**	0.96	0.91	0.87	0.88	1.14
3PRF	1.04	1.12	1.13	1.15	1.09	0.91	0.89	0.94	0.98	0.99
HardThres1	1.03	1.04	1.11	1.05	1.03	1.02	0.98	0.98	0.99	1.01
HardThres2	1.05	1.08	1.10	1.08	1.06	1.06	0.98	0.99	0.99	1.00
SoftThres1	1.00	1.01	1.01	0.98	0.96	0.95	0.97	1.06	1.13	1.03
SoftThres2	1.00	1.01	1.01	0.98	0.97	0.96	0.93	1.05	1.00	1.04
TVP-3PRF	**0.94**	**0.96**	**0.96**	**0.95**	0.96	**0.81**	**0.77**	**0.78**	**0.76**	**0.98**
	Import					FedFund				
PLS	**0.56**	**0.82**	0.92	1.07	1.10	0.61	0.84	0.94	1.15	1.78
3PRF	0.71	0.84	0.89	0.89	0.93	0.78	0.88	1.00	1.09	1.03
HardThres1	0.99	1.05	1.11	1.16	1.03	0.92	0.95	0.99	1.03	1.00
HardThres2	0.96	1.06	1.12	1.17	1.05	0.91	0.97	1.01	1.05	**0.98**
SoftThres1	0.82	1.24	1.06	1.39	1.02	1.26	1.44	1.11	1.12	1.26
SoftThres2	0.81	0.93	1.16	1.24	0.93	1.16	1.22	1.10	1.10	1.30
TVP-3PRF	0.60	0.89	**0.82**	**0.82**	**0.82**	**0.56**	**0.73**	**0.76**	**0.92**	1.23

Note: Entries in this table are root mean squared forecast errors for $h = 1, 2, 3, 4, 8$-steps ahead. TVP-3PRF and 3PRF use a single automatic proxy. PLS uses only the first common (Wold (1966)). HardThres1, HardThres2, SoftThres1 and SoftThres2, use t statistic 1.28 and 1.65, lasso, and lars respectively, to select a predictor subset from which principal components are extracted and used to forecast, in accordance with Bai and Ng (2008). The benchmark model is PCA regression using the first principal component. Bold numbers denote the best-performing procedure for each forecast horizon.

Table 3.5 presents the out-of-sample forecasting results. All results are reported relative to the forecasts obtained from PCA. Hence, a number below 1 indicates that a given approach outperforms PCA. Overall, across all 5 forecast horizons and all 12 predicted variables (60 cases), the TVP-3PRF yields the best forecasting results in 49 cases, PLS in 4 cases in which one case performs equally as well as TVP-3PRF, the linear 3PRF in 3 cases in which one case performs equally as well as TVP-3PRF, and hard-threshold in 5 and 1 case(s) for t=1.65 and t=1.28, respectively. For all the predicted variables, TVP-3PRF only has poor performance in the short- and medium-term forecasting of consumption, whereas for almost all other variables, TVP-3PRF can be said to be the best prediction method. Specifically, TVP-3PRF provides the best prediction for all the horizons for GDP, industrial production, CPI, investment, and export, and it also gives the smallest mean square forecasting error in most prediction horizon for the remainder of the predicted variables. Finally, for import, TVP-3PRF still provides the most accurate results for h=3,4, and 8.

To explore how TVP-3PRF plays a role in predicting macroeconomic variables, we chose the first six predicted macroeconomic variables (GDP, consumption, Industrial Production, Employment, Hours, CPI) as representative examples to illustrate the time-varying factor loading in the one-step prediction of the target variable at a certain time point in recursive forecasting. Figure 3.1 depicts this using a heatmap.

The title of each subplot displays the predicted variable and the time point of the one-step forecasting. In each subplot, the horizontal axis represents different predictors, and the vertical axis represents different in-sample time points. Therefore, the color shade of each color bar represents the size of the corresponding factor loading for each predictor at different times. As illustrated in Figure 3.1, for all target variables, the factor loading of most predictors at most times is zero, and only a few variables contribute to the final common factor. More importantly, for the same predictor, its contribution to common factors is not constant with time. For example, when predicting the consumption in 1990Q1, the 35th predictor is more important in the middle of the sample than at the beginning and end of the sample because the factor loading is larger, whereas when forecasting the Hours in 2021Q1, the 90th variable is less important in the middle stages of the sample than in the early and later stages. More obvious examples are the 97th and 100th predictors in forecasting the value of Industrial Production in 2021Q4 and Hours in 2021Q4, respectively, which are ineffective in almost half of the period and extremely important in the other half. Thus, there is a variable that enters or exits the variable set used in extracting factors at a certain time, which fully illustrates the previously mentioned fact that economic variables are short-lived in macroeconomic forecasting.

3.5 Conclusion

In this paper, we proposed a factor model estimation method used in high-dimensional data analysis, which takes into account the factor extracting process

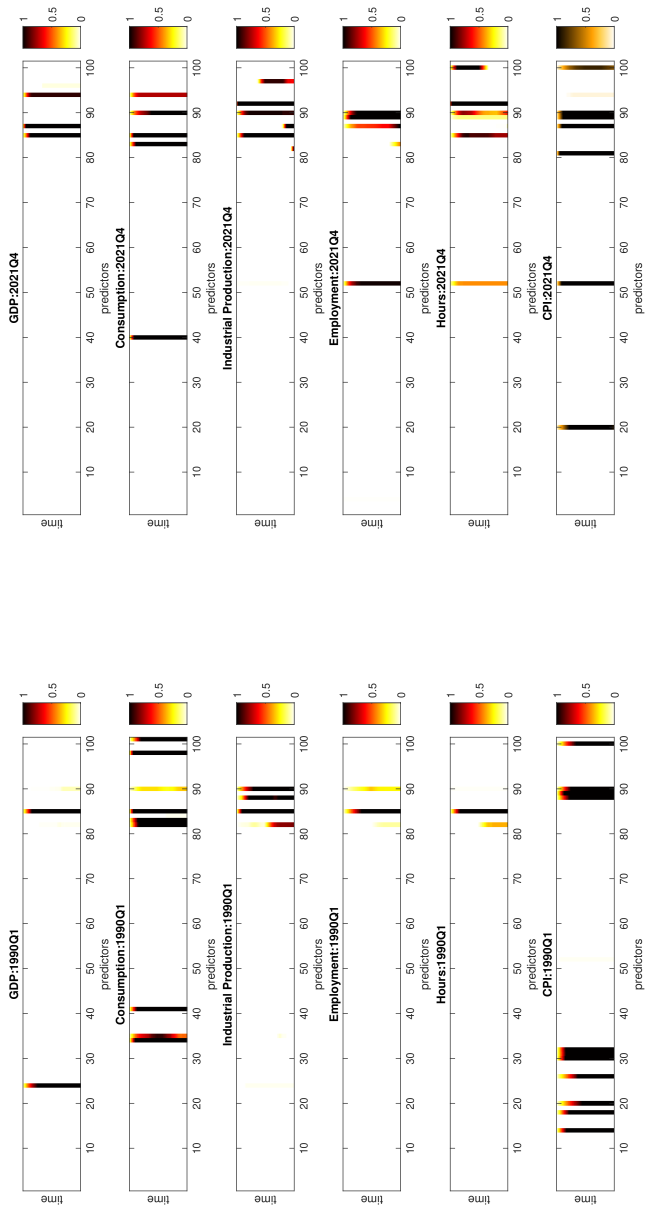

Figure 3.1: *Heatmap of the time-varying factor loadings for six targeted variables. The title of each subplot shows the variable and the time point we are going to forecast, and the forecasting horizon is 1.The x-axis denotes the predictors used in the extracting process, and the y-axis is the time. The shade represents the value of factor loading in each time*

should both target on predicted variable and have time-varying factor loading. Targeting on the predicted variable, TVP-3PRF can effectively identify the subset of factors that is useful for forecasting a given target variable while discarding factors that are irrelevant but may be pervasive among predictors. With the benefit of time-varying factor loading, the short-lived characteristics of economic variables were considered in the process of common factor estimation, which makes the estimation of relevant factors more accurate. Simulation and the empirical study both demonstrated the ability of TVP-3PRF. From the Monte Carlo simulation, it was clear that TVP-3PRF exhibits the best fitting on the true factor when the factor loading is time-varying or time-varying and sparse, while its performance in constant factor loading setting was similar to linear 3PRF. In empirical research, compared with other methods (PCA, PLS, linear 3PRF, hard-threshold based on t =1.28 and 1.64, soft-threshold based on Lasso and Lars), TVP-3PRF performed best in short-term, middle-term and long-term prediction for most macroeconomic variables.

Chapter 4

Sparse or Dense? High-dimensional macroeconomic forecasting in a multi-country setting

4.1 Introduction

The development of digital technology typically involves the generation of high dimensional data, and poses challenges in statistical learning and modeling. There are generally two categories of methods for processing this kind of data, depending on whether the information set used in the final model and algorithm is sparse or dense. What we are interested in is what the real data structure is like in the forecasting model.

According to the dense structure viewpoint, all explanatory variables are useful and contain information for the prediction of a target variable. Although some variables may provide little information, it is beneficial to enhance the prediction ability to include them. For example, the widely used factor model and ridge regression is the product of this dense-structured view. Several studies employ this kind of technique in economics and finance research (De Mol et al., 2008; Hoerl and Kennard, 1970; Lawley and Maxwell, 1962; Leamer, 1973; Pearson, 1901; Spearman, 1961; Stock and Watson, 2002a,b; Tihonov, 1963). As another technique to avoid overfitting, those who use sparse modelling not only have the perspective that the information used for forecasting the predicted variable only exists in a few important explanatory variables, but also believe that the inclusion of irrelevant variables can even jeopardise the predictive accuracy. Therefore, such a standpoint leads to the conclusion that only the relevant variables should be identified and used in the model. The final data set used in the model is low-dimensional and sparse following the principle of sparse modelling. The most famous mathematical procedure in this class is the Lasso(least absolute shrinkage and selection operator, Tibshirani, 1996). To summarize, in this study, we use the terms 'dense' and 'sparse' to refer to different approaches for estimating and predicting target variables. The dense approach involves utilizing the entire set of explanatory variables, whereas the sparse approach involves selecting only the most useful variables for estimation and prediction purposes.

We aim to explore the real structure of the data in the prediction model. Although the performance of these two classes of methods is both good in various cases, in fact, if the data is not sparse, the significance of sparse modelling becomes unclear. In this case, researchers use sparse models not because the real model is sparse, but rather because doing so reduces estimation uncertainty, eases the curse of dimensionality and improves the predictive accuracy. This artificial assumption of sparsity leads to the consequence that the selected explanatory variables no longer represent the fact that they are important for predicting the target variable.

The method employed in this paper is the spike-and-slab prior proposed by Giannone et al. (2021). The greatest feature of this prior is to allow the model to be sparse, but not to assume it, which is attributed to the design of two solutions to avoiding the curse of dimensionality in the same prior. In the past, much existing literature using spike-and-slab prior tends to set the probability of inclusion for the predictors in the model to be 0.5, that is, they assume that the indicator employed to assess whether a variable is included in the model is a Bernoulli distribution with a parameter of 0.5. However, the spike-and-slab prior proposed by Giannone et al. (2021) does not make such an assumption, but sets a prior distribution to Bernoulli parameters, which greatly improves the flexibility of the model, achieving the so-called "allow the model to be sparse but not assume it"(Giannone et al., 2018, p.2). At the same time, when a variable is included in the model, the shrinkage degree of its coefficient, i.e. the variance of the coefficient distribution, is also set as a hyperparameter. Under this setting, the solution to the over-fitting problem of the high-dimensional data is not only variable selection but also shrinkage. This method is briefly introduced in the next section.

Working with this approach, we test the performance of six large economies when forecasting production growth. these economies are the UK, the EU and the four largest countries in the EU (Germany, France, Italy, Spain). For each country, we have exactly the same 69 variables. The detailed variable list and its transform method can be found in the appendix. The results indicate that all of the economies display the favor of sparse model since the probability of variables inclusion of each economy is distributed mainly within the interval [0,0.5] and peak at a point that even lower than 0.2. Certain variables are important for all the economies according to the posterior, such as the industrial production index in manufacturing, Dow Jones industrial share price index, and U.S. unemployment. Nevertheless, uncertainty still exists because there is a high degree of collinearity among many predictors, which means that the choice of relevant explanatory variables remains unclear. UK, France and Italy exhibit less uncertainty than other economies, thus models with few explanatory variables should have a better performance in the process of forecasting. For the economies where the degress of uncertainty degree is high i.e. EU, Germany and Spain, model averaging techniques with different sets of regressors would be very helpful for achieving better prediction. However, the most surprising result of this research is caused by the COVID-19 pandemic. Since the pandemic started, the dense model has surpassed all the other methods becoming the best for all the economies, which indicates that in a period with chaos and disorder, combining all the information could be the best choice.

In the second section, the spike and slab prior is introduced. The empirical analysis is presented in the third section, and includes the illustration of data, posterior exploring and out-of-sample forecasting research. The final section is the conclusion.

4.2 Methodology

The model used to forecast a response variable y_t is,

$$y_t = u_t'\phi + x_t'\beta + \varepsilon_t \tag{4.1}$$

where ε_t is an i.i.d. normal error term with zero mean and variance equal to σ^2, and u_t and x_t are two vectors of regressors of dimensions l and k respectively, typically with $k \gg l$, and whose variance has been normalized to one. Without loss of generality, the vector u_t represents the set of explanatory variables a researcher always wants to include in the model; for instance, a constant term or fixed effects in a panel regression. Therefore, the corresponding regression coefficients ϕ are never identically zero. Instead, the variables in x_t represent possible, but not necessarily useful predictors of y_t, as some elements of β might be zero.

To capture these ideas, and address the question of whether sparse or dense representations of economic predictive models fit the data better, the following prior distribution is set for the unknown coefficients (σ^2, ϕ, β),

$$p\left(\sigma^2\right) \propto \frac{1}{\sigma^2}$$
$$\phi \sim \text{ flat}$$
$$\beta_i \mid \sigma^2, \gamma^2, q \sim \begin{cases} \mathcal{N}\left(0, \sigma^2\gamma^2\right) & \text{with pr. } q \\ 0 & \text{with pr. } 1-q \end{cases} \qquad i = 1, \ldots, k \tag{4.2}$$

The priors for σ^2 and the low-dimensional parameter vector ϕ are rather standard and designed to be uninformative. Instead, the elements of the vector β are either zero, with probability $1-q$, or normally distributed with the same variance, given the standardization of the regressors. The hyperparameter γ^2 plays a crucial role as it controls the variance of this Gaussian density, and thus the degree of shrinkage when a regressor is included in the model.

To specify a hyperprior on q and γ^2, defining the mapping $R^2\left(\gamma^2, q\right) \equiv \frac{qk\gamma^2\bar{v}_x}{qk\gamma^2\bar{v}_x+1}$ where $\bar{v}_x$ is the average sample variance of the predictors (equal to 1 in our case, given standardization of the x 's). We then place the following independent priors on q and R^2:

$$\begin{aligned} q &\sim \text{Beta}(a, b) \\ R^2 &\sim \text{Beta}(A, B) \end{aligned} \tag{4.3}$$

The marginal prior for q is a Beta distribution, with support $[0, 1]$, and shape coefficients a and b. In our empirical applications, we work with $a = b = 1$, which corresponds to a uniform prior. Turning to γ^2, it is difficult to elicit a prior directly on this hyperparameter. Instead, the function $R^2(\gamma^2, q)$ offer an intuitive interpretation of the share of the expected sample variance of y_t due to the $x_t'\beta$ term relative to the error. We model this ratio as a Beta distribution with shape coefficients A and B, and base our inference on the uninformative case with $A = B = 1$. The appeal of this hyperprior is that it can be used for models of potentially very different size, because it has the interpretation of a prior on the R^2 of the regression. Another attractive feature is that it is agnostic about whether to deal with the curse of dimensionality using sparsity or shrinkage.

4.3 Empirical analysis

4.3.1 Data

This paper uses the monthly data of six economies to forecast production growth. The variables are mainly selected based on Caggiano et al. (2011).

Given that the variable set should be exactly the same for comparison purpose, the final number of the variables for each county is 69, including the target variable which is production growth. According to the classification of data by Stock and Watson (2002a,b), these data cover nine categories, namely: Money, Financial Variables, Surveys, Industrial Production, Other indicators of real activity, labour market, HICP, PPI, and international variables. All data are processed to stationarity. A detailed list of variables and the transformation code used to make series stationery can be found in the Appendix.

4.3.2 Analysis of Posterior Distributions

The main results are based on the two important parameters mentioned above: the probability of inclusion q and shrinkage degree γ.

Figure 4.1 presents the marginal posterior distribution of the parameter probability of inclusion q, which is obtained by integrating the γ^2 out from the joint posterior distribution.

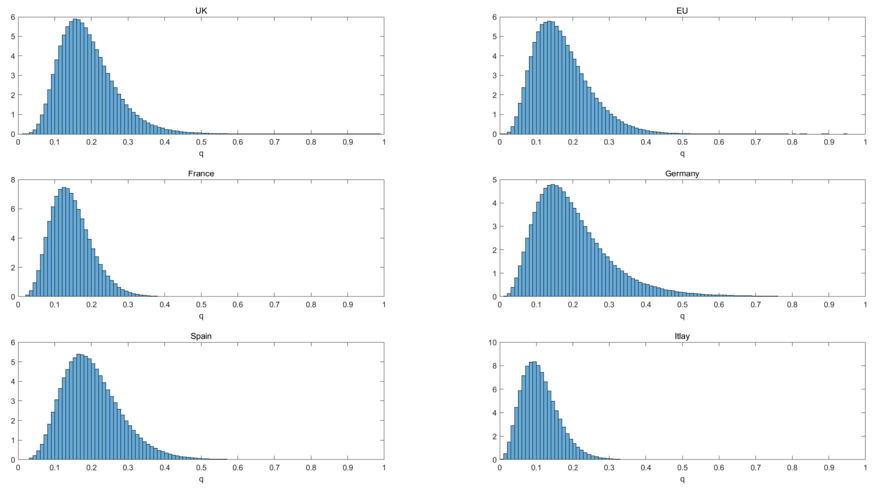

Figure 4.1: *Posterior density of q*

As Figure 4.1 indicates, there is little difference in the distribution of q in the six economies because they are all concentrated in areas less than 0.5, and the peak values are between 0.1 and 0.2 except Italy which has a mode less than 0.1. If distinguishing specifically, the assumption of sparsity might be more reasonable for the two economies, Italy and France, while Germany may be better suited to the dense model the most, because the distribution of q even spread to above 0.7. Based on the results up until this step, the key question becomes which variables are the most important and should be included in the model and whether these relevant predictors can be identified. This question can be intuitively addresses by viewing the heatmap of the probability that each coefficient is included in the model.

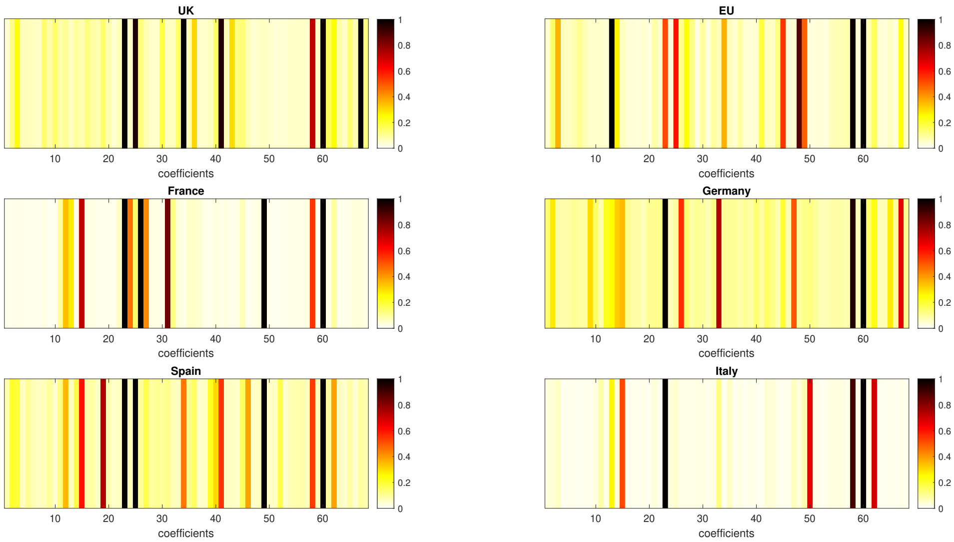

Figure 4.2: *Heatmap of the probabilities of inclusion of each predictor*

Although the overall probability of inclusion, i.e. q, is fixed for a model of an economy, the probability of inclusion for a single explanatory variable, which can be measured by the frequency of inclusion among all the posterior sampling, is different. The different probability of inclusion of different variables is drawn in the heatmap where each longitudinal band represent a possible predictor, and the darker the colour, the higher the probability that this variable is included in the model. Notably, the average probability of inclusion across all possible predictors is q.

Figure 4.2 presents the heatmap for each country. Firstly, no matter how many dark vertical bars there are in each picture, if we want to measure the uncertainty of the variable selection for this economy, we need to check the overall depth of the background color except for a few exceptionally dark vertical bars. Only when the background color is light, and the dark bar is clearly defined, can it be explained that it is generally straightforward to select relevant and important variables in the analysis and forecast process for this economy. This means that the uncertainty of variable selection is relatively low. On the contrary, if the background color is not light, or the difference between the color bars is not significant, this means that every variable has a similar probability to be included in the model, therefore the uncertainty of variables contained in the sparse model is high. Under these circumstances, it may be that the weighted average of many possible models with different possible predictors will provide the best prediction.

According to this principle, it can be seen that in Figure 4.2 that most areas of Italy, France and the UK are lighter than those of the other three countries. This indicates that the uncertainty of these three countries is relatively low, and it is easy to distinguish which variables should be included in the model. Taking Italy, which has the lowest uncertainty, as an example. If all series with a probability greater than 0.08 are included in the model, this gives eight variables: trade balance, exports of goods, IPMAN(industrial production: manufacturing), manufacturing production: future tendency, unemployment rate: females, Dow Jones index, U.S. unemployment, U.S. total civilian employment. Conversely, the other three economies, Germany, EU and Spain, especially Germany, have extremely high uncertainty in variable selection. In line with the same standard, there are 58 variables whose probability of inclusion is greater than 0.08 in Germany, thus only 10 explanatory variables can be determined to be excluded from the model.

However, there are also several sparsity models that have successfully selected a certain number of predictors and exhibit good forecasting performance when forecasting the production growth of Germany. According to Giannone et al. (2021), this is because those kinds of literature that exclude a priori possibility that the real model is high-dimensional, even though the real model should contain many potential variables. This can be seen explicitly in the heatmap (Figure 4.3) of the probability of inclusion of each variable conditional on all possible values of q.

As illustrated in Figure 4.3, the vertical axis indicates the size of the probability of inclusion q. From bottom to top, the probability of inclusion becomes larger, so that the probability of inclusion of each variable (horizontal axis) in the model increases accordingly. Taking Germany as an example, when q is very small, only around four of the most important variables are chosen and put into the model for forecasting production growth, and with the increase of q, the predictors that should be selected are increasingly uncertain. Especially when q is very high, almost all explanatory variables have a high probability of being added to the model. However, this situation does not exist in Italy and France, because the full distribution of q even does not contain a part greater than 0.6. Even at the right end of the q distribution, a high degree of sparsity remains, which means a clear preference for the sparse model with little uncertainty.

The black-and-white line in Figure 4.3 represents the position of the posterior mode of q. For all the economies, the uncertainties of identifying relevant variables

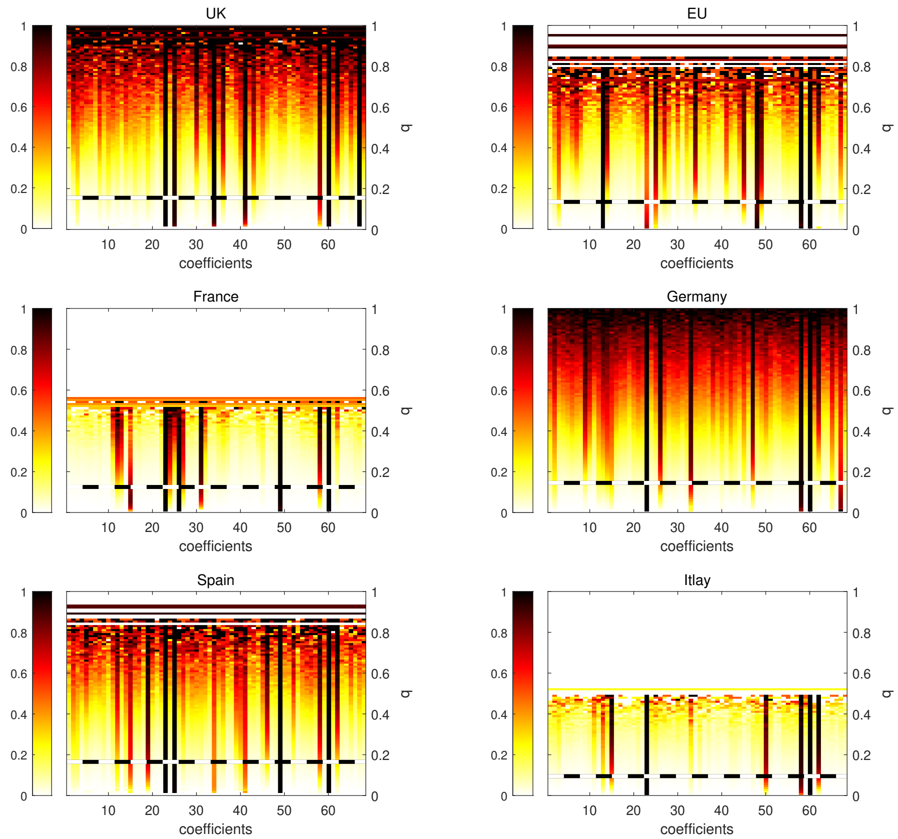

Figure 4.3: *Heatmap of the probabilities of inclusion of each predictor, conditional on q*

conditional on the mode of q are not all high. The uncertainty is higher only when q is at a higher position in the posterior distribution. But according to Giannone et al. (2021), when the low-dimensional model is used for prediction, only a small degree of uncertainty may still cause some major consequences, hence the model averaging technique might be still extremely useful and hence preferred. The improvement in the forecasting accuracy of the model using the model average is studied in the next section through a comparison with the sparsity and dense model.

4.4 Out-of-sample forecasting

The previous section illustrates that although all the countries prefer sparse setting based on the posterior of q, uncertainty is more substantial in the relevant variable selection process in the EU, Germany and Spain than in the UK, France and Italy. In these cases, sparsity-based methods may cause the loss of predictability. While it seems that for the UK, France and Italy data, the sparsity assumption is appropriate and relevant important predictors can be selected. This section identifies the consequences of ignoring model uncertainties by using the model average technique and comparing them with low-dimensional model predictions.

The question to be explored is that when the inclusion of probability clearly favors the sparse model, and there is a certain degree of uncertainty in the specific choice of variables, will this provide better prediction than the dense model and model average technology by restricting the model space in a variety of informative ways? Similar to the study by Giannone et al. (2021), in this exercise we re-estimate the model on many training samples, obtained as subsets of the full sample. We then evaluate the predictive performance of our model on several corresponding test samples. Specifically, the full sample is 1993M1 to 2021M8. The in-sample period is set as 1993M1-2000M8 at the beginning, and the corresponding out-of-sample period is 2000M9-2001M8. We then repeat this exercise 21 times, each time expanding the training sample by one year and shifting the evaluation sample accordingly. This is in fact consistent with recursively regression, except that adding 12 steps each time is more time-saving than adding 1 step each time, under the condition that the conclusion is still credible and valuable.

The models we compare include three models with Bayesian model averaging, which considers the uncertainty of variable selection. They are BMA-all, which is a full model that combines all the possible individual models, weighted by their posterior probability; BMA-5 and BMA-10, which restrict the model space to the combinations of individual models with up to five and ten predictors respectively, weighted by their relative posterior probability. And those of the single best models with up to five and ten predictors, SS-5 and SS-10, which are selected as the models with the highest posterior probability in the set of those with up to five and ten predictors are also included in the comparison. Moreover, the dense model with all the predictors, SS-k, i.e. the ridge regression is considered. Other sparse models include:

1. Lasso(Tibshirani, 1996)-5, 10, cv2, cv5, cv10 and *asy*.

2. Post-Lasso(Belloni and Chernozhukov, 2013)-5, 10, cv2, cv5, cv10 and *asy*.

3. SBR(Polson and Sun, 2019a)-5, 10, cv2, cv5 and cv10.

   (These are models with five or ten predictors, and with selection based 2-,5-, and 10-fold cross-validation. *asy* denote the penalty parameter based on the asymptotic criterion proposed by Bickel et al. (2009).)

4. TBFMS-I or II or III or IV: four version of TBFMS (Kozbur (2020))

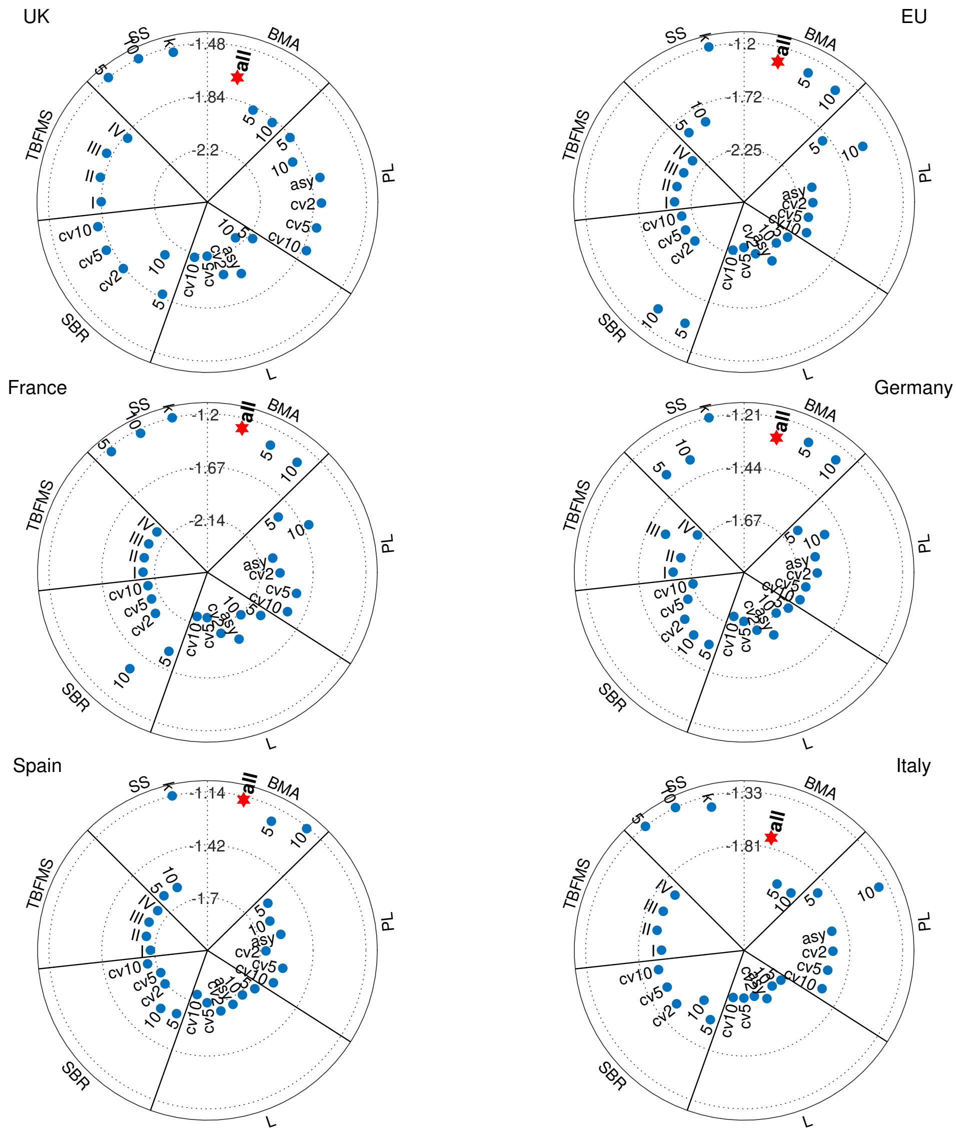

Figure 4.4: *Out-of-sample average log predictive score of different models*

Compared with tables, polar scatter plots are extremely intuitive graphs that can illustrate and compare the prediction accuracy of different estimation methods. Figure 4.4 and 4.5 are polar scatter plots of mean square prediction error(MSPE) and average log predictive score(LPS), respectively, for all the above estimation methods. To analyze these two figures more conveniently, all the results have been processed such that the closer the position is to the outer circle, the better the prediction results. This means the outer circle number of out-of-sample average log predictive scores is larger than the number in the inner circle, while the outer circle number of out-of-sample of mean squared forecasting error is smaller than its counterpart in the inner circle.

The results of the LPS presented in Figure 4.4 are highly consistent with the previous analysis. For the countries with a small uncertainty, Italy, France and the UK, SS-5 and SS-10 perform well in prediction. Given that LPS is an indicator to measure whether the density forecasting is accurate, it is unsurprising that using the variables selected by the highest posterior can give the best prediction for the target variable. Moreover, excluding these two sparse models, the BMA-all is almost better than all other models, which suggests that when the statistical results support that models are low-dimensional, even if there is a relatively small degree of model uncertainty, the researcher cannot use only one sparse model, and a model that integrates all the variable sets with different sizes is still highly advantageous in density forecasting. For the three countries with relatively high uncertainty, BMA-all is obviously the best prediction model except for SS-k. At the same time, BMA-5 and BMA-10 also performed very well, which highlights

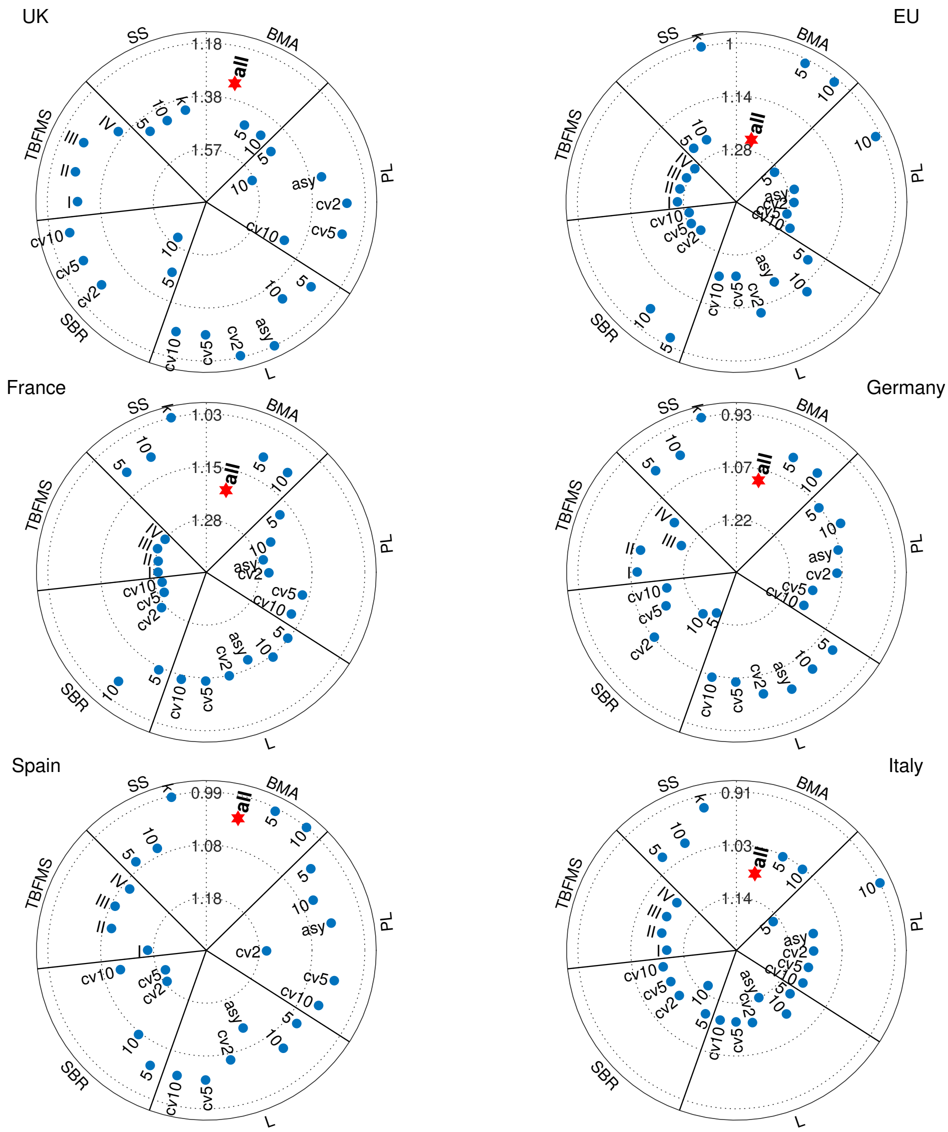

Figure 4.5: *Out-of-sample mean squared prediction error of different models*

the importance of considering uncertainty when data prefer the sparse setting. In such a scenario, considering all possible small variable sets, or a model with variable sets of all different sizes, can give better prediction results on average as they cover all the useful idiosyncratic information. The most unexpected result is that SS-k gives the best density prediction in all countries. To explore the reason for this, cumulative plots are a useful tool that will be analyzed in the next section.

Figure 4.5 for MSPE does not show as obvious a pattern as Figure 4.4 for L PS, nevertheless it can be seen that BMA-5 and BMA-10 outperform most other estimation methods except for the UK. This also confirms that all countries prefer low-dimensional estimation, but the uncertainty of variable selection will cause many sparse models based on variable selection to fail, and the model averaging technique the consider uncertainty solves this problem very well. Similar to the results shown in LPS, the excellence of the performance of SS-k inspires us to use the cumulative graph to make a specific analysis.

Figure 4.6 presents a cumulative log predictive score for each economy. To clearly illustrate the change of each line and the contrast between different lines, we only selected BMA-all, SS-k and a representative sparse estimation method for each country. The polar scatter plot of LPS in Figure 4.4 indicates that the LPS of many estimation methods are on the same radius, suggesting these methods are highly correlated in the density prediction of target variables, which means it is unnecessary to include all the methods in the plot. To highlight the significant influence of the pandemic on the prediction ability of the estimation algorithm,

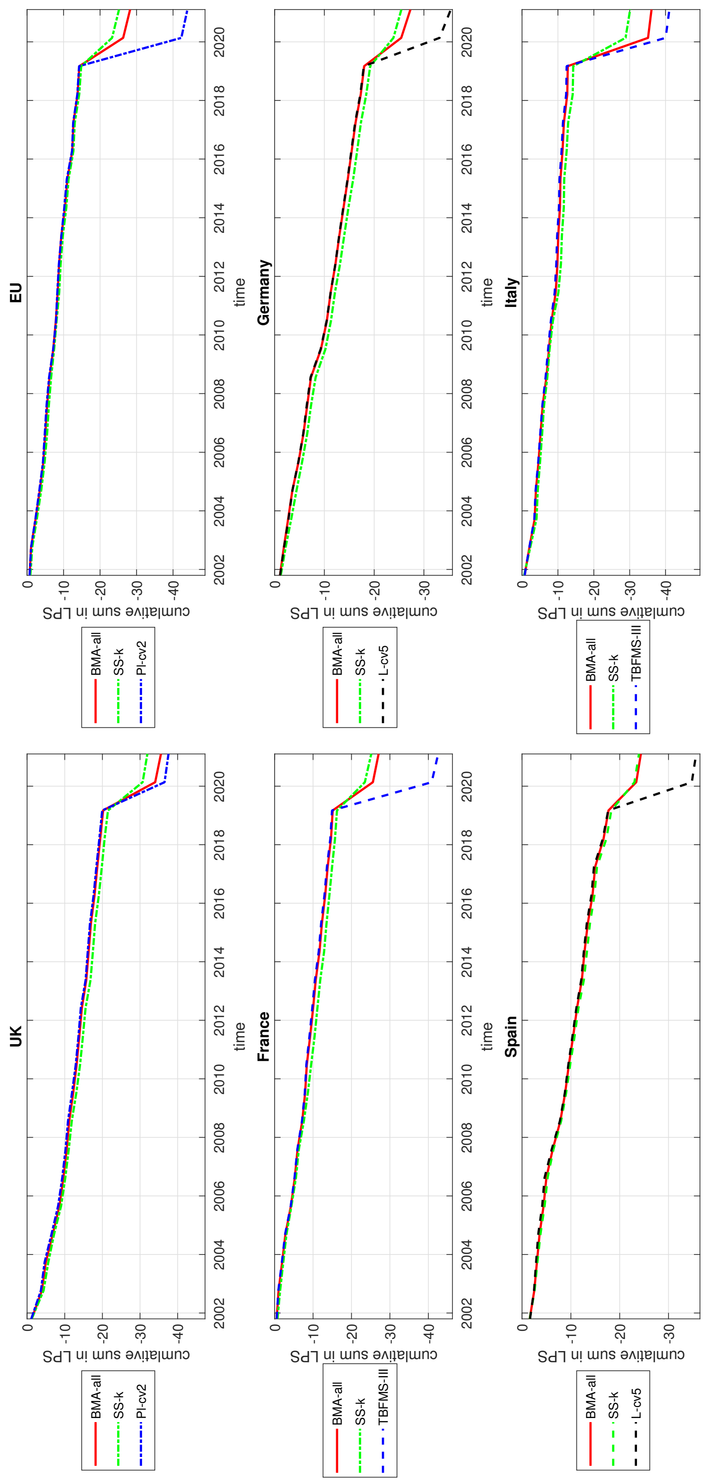

Figure 4.6: *Cumulative log predictive score of different models*

the methods to be displayed are those that are slightly better than BMA-all before the pandemic, but ultimately far worse than BMA-all due to their poor performance during the pandemic. As depicted in Figure 4.6, for all countries, SS-k is the worst before the pandemic (i.e. LPS is the smallest), which is lower than BMA-all obviously, but has a significant improvement after the pandemic, becoming even much higher than BMA-all. This is why SS-k eventually overtakes BMA-all on average. Other sparse model estimation methods, regardless of whether they exceeded BMA-all or were inferior to BMA-all before the pandemic, were almost wiped out after the pandemic. This demonstrates that, because of the occurrence of special events such as the pandemic, the important variable that determines target variable may vary greatly, while SS-k has enough information to predict the explained variable, even in the period of change, because of its wide coverage. The other sparse models, which select variables based on certain information criteria or penalties, cannot cope with large changes in the economy. The BMA-all performs relatively well during the pandemic because it considers a set of variables of all sizes.

Figure 4.7 depicts the cumulative square error for each economy. In addition to the BMA-all, which considers the uncertainty, and the SS-k, which exhibits exceptional performance, some other representative methods are also selected to demonstrate. For the UK, L-cv2 and L-asy are two good representatives. As indicates in Figure 4.5, L-cv2 and L-asy are the two methods with the best performance on average, and are both better than BMA-all in point forecasting of the growth rate of industrial production. However, according to the cumulative

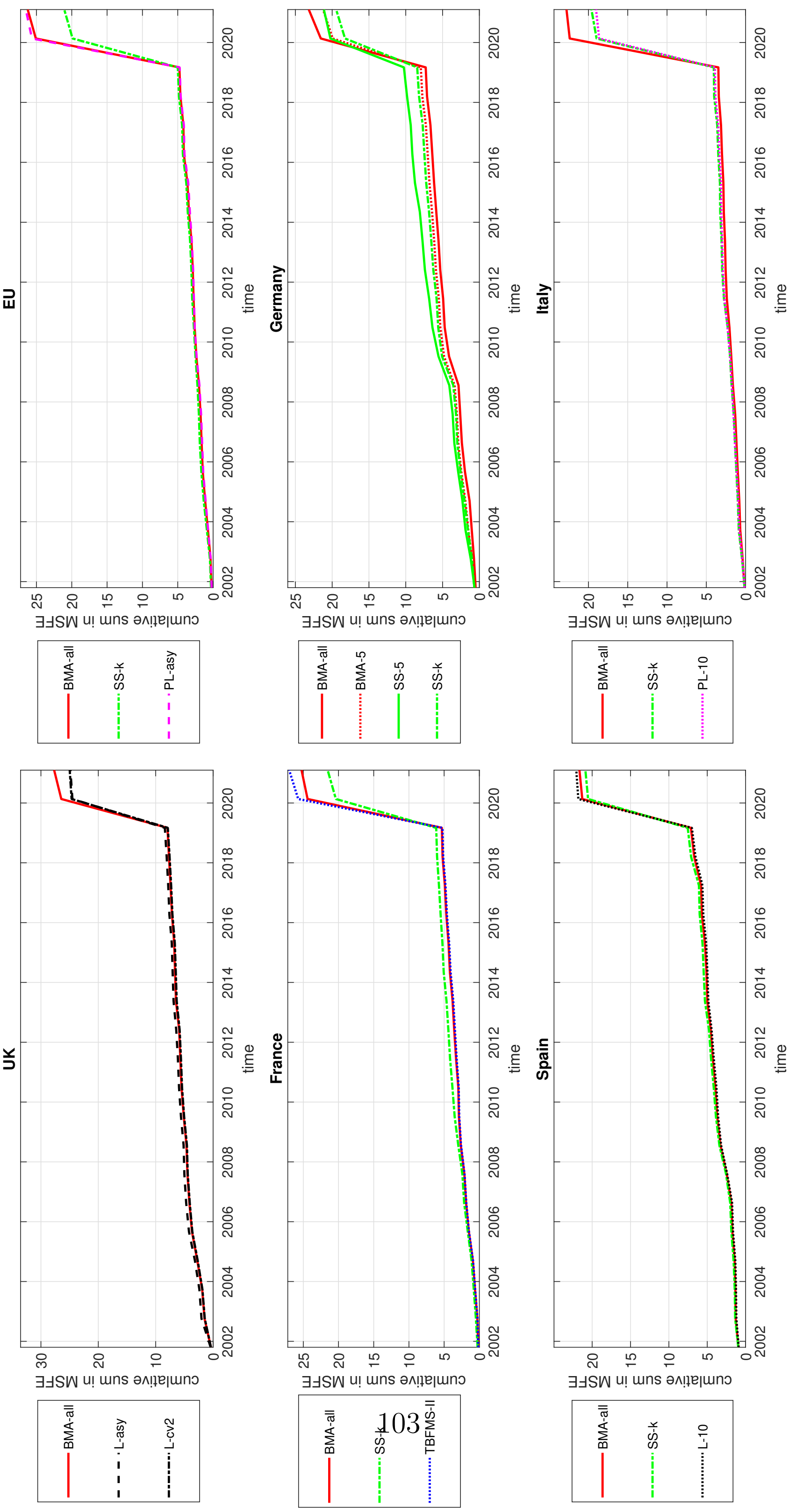

Figure 4.7: *Cumulative mean squared prediction error of different models*

graph, the performance of L-cv2 is relatively stable as it performs well before and after the pandemic. By contrast, the performance of L-asy is not that good before the pandemic, but provides the best prediction from when the pandemic began. For the EU, the performance of BMA-all was almost the best before the pandemic, but was surpassed by more than half of the methods after the pandemic, of which SS-k is an example. However, at the same time, a similar approach that performs well before the outbreak but deteriorated significantly after the outbreak is PL-asy. The results for France are similar to those of the EU. Germany displays several lines including SS-k because almost all the methods superior to BMA-all reversed the situation during the pandemic. The excellent performance of BMA-all in the pre-pandemic data of Germany, the country with the highest level of uncertainty, also illustrates the non-negligibility of uncertainty. Spain is the only country in which the forecast performance of BMA-all during the pandemic improved its ranking compared to the other models, but it is clear that SS-k, as a more informative model, performed too well during the pandemic to provide a smaller MSE on average. For Italy, a country with low uncertainty and obvious sparse features, the results given by most methods are, in general highly correlated. Therefore, the sparse model actually performed well before the outbreak. However, SS-k also dominated after the pandemic.

4.5 Conclusion

An increasing number of studies emphasize that it is not that more variables are more helpful in improving prediction accuracy, but that it is not rational to use the sparse model based on variable selection without statistical evidence or the support of an economic theory. In this paper, we explored the "dense or sparse" issue by examining macroeconomic data from different economies. Our main finding was that although all the economies show a preference for sparse models, the macroeconomic data itself does not provide enough information to clearly identify important variables from a large set of variables. Among all six economies, Germany presented the highest uncertainty in relevant variable selection, while the reverse was the case for Italy. The empirical results revealed that the higher the uncertainty of variable selection, the worse the low-dimensional model will perform most of the time, while the model averaging technique (BMA-all, BMA-5 and BMA-10), which takes the uncertainty into consideration, performed well. More precisely, in Germany, BMA was far superior to other sparse estimation methods, and in Italy, had certain advantages over other algorithms according to LPS. Another valuable finding was that the dense model based on ridge regression turned the tables perfectly after the outbreak and performed exceptionally well during a period of significant economic shocks such as the COVID-19 pandemic. Corresponding to the findings of Abadie and Kasy (2019), the circumstance of excellent ridge performance is that much of the probability mass of the distribution of the real effect of predictors is relatively tightly concentrated around zero. Hence the outstanding performance of ridge regression during the pandemic may be because the target economic variable was more sensitive to changes in other economic variables during this special period, but at the same time, the response to each economic variable was relatively average and small. Therefore, SS-k, as a method containing all the useful idiosyncratic information, became the best performing model. Further research direction is to consider the time-varying framework based on the variability of the economic structure and the short-lived nature of important variables.

Chapter 5

Conclusion

5.1 Summary & policy implications

In this era, central banks in all countries are confronted with a huge amount of data and hundreds of variables on a daily basis. Data dimensionality reduction has become an important tool for keeping abreast of economic conditions and predicting key economic variables more accurately in order to make appropriate economic decisions. Overall, this paper makes a contribution to both data dimensionality reduction and policy support in three distinct chapters. Firstly, in Chapter 2 we presented six indices of consumer attitudes that reflect different aspects of consumer attitudes and respond to different economic and financial shocks. These were based on a 61-year survey by the University of Michigan that adopted the theoretical perspective of behavioural or psychological economics. By observing these indices, policymakers and researchers can understand the nature of consumers' sentiments, whether they are pessimistic, how their views differ, whether the purchase pressure stems from interest rates or prices, and how uncertain consumers are about the economy in general. At the same time, using the time-varying parameter model with stochastic volatility, consumer indices can also be useful in predicting important real economic variables such as GDP, consumption, and investment.

Given the gradual change over time in real-world economic conditions, and hence the short-lived nature of the variables associated with the model estimation and forecasting targets, in Chapter 3 we proposed a common factor estimation method. The advantage of this approach is that it considers the predicted variables in the factor-extracting process and time-varying factor loadings that allow multiple change points, thereby delivering more accurate common factors used in forecasting macroeconomic variables. The empirical evidence demonstrated the relative accuracy of the TVP-3PRF algorithm in forecasting US macroeconomic variables compared to other algorithms. This method can therefore be utilized for forecasting target variables based on the presence of a large number of potential explanatory variables, not only in the macroeconomic field but also for forecasting exchange rate market and stock market data.

Last but not least, researchers and policymakers cannot concentrate solely on statistical predictive effects, i.e. taking into account which variables are picked to improve the predictive accuracy of the model, they must also focus on the structure in which variables enter the model in practice. The sparsity or density of the model should be examined so as to make forecasts more economically relevant and ensure information in the various variables can be utilised more precisely.

This was then discussed in Chapter 4. It was discovered that uncertainty in the choice of variables differed between countries after testing sparsity modeling using macroeconomic variables from several countries. For countries with low uncertainty of variable selection, it is safe to adopt a sparse modeling strategy, that is, model estimation based on variable selection. By contrast, averaging strategies based on models with different sets of regressors are the superior choice for countries with considerable uncertainty in the choice of variables. Simultaneously, evidence from the pandemic period suggests that a suitable approach might be to include all potential variables in the model when the economy suffers from external shocks and turbulence. To sum up, these findings are helpful in guiding macroeconomic policymakers in the right direction when choosing models.

5.2 Further research

While each of the self-contained chapters in this paper has produced reasonably robust results, there is still room for expansion. In the first chapter, we discussed how to use qualitative survey data to construct the index, which is a good attempt. Our findings suggest that the application of quantitative measures to qualitative consumer information can capture consumer attitudes from a more comprehensive perspective and further contribute to the prediction of macroeconomic variables. Therefore, we may consider further optimization of information acquisition in the measurement of consumer attitudes in the future, such as quantifying the textual data using text analysis. The reason for considering the use of textual data is that although the consumer survey index provides useful information on macroeconomic variables forecasting, it have several limitations. Time-consuming is one potential weakness as data collecting and aggregating process all takes time. Another obvious flaw is that the survey only comprises data obtained from a questionnaire, therefore it can be difficult to keep pace with the direct impact of newly occurring issues (Song and Shin, 2019). As a result, a strategy to supplement the survey-based method is to use text-based data. Numerous literature on sentiment analysis have been conducted on subjective texts, including blogs and product reviews, such as Daniel et al. (2017) and Kontopoulos et al. (2013). The most commonly employed techniques in this field include the lexicon-based approach, which is an unsupervised learning method, and the supervised learning method, which entails creating classifiers from labelled documents or sentences. More advanced textual data analysis techniques can be employed or even proposed to assess consumer attitudes.

In the second chapter, LASSO was used in the first-pass of the TVP-3PRF. One potential improvement this provided is to consider the univariate forgetting factor least-squares (FFLS) algorithm, which might have a more accurate filtering effect because it also takes the changing volatility into account. Specifically, this involves running the univariate regression with the forgetting factor in the first-pass, then checking the predictive likelihood of whether this coefficient should be zero. The second potential improvement direction is to consider adding another pass in the whole estimation process. The reason for doing this is that the omitted-variable bias can cause the inaccurate estimation of $\phi_{f,t}$. The first-pass omit the proxy for irrelevant factors $\mathbf{g}_t$. The way to solve this problem is by using the principal component of $\mathbf{x}_t$ as the dependent variable, regressing it with the proxy of $\mathbf{f}_t$, which is the targeted variable y_t, and then obtain the residual. This residual can be used as the proxy of irrelevant factors $\mathbf{g}_t$ because the information of f_t is removed from the $\mathbf{x}_t$, and the remainder should be the information in $\mathbf{g}_t$. This approach has been proven effective at simulation and empirical levels, therefore it would be worthwhile to proceed and prove the statistical nature of this algorithm in future studies.

Based on the research objective of the third chapter, proposing a sparse and dense classification prior with time-varying parameters can be considered given that economic structure changes smoothly and the essential variables are short-lived. In the first step, as with the time-varying coefficient estimation method employed in chapters 1 and 2, we can estimate the coefficient based on Chan et al. (2014) and Korobilis (2021) and write the TVP regression as a linear regression in stacked form. Accoding to Equation 2.16 and 3.4, β^{Δ} is the result of a trivial rotation of β by the matrix H, that is $\beta^{\Delta} = H\beta$. By doing so we have the benefit that $\beta^{\Delta} \sim N(0, S)$ where S can be typically diagonal or block-diagonal. This can result in efficient sampling relative to the non-diagonal covariance matrix $H^{-1}SH^{-1}$ of equation (5) proposed by Chan and Jeliazkov (2009). Therefore, we can specify Giannone et al. (2021)'s spike-and-slab prior for the coefficients. In this way, the crucial parameter that affects how much β_t varies from period $t = s$ to period $t = s + 1$ is the prior covariance matrix S. S can be treated as a diagonal prior covariance matrix and allows each of its Tp elements to be shrunk independently (adaptive shrinkage). This means that any of the p predictors in x_t is shrunk in any of the time periods t, independently. Moreover, it will be beneficial to take some other prior with imposing structure into account since group structures of variables may be introduced into a model to make use of some domain specific prior knowledge. As in our cases, all Tp variables form a natural group, which is p groups with T elements in each group. Therefore, I consider a hierarchical structured variable selection (HSVS) prior proposed by Zhang et al. (2014), that can simultaneously select both group and within-group variables, to be of interest. The HSVS prior utilizes a discrete mixture prior distribution for group selection and group specific Bayesian lasso hierarchies for variable selection within groups. The methods can also take serial correlation within groups into account. Therefore, a combination of TVP modeling and HSVS prior introduced above can be used to test the sparsity of the time-varying coefficient model in future studies.

Bibliography

K. A. Aastveit, G. J. Natvik, and S. Sola. Economic uncertainty and the influence of monetary policy. *Journal of International Money and Finance*, 76:50–67, 2017.

A. Abadie and M. Kasy. Choosing among regularized estimators in empirical economics: The risk of machine learning. *Review of Economics and Statistics*, 101(5):743–762, 2019.

F. Alvarez-Parra, L. Brandao-Marques, and M. Toledo. Durable goods, financial frictions, and business cycles in emerging economies. *Journal of Monetary Economics*, 60(6):720–736, 2013.

O. Armantier, G. Topa, W. Van der Klaauw, and B. Zafar. An overview of the survey of consumer expectations. *Economic Policy Review*, (23-2):51–72, 2017.

J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

J. Bai and S. Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150, 2006.

J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

E. Bair, T. Hastie, D. Paul, and R. Tibshirani. Prediction by supervised principal components. *Journal of the American Statistical Association*, 101(473):119–137, 2006.

B. J. Bates, M. Plagborg-Møller, J. H. Stock, and M. W. Watson. Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, 177(2):289–304, 2013.

E. Beale, M. Kendall, and D. Mann. The discarding of variables in multivariate analysis. *Biometrika*, 54(3-4):357–366, 1967.

A. Belloni and V. Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547, 2013.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*, 2011.

A. Bhattacharya, A. Chakraborty, and B. K. Mallick. Fast sampling with gaussian scale mixture priors in high-dimensional regression. *Biometrika*, page asw042, 2016.

P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, 37(4):1705–1732, 2009.

N. Bloom. The impact of uncertainty shocks. *econometrica*, 77(3):623–685, 2009.

J. Boivin and S. Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194, 2006.

W. H. Boshoff and L. H. Binge. Alternative cycle indicators for the south african business cycle. *Business cycles in BRICS*, pages 447–464, 2019.

Z. I. Botev. The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):125–148, 2017.

P. Bühlmann and S. Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

G. Bulligan, M. Marcellino, and F. Venditti. Forecasting economic activity with targeted predictors. *International Journal of Forecasting*, 31(1):188–206, 2015.

G. Caggiano, G. Kapetanios, and V. Labhard. Are more data always better for factor analysis? results for the euro area, the six largest euro area countries and the uk. *Journal of Forecasting*, 30(8):736–752, 2011.

C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets, 1982.

J. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.

J. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120, 2009.

N. H. Chan, C. Y. Yau, and R.-M. Zhang. Group lasso for structural break time series. *Journal of the American Statistical Association*, 109(506):590–599, 2014. ISSN 01621459. URL `http://www.jstor.org/stable/24247188`.

X. Cheng, Z. Liao, and F. Schorfheide. Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4):1511–1543, 2016.

V. Chernozhukov, C. Hansen, and M. Spindler. Valid post-selection and post-regularization inference: An elementary, general approach. *Annu. Rev. Econ.*, 7(1):649–688, 2015.

M. Daniel, R. F. Neves, and N. Horta. Company event popularity for financial markets using twitter and sentiment analysis. *Expert Systems with Applications*, 71:111–124, 2017.

C. De Mol, D. Giannone, and L. Reichlin. Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328, 2008.

J. Dovern, U. Fritsche, and J. Slacalek. Disagreement among forecasters in g7 countries. *Review of Economics and Statistics*, 94(4):1081–1096, 2012.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82 (4):540–554, 2000.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840, 2005.

J. Fuentes, P. Poncela, and J. Rodríguez. Sparse partial least squares in time series for macroeconomic forecasting. *Journal of Applied Econometrics*, 30(4): 576–595, 2015.

A. Gehringer and T. Mayer. Measuring the business cycle chronology with a novel business cycle indicator for germany. *Journal of Business Cycle Research*, 17 (1):71–89, 2021.

J. F. Geweke. Bayesian inference for linear models subject to linear inequality constraints. In *Modelling and Prediction Honoring Seymour Geisser*, pages 248–263. Springer, 1996.

D. Giannone, M. Lenza, and G. E. Primiceri. Economic predictions with big data: The illusion of sparsity. 2018.

D. Giannone, M. Lenza, and G. E. Primiceri. Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437, 2021.

C. Gillitzer and N. Prasad. The effect of consumer sentiment on consumption: cross-sectional evidence from elections. *American Economic Journal: Macroeconomics*, 10(4):234–69, 2018.

T. Hastie, R. Tibshirani, and M. Wainwright. Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143:143, 2015.

R. R. Hocking and R. Leslie. Selection of the best subset in regression analysis. *Technometrics*, 9(4):531–540, 1967.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968.

Y. Jiang, Y. He, and H. Zhang. Variable selection with prior information for generalized linear models via the prior lasso method. *Journal of the American Statistical Association*, 111(513):355–376, 2016.

K. Jurado, S. C. Ludvigson, and S. Ng. Measuring uncertainty. *American Economic Review*, 105(3):1177–1216, 2015.

B. Kelly and S. Pruitt. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316, 2015.

L. Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69, 2009.

H. H. Kim and N. R. Swanson. Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178:352–367, 2014.

E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades. Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*, 40(10): 4065–4074, 2013.

G. Koop and D. Korobilis. A new index of financial conditions. *European Economic Review*, 71:101–116, 2014.

G. Koop, D. Korobilis, and D. Pettenuzzo. Bayesian compressed vector autoregressions. *Journal of Econometrics*, 210(1):135–154, 2019. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2018.11.009. URL https://www.sciencedirect.com/science/article/pii/S0304407618302100. Annals Issue in Honor of John Geweke "Complexity and Big Data in Economics and Finance: Recent Developments from a Bayesian Perspective".

D. Korobilis. A new algorithm for sign restrictions in vector autoregressions. *Available at SSRN 3557911*, 2020.

D. Korobilis. High-dimensional macroeconomic forecasting using message passing algorithms. *Journal of Business & Economic Statistics*, 39(2):493–504, 2021. doi: 10.1080/07350015.2019.1677472.

D. Korobilis, B. Landau, A. Musso, and A. Phella. The time-varying evolution of inflation risks. 2021.

D. Kozbur. Analysis of testing-based forward model selection. *Econometrica*, 88 (5):2147–2173, 2020.

F. E. Kydland, P. Rupert, and R. Šustek. Housing dynamics over the business cycle. *International Economic Review*, 57(4):1149–1177, 2016.

A. Lanteri. The market for used capital: Endogenous irreversibility and reallocation over the business cycle. *American Economic Review*, 108(9):2383–2419, 2018.

D. N. Lawley and A. E. Maxwell. Factor analysis as a statistical method. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 12(3):209–229, 1962.

E. E. Leamer. Multicollinearity: a bayesian interpretation. *The review of economics and statistics*, pages 371–380, 1973.

H. Leeb and B. M. Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59, 2005.

H. Leeb and B. M. Pötscher. Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, 24(2):338–376, 2008a.

H. Leeb and B. M. Pötscher. Sparse estimators and the oracle property, or the return of hodges' estimator. *Journal of Econometrics*, 142(1):201–211, 2008b.

S. C. Ludvigson, S. Ma, and S. Ng. Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, 13(4):369–410, 2021.

R. Moessner, F. Zhu, and C. Ellis. Measuring disagreement in uk consumer and central bank inflation forecasts. 2011.

R. M. Neal. Slice sampling. *The annals of statistics*, 31(3):705–767, 2003.

K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

N. G. Polson and L. Sun. Bayesian l0-regularized least squares. *Applied Stochastic Models in Business and Industry*, 35(3):717–731, 2019a. doi: https://doi.org/10.1002/asmb.2381. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2381`.

N. G. Polson and L. Sun. Bayesian l 0-regularized least squares. *Applied Stochastic Models in Business and Industry*, 35(3):717–731, 2019b.

J. Qian and L. Su. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191(1):86–109, 2016.

D. E. Rapach, J. K. Strauss, and G. Zhou. International stock return predictability: What is the role of the united states? *The Journal of Finance*, 68(4): 1633–1662, 2013.

C. Sammut and G. I. Webb. *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.

A. H. Shapiro, M. Sudhof, and D. J. Wilson. Measuring news sentiment. *Journal of Econometrics*, 2020.

M. Song and K.-s. Shin. Forecasting economic indicators using a consumer sentiment index: Survey-based versus text-based data. *Journal of forecasting*, 38 (6):504–518, 2019.

C. Soussen, J. Idier, D. Brie, and J. Duan. From bernoulli–gaussian deconvolution to sparse signal restoration. *IEEE Transactions on Signal Processing*, 59(10): 4572–4584, 2011.

C. Spearman. " general intelligence" objectively determined and measured. 1961.

Steve, Lohr. The origins of 'big data': An etymological detective story. February 2013.

J. H. Stock and M. Watson. Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205, 2009.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97 (460):1167–1179, 2002a.

J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002b.

J. H. Stock and M. W. Watson. Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493, 2012.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.

A. N. Tihonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math.*, 4:1035–1038, 1963.

R. P. Velu, G. C. Reinsel, and D. W. Wichern. Reduced rank models for multiple time series. *Biometrika*, 73(1):105–118, 1986. ISSN 00063444. URL `http://www.jstor.org/stable/2336276`.

H. Wold. Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*, pages 391–420, 1966.

S. S.-L. Wong, P. Chin-Hong, S. A. Mansor, and V. K.-S. Liew. Measuring business cycle fluctuations: An alternative precursor to economic crises. In *ACRN Proceedings in Finance and Risk Series '13: Proceedings of the 13th FRAP Conference in Cambridge*, volume 2, page 33. ACRN Publishing House, 2014.

H. Ye and G. Sugihara. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. *Science*, 353(6302):922–925, 2016.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

L. Zhang, V. Baladandayuthapani, B. K. Mallick, G. C. Manyam, P. A. Thompson, M. L. Bondy, and K.-A. Do. Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(4):595–620, 2014.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67 (2):301–320, 2005.

Appendix A

Chapter 2 Appendix

A.1 Data Appendix

In the study of this paper, there are four data sources.

The data used in the forecasting part is consumer survey data from the University of Michigan surveys of consumer (https://data.sca.isr.umich.edu/data-archive/mine.php.) and FRED quarterly large macroeconomic databases, which is database is provided by Michael W. McCracken who update it though the St Louis Federal Reserve Economic Database (FRED) in real-time (https://research.stlouisfed.org/econ/mccracken/fred-databases/). Our predicted variables, GDP, consumption, and investment and also their transform codes are all from FRED quarterly large macroeconomic databases. The data used to extract the macroeconomic common factors is the variables in FRED quarterly large macroeconomic databases with the factor code equals 1.

There are two additional data used in the VAR part, and they are JLNF12 and SVAR. JLNF12 is the uncertainty index proposed by Jurado et al. (2015), SVAR is the stock market volatility index proposed by Bloom (2009). Other variables and their corresponding transform codes can still be found in FRED quarterly large macroeconomic databases.

A.2 Full Gibbs sampler for VARs with sign restrictions

The Gibbs sampler presented in the main paper is a quite accurate description of the core algorithmic steps required in order to estimate the VAR model with factor structure in the residuals. Nevertheless, in order to produce empirical and other results, we have relied on the hierarchical horseshoe prior of Carvalho et al. (2010), two fast algorithms from drawing from the Normal (Bhattacharya et al., 2016) and truncated Normal (Botev, 2017) distributions, respectively, and the slice sampler of Neal (2003) in order to update the horseshoe prior parameters. Therefore, it is important to rewrite the Gibbs sampling algorithm in full, and give further explanations about the three enhancements that guarantee a fast and reliable algorithm in high dimensions.

I repeat the full prior specification, which now includes the hierarchical horseshoe prior on ϕ_i. The priors for the i^{th} VAR equation, $i = 1, ..., n$ is:

$$\phi_i|\sigma_i^2, \tau_i^2, \mathbf{\Psi}_i^2 \sim N_k\left(\mathbf{0}, \sigma_i^2\tau_i^2\mathbf{\Psi}_i^2\right), \quad \mathbf{\Psi}_i^2 = diag\left(\psi_{i,1}^2, ..., \psi_{i,k}^2\right), \tag{A.1}$$

$$\psi_{i,j} \sim Cauchy^+(0,1), \quad j = 1, ..., k, \tag{A.2}$$

$$\tau_i \sim Cauchy^+(0,1), \tag{A.3}$$

$$\mathbf{\Lambda}_{ij} \sim \begin{cases} N\left(0, \underline{h}_{ij}\right) I(\Lambda_{ij} > 0), & \text{if } S_{ij} = 1, \\ N\left(0, \underline{h}_{ij}\right) I(\Lambda_{ij} < 0), & \text{if } S_{ij} = -1, \\ \delta_0\left(\mathbf{\Lambda}_{ij}\right), & \text{if } S_{ij} = 0, \\ N\left(0, \underline{h}_{ij}\right), & \text{otherwise,} \end{cases} \quad j = 1, ..., r, \tag{A.4}$$

$$\mathbf{f}_t \sim N_r(\mathbf{0}, \mathbf{I}), \tag{A.5}$$

$$\sigma_i^2 \sim inv - Gamma\left(\underline{\rho}_i, \underline{\kappa}_i\right), \tag{A.6}$$

where we set $\underline{h}_{ij} = 4$, $\underline{\rho}_i = 1$ and $\underline{\kappa}_i = 0.01$.

Under these priors, the full factor sign restrictions algorithm takes the following form **Factor sign restrictions (FSR) algorithm**

1. Sample ϕ_i for $i = 1, ..., n$ from

$$\phi_i|\mathbf{\Sigma}, \mathbf{\Lambda}, \mathbf{f}, \mathbf{y} \sim N_k\left(\overline{\mathbf{V}}_i\left(\sum_{t=1}^{T}\sigma_i^{-2}\mathbf{x}_t'\widetilde{\mathbf{y}}_{it}\right), \overline{\mathbf{V}}_i\right), \tag{A.7}$$

   where $\widetilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \mathbf{\Lambda}_i\mathbf{f}_t$ and $\overline{\mathbf{V}}_i^{-1} = \left(\underline{\mathbf{V}}_i^{-1} + \sum_{t=1}^{T}\sigma_i^{-2}\mathbf{x}_t'\mathbf{x}_t\right)$. We use the efficient sampler of Bhattacharya et al. (2016) in order to sample these elements.

2. Sample ψ_{ij} using slice sampling (Neal, 2003)

a. Set $\eta_{ij} = 1/\psi_{ij}^2$ using the last available sample of ψ_{ij}^2.

b. Sample a random variable u from

$$u|\eta_{ij} \sim Uniform\left(0, \frac{1}{1+\eta_{ij}}\right). \tag{A.8}$$

c. Sample η_{ij} from

$$\eta_{ij} \sim e^{\frac{\phi_{ij}^2}{2\sigma_i^2}\eta_{ij}} I\left(\frac{u}{1-u} > \eta_{ij}\right) \tag{A.9}$$

and set $\psi_{ij} = 1/\sqrt{\eta_{ij}}$.

3. Sample τ_i using slice sampling (Neal, 2003)

a. Set $\xi_i = 1/\tau_i^2$ using the last available sample of τ_i^2.

b. Sample a random variable u from

$$v|\xi_{ij} \sim Uniform\left(0, \frac{1}{1+\xi_{ij}}\right). \tag{A.10}$$

c. Sample ξ_i from

$$\xi_i \sim \gamma\left((k+1)/2, v\frac{2\sigma^2}{\sum\left(\frac{\phi_{ij}}{\psi_{ij}}\right)^2}\right), \tag{A.11}$$

where $\gamma(\bullet)$ is the lower incomplete gamma function, and set $\tau_i = 1/\sqrt{\xi_i}$

4. Sample $\mathbf{\Lambda}_{ij}$ from univariate conditional posteriors (Geweke, 1996) of the

form

$$\mathbf{\Lambda}_{ij}|\mathbf{\Lambda}_{-ij}\mathbf{\Phi},\mathbf{\Sigma},\mathbf{f},\mathbf{y}\sim TN_{(\mathbf{a}_{ij},\mathbf{b}_{ij})}\left(\overline{\lambda}_{ij}-\overline{h}_{ij}\sum_{l\neq j}\overline{h}_{il}^{-1}\left(\mathbf{\Lambda}_{il}-\overline{\lambda}_{il}\right),\overline{h}_{ij}\right), \tag{A.12}$$

where $\overline{\lambda}_{ij}$ and $\overline{h}_{ij}$ denote the ij^{th} elements of the joint posterior mean and variance, respectively, of $\mathbf{\Lambda}_i$. The joint posterior variance is $\overline{\mathbf{H}}^{-1} = \left(\underline{\mathbf{H}}^{-1}+\sum_{t=1}^{T}\sigma_i^{-2}\mathbf{f}_t'\mathbf{f}_t\right)$ and the joint posterior mean is $\overline{\mathbf{H}}\left(\sum_{t=1}^{T}\sigma_i^{-2}\mathbf{f}_t'\widehat{\mathbf{y}}_{it}\right)$ with $\widehat{\mathbf{y}}_{it}\equiv\varepsilon_{it}=\mathbf{y}_{it}-\phi_i\mathbf{x}_t$. Here $TN_{(\mathbf{a}_{ij},\mathbf{b}_{ij})}(\bullet)$ denotes the **univariate** truncated Normal distribution with bounds:

$$(\mathbf{a}_{ij},\mathbf{b}_{ij})=\begin{cases}(-\infty,0) & \text{if } S_{ij}=-1,\\ (0,\infty) & \text{if } S_{ij}=1,\\ (0,0) & \text{if } S_{ij}=0,\\ (-\infty,\infty) & \text{otherwise,}\end{cases} \tag{A.13}$$

We use the efficient univariate truncated Normal generator provided by Botev (2017) in order to sample these elements.

5. Sample $\mathbf{f}_t$ for $t=1,...,T$ from

$$\mathbf{f}_t|\mathbf{\Lambda},\mathbf{\Sigma},\mathbf{\Phi},\mathbf{y}\sim N\left(\overline{\mathbf{G}}\left(\mathbf{\Lambda}\mathbf{\Sigma}^{-1}\widehat{\mathbf{y}}_t\right),\overline{\mathbf{G}}\right), \tag{A.14}$$

where $\overline{\mathbf{G}}^{-1}=(\mathbf{I}_r+\mathbf{\Lambda}'\mathbf{\Sigma}\mathbf{\Lambda})$. Post-process the draws of the $T\times r$ matrix $\mathbf{f}=(\mathbf{f}_1,...,\mathbf{f}_T)'$ such that its r columns (corresponding to structural shocks) are uncorrelated and standardized to unit variance. This is done by applying first the Gram-Schmidt procedure and subsequently dividing each column of $\mathbf{f}$ with its standard deviation.

6. Sample σ_i^2 for $i = 1, ..., n$ from

$$\sigma_i^2|\mathbf{\Lambda}, \mathbf{f}, \mathbf{\Phi}, \mathbf{y} \sim inv{-}Gamma\left(\frac{T}{2} + \underline{\rho}_i, \left[\underline{\kappa}_i^{-1} + \sum_{t=1}^{T}\left(\mathbf{y}_{it} - \phi_i \mathbf{x}_t - \mathbf{\Lambda}_i \mathbf{f}_t\right)' \left(\mathbf{y}_{it} - \phi_i \mathbf{x}_t - \mathbf{\Lambda}_i \mathbf{f}_t\right)\right]^{-1}\right) \tag{A.15}$$

Appendix B

Chapter 3 Appendix

B.1 Additional Monte-Carlo Result

B.1.1 Out-of-Sample Forecasting

Table B.1: RMSE for DGP1

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	Hard1	Hard2	Soft1	Soft2
					Normal Factors					
0	0	0	0	1.0116	1.0175	1.0177	1.0201	1.0189	1.0023	1.0030
0.3	0.9	0.3	0	**0.9987**	1.0244	1.0252	1.0226	1.0232	1.0271	1.0174
0.3	0.9	0.3	1	1.0075	**0.9702**	**0.9721**	**0.9883**	**0.9952**	1.0025	**0.9977**
0.3	0.9	0.9	0	**0.9491**	1.0045	1.0040	1.0054	1.0123	1.0164	**0.9954**
0.3	0.9	0.9	1	**0.9791**	1.0052	1.0035	**0.9970**	1.0148	1.0019	1.0059
0.9	0.3	0.3	0	**0.8053**	**0.8222**	**0.8241**	**0.9752**	**0.8962**	**0.9453**	**0.9515**
0.9	0.3	0.3	1	**0.7370**	**0.7022**	**0.7038**	**0.8390**	**0.7963**	**0.9389**	**0.9874**
0.9	0.3	0.9	0	**0.6929**	**0.7559**	**0.7540**	**0.8193**	**0.7967**	**0.8779**	**0.9282**
0.9	0.3	0.9	1	**0.7639**	**0.8278**	**0.8139**	**0.8627**	**0.8485**	**0.8992**	1.0250
					Moderately Weak Factors					
0	0	0	0	**0.9924**	**0.9900**	**0.9907**	0.9899	**0.9901**	1.0186	1.0130
0.3	0.9	0.3	0	**0.9753**	**0.9661**	**0.9676**	**0.9740**	0.9489	1.0345	1.0322

continued

Table B.1: RMSE for DGP1

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	Hard1	Hard2	Soft1	Soft2
0.3	0.9	0.3	1	1.0095	1.0172	1.0249	1.0381	1.0175	1.0400	1.0428
0.3	0.9	0.9	0	1.0245	1.0663	1.0704	1.0323	1.0191	1.0072	1.0014
0.3	0.9	0.9	1	1.0226	1.0556	1.0563	1.0205	1.0297	**0.9914**	1.0066
0.9	0.3	0.3	0	**0.7037**	**0.7230**	**0.7253**	**0.7961**	**0.7534**	**0.9351**	**0.9611**
0.9	0.3	0.3	1	0.7440	**0.7438**	**0.7463**	**0.8150**	**0.7821**	**0.9575**	**0.9983**
0.9	0.3	0.9	0	0.7723	**0.7780**	**0.7778**	**0.7846**	**0.7692**	**0.9708**	**0.9660**
0.9	0.3	0.9	1	0.8173	**0.8179**	**0.8058**	**0.8944**	**0.8752**	**0.9413**	1.0068
						Weak Factors				
0	0	0	0	**0.9969**	**0.9962**	**0.9955**	**0.9756**	**0.9791**	**0.9919**	**0.9875**
0.3	0.9	0.3	0	1.0030	**0.9712**	**0.9718**	1.0022	1.0168	1.0057	1.0042
0.3	0.9	0.3	1	**0.9677**	**0.9857**	**0.9848**	**0.9862**	**0.9821**	1.0139	1.0035
0.3	0.9	0.9	0	**0.9726**	1.0021	1.0030	1.0015	1.0100	**0.9949**	**0.9889**
0.3	0.9	0.9	1	**0.9949**	1.1082	1.1001	1.0353	1.0265	**0.9966**	**0.9939**
0.9	0.3	0.3	0	**0.7564**	**0.8310**	**0.8270**	**0.9055**	**0.8992**	**0.9300**	**0.9505**
0.9	0.3	0.3	1	**0.8092**	**0.8292**	**0.8294**	**0.8974**	**0.8635**	**0.9699**	1.0266
0.9	0.3	0.9	0	**0.7406**	**0.7650**	**0.7678**	**0.8598**	**0.8267**	**0.9373**	**0.9985**
0.9	0.3	0.9	1	**0.7660**	**0.7960**	**0.8033**	**0.8129**	**0.7604**	**0.9147**	**0.9934**
						Moderately Weak and Non-Pervasive Factors				
0	0	0	0	1.0112	1.0133	1.0154	1.0181	1.0198	1.0003	**0.9993**
0.3	0.9	0.3	0	**0.9593**	**0.9989**	**0.9985**	**0.9968**	**0.9984**	**0.9938**	**0.9939**
0.3	0.9	0.3	1	1.0034	1.0257	1.0282	1.0215	1.0107	**0.9890**	1.0041
0.3	0.9	0.9	0	1.0388	1.0803	1.0776	1.0339	1.0680	1.0139	**0.9831**
0.3	0.9	0.9	1	1.0161	1.0580	1.0585	1.0519	1.0540	1.0380	1.0170
0.9	0.3	0.3	0	**0.7261**	**0.7275**	**0.7338**	**0.8374**	**0.7785**	**0.9053**	**0.9639**
0.9	0.3	0.3	1	**0.8294**	**0.8128**	**0.8128**	**0.8561**	**0.8344**	**0.8792**	**0.8993**
0.9	0.3	0.9	0	**0.8045**	**0.8425**	**0.8442**	**0.9084**	**0.9185**	**0.9034**	**0.9413**
0.9	0.3	0.9	1	**0.7921**	**0.8359**	**0.8372**	**0.8644**	**0.8311**	**0.9097**	1.0071

Note: The table reports the out-of-sample forecasting median RMSE based on 1000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation in the predictors' residuals, respectively. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

Table B.2: RMSE for DGP2

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	Hard1	Hard2	Soft1	Soft2
						Normal Factors				
0	0	0	0	1.0024	1.0096	**0.9991**	**0.9964**	**0.9978**	**0.9956**	**<u>0.9951</u>**
0.3	0.9	0.3	0	1.0043	1.0610	1.0422	1.0331	1.0436	1.0136	1.0137
0.3	0.9	0.3	1	**<u>0.9800</u>**	**0.9837**	1.0161	1.0112	1.0266	**0.9987**	**0.9957**
0.3	0.9	0.9	0	**0.9931**	1.0552	1.0444	1.0341	1.0207	1.0012	**<u>0.9904</u>**
0.3	0.9	0.9	1	**<u>0.9620</u>**	1.0425	1.0595	1.0107	1.0140	1.0035	**0.9979**
0.9	0.3	0.3	0	**0.7660**	**<u>0.7578</u>**	**0.9426**	**0.9906**	**0.9732**	**0.9989**	**0.9983**
0.9	0.3	0.3	1	**<u>0.7719</u>**	**0.8786**	**0.9627**	**0.9716**	**0.9698**	**0.9941**	**0.9952**
0.9	0.3	0.9	0	**<u>0.6840</u>**	**0.8008**	**0.8707**	**0.9807**	**0.9716**	**0.9915**	**0.9853**
0.9	0.3	0.9	1	**<u>0.7430</u>**	**0.7436**	**0.7841**	1.0141	**0.9324**	1.0002	1.0040
						Moderately Weak Factors				
0	0	0	0	**<u>0.9804</u>**	**0.9888**	**0.9916**	**0.9999**	**0.9988**	1.0001	**0.9978**
0.3	0.9	0.3	0	**<u>0.9593</u>**	**0.9939**	**0.9893**	1.0022	**0.9937**	**0.9923**	**0.9895**
0.3	0.9	0.3	1	1.0263	1.0614	1.0323	1.0121	**<u>0.9983</u>**	1.0048	1.0050
0.3	0.9	0.9	0	1.0211	1.0664	1.0829	1.0416	1.0568	1.0224	1.0086
0.3	0.9	0.9	1	1.0014	1.0594	1.0481	1.0134	1.0097	1.0076	1.0042
0.9	0.3	0.3	0	**<u>0.7576</u>**	**0.7977**	**0.9002**	**0.9855**	**0.9834**	1.0020	**0.9985**
0.9	0.3	0.3	1	**<u>0.8091</u>**	**0.8552**	**0.8861**	**0.9713**	**0.9793**	**0.9869**	**0.9971**
0.9	0.3	0.9	0	**0.8289**	**<u>0.8135</u>**	**0.8315**	**0.9460**	**0.9084**	**0.9534**	1.0112
0.9	0.3	0.9	1	**<u>0.8119</u>**	**0.8993**	**0.9102**	**0.9516**	**0.9306**	1.0219	1.0074
						Weak Factors				
0	0	0	0	**<u>0.9861</u>**	1.0005	**0.9976**	**0.9927**	1.0046	**0.9974**	1.0072
0.3	0.9	0.3	0	1.0199	1.0525	1.0357	1.0439	1.0285	1.0021	**<u>0.9972</u>**
0.3	0.9	0.3	1	**<u>0.9847</u>**	1.0180	**0.9943**	1.0213	1.0381	1.0063	1.0023
0.3	0.9	0.9	0	**<u>0.9765</u>**	1.1012	1.1146	1.0828	1.0998	1.0228	**0.9963**
0.3	0.9	0.9	1	1.0051	1.1025	1.0817	1.0199	1.0428	1.0252	**<u>0.9988</u>**
0.9	0.3	0.3	0	**<u>0.7831</u>**	**0.8121**	**0.8428**	**0.9530**	**0.9688**	1.0094	1.0007
0.9	0.3	0.3	1	**<u>0.7101</u>**	**0.8712**	**0.8739**	**0.9598**	**0.9555**	1.0083	**0.9976**
0.9	0.3	0.9	0	**<u>0.8328</u>**	**0.8597**	**0.8470**	**0.8681**	**0.8563**	**0.9341**	**0.9943**
0.9	0.3	0.9	1	**<u>0.7544</u>**	**0.7554**	**0.7725**	**0.8385**	**0.8333**	**0.9765**	1.0332
						Moderately Weak and Non-Pervasive Factors				
0	0	0	0	**0.9954**	**0.9949**	1.0023	**0.9923**	**<u>0.9919</u>**	1.0014	**0.9967**
0.3	0.9	0.3	0	1.0130	1.0685	1.0374	1.0013	**<u>0.9978</u>**	1.0063	1.0039
0.3	0.9	0.3	1	1.0155	1.0488	1.0111	1.0037	**<u>0.9943</u>**	1.0058	**0.9965**
0.3	0.9	0.9	0	**<u>0.9582</u>**	1.0224	1.0562	1.0127	1.0252	1.0083	**0.9835**
0.3	0.9	0.9	1	1.0521	1.1554	1.1287	1.0691	1.1379	1.0027	**<u>0.9962</u>**

continued

Table B.2: RMSE for DGP2

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	Hard1	Hard2	Soft1	Soft2
0.9	0.3	0.3	0	**0.7006**	**0.7648**	**0.8488**	**0.9848**	**0.9705**	1.0006	1.0038
0.9	0.3	0.3	1	**0.7507**	**0.8705**	**0.9192**	**0.9717**	**0.9618**	1.0014	**0.9966**
0.9	0.3	0.9	0	**0.7834**	**0.8232**	**0.8156**	**0.9107**	**0.8808**	**0.9536**	1.0081
0.9	0.3	0.9	1	**0.7064**	**0.6654**	**0.7501**	**0.8688**	**0.8109**	1.0100	**0.9864**

Note: The table reports the out-of-sample forecasting median RMSE based on 1000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation in the predictors' residuals, respectively. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

Table B.3: RMSE for DGP3

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	Hard1	Hard2	Soft1	Soft2
					Normal Factors					
0	0	0	0	1.0068	1.0111	**0.9976**	**0.9948**	**0.9932**	**0.9969**	**0.9965**
0.3	0.9	0.3	0	**0.9790**	1.0677	1.0421	1.0024	1.0175	1.0059	1.0060
0.3	0.9	0.3	1	1.0007	1.0223	1.0228	1.0030	1.0150	**0.9986**	**0.9974**
0.3	0.9	0.9	0	1.0374	1.1101	1.0663	1.0495	1.0180	1.0145	**0.9982**
0.3	0.9	0.9	1	1.0747	1.1167	1.0986	1.0437	1.0418	**0.9902**	**0.9967**
0.9	0.3	0.3	0	**0.7922**	**0.8972**	**0.8844**	**0.9734**	**0.9794**	**0.9949**	**0.9966**
0.9	0.3	0.3	1	**0.7874**	**0.8780**	**0.9221**	**0.9792**	**0.9937**	**0.9899**	**0.9936**
0.9	0.3	0.9	0	**0.7351**	**0.7785**	**0.7875**	**0.8785**	**0.8393**	**0.9248**	**0.9551**
0.9	0.3	0.9	1	**0.7894**	**0.8710**	**0.8707**	**0.9287**	**0.9182**	**0.9759**	1.0208
					Moderately Weak Factors					
0	0	0	0	1.0056	1.0187	1.0137	**0.9998**	1.0004	1.0147	1.0071
0.3	0.9	0.3	0	1.0099	1.0147	1.0165	**0.9990**	**0.9985**	**0.9809**	**0.9773**
0.3	0.9	0.3	1	1.0116	1.0541	1.0493	1.0175	1.0207	**1.0057**	1.0083
0.3	0.9	0.9	0	1.0095	1.0498	1.0560	1.0172	1.0268	**0.9831**	**0.9855**
0.3	0.9	0.9	1	1.0008	1.0782	1.0767	1.0725	1.0852	**0.9911**	**0.9955**
0.9	0.3	0.3	0	**0.7623**	**0.8563**	**0.8736**	**0.9573**	**0.9356**	1.0071	1.0067
0.9	0.3	0.3	1	**0.7613**	**0.8195**	**0.8476**	**0.9779**	**0.9574**	**0.9979**	**0.9945**
0.9	0.3	0.9	0	**0.7987**	**0.8699**	**0.8870**	**0.9616**	**0.9250**	1.0176	1.0212
0.9	0.3	0.9	1	**0.7401**	**0.8705**	**0.8733**	**0.9204**	**0.9220**	**0.9705**	**0.9738**
					Weak Factors					
0	0	0	0	**0.9927**	1.0077	1.0025	**0.9982**	**0.9928**	**0.9941**	**0.9933**

continued

Table B.3: RMSE for DGP3

ρ_f	ρ_g	a_0	d_0	TVP-3PRF	3PRF	PLS	Hard1	Hard2	Soft1	Soft2
0.3	0.9	0.3	0	**<u>0.9435</u>**	**0.9819**	**0.9915**	**0.9999**	**0.9877**	1.0062	**0.9994**
0.3	0.9	0.3	1	1.0176	1.0361	1.0280	1.0255	1.0004	1.0113	1.0168
0.3	0.9	0.9	0	**<u>0.9470</u>**	1.0648	1.0656	1.0306	1.0643	1.0195	1.0026
0.3	0.9	0.9	1	1.0477	1.0621	1.0606	1.0241	1.0446	1.0225	1.0020
0.9	0.3	0.3	0	**<u>0.7766</u>**	**0.8143**	**0.8022**	**0.9580**	**0.8799**	1.0033	**0.9802**
0.9	0.3	0.3	1	**<u>0.8072</u>**	**0.8404**	**0.8744**	**0.9863**	**0.9618**	1.0018	**0.9907**
0.9	0.3	0.9	0	**0.9038**	**0.9007**	**0.9017**	**0.9084**	**<u>0.9006</u>**	**0.9949**	**0.9779**
0.9	0.3	0.9	1	**<u>0.8244</u>**	**0.8933**	**0.8896**	**0.9999**	1.0088	**0.9562**	1.0024
				Moderately Weak and Non-Pervasive Factors						
0	0	0	0	1.0109	1.0153	1.0107	1.0037	**<u>0.9978</u>**	1.0074	1.0023
0.3	0.9	0.3	0	**<u>0.9633</u>**	1.0161	1.0257	1.0089	**0.9947**	**0.9818**	**0.9788**
0.3	0.9	0.3	1	**0.9884**	**0.9934**	**<u>0.9819</u>**	1.0219	1.0154	1.0010	1.0001
0.3	0.9	0.9	0	1.0284	1.1023	1.0714	1.0260	1.0331	1.0049	1.0040
0.3	0.9	0.9	1	1.0327	1.0124	1.0232	1.0004	**<u>0.9847</u>**	1.0134	1.0047
0.9	0.3	0.3	0	**<u>0.8315</u>**	**0.8778**	**0.8888**	**0.9769**	**0.9344**	**0.9989**	**0.9925**
0.9	0.3	0.3	1	**<u>0.8491</u>**	**0.9372**	**0.9206**	1.0094	**0.9942**	1.0020	1.0058
0.9	0.3	0.9	0	**<u>0.7480</u>**	**0.7572**	**0.7535**	**0.8112**	**0.7784**	**0.9126**	**0.9497**
0.9	0.3	0.9	1	**<u>0.7912</u>**	**0.8854**	**0.8551**	**0.9425**	**0.9469**	1.0239	1.0467

Note: The table reports the out-of-sample forecasting median RMSE based on 1000 replications. Serial correlation in the factors is governed by ρ_f and ρ_g, while a_0 and d_0 govern serial and cross sectional correlation in the predictors' residuals, respectively. Entries in bold represent the lowest median MSE for each specification. See text for additional details.

B.1.2 Factor Fitting

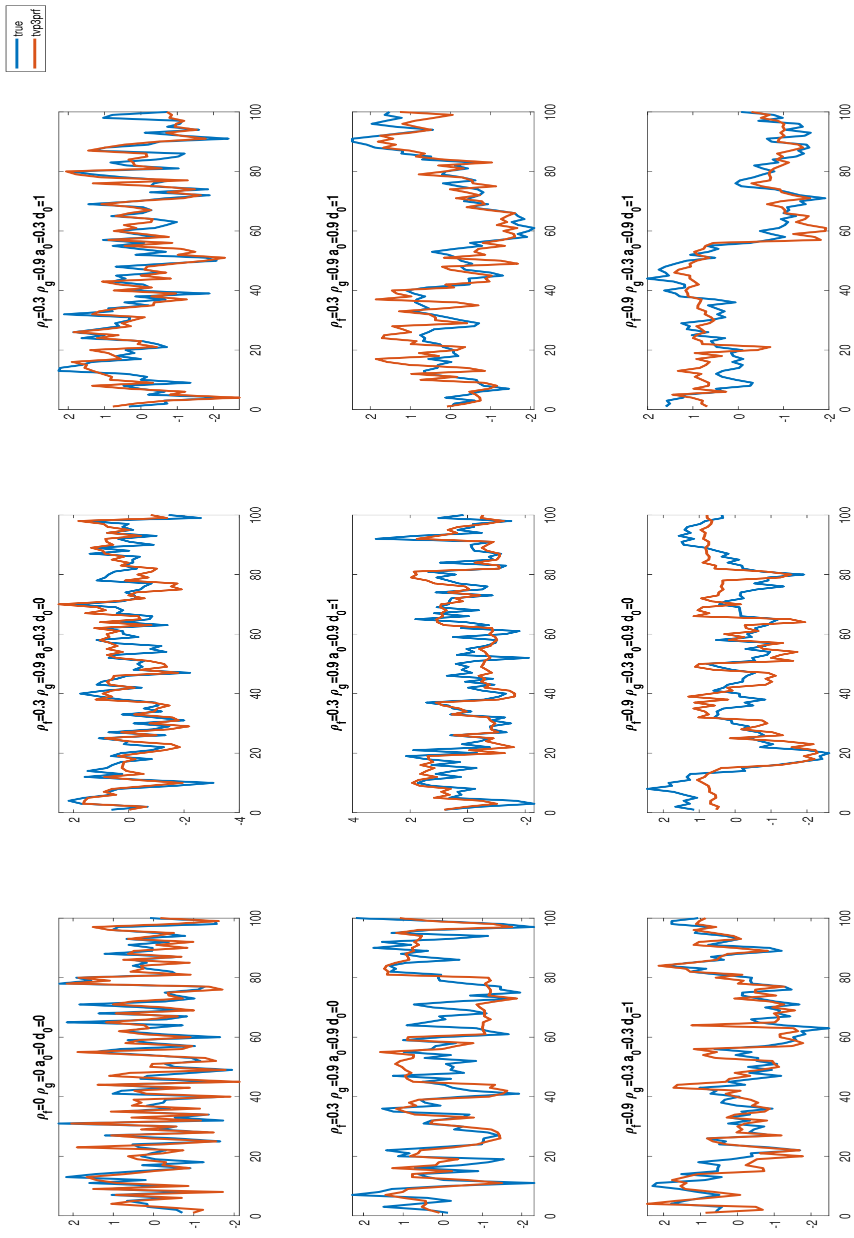

Figure B.1: *Fitted factor using TVP-3PRF for DGP1(Normal Factor)*

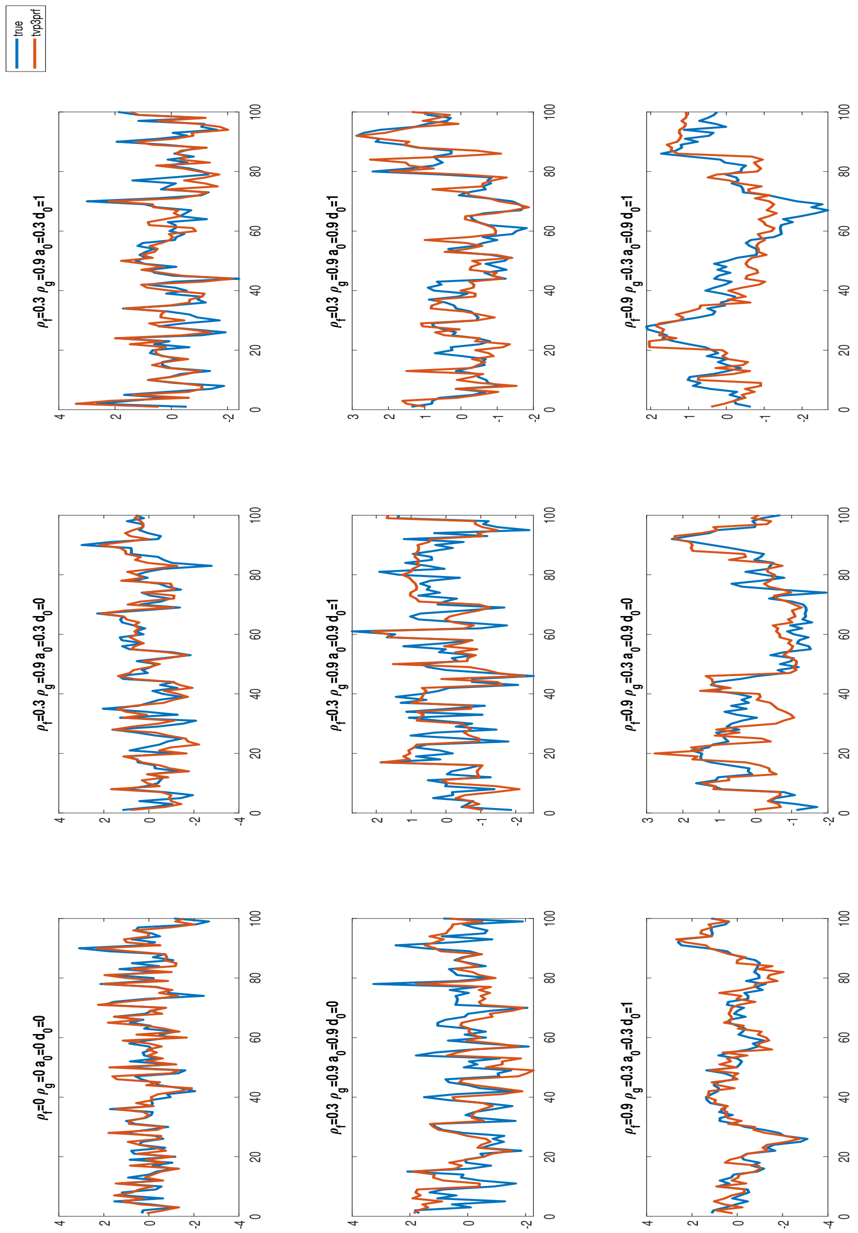

Figure B.2: *Fitted factor using TVP-3PRF for DGP1(Moderately Weak Factors)*

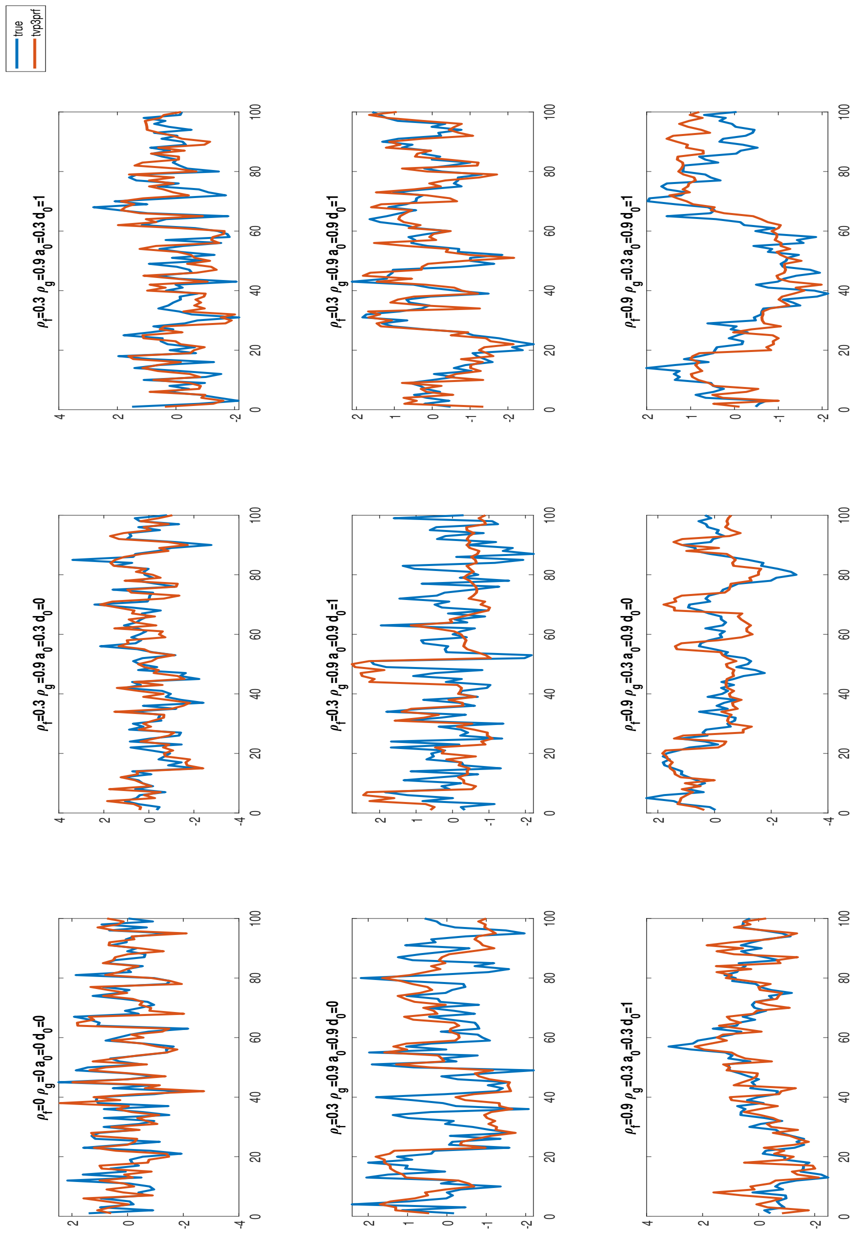

Figure B.3: *Fitted factor using TVP-3PRF for DGP1(Weak Factors)*

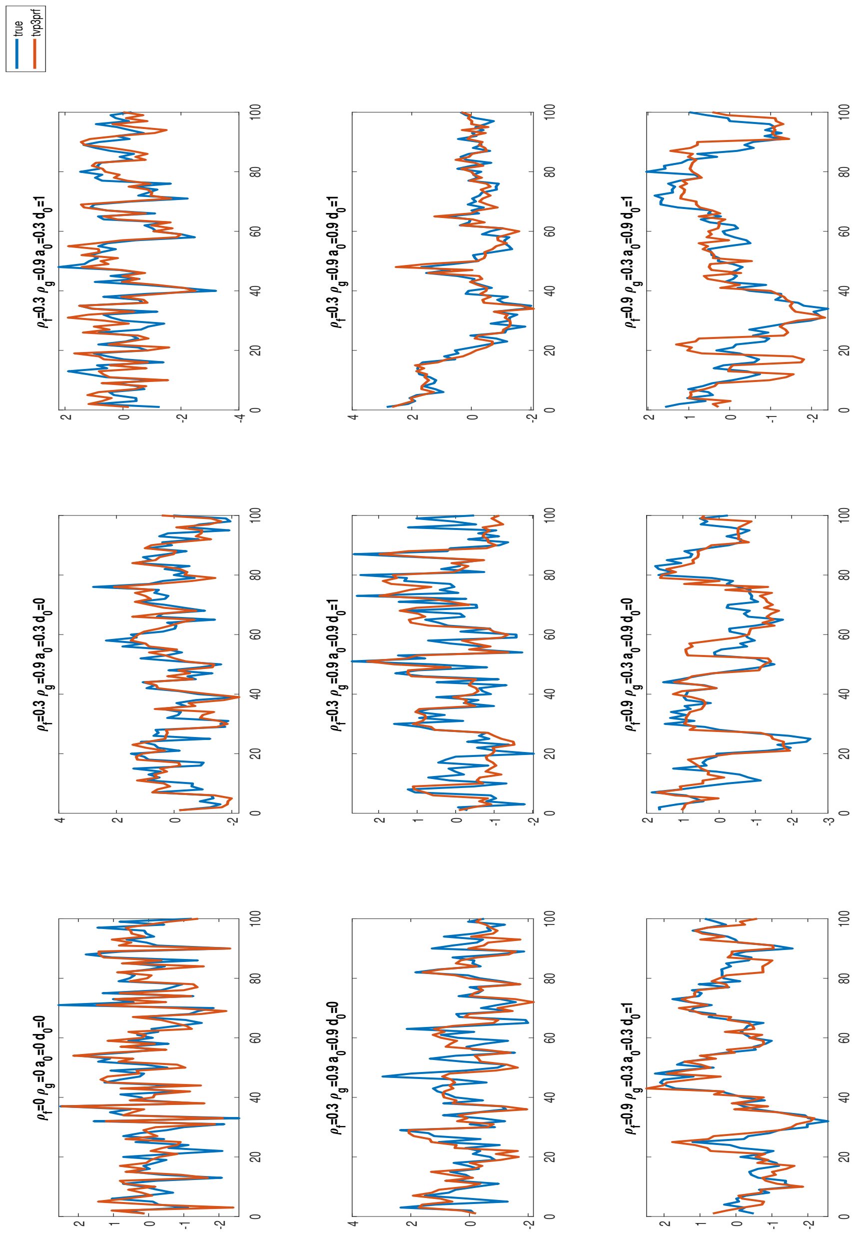

Figure B.4: *Fitted factor using TVP-3PRF for DGP1(Moderately Weak and Non-pervasive Factors)*

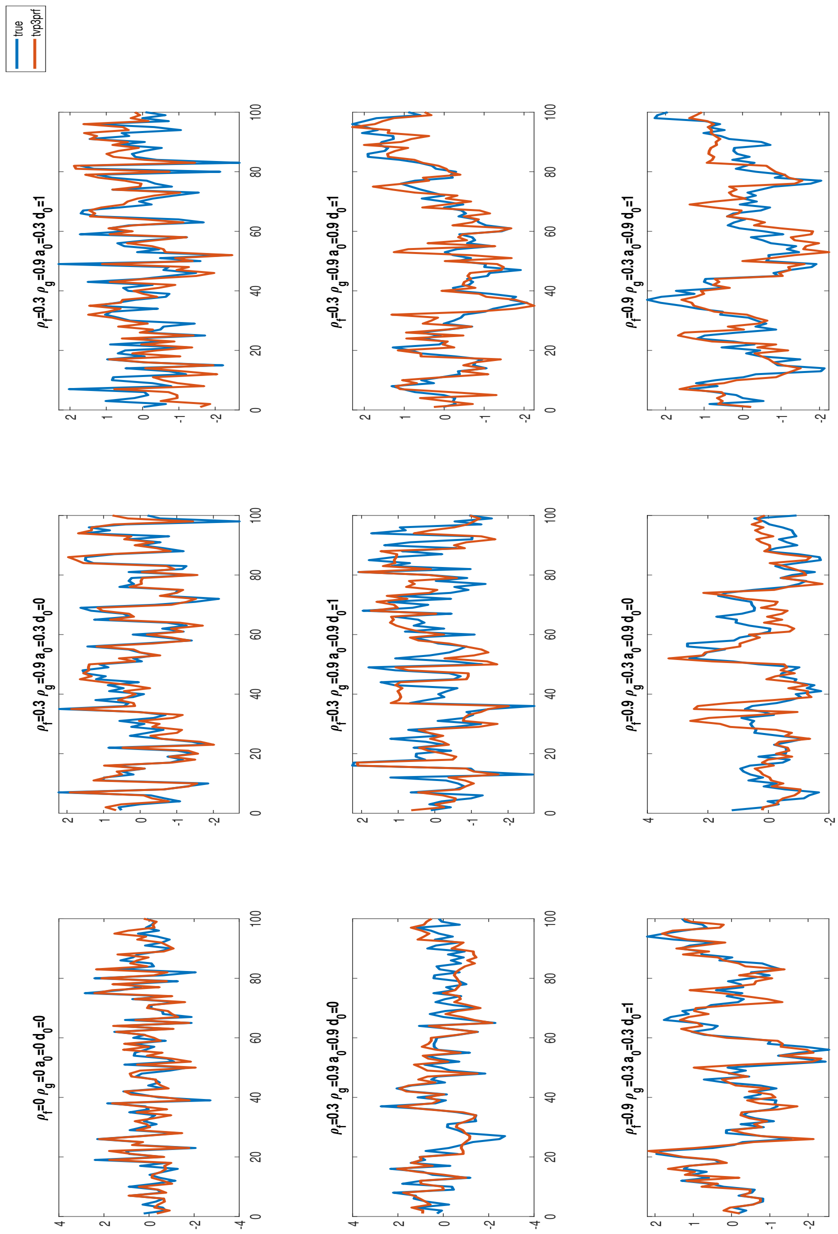

Figure B.5: *Fitted factor using TVP-3PRF for DGP1(No distinguish for factor strength)*

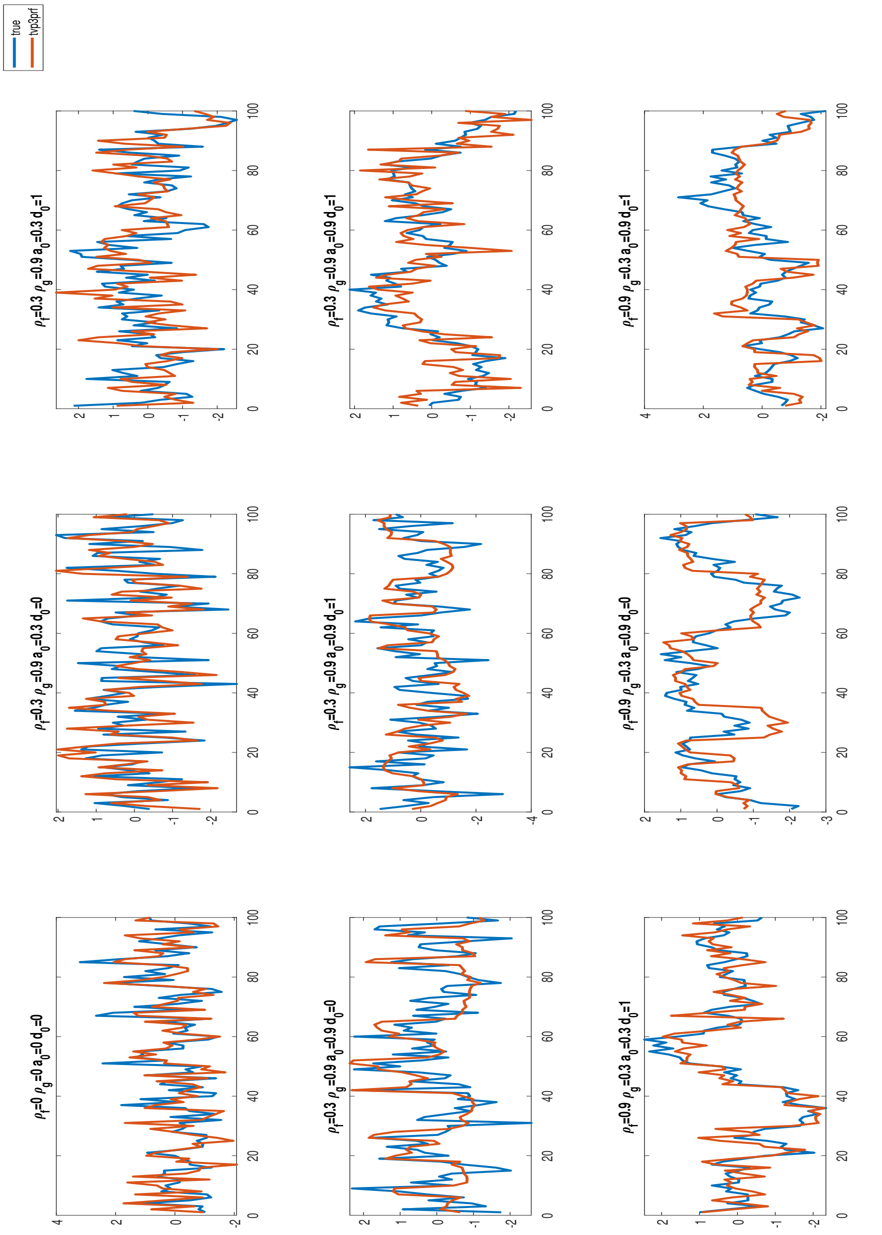

Figure B.6: *Fitted factor using TVP-3PRF for DGP2(Normal Factor)*

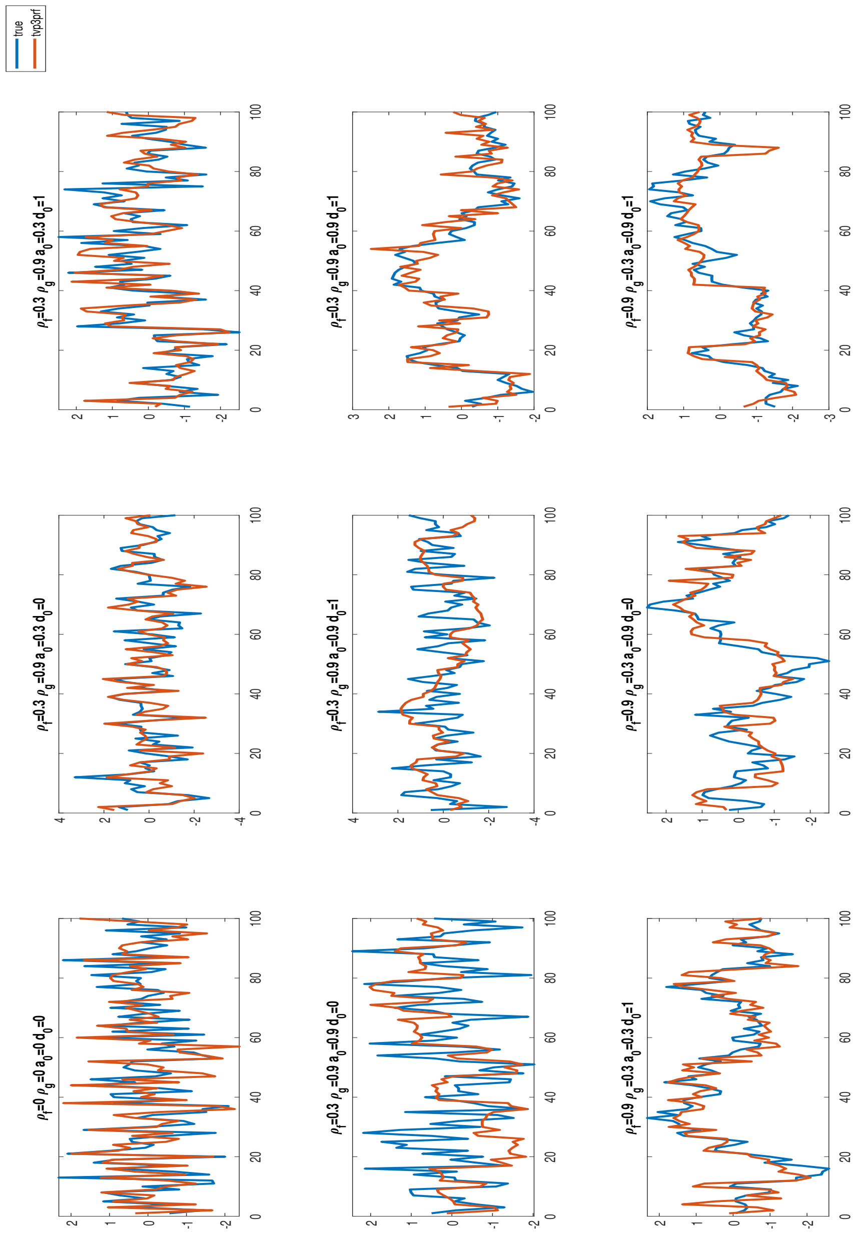

Figure B.7: *Fitted factor using TVP-3PRF for DGP2(Moderately Weak Factors)*

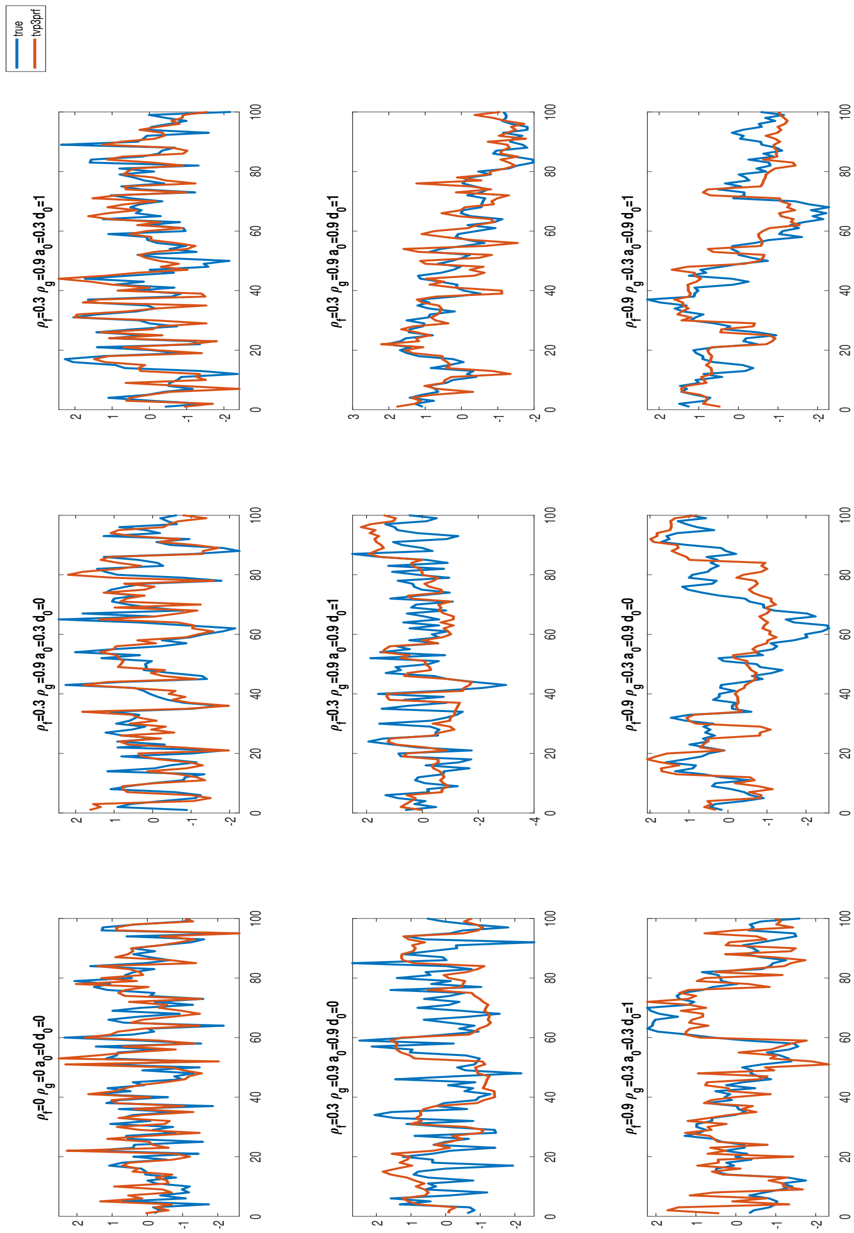

Figure B.8: *Fitted factor using TVP-3PRF for DGP2(Weak Factors)*

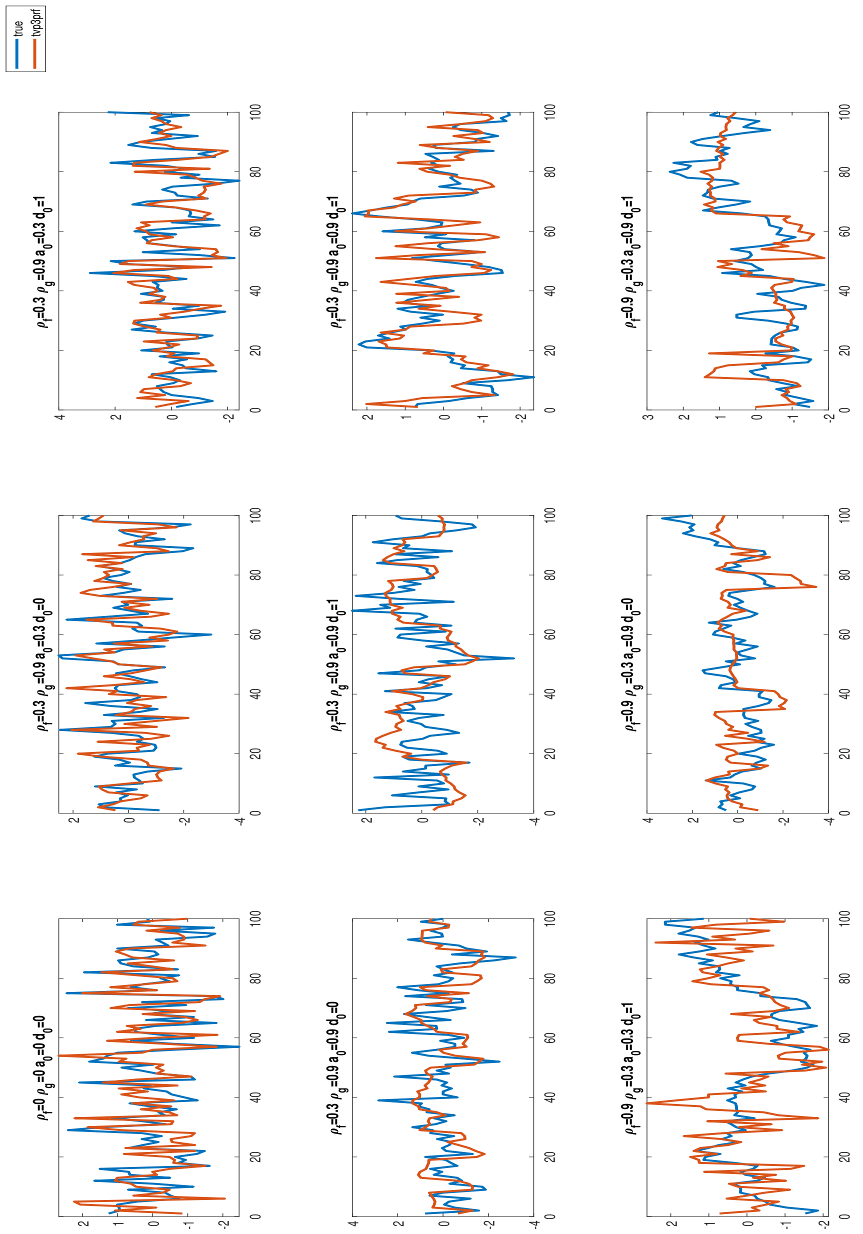

Figure B.9: *Fitted factor using TVP-3PRF for DGP2(Moderately Weak and Non-pervasive Factors)*

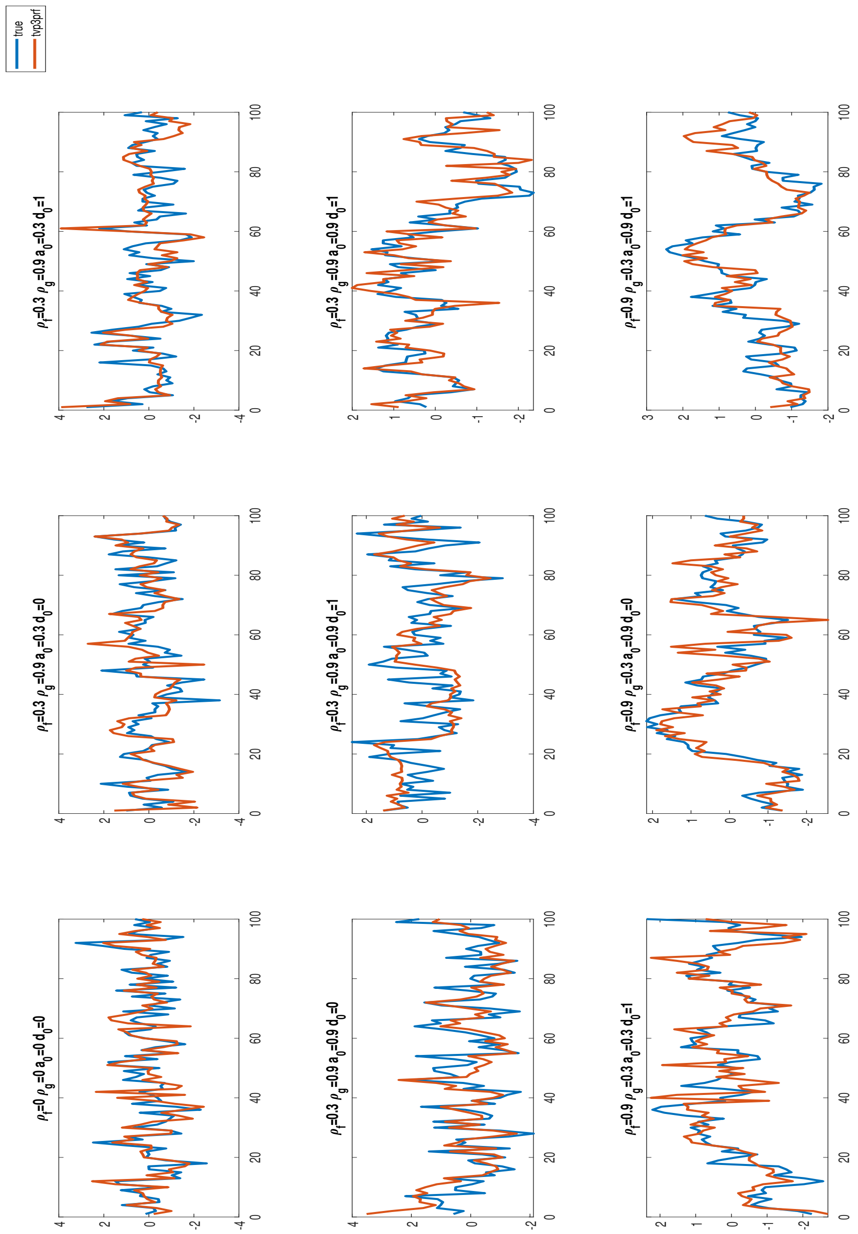

Figure B.10: *Fitted factor using TVP-3PRF for DGP2(No distinguish for factor strength)*

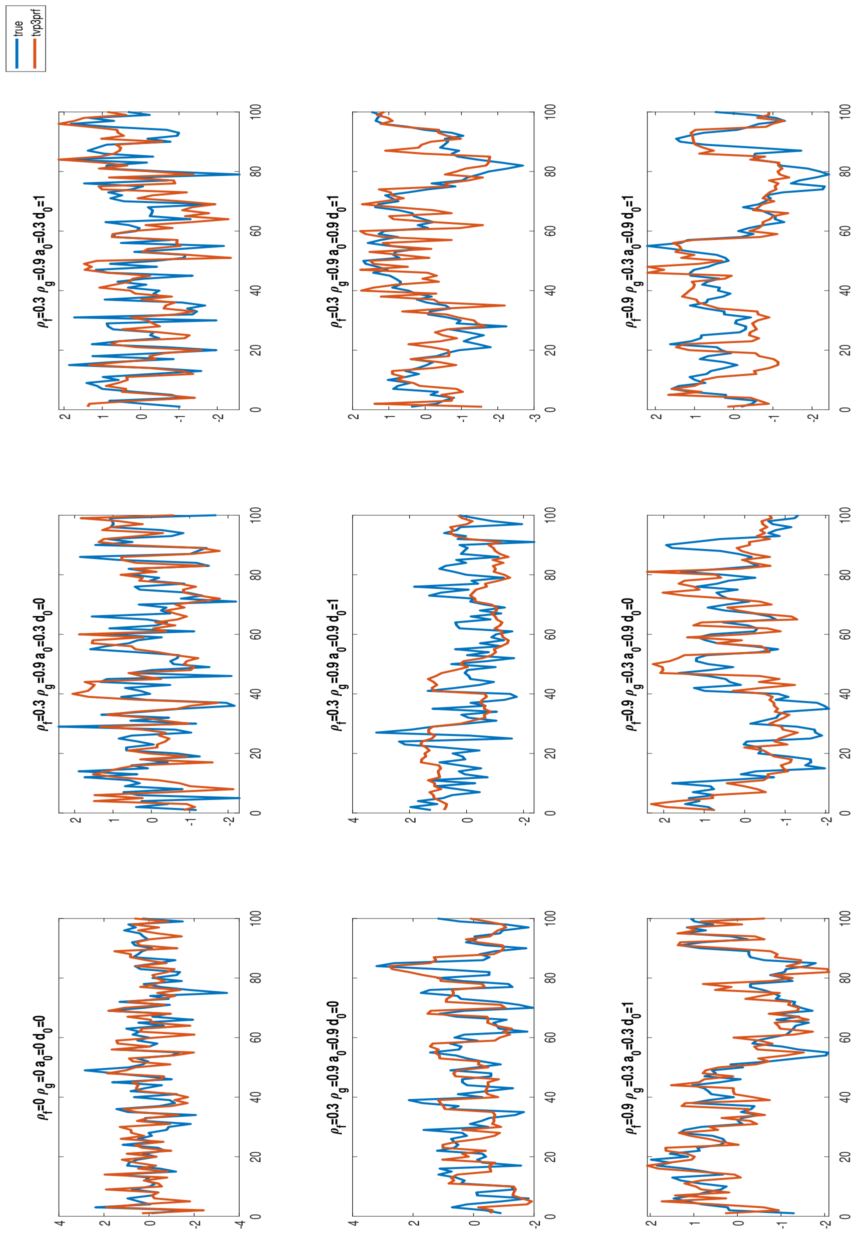

Figure B.11: *Fitted factor using TVP-3PRF for DGP3(Normal Factor)*

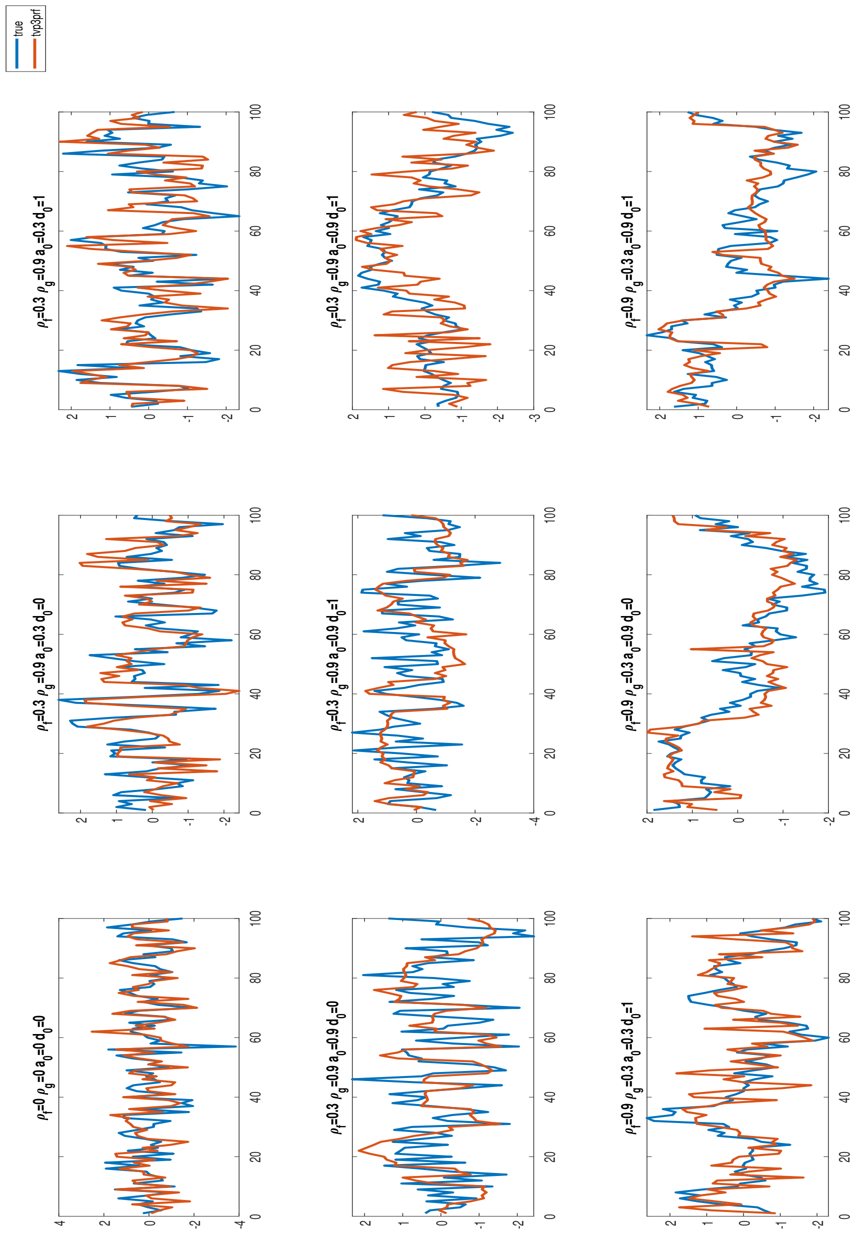

Figure B.12: *Fitted factor using TVP-3PRF for DGP3(Moderately Weak Factors)*

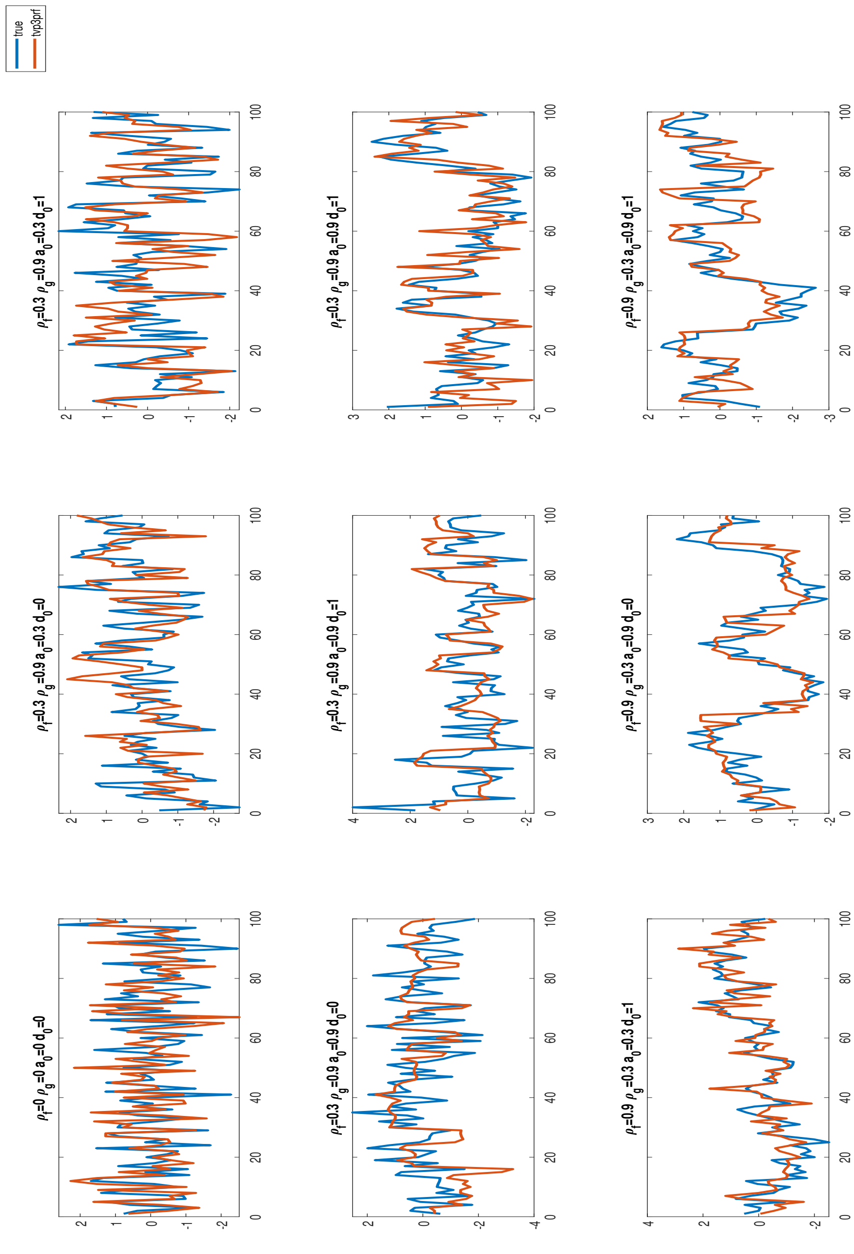

Figure B.13: *Fitted factor using TVP-3PRF for DGP3(Weak Factors)*

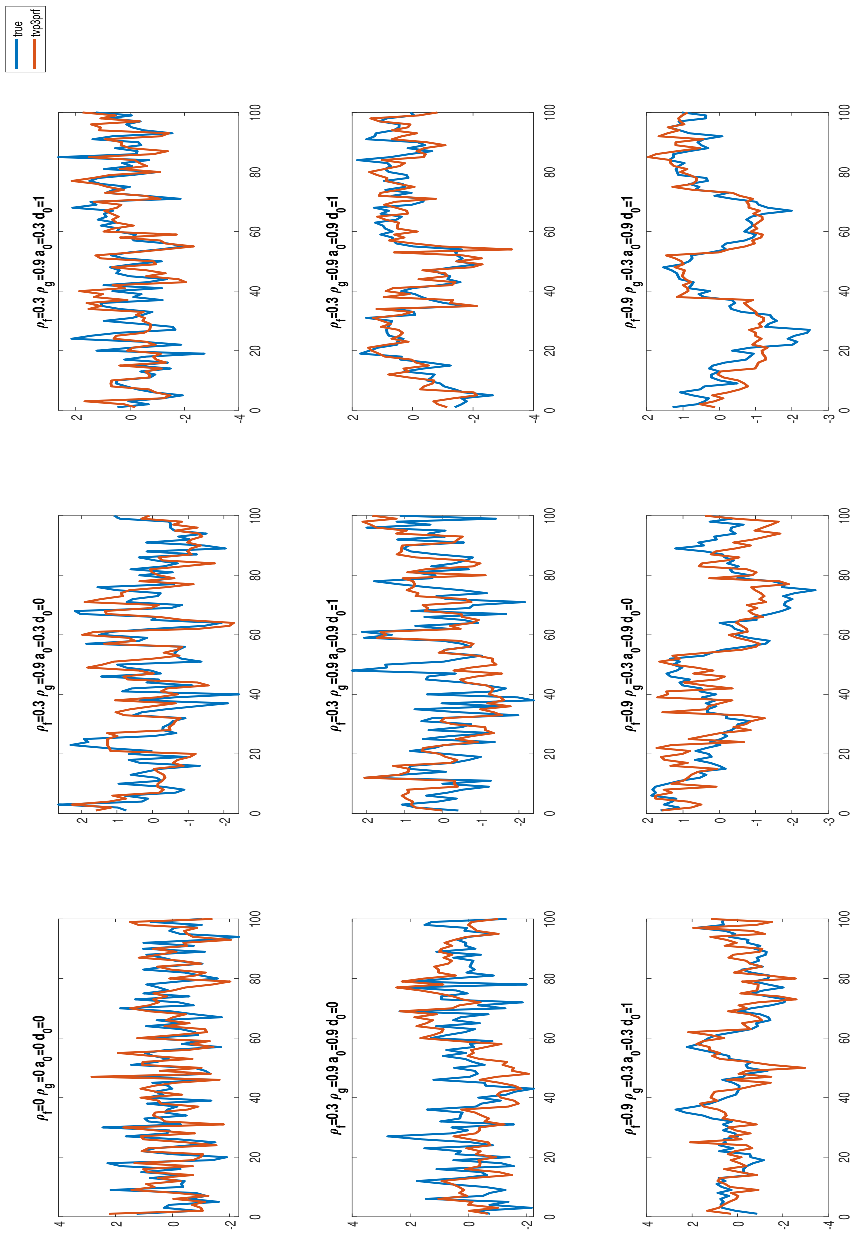

Figure B.14: *Fitted factor using TVP-3PRF for DGP3(No distinguish for factor strength)*

B.2 Additional Empirical Result for $L = 2$

Table B.4: *RMSFE for TVP-3PRF(L=2)*

RMSPE	h=1	h=2	h=3	h=4	h=8
GDP	1.01	**1.00**	**1.00**	1.03	1.03
Consumption	**0.98**	1.01	1.00	**1.00**	1.03
Industrial Production	**0.99**	1.02	**1.00**	1.02	1.01
Employment	**1.00**	1.01	1.04	1.01	1.01
Hours	**0.98**	**0.99**	1.01	1.00	1.04
CPI	1.03	1.01	1.05	**0.99**	1.03
GS10	1.00	**0.99**	**1.00**	**0.99**	**0.96**
Investment	**0.99**	**0.98**	**1.00**	**0.99**	1.02
S&P500	1.02	1.01	1.03	1.02	**1.00**
Export	1.01	**0.93**	**0.96**	**0.98**	**0.95**
Import	**0.81**	**0.69**	**0.90**	1.00	1.13
FedFund	**0.51**	**0.63**	**0.69**	**0.70**	**0.85**

Notes: This table shows the empirical forecasting results for TVP-3PRF with two proxies(TVP-3PRF(L=2)). The benchmark model is TVP-3PRF with one proxy(TVP-3PRF(L=1)).

Table B.4 compares the results of using TVP-3PRF (L=2) and TVP-PRF (L=1) to predict macroeconomic variables by relative mean squared predictive error(RMSPE). A value larger than 1 denotes that TVP-3PRF (L=2) is superior than the benchmark model, TVP-PRF (L=1), and vice versa. This finding demonstrates that, for the majority of the selected macroeconomic variables, the predictive power is not greatly changed by the addition of another proxy, or common factor. However, for variables like the Federal Funds Rate, adding one more common factors can considerably enhance forecast ability of the model.

B.3 Assumption of the 3PRF algorithm

The Assumptions that provide a groundwork for developing asymptotic properties of the 3PRF is as follow, which can also be found in Kelly and Pruitt (2015).

Assumption 1 *(Factor Structure). The data are generated by the following:*

$$\begin{aligned} \boldsymbol{x}_t &= \phi_0 + \boldsymbol{\Phi} \boldsymbol{F}_t + \varepsilon_t & y_{t+1} &= \beta_0 + \boldsymbol{\beta}' \boldsymbol{F}_t + \eta_{t+1} & z_t &= \lambda_0 + \boldsymbol{\Lambda} \boldsymbol{F}_t + \omega_t \\ \boldsymbol{X} &= \boldsymbol{\iota} \phi_0' + \boldsymbol{F} \Phi' + \varepsilon & \boldsymbol{y} &= \boldsymbol{\imath} \beta_0 + \boldsymbol{F} \boldsymbol{\beta} + \eta & \boldsymbol{Z} &= \boldsymbol{\imath} \boldsymbol{\lambda}_0' + \boldsymbol{F} \boldsymbol{\Lambda}' + \omega \end{aligned} \tag{B.1}$$

where $\boldsymbol{F}_t = (\boldsymbol{f}_t', \boldsymbol{g}_t')'$, $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_f, \boldsymbol{\Phi}_g)$, $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_f, \boldsymbol{\Lambda}_g)$, *and* $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_f', \mathbf{0}'\right)'$ *with* $\left|\boldsymbol{\beta}_f\right| > \mathbf{0}$.$K_f > 0$ *is the dimension of vector* $\boldsymbol{f}_t$, $K_g \geq 0$ *is the dimension of vector* g_t, L *is the dimension of vector* $z_t (0 < L < \min(N, T))$, *and* $K = K_f + K_g$.

Assumption 1 defines the factor structure. According to the definition of factor loadings of the target variable, the target rely on a certain subset of the factors what also drive the predictors. We refer to this subset as the relevant factors, which are denoted $\boldsymbol{f}_t$. In contrast, irrelevant factors, $\mathbf{g}_t$, do not influence the forecast target but may drive the cross section of predictive information $\boldsymbol{x}_t$. The proxies $\boldsymbol{z}_t$ are driven by factors and proxy noise.

Assumption 2 *(Factors, Loadings and Residuals). Let* $M < \infty$. *For any* i, s, t

1. $\mathbb{E} \left\|\boldsymbol{F}_t\right\|^4 < M, T^{-1} \sum_{s=1}^{T} \boldsymbol{F}_S \xrightarrow[T \to \infty]{p} \boldsymbol{\mu}$ *and* $T^{-1} \boldsymbol{F}' \boldsymbol{J}_T \boldsymbol{F} \xrightarrow[T \to \infty]{p} \boldsymbol{\Delta}_F$
2. $\mathbb{E} \left\|\phi_i\right\|^4 \leq M, N^{-1} \sum_{j=1}^{N} \phi_j \xrightarrow[T \to \infty]{p} \bar{\phi}, N^{-1} \Phi' J_N \Phi \xrightarrow[N \to \infty]{p} \mathcal{P}$ *and* $N^{-1} \boldsymbol{\Phi}' \boldsymbol{J}_N \boldsymbol{\phi}_0 \xrightarrow[N \to \infty]{p} \boldsymbol{P}_1^6$
3. $\mathbb{E}\left(\varepsilon_{it}\right) = 0, \mathbb{E}\left|\varepsilon_{it}\right|^8 \leq M$

4. $\mathbb{E}(\omega_t) = \mathbf{0}, \mathbb{E}\left\|\boldsymbol{\omega}_t\right\|^4 \leq M, T^{-1/2}\sum_{s=1}^{T}\boldsymbol{\omega}_s = \boldsymbol{O}_p(1)$ *and* $T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\omega} \xrightarrow[N\to\infty]{p} \boldsymbol{\Delta}_\omega$

5. $\mathbb{E}_t(\eta_{t+1}) = \mathbb{E}(\eta_{t+1} \mid y_t, F_t, y_{t-1}, F_{t-1}, \ldots) = 0, \mathbb{E}\left(\eta_{t+1}^4\right) \leq M$, *and* η_{t+1} *is independent of* $\phi_i(m)$ *and* $\varepsilon_{i,t}$.

Since η_{t+1} is a martingale difference sequence with respect to all information known at time t, $\beta_0 + \boldsymbol{\beta}_f'\boldsymbol{f}_t$ gives the best time t forecast. But it is infeasible since the relevant factors $\boldsymbol{f}_t$ are unobserved.

We require factors and loadings to be cross-sectionally regular in that they have well-behaved covariance matrices for large T and N, respectively. Assumption 2 does not exist in the work of Stock and Watson or Bai and Ng, and is required because the 3PRF uses proxies to extract factors. We bound the moments of proxy noise ω_t in the same manner as the bounds on factor moments.

Assumption 3 *(Dependence). Let* $x(m)$ *denote the* m *th element of* $\boldsymbol{x}$. *For* $M < \infty$ *and any* i, j, t, s, m_1, m_2

1. $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij,ts}, |\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ *and* $|\sigma_{ij,ts}| \leq \tau_{ts}$, *and*

   (a) $N^{-1}\sum_{i,j=1}^{N}\bar{\sigma}_{ij} \leq M$

   (b) $T^{-1}\sum_{t,s=1}^{T}\tau_{ts} \leq M$

   (c) $N^{-1}\sum_{i,s}|\sigma_{ii,ts}| \leq M$

   (d) $N^{-1}T^{-1}\sum_{i,j,t,s}|\sigma_{ij,ts}| \leq M$

2. $\mathbb{E}\left|N^{-1/2}T^{-1/2}\sum_{s=1}^{T}\sum_{i=1}^{N}\left[\varepsilon_{is}\varepsilon_{it} - \mathbb{E}(\varepsilon_{is}\varepsilon_{it})\right]\right|^2 \leq M$

3. $\mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}F_t(m_1)\,\omega_t(m_2)\right|^2 \leq M$

4. $\mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}\omega_t(m_1)\varepsilon_{it}\right|^2 \leq M.$

Assumption 4 *(Central Limit Theorems). For any i, t*

1. $N^{-1/2}\sum_{i=1}^{N}\phi_i\varepsilon_{it} \xrightarrow{d} \mathcal{N}(0, \Gamma_{\Phi\varepsilon})$*, where* $\Gamma_{\Phi\varepsilon} = \text{plim}_{N\to\infty} N^{-1}\sum_{i,j=1}^{N}\mathbb{E}\left[\phi_i\phi_j'\varepsilon_{it}\varepsilon_{jt}\right]$

2. $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{F}_t\eta_{t+1} \xrightarrow{d} \mathcal{N}(0, \Gamma_{F\eta})$*, where* $\Gamma_{F\eta} = \text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{t+1}^2\boldsymbol{F}_t\boldsymbol{F}_t'\right] > 0$

3. $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{F}_t\varepsilon_{it} \xrightarrow{d} \mathcal{N}(0, \Gamma_{F\varepsilon,i})$*, where* $\Gamma_{F\varepsilon,i} = \text{plim}_{T\to\infty} T^{-1}\sum_{t,s=1}^{T}\mathbb{E}\left[\boldsymbol{F}_t\boldsymbol{F}_s'\varepsilon_{it}\varepsilon_{is}\right] > 0.$

Assumption 3 allows the factor structure to be approximate in the sense that some cross section correlation among ε_{it} is permitted, following Chamberlain and Rothschild (1982). Similarly, we allow for serial dependence among ε_{it} (including GARCH) as in Stock and Watson (2002a). In addition, we allow some proxy noise dependence with factors and idiosyncratic shocks. Assumption 4 requires that central limit theorems apply, and is satisfied when various mixing conditions hold among factors, loadings and shocks.

Assumption 5 *(Normalization).* $\mathcal{P} = \boldsymbol{I}, \boldsymbol{P}_1 = \boldsymbol{0}$ *and* $\boldsymbol{\Delta}_{\text{F}}$ *is diagonal, positive definite, and each diagonal element is unique.*

Assumption 5 recognizes that there exists an inherent unidentification between the factors and factor loadings. It therefore selects a normalization in which the covariance of predictor loadings is the identity matrix, and in which factors are orthogonal to one another. As with principal components, the particular normalization is unimportant. We ultimately estimate a vector space spanned by the factors, and this space does not depend upon the choice of normalization.

Assumption 6 *(Relevant Proxies).* $\boldsymbol{\Lambda} = [\boldsymbol{\Lambda}_f, \mathbf{0}]$ *and* $\boldsymbol{\Lambda}_f$ *is nonsingular.*

Assumption 6 states that proxies (i) have zero loading on irrelevant factors, (ii) have linearly independent loadings on the relevant factors, and (iii) number equal to the number of relevant factors. Combined with the normalization assumption, this says that the common component of proxies spans the relevant factor space, and that none of the proxy variation is due to irrelevant factors.

Note that Assumptions 2.4, 3.3, 3.4 and 6 are the only conditions involving the proxy variables. We prove in Theorem 7 that automatic proxies, which are generally constructable using $\boldsymbol{X}$ and $\boldsymbol{y}$, are guaranteed to satisfy these proxy assumptions.

With these assumptions in place, the asymptotic properties of the three-pass regression filte is derived. Our proofs build upon the seminal theory of Stock and Watson (2002a), Bai (2003) and Bai and Ng (2002, 2006).

B.4 Consistency of 3PRF estimation

Theorem 1 *Let Assumptions 1-6 hold. The three-pass regression filter forecast is consistent for the infeasible best forecast,* $\hat{y}_{t+1} \xrightarrow[T,N\rightarrow\infty]{p} \beta_0 + \boldsymbol{F}_t'\boldsymbol{\beta}$

The following result builds on the previous lemma. It identifies finite-dimensional matrices that appear in the expression for the 3PRF, and then looks to find the stochastic order of any generic element of the matrix.

Lemma 1 *Let Assumptions 1-4 hold. Then*

1. $T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p(1)$
2. $T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\eta = \boldsymbol{O}_p\ (1)$
3. $T^{-1/2}\varepsilon'\boldsymbol{J}_T\eta = \boldsymbol{O}_p(1)$
4. $N^{-1/2}\varepsilon_t'\boldsymbol{J}_N\boldsymbol{\Phi} = \boldsymbol{O}_p(1)$
5. $N^{-1}T^{-1}\Phi'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
6. $N^{-1}T^{-1/2}\Phi'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p(1)$
7. $N^{-1/2}T^{-1/2}\boldsymbol{\Phi}\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\eta = \boldsymbol{O}_p(1)$
8. $N^{-1}T^{-3/2}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
9. $N^{-1}T^{-3/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
10. $N^{-1}T^{-3/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
11. $N^{-1}T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon_t = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
12. $N^{-1}T^{-1/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon_t = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
13. $N^{-1}T^{-3/2}\eta'J_T\varepsilon\boldsymbol{J}_N\varepsilon'J_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$
14. $N^{-1}T^{-3/2}\eta'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'J_T\boldsymbol{\omega} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

The stochastic order is understood to hold as $N, T \to \infty$*, stochastic orders of matrices are understood to apply to each entry, and* $\delta_{NT} \equiv \min(\sqrt{N}, \sqrt{T})$

The following lemma 2 finds the probability limit for our factor estimator $\hat{\boldsymbol{F}}$. It expands out this expression to find terms involving X, Z, y that can then be expressed using Assumption 1 as matrices appearing in Lemma 1.

Lemma 2 *Let Assumptions 1–4 hold. Then the probability limits of* $\hat{\boldsymbol{\Phi}}$ *and* $\hat{\boldsymbol{F}}_t$ *are*

$$\hat{\boldsymbol{\Phi}} \xrightarrow[T\to\infty]{p} \left(\boldsymbol{\Lambda}\boldsymbol{\Delta}_F\boldsymbol{\Lambda}' + \boldsymbol{\Delta}_\omega\right)^{-1} \boldsymbol{\Lambda}\boldsymbol{\Delta}_F\boldsymbol{\Phi}' \tag{B.2}$$

and

$$\hat{\boldsymbol{F}}_t \xrightarrow[T,N\to\infty]{p} \left(\boldsymbol{\Lambda}\boldsymbol{\Delta}_F\boldsymbol{\Lambda}' + \boldsymbol{\Delta}_\omega\right)\left(\boldsymbol{\Lambda}\boldsymbol{\Delta}_F\mathcal{P}\boldsymbol{\Delta}_F\boldsymbol{\Lambda}'\right)^{-1}\left(\boldsymbol{\Lambda}\boldsymbol{\Delta}_F\boldsymbol{P}_1 + \boldsymbol{\Lambda}\boldsymbol{\Delta}_F\mathcal{P}\boldsymbol{F}_t\right) \tag{B.3}$$

The following lemma3 finds the probability limit for our factor estimator $\hat{\beta}$. It expands out this expression to find terms involving X, Z, y that can then be expressed using Assumption 1 as matrices appearing in Lemma 1.

Lemma 3 *Let Assumptions 1–4 hold. Then the probability limit of estimated third stage predictive coefficients* $\hat{\beta}$ *is*

$$\begin{aligned}\hat{\boldsymbol{\beta}} \xrightarrow[T,N\to\infty]{p} & \left(\Lambda\Delta_F\Lambda' + \Delta_\omega\right)^{-1}\Lambda\Delta_F\mathscr{P}\Delta_F\Lambda' \\ & \times\left(\Lambda\Delta_F\mathcal{P}\Delta_F\mathcal{P}\Delta_F\Lambda'\right)^{-1}\Lambda\Delta_F\mathcal{P}\Delta_F\beta\end{aligned} \tag{B.4}$$

This lemma finds the probability limit for our factor estimator $\hat{y}$, but is immediate from the two preceding proofs.

Lemma 4 *Let Assumptions 1–3 hold. Then the three pass regression filter fore-*

cast satisfies

$$\hat{y}_{t+1} \frac{p}{T, N \to \infty} \beta_0 + \mu' \boldsymbol{\beta} + (\boldsymbol{F}_t - \boldsymbol{\mu})' \mathcal{P} \boldsymbol{\Delta}_F \boldsymbol{\Lambda}' \\ \times [\boldsymbol{\Lambda} \boldsymbol{\Delta}_F \mathcal{P} \boldsymbol{\Delta}_F \mathcal{P}_{\mathrm{F}} \boldsymbol{\Lambda}']^{-1} \boldsymbol{\Lambda} \boldsymbol{\Delta}_F \mathcal{P}_{\mathrm{F}} \boldsymbol{\beta} \tag{B.5}$$

Until here, the auxiliary lemmas are enough to proof Theorem1.

Given Assumptions 1–3, Lemma 4 holds and we can therefore manipulate Equation B.5. Partition $\mathcal{P}$ and Δ_F as

$$\mathcal{P} = \begin{bmatrix} \mathcal{P}_1 & \mathcal{P}_{12} \\ \mathcal{P}'_{12} & \mathcal{P}_2 \end{bmatrix}, \quad \boldsymbol{\Delta}_F = \begin{bmatrix} \boldsymbol{\Delta}_{F,1} & \boldsymbol{\Delta}_{F,12} \\ \boldsymbol{\Delta}'_{F,12} & \boldsymbol{\Delta}_{F,2} \end{bmatrix} \tag{B.6}$$

such that the block dimensions of $\mathcal{P}$ and Δ_F coincide. By Assumption 5, the off-diagonal blocks, $\mathcal{P}$ and $\Delta_{F,12}$, are zero. As a result, the first diagonal block of the term $\Delta_F \mathscr{P} \Delta_F \mathscr{P} \Delta_F$ in Eq.B.5 is $\Delta_{F,1} \mathscr{P}_1 \Delta_{F,1} \mathscr{P}_1 \Delta_{F,1}$. y Assumption 6, pre- and post- multiplying by $\Lambda = [\boldsymbol{\Lambda}_f, \mathbf{0}]$ reduces the term in square brackets to $\boldsymbol{\Lambda}_f \Delta_{F,1} \mathcal{P}_1 \Delta_{F,1} \mathscr{P}_1 \Delta_{F,1} \boldsymbol{\Lambda}_f$. Similarly, $\mathscr{P} \Delta_F \Lambda' = [\boldsymbol{\Lambda}_f \mathscr{P}_1 \Delta_{F,1}, \mathbf{0}]'$ and $\boldsymbol{\Lambda} \boldsymbol{\Delta}_F \mathscr{P} \boldsymbol{\Delta}_F = [\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_{F,1} \mathscr{P}_1 \boldsymbol{\Delta}_{F,1}, \mathbf{0}]$. By Assumption 6, $\boldsymbol{\Lambda}_f$ is invertible and therefore the expression for $\hat{y}_{t+1}$ reduces to $\beta_0 + \boldsymbol{F}'_t \beta$. ■

Theorem 1 says that the 3PRF is consistent so that for large N and T the difference between this feasible forecast and the infeasible best vanishes. This and other asymptotic results related to 3PRF are based on simultaneous N and T limits. As discussed by Bai (2003), the existence of a simultaneous limit implies the existence of coinciding sequential and pathwise limits, but the converse is not true.

The estimated loadings on individual predictors, $\hat{\alpha}$, play an important role in the interpretation of the 3PRF. The next theorem provides the probability limit for the loading on each predictor i.

Theorem 2 *Let $\hat{\alpha}_i$ denote the ith element of $\hat{\alpha}$, and let Assumptions 1–6 hold. Then for any i,*

$$N\hat{\alpha}_i \xrightarrow[T,N\to\infty]{p} \left(\phi_i - \bar{\phi}\right)' \beta. \tag{B.7}$$

Proof. Rewrite $\hat{\alpha}_i = S_i\hat{\alpha}$, where $\boldsymbol{S}_i$ is the $(1 \times N)$ selector vector with *ith* element equal to one and remaining elements zero. An alternative way to write the forecast is

$$\begin{aligned} \hat{\boldsymbol{y}} &= \imath + \boldsymbol{J}_T\boldsymbol{X}\hat{\alpha} \\ \hat{\alpha} &= \boldsymbol{W}_{XZ}\left(\boldsymbol{W}'_{XZ}\boldsymbol{S}_{XX}\boldsymbol{W}'_{XZ}\right)^{-1}\boldsymbol{W}'_{XZ}\boldsymbol{s}_{Xy} \end{aligned} \tag{B.8}$$

Expanding the expression for $\hat{\alpha}$ in Eq. (B.8), the first term in $S_i\hat{\alpha}$ is the $(1 \times K)$ matrix $S_iJ_N\Phi$, which has probability limit $(\phi_i - \boldsymbol{\phi})$ as $N, T \to \infty$. It then follows directly from previous results that

$$N\hat{\alpha}_i \xrightarrow[T,N\to\infty]{p} \left(\phi_i - \bar{\phi}\right)' \Delta_F\Lambda'\left(\Lambda\Delta_F\mathscr{P}\Delta_F\mathcal{P}\Delta_F\Lambda'\right)^{-1}\Lambda\Delta_F\mathcal{P}\Delta_F\beta \tag{B.9}$$

Under Assumptions 5 and 6, this reduces to $\left(\phi_i - \bar{\phi}\right)'\beta$. ■

The coefficient α maps underlying factors to the forecast target via the observable predictors. As a result the probability limit of $\hat{\alpha}$ is a product of the loadings of $\boldsymbol{X}$ and $\boldsymbol{y}$ on the relevant factors $\boldsymbol{f}$. This arises from the interpretation of $\hat{\alpha}$ as a constrained least squares coefficient estimate, which we elaborate on in the next section. Note that $\hat{\alpha}$ is multiplied by N in order to derive its limit. This is because the dimension of $\hat{\alpha}$ grows with the number of predictors. As N grows, the predictive information in $\boldsymbol{f}$ is spread across a larger number of predictors so each predictor's contribution approaches zero. Standardizing by N is necessary to identify the non-degenerate limit.

What distinguishes these results from previous work using PCR is the fact that the 3PRF uses only as many predictive factors as the number of factors relevant to y_{t+1}. In contrast, the PCR forecast is asymptotically efficient when there are as many predictive factors as the total number of factors driving $\boldsymbol{x}_t$ ((Stock and Watson, 2002a)). This distinction is especially important when the number of relevant factors is strictly less than the number of total factors in the predictor data and the target-relevant principal components are dominated by other components in $\boldsymbol{x}_t$. In particular, if the factors driving the target are weak in the sense that they contribute a only small fraction of the total variability in the predictors, then principal components may have difficulty identifying them. Said another way, there is no sense in which the method of principal components is assured to first extract predictive factors that are relevant to y_{t+1}. This point has in part motivated recent econometric work on thresholding ((Bai and Ng, 2008)) and shrinking ((Stock and Watson, 2012)) principal components for the purposes of forecasting.

On the other hand, the 3PRF identifies exactly those relevant factors in its second pass factor estimation. This step extracts leading indicators-estimated factors that are specifically valuable for forecasting a given target. To illustrate how this works, consider the special case in which there is only one relevant factor, and the sole proxy is the target variable y_{t+1} itself. We refer to this case as the target-proxy three-pass regression filter. The following corollary is immediate from Theorem 1

Corollary 1 *Let Assumptions 1-5 hold with the exception of Assumptions 2.4, 3.3 and 3.4. Additionally, assume that there is only one relevant factor. Then the target-proxy three-pass regression filter forecaster is consistent for the infeasible best forecast.*

Proof: It follows directly from previous result by noting that the loading of $\boldsymbol{y}$ on $\boldsymbol{F}$ is $\boldsymbol{\beta} = (\beta_1, \mathbf{0}')'$ with $\beta_1 \neq 0$. Therefore the target satisfies the condition of Assumption 6. □

Corollary 1 holds regardless of the number of irrelevant factors driving $\boldsymbol{X}$ and regardless of where the relevant factor stands in the principal component ordering for $\boldsymbol{X}$. Compare this to PCR, whose first predictive factor is ensured to be the one that explains most of the covariance among $\boldsymbol{x}_t$, regardless of that factor's relationship to y_{t+1}. Only if the relevant factor happens to also drive most of the variation within the predictors does the first component achieve the infeasible best. It is in this sense that the forecast performance of the 3PRF is robust to the presence of irrelevant factors.

B.5 Asymptotic distributions of each pass

Not only is the 3PRF consistent for the infeasible best forecast, each forecast has a normal asymptotic distribution. We first derive the asymptotic distribution for

$\hat{\alpha}$ since this is useful for establishing the asymptotic distribution of forecasts.

Theorem 3 *Under Assumptions 1-6, as $N, T \to \infty$ we have*

$$\frac{\sqrt{T}N\left(\hat{\alpha}_i - \tilde{\alpha}_i\right)}{A_i} \xrightarrow{d} \mathcal{N}(0,1)$$

where A_i^2 is the ith diagonal element of

$$\widehat{A\,\text{var}}(\hat{\boldsymbol{\alpha}}) = \boldsymbol{\Omega}_\alpha \left(\tfrac{1}{T}\sum_t \hat{\eta}_{t+1}^2 \left(\boldsymbol{X}_t - \overline{\boldsymbol{X}}\right)\left(\boldsymbol{X}_t - \overline{\boldsymbol{X}}\right)'\right)\boldsymbol{\Omega}_\alpha',$$

$\hat{\eta}_{t+1}$ is the estimated 3PRF forecast error, $\tilde{\alpha}_i \equiv \boldsymbol{S}_i \boldsymbol{G}_\alpha \beta$, where $\boldsymbol{S}_i$ is selects the ith element of vector $\boldsymbol{G}_\alpha \boldsymbol{\beta}$ and

$$\boldsymbol{G}_\alpha = \boldsymbol{J}_N \left(T^{-1}\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}\right)\left(T^{-3}N^{-2}\boldsymbol{W'}_{XZ}\boldsymbol{S}_{XX}\boldsymbol{W}_{XZ}\right)^{-1}\left(N^{-1}T^{-2}\boldsymbol{W'}_{XZ}\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{F}\right)$$

and

$$\boldsymbol{\Omega}_\alpha = \boldsymbol{J}_N\left(\frac{1}{T}\boldsymbol{S}_{XZ}\right)\left(\frac{1}{T^3N^2}\boldsymbol{W'}_{XZ}\boldsymbol{S}_{XX}\boldsymbol{W}_{XZ}\right)^{-1}\left(\frac{1}{TN}\boldsymbol{W'}_{XZ}\right)$$

Proof. The lemmas used here are as follow. They finds the probability limit of the predictors' "residuals" that are unexplained by the factor estimator $\hat{\boldsymbol{F}}$ in the limit. Notice that $\hat{\boldsymbol{\varepsilon}}$ is consistent for the true idiosyncratic errors (for which cross-sectional dependence is limited by Assumption 3) and a linear combination of the irrelevant factors g which can be pervasive across predictors. This fact complicates the construction of a consistent estimator for the asymptotic variance of $\hat{\boldsymbol{F}}_t$.

Lemma 5 *Define $\hat{\varepsilon} = \boldsymbol{X} - \imath\hat{\boldsymbol{\phi}}_0 - \hat{\boldsymbol{F}}\hat{\boldsymbol{\Phi}}'$, where $\hat{\boldsymbol{\phi}}_0 = T^{-1}\sum_t \boldsymbol{x}_t - \hat{\boldsymbol{\Phi}}\left(T^{-1}\sum_t \hat{\boldsymbol{F}}_t\right)$. Under Assumptions 1-6, $\hat{\boldsymbol{F}}\hat{\Phi}'\ T, N \xrightarrow{p} \infty\ \boldsymbol{f}\Phi_f'$ and $\hat{\varepsilon} \xrightarrow[T,N\to\infty]{p} \varepsilon + g\Phi_g'$.*

The following lemma establishes the asymptotic independence of $\hat{\boldsymbol{F}}_t$ and η_{t+1}, which is used to find the asymptotic distribution of $\hat{\alpha}$.

Lemma 6 *Under Assumptions 1-4,* $\text{plim}_{N,T\to\infty} T^{-1}\sum_t \hat{\boldsymbol{F}}_t\eta_{t+1} = 0$ *for all h.*

Lemma 7 *Under Assumptions 1-4, as* $N, T \to \infty$ *we have*

$$N^{-1}T^{-3/2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\eta \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Lambda}\boldsymbol{\Delta}_{\mathbf{F}}\mathcal{P}\boldsymbol{\Gamma}_{F\eta}\mathcal{P}\boldsymbol{\Delta}_{\mathrm{F}}\boldsymbol{\Lambda}'\right)$$

Until here, the auxiliary lemmas are enough to proof Theorem2.

Given the definition of $\tilde{\alpha}_i$, note that

$$N\hat{\alpha}_i - N\tilde{\alpha}_i \stackrel{d}{=} \boldsymbol{S}_iT^{-1}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}\left(T^{-3}N^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}\right)^{-1}$$
$$\times T^{-2}N^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\eta$$

The asymptotic distribution and consistent variance estimator follow directly from Lemma 7 and previously derived limits, Assumptions 5 and 6, and noting that $\hat{\eta}_{t+1} = \eta_{t+1} + o_p(1)$ by Theorem 1 ■

While Theorem 2 demonstrates that $\hat{\alpha}$ may be used to measure the relative forecast contribution of each predictor, Theorem 3 offers a distribution theory, including feasible t-statistics, for inference. The $\boldsymbol{G}_\alpha$ matrix appears here because the factors are only identified up to an orthonormal rotation.

From here, we derive the asymptotic distribution of the 3PRF forecasts.

Theorem 4 *Theorem 4. Under Assumptions 1-6, as* $N, T \to \infty$ *we have*

$$\frac{\sqrt{T}\left(\hat{y}_{t+1} - \mathbb{E}_t y_{t+1}\right)}{Q_t} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\mathbb{E}_t y_{t+1} = \beta_0 + \beta' \boldsymbol{F}_t$ *and* Q_t^2 *is the th diagonal element of* $\frac{1}{N^2}\boldsymbol{J}_T\boldsymbol{X}\widehat{\text{Avar}}(\hat{\boldsymbol{\alpha}})\boldsymbol{X}'\boldsymbol{J}_T$.

Proof: The result follows directly from Theorems 2 and 3 . Note that the theorem may be restated replacing $\tilde{y}_{t+1}$ with $\mathbb{E}_t y_{t+1}$ since the argument leading up to Theorem 1 implies that $\sqrt{T}\tilde{y}_{t+1} \xrightarrow[T,N\to\infty]{p} \mathbb{E}_t y_{t+1}$. By Slutsky's theorem convergence in distribution follows, yielding the theorem statement in the paper's text. ■

This result shows that besides being consistent for the infeasible best forecast $\mathbb{E}_t\left(y_{t+1}\right) \equiv \beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t$, the 3PRF forecast is asymptotically normal and provides a standard error estimator for constructing forecast confidence intervals. A subtle but interesting feature of this result is that we only need the asymptotic variance of individual predictor loadings $\widehat{Avar}(\hat{\alpha})$ for the prediction intervals. This differs from the confidence intervals of PCR forecasts in Bai and Ng (2006), which require an estimate of the asymptotic variance for the predictive factor loadings (the analogue of our $\widehat{Avar}(\hat{\boldsymbol{\beta}})$ below) as well as an estimate for the asymptotic variance of the fitted latent factors, $\widehat{Avar}(\hat{\boldsymbol{F}})$. Unlike PCR, our framework allows us to represent loadings on individual predictors in a convenient algebraic form, $\hat{\alpha}$. Inspection of $\hat{\alpha}$ reveals why variability in both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{F}}$ is captured by $\widehat{\text{Avar}}(\hat{\alpha})$.

Next, we provide the asymptotic distribution of predictive loadings on the latent factors and a consistent estimator of their asymptotic covariance matrix.

Theorem 5 *Under Assumptions 1-6, as* $N, T \to \infty$ *we have*

$$\sqrt{T}\left(\hat{\beta} - G_\beta \beta\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma_\beta\right)$$

where $\Sigma_\beta = \Sigma_z^{-1}\Gamma_{F\eta}\Sigma_z^{-1}$ *and* $\Sigma_z = \Lambda\Delta_F\Lambda' + \Delta_\omega$. *Furthermore,*

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}) = \left(T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\hat{\boldsymbol{F}}\right)^{-1} T^{-1}\sum_t \hat{\eta}_{t+1}^2 \left(\hat{\boldsymbol{F}}_t - \hat{\boldsymbol{\mu}}\right)\left(\hat{\boldsymbol{F}}_t - \hat{\boldsymbol{\mu}}\right)' \times \left(T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\hat{\boldsymbol{F}}\right)^{-1}$$

is a consistent estimator of $\boldsymbol{\Sigma}_\beta$.$\boldsymbol{G}_\beta$ *is defined in the proof part.*

Proof: Define $\boldsymbol{G}_\beta = \hat{\boldsymbol{\beta}}_1^{-1}\hat{\boldsymbol{\beta}}_2\hat{\boldsymbol{\beta}}_3^{-1}\left(N^{-1}T^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{F}\right)$. The asymptotic distribution follows directly from Lemma 7 noting that

$$\hat{\boldsymbol{\beta}} - G_\beta\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_1^{-1}\hat{\boldsymbol{\beta}}_2\hat{\boldsymbol{\beta}}_3^{-1}\left(N^{-1}T^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\eta\right).$$

The asymptotic covariance matrix (before employing Assumptions 5 and 6) is $\boldsymbol{\Sigma}_\beta = \boldsymbol{\Psi}_\beta\boldsymbol{\Gamma}_{\mathrm{F}\eta}\boldsymbol{\Psi}_\beta'$, where $\boldsymbol{\Psi}_\beta = \boldsymbol{\Sigma}_z^{-1}\Lambda\boldsymbol{\Delta}_{\mathrm{F}}\mathcal{P}\Delta_F\boldsymbol{\Lambda}'\left(\boldsymbol{\Lambda}\boldsymbol{\Delta}_F\mathcal{P}\boldsymbol{\Delta}_F\mathcal{P}\boldsymbol{\Delta}_F\boldsymbol{\Lambda}'\right)^{-1}\boldsymbol{\Lambda}\boldsymbol{\Delta}_{\mathrm{F}}\mathcal{P}$. This expression follows from Lemma 7 and the probability limits derived in the proof of Lemma 3. Assumptions 5 and 6 together with the derivation in the proof of Theorem 1 reduces Σ_β to the stated form.

To show consistency of $\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}})$, note that $\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{G}_\beta\beta\right) = \left(T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\hat{\boldsymbol{F}}\right)^{-1}T^{-1/2}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\eta$, which implies that the asymptotic variance of $\hat{\beta}$ is equal to the probability limit of

$$\left(T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\hat{\boldsymbol{F}}\right)^{-1}T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\eta\eta'\boldsymbol{J}_T\hat{\boldsymbol{F}}\left(T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\hat{\boldsymbol{F}}\right)^{-1} \tag{B.10}$$

Assumption 2.5 and Lemma 6 imply that $\text{plim}_{T,N\to\infty}T^{-1}\hat{\boldsymbol{F}}'\boldsymbol{J}_T\eta\eta'\boldsymbol{J}_T\hat{\boldsymbol{F}} = \text{plim}_{T,N\to\infty}T^{-1}\sum_t\eta_{t+1}^2\left(\hat{\boldsymbol{F}}_t - \hat{\boldsymbol{\mu}}\right)\left(\hat{\boldsymbol{F}}_t - \hat{\boldsymbol{\mu}}\right)'$. By Theorem 1, $\eta_{t+1} = \hat{\eta}_{t+1} + o_p(1)$, which implies that $\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}})$ and Eq.B.10 share the same probability limit, therefore $\widehat{A\,\text{var}}(\hat{\boldsymbol{\beta}})$ is a consistent estimator of Σ_β. ■

The asymptotic distribution of the estimated relevant latent factor rotation is described in Theorem 6.

Theorem 6 *Under Assumptions1-6, as $N, T \to \infty$ we have for every t*

1. *if $\sqrt{N}/T \to 0$, then*

$$\sqrt{N}\left[\hat{\boldsymbol{F}}_t - (\boldsymbol{H}_0 + \boldsymbol{H}\boldsymbol{F}_t)\right] \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_F\right)$$

2. *if $\liminf \sqrt{N}/T \geq \tau \geq 0$, then*

$$T\left[\hat{\boldsymbol{F}}_t - (\boldsymbol{H}_0 + \boldsymbol{H}\boldsymbol{F}_t)\right] = \boldsymbol{O}_p(1)$$

where $\Sigma_F = (\Lambda\Delta_F\Lambda' + \Delta_\omega)\left(\Lambda\Delta_F^2\Lambda'\right)^{-1}\Lambda\Delta_F\Gamma_{\Phi\varepsilon}\Delta_F\Lambda'\left(\Lambda\Delta_F^2\boldsymbol{\Lambda}'\right)^{-1}(\boldsymbol{\Lambda}\boldsymbol{\Delta}_F\boldsymbol{\Lambda}' + \boldsymbol{\Delta}_\omega)$ *.$\boldsymbol{H}_0$ and $\boldsymbol{H}$ are defined in the proof part.*

Proof. The lemma used here is lemma 8

Lemma 8 *Under Assumptions 1–4, as $N, T \to \infty$ we have*

1. *if $\sqrt{N}/T \to 0$, then for every t*

$$N^{-1/2}T^{-1}Z'J_TXJ_N\varepsilon_t \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Lambda\Delta_F\Gamma_{\Phi\varepsilon}\Delta_F\Lambda'\right)$$

2. *if* $\liminf \sqrt{N}/T \geq \tau \geq 0$, *then*

$$N^{-1}Z'J_TXJ_N\varepsilon_t = \boldsymbol{O}_p(1)$$

Define

$$\begin{aligned}\boldsymbol{H}_0 &= \hat{\boldsymbol{F}}_A\hat{\boldsymbol{F}}_B^{-1}N^{-1}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\phi_0 \quad \text{and} \\ \boldsymbol{H} &= \hat{\boldsymbol{F}}_A\hat{\boldsymbol{F}}_B^{-1}N^{-1}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\mathbf{X}\boldsymbol{J}_N\boldsymbol{\Phi}\end{aligned} \tag{B.11}$$

The result of Theorem 6 follows directly from Lemma 8, noting that $\hat{\boldsymbol{F}}_t - (\boldsymbol{H}_0 + \boldsymbol{H}\boldsymbol{F}_t) = \hat{\boldsymbol{F}}_A\hat{\boldsymbol{F}}_B^{-1}N^{-1}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\varepsilon_t$. The asymptotic covariance matrix Σ_F is found from Lemma 8, the probability limits derived in the proof of Lemma 2, and by Assumption 5 (which sets $\mathscr{P} = I$). ■

The matrices $\boldsymbol{G}_\beta$ and $\boldsymbol{H}$ are present since we are in effect estimating a vector space. Quoting Bai and Ng (2006), Theorems 5 and 6 in fact "pertain to the difference between $\left[\hat{\boldsymbol{F}}_t \mid \hat{\boldsymbol{\beta}}\right]$ and the space spanned by $[\boldsymbol{F}_t/\boldsymbol{\beta}]$ ". Note that we do not provide an estimator the asymptotic variance of $\hat{\boldsymbol{F}}$. While under some circumstances such an estimator is available, this is not generally the case. In particular, when there exist irrelevant factors driving the predictors, the 3PRF only estimates the relevant factor subspace. This complicates the construction of a consistent estimator of $\text{Avar}(\hat{\boldsymbol{F}})$. Estimators for the asymptotic variance of $\hat{\alpha}, \hat{\boldsymbol{\beta}}$ and $\hat{y}_{t+1}$ do not confront this difficulty for reasons discussed following Theorem 4.

Appendix C

Chapter 4 Appendix

C.1 Algorithm for Posterior Inference

In order to do the posterior inference of the model, it is useful to introduce a set of latent variables $z = [z_1, ..., z_k, ..., z_{Tp}]'$. The kth elements of z is equal to 1 if the coefficient of the corresponding kth regressor is non-zero which means it's included in the model. Let us denote by $Y = [y_1, \dots, y_T]'$, $U = [u_1, \dots, u_T]'$, $\mathcal{Z}$ is shown in equation (8), and $\beta^{\Delta} = \{\beta_i, i = 1, ..., Tp\}$. For simplicity of expression, we use notation β instead of β^{Δ} in the following process. T is the number of observations. The posterior of the unknown objects of the model is given by

$$\begin{aligned}
p\left(\phi, \beta, \sigma^2, R^2, z, q \mid Y, \mathcal{Z}\right) &\propto p\left(Y \mid \mathcal{Z}, \phi, \beta, \sigma^2, R^2, z, q\right) \cdot p\left(\phi, \beta, \sigma^2, R^2, z, q\right) \\
&\propto p\left(Y \mid \mathcal{Z}, \phi, \beta, \sigma^2\right) \cdot p\left(\beta \mid \sigma^2, R^2, z, q\right) \cdot p\left(z \mid q, \sigma^2, R^2\right) \cdot p(q) \cdot p\left(\sigma^2\right) \cdot p\left(R^2\right) \\
&\propto \left(\tfrac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\sigma^2}(Y-U\phi-\mathcal{Z}\beta)'(Y-U\phi-\mathcal{Z}\beta)} \\
&\cdot \textstyle\prod_{i=1}^{Tp} \left[\left(\frac{1}{2\pi\sigma^2\gamma^2}\right)^{\frac{1}{2}} e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \left[\delta\left(\beta_i\right)\right]^{1-z_i} \\
&\cdot \textstyle\prod_{i=1}^{Tp} q^{z_i}(1-q)^{1-z_i} \\
&\cdot q^{a-1}(1-q)^{b-1} \\
&\cdot \left(\tfrac{1}{\sigma^2}\right) \\
&\cdot \left(R^2\right)^{A-1}\left(1-R^2\right)^{B-1}
\end{aligned} \tag{C.1}$$

where $\gamma^2 = \frac{1}{Tp\bar{v}_x q} \cdot \frac{R^2}{1-R^2}$, and $\delta(\cdot)$ is the Dirac-delta function, which is used to make sure the probability density function of β is integrated to 1.

We can sample from the posterior of $(\phi, \beta, \sigma^2, R^2, z, q)$ using a Gibbs sampling algorithm with blocks (i) R^2 and q, (ii) ϕ, and (iii) (z, β, σ^2).

- The conditional posterior of R^2 and q is given by

$$\begin{aligned}
&p\left(R^2, q \mid Y, \phi, \beta, \sigma^2, z\right) \propto \prod_{i=1}^{Tp} \left[\left(\frac{1}{2\pi\sigma^2\gamma^2}\right)^{\frac{1}{2}} e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \left[\delta\left(\beta_i\right)\right]^{1-z_i} \\
&\cdot \prod_{i=1}^{Tp} q^{z_i}(1-q)^{1-z_i} \cdot q^{a-1}(1-q)^{b-1} \cdot \left(R^2\right)^{A-1}\left(1-R^2\right)^{B-1} \\
&\propto \prod_{i=1}^{Tp} \left[\frac{(Tpv_xq)(1-R^2)}{R^2}\right]^{\frac{z_i}{2}} \left[e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \\
&\cdot q^{\tau(z)+a-1}(1-q)^{Tp-\tau(z)+b-1} \cdot \left(R^2\right)^{A-1}\left(1-R^2\right)^{B-1} \\
&\propto \prod_{i=1}^{Tp} \left[\frac{q(1-R^2)}{R^2}\right]^{\frac{z_i}{2}} \left[e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \\
&\cdot q^{\tau(z)+a-1}(1-q)^{Tp-\tau(z)+b-1} \cdot \left(R^2\right)^{A-1}\left(1-R^2\right)^{B-1} \\
&\propto \prod_{i=1}^{Tp} \left[e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \\
&\cdot q^{\tau(z)+\frac{\tau(z)}{2}+a-1}(1-q)^{Tp-\tau(z)+b-1} \cdot \left(R^2\right)^{A-1-\frac{\tau(z)}{2}}\left(1-R^2\right)^{\frac{\tau(z)}{2}+B-1} \\
&\propto \left[e^{-\frac{1}{2\sigma^2}\frac{Tp\bar{v}_xq\left(1-R^2\right)}{R^2}} \beta' \operatorname{diag}(z)\beta\right] \\
&\cdot q^{\tau(z)+\frac{\tau(z)}{2}+a-1}(1-q)^{Tp-\tau(z)+b-1} \cdot \left(R^2\right)^{A-1-\frac{\tau(z)}{2}}\left(1-R^2\right)^{\frac{\tau(z)}{2}+B-1}
\end{aligned} \tag{C.2}$$

where $\tau(z) \equiv \sum_{i=1}^{Tp} z_i$. To sampling from this continuous distribution, the discretization process is used to the [0,1] support of R^2 and q. More specifically, for both R^2 and q we define a grid with increments of 0.01, and finer increments of 0.001 near the boundaries of the support.

- The conditional posterior of ϕ is given by

$$\begin{aligned}
&p\left(\phi \mid Y, z, \beta, R^2, q, \sigma\right) \propto \left(\tfrac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} exp(-\tfrac{1}{2\sigma^2}(Y-U\phi-X\beta)'(Y-U\phi-X\beta)) \\
&\propto exp(-\tfrac{1}{2\sigma^2}(Y-U\phi-X\beta)'(Y-U\phi-X\beta)) \\
&\propto exp(-\tfrac{1}{2\sigma^2}\tfrac{(Y-U\phi-X\beta)'((U'U)^{-1}U')'(U'U)^{-1}U'(Y-U\phi-X\beta)}{(U'U)^{-1}U'((U'U)^{-1}U')'}) \\
&\propto ((2\pi)^l|\sigma^2\left(U'U\right)^{-1}|)^{-\frac{1}{2}} \\
&\cdot exp(-\tfrac{1}{2\sigma^2}\tfrac{(Y-U\phi-X\beta)'((U'U)^{-1}U')'(U'U)^{-1}U'(Y-U\phi-X\beta)}{(U'U)^{-1}}) \\
&\propto ((2\pi)^l|\sigma^2\left(U'U\right)^{-1}|)^{-\frac{1}{2}} \\
&\cdot exp(-\tfrac{1}{2\sigma^2}\tfrac{(\phi-(U'U)^{-1}U(Y-X\beta))\prime(\phi-(U'U)^{-1}U(Y-X\beta))}{(U'U)^{-1}})
\end{aligned} \tag{C.3}$$

which implies

$$\phi \mid Y, z, \beta, \gamma, q, \sigma \sim N\left(\left(U'U\right)^{-1}U'(Y-X\beta), \sigma^2\left(U'U\right)^{-1}\right) \tag{C.4}$$

- To draw from the posterior of $z, \beta, \sigma^2 \mid Y, \phi, R^2, q$, since $p(z, \beta, \sigma^2 \mid Y, \phi, R^2, q) = p\left(\beta, \sigma^2 \mid Y, \phi, R^2, q, z\right) \cdot p\left(z \mid Y, \phi, R^2, q\right) = p(\tilde{\beta} \mid Y, \phi, \sigma^2, R^2, q, z) \cdot p\left(\sigma^2 \mid Y, \phi, R^2, q, z\right) \cdot p\left(z \mid Y, \phi, R^2, q\right)$, we will first draw from $p\left(z \mid Y, \phi, R^2, q\right)$, and $p\left(\sigma^2 \mid Y, \phi, R^2, q, z\right)$, then $p\left(\beta \mid Y, \phi, R^2, q, z, \sigma^2\right)$. Then z, β, σ^2 are valid draws from $p\left(z, \beta, \sigma^2 \mid Y, \phi, R^2, q\right)$. To draw from the posterior $z \mid Y, \phi, R^2, q$, observe that

$$
\begin{aligned}
& p\left(z \mid Y, \phi, R^2, q\right)=\int p\left(z, \beta, \sigma^2 \mid Y, \phi, R^2, q\right) d\left(\beta, \sigma^2\right) \\
& \propto \int\left(\frac{1}{2 \pi \sigma^2}\right)^{\frac{T}{2}} exp(-\frac{1}{2 \sigma^2}(Y-U \phi-X \beta)^{\prime}(Y-U \phi-X \beta)) \\
& \cdot \prod_{i=1}^{Tp}\left[\left(\frac{1}{2 \pi \sigma^2 \gamma^2}\right)^{\frac{1}{2}} exp(-\frac{\beta_i^2}{2 \sigma^2 \gamma^2})\right]^{z_i}\left[\delta\left(\beta_i\right)\right]^{1-z_i} \\
& \cdot \prod_{i=1}^{Tp} q^{z_i}(1-q)^{1-z_i} \\
& \cdot\left(\frac{1}{\sigma^2}\right) d\left(\beta, \sigma^2\right) \\
& \propto q^{\tau(z)}(1-q)^{Tp-\tau(z)}\left(\frac{1}{2 \pi \gamma^2}\right)^{\frac{\tau(z)}{2}} \int\left(\frac{1}{\sigma^2}\right)^{\frac{T+\tau(z)}{2}+1} \\
& \cdot exp(-\frac{1}{2 \sigma^2}\left[(Y-U \phi-\tilde{X} \tilde{\beta})^{\prime}(Y-U \phi-\tilde{X} \tilde{\beta})+\frac{\tilde{\beta}^{\prime} \tilde{\beta}}{\gamma^2}\right]) d\left(\tilde{\beta}, \sigma^2\right) \\
& \propto q^{\tau(z)}(1-q)^{Tp-\tau(z)}\left(\frac{1}{2 \pi \gamma^2}\right)^{\frac{\tau(z)}{2}} \int\left(\frac{1}{\sigma^2}\right)^{\frac{T+\tau(z)}{2}+1} \\
& \cdot exp(-\frac{1}{2 \sigma^2}[(Y-U \phi)^{\prime}(Y-U \phi)-(Y-U \phi)^{\prime} \tilde{X} \tilde{\beta}-\tilde{\beta}^{\prime} \tilde{X}^{\prime}(Y-U \phi)+\tilde{\beta}^{\prime} \tilde{X}^{\prime} \tilde{X} \tilde{\beta}+\frac{\tilde{\beta}^{\prime} \tilde{\beta}}{\gamma^2}]) \\
& \cdot d\left(\tilde{\beta}, \sigma^2\right) \\
& \propto q^{\tau(z)}(1-q)^{Tp-\tau(z)}\left(\frac{1}{2 \pi \gamma^2}\right)^{\frac{\tau(z)}{2}} \int\left(\frac{1}{\sigma^2}\right)^{\frac{T+\tau(z)}{2}+1} \\
& \cdot exp(-\frac{1}{2 \sigma^2}[\tilde{Y}^{\prime} \tilde{Y}-\tilde{Y}^{\prime} \tilde{X} \tilde{\beta}-\tilde{\beta}^{\prime} \tilde{X}^{\prime} \tilde{Y}+\tilde{\beta}^{\prime} \tilde{W} \tilde{\beta}]) d\left(\tilde{\beta}, \sigma^2\right)
\end{aligned} \tag{C.5}
$$

where $\tilde{W} = \left(\tilde{X}'\tilde{X} + I_{\tau(z)}/\gamma^2\right)$. Since $\tilde{W}$ is a symmetric matrix, we can denote it as $\tilde{W} = C'C$, then the exponent part of the equation (21) is

$$\begin{aligned}
&\tilde{Y}'\tilde{Y} - 2\tilde{\beta}'\tilde{X}'\tilde{Y} + \tilde{\beta}'\tilde{W}\tilde{\beta} \\
&= \tilde{Y}'\tilde{Y} - 2\tilde{\beta}'C'C'^{-1}\tilde{X}'\tilde{Y} + \tilde{\beta}'C'C\tilde{\beta} \\
&= \tilde{Y}'\tilde{Y} - 2(C\tilde{\beta})'C'^{-1}\tilde{X}'\tilde{Y} + (C\tilde{\beta})'C\tilde{\beta} \\
&= \tilde{Y}'\tilde{Y} - 2(C\tilde{\beta})'C'^{-1}\tilde{X}'\tilde{Y} + (C\tilde{\beta})'C\tilde{\beta} \\
&\quad +(C'^{-1}\tilde{X}'\tilde{Y})'C'^{-1}\tilde{X}'\tilde{Y} - (C'^{-1}\tilde{X}'\tilde{Y})'C'^{-1}\tilde{X}'\tilde{Y} \\
&= \tilde{Y}'\tilde{Y} + (C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y})'(C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y}) - (C'^{-1}\tilde{X}'\tilde{Y})'C'^{-1}\tilde{X}'\tilde{Y} \\
&= \tilde{Y}'\tilde{Y} + (C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y})'(C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y}) - \tilde{Y}'\tilde{X}C^{-1}C'^{-1}\tilde{X}'\tilde{Y} \\
&= \tilde{Y}'\tilde{Y} + (C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y})'(C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y}) - \tilde{Y}'\tilde{X}\tilde{W}^{-1}\tilde{X}'\tilde{Y} \\
&= \tilde{Y}'\tilde{Y} + (C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y})'(C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y}) - \tilde{Y}'\tilde{X}\tilde{W}^{-1}\tilde{W}\tilde{W}^{-1}\tilde{X}'\tilde{Y} \\
&= \tilde{Y}'\tilde{Y} + (C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y})'(C\tilde{\beta} - C'^{-1}\tilde{X}'\tilde{Y}) - \hat{\tilde{\beta}}'\tilde{W}\hat{\tilde{\beta}}
\end{aligned} \tag{C.6}$$

where $\hat{\tilde{\beta}} = \tilde{W}^{-1}\tilde{X}'\tilde{Y}$. If we write $\beta^* = C\tilde{\beta}$, notice that

$$\begin{aligned}
&\int exp(-\tfrac{1}{2\sigma^2}[\left(\beta^* - C'^{-1}\tilde{X}'\tilde{Y}\right)'\left(\beta^* - C'^{-1}\tilde{X}'\tilde{Y}\right)])d\tilde{\beta} \\
&= \int exp(-\tfrac{1}{2\sigma^2}[\left(\beta^* - C'^{-1}\tilde{X}'\tilde{Y}\right)'\left(\beta^* - C'^{-1}\tilde{X}'\tilde{Y}\right)])dC^{-1}\beta^* \\
&= |C^{-1}| \int (\tfrac{1}{\sqrt{2\pi\sigma^2}})^{\tau(z)}(\sqrt{2\pi\sigma^2})^{\tau(z)} \\
&\quad \cdot exp(-\tfrac{1}{2\sigma^2}[\left(\beta^* - C'^{-1}\tilde{X}'\tilde{Y}\right)'\left(\beta^* - C'^{-1}\tilde{X}'\tilde{Y}\right)])d\beta^* \\
&= |C^{-1}|(\sqrt{2\pi\sigma^2})^{\tau(z)}(\sqrt{2\pi\sigma^2})^{\tau(z)} \\
&= |\tilde{W}|^{-\frac{1}{2}}(2\pi)^{\frac{\tau(z)}{2}}(\sigma^2)^{\frac{\tau(z)}{2}}
\end{aligned} \tag{C.7}$$

Hence the posterior of $z \mid Y, \phi, R^2, q$ is

$$
\begin{aligned}
& p\left(z \mid Y, \phi, R^2, q\right) \\
\propto\ & q^{\tau(z)}(1-q)^{Tp-\tau(z)}\left(\frac{1}{2\pi\gamma^2}\right)^{\frac{\tau(z)}{2}}(2\pi)^{\frac{\tau(z)}{2}}|\tilde{W}|^{-\frac{1}{2}}\int\left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\tilde{W}\hat{\tilde{\beta}}\right]}d\sigma^2 \\
\propto\ & q^{\tau(z)}(1-q)^{Tp-\tau(z)}\left(\frac{1}{\gamma^2}\right)^{\frac{\gamma(z)}{2}}|\tilde{W}|^{-\frac{1}{2}}\int\left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\tilde{W}\hat{\tilde{\beta}}\right]}d\sigma^2
\end{aligned}
\tag{C.8}
$$

Denote $\frac{\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\tilde{W}\hat{\tilde{\beta}}}{2}$ as V, and the integral part can be written as

$$
\int\left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}e^{-\frac{V}{\sigma^2}}d\sigma^2 \tag{C.9}
$$

then denote $\frac{V}{\sigma^2}=t$, we have $dt=\frac{V}{\sigma^4}d\sigma^2$, hence $d\sigma^2=\frac{\sigma^4}{V}dt$

$$
\begin{aligned}
& \int\left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}e^{-\frac{V}{\sigma^2}}d\sigma^2 \\
& =\frac{1}{V}\int\left(\sigma^4\right)\left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}+1}e^{-t}d\sigma^2 \\
& =\frac{1}{V}\int\left(\frac{1}{\sigma^2}\right)^{\frac{T}{2}-1}e^{-t}d\sigma^2 \\
& =\frac{1}{V}V^{-(\frac{T}{2}-1)}\int\left(\frac{V}{\sigma^2}\right)^{\frac{T}{2}-1}e^{-t}d\sigma^2 \\
& =V^{-\frac{T}{2}}\int t^{\frac{T}{2}-1}e^{-t}d\sigma^2 \\
& =\left[\frac{\tilde{Y}'\tilde{Y}-\hat{\beta}'\tilde{W}\hat{\beta}}{2}\right]^{-\frac{T}{2}}\Gamma\left(\frac{T}{2}\right)
\end{aligned}
\tag{C.10}
$$

Therefore, the posterior is

$$
p\left(z \mid Y, \phi, R^2, q\right) \propto q^{\tau(z)}(1-q)^{Tp-\tau(z)}\left(\frac{1}{\gamma^2}\right)^{\frac{\tau(z)}{2}}|\tilde{W}|^{-\frac{1}{2}}\left[\frac{\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\tilde{W}\hat{\tilde{\beta}}}{2}\right]^{-\frac{T}{2}}\Gamma\left(\frac{T}{2}\right) \tag{C.11}
$$

Finally, to draw from the posterior of $\beta, \sigma^2 \mid Y, \phi, R^2, q, z$, observe that

$$\begin{aligned}
& p(\sigma^2 \mid Y, \phi, R^2, q, z) \\
&= \int p(\beta, \sigma^2 \mid Y, \phi, R^2, q, z) d\beta \\
&\propto \int \left(\tfrac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\sigma^2}(Y-U\phi-X\beta)'(Y-U\phi-X\beta)} \\
&\cdot \prod_{i=1}^{Tp} \left[\left(\tfrac{1}{2\pi\sigma^2\gamma^2}\right)^{\frac{1}{2}} e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \left[\delta\left(\beta_i\right)\right]^{1-z_i} \cdot \left(\tfrac{1}{\sigma^2}\right) d\left(\beta\right) \\
&\propto \left(\tfrac{1}{\sigma^2}\right)^{\frac{T+\tau(z)}{2}+1} \int e^{-\frac{1}{2\sigma^2}\left[(Y-U\phi-\tilde{X}\tilde{\beta})'(Y-U\phi-\tilde{X}\tilde{\beta})+\frac{\tilde{\beta}'\tilde{\beta}}{\gamma^2}\right]} d(\tilde{\beta}) \qquad \text{(C.12)} \\
&\propto \left(\tfrac{1}{\sigma^2}\right)^{\frac{T+\tau(z)}{2}+1} \int e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\tilde{Y}'\tilde{X}\tilde{\beta}-\tilde{\beta}'\tilde{X}'\tilde{Y}+\tilde{\beta}'\bar{W}\tilde{\beta}\right]} d(\tilde{\beta}) \\
&\propto \left(\tfrac{1}{\sigma^2}\right)^{\frac{T+\tau(z)}{2}+1} |\tilde{W}|^{-\frac{1}{2}} (2\pi)^{\frac{\tau(z)}{2}} \left(\sigma^2\right)^{\frac{\tau(z)}{2}} e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\bar{W}\hat{\tilde{\beta}}\right]} \\
&\propto \left(\tfrac{1}{\sigma^2}\right)^{\frac{T}{2}+1} e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\bar{W}\hat{\tilde{\beta}}\right]} \\
&\propto \left[\Gamma\left(\tfrac{T}{2}\right) \tfrac{2}{\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\tilde{W}\hat{\tilde{\beta}}}\right]^{-1} \left(\tfrac{1}{\sigma^2}\right)^{\frac{T}{2}+1} e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\hat{\tilde{\beta}}'\bar{W}\hat{\tilde{\beta}}\right]}
\end{aligned}$$

Therefore, the conditional posterior distribution of $\sigma^2 \mid Y, \phi, R^2, q, z)$ is

$$\sigma^2 \mid Y, \phi, R^2, q, z \sim IG\left(\frac{T}{2}, \frac{\tilde{Y}'\tilde{Y} - \hat{\tilde{\beta}}'\left(\tilde{X}'\tilde{X} + I_{\tau(z)}/\gamma^2\right)\hat{\tilde{\beta}}}{2}\right) \qquad \text{(C.13)}$$

Moreover,

$$
\begin{aligned}
& p(\tilde{\beta} \mid Y, \phi, \sigma^2, R^2, q, z) \\
\propto \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\sigma^2}(Y-U\phi-X\beta)'(Y-U\phi-X\beta)} \cdot \prod_{i=1}^{Tp} & \left[\left(\frac{1}{2\pi\sigma^2\gamma^2}\right)^{\frac{1}{2}} e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}}\right]^{z_i} \left[\delta\left(\beta_i\right)\right]^{1-z_i} \\
& \propto e^{-\frac{1}{2\sigma^2}\left[(Y-U\phi-\tilde{X}\tilde{\beta})'(Y-U\phi-\tilde{X}\tilde{\beta})+\frac{\tilde{\beta}'\tilde{\beta}}{\gamma^2}\right]} \\
& \propto e^{-\frac{1}{2\sigma^2}\left[\tilde{Y}'\tilde{Y}-\tilde{Y}'\tilde{X}\tilde{\beta}-\tilde{\beta}'\tilde{X}'\tilde{Y}+\tilde{\beta}'\tilde{W}\tilde{\beta}\right]} \\
& \propto e^{-\frac{1}{2\sigma^2}\left(C\tilde{\beta}-C'^{-1}\tilde{X}'\tilde{Y}\right)'\left(C\tilde{\beta}-C'^{-1}\tilde{X}'\tilde{Y}\right)} \\
& \propto e^{-\frac{1}{2\sigma^2}\frac{\left(C\tilde{\beta}-C'^{-1}\tilde{X}'\tilde{Y}\right)'C'^{-1}C^{-1}\left(C\tilde{\beta}-C'^{-1}\tilde{X}'\tilde{Y}\right)}{C^{-1}C'^{-1}}} \\
& \propto e^{-\frac{1}{2\sigma^2}\frac{\left(\tilde{\beta}-\tilde{W}'^{-1}\tilde{X}'\tilde{Y}\right)'\left(\tilde{\beta}-\tilde{W}'^{-1}\tilde{X}'\tilde{Y}\right)}{\tilde{W}^{-1}}} \\
& \propto e^{-\frac{1}{2\sigma^2}\frac{\left(\tilde{\beta}-\hat{\tilde{\beta}}\right)'\left(\tilde{\beta}-\hat{\tilde{\beta}}\right)}{\tilde{W}^{-1}}}
\end{aligned} \tag{C.14}
$$

Therefore,

$$
\tilde{\beta} \mid Y, \phi, \sigma^2, R^2, q, z \sim N\left(\hat{\tilde{\beta}}, \sigma^2\left(\tilde{X}'\tilde{X} + I_{\tau(z)}/\gamma^2\right)^{-1}\right) \tag{C.15}
$$

and the other β_i's are equal to 0.

C.2 Details of the Out-of-Sample Prediction Exercise

This appendix provides the details of the out-of-sample exercise presented in the main text. This exercise is designed as a standard forecasting exercise for applications with time series data.

The measures of forecasting accuracy reported in the main text are computed by averaging the log-predictive scores and the squared forecast errors over the elements of a test sample, and across all test samples.

We evaluate the prediction accuracy of the following baseline and restricted versions of our model: BMA-all, which is our full model that combines all the possible individual models, weighted by their posterior probability; BMA-5 and BMA-10, which restrict the model space to the combinations of individual models with up to five and ten predictors respectively, weighted by their relative posterior probability; and SS-k, which is the dense model including all the predictors. The predictive density of y_{T+1} implied by these models is a mixture of Gaussian densities with means $u'_{T+1}\phi^{(j)} + x'_{T+1}\beta(j)$ and variances $\sigma^{2(j)}$, where $\phi^{(j)}, \beta^{(j)}$ and $\sigma^{2(j)}, j = 1, \ldots, M$, are draws from their posterior distribution. The predictive score is computed as the value of this density at the actual realization of y_{T+1}.

To select the "best" individual models for each training sample, we employ three different sparse modeling strategies:

- **Spike-and-slab (SS).** Within our spike-and-slab framework, we select SS-5 and SS-10 as the individual models with the highest posterior probability in the set of those with up to five and ten predictors. To robustify the procedure, instead of simply counting the number of times an individual model is visited by the MCMC algorithm, we numerically compute the posterior model probability of all models that are visited at least once, and pick the model with the highest.[1] The predictive density of y_{T+1} implied by these

[1] If models with less than 5 or 10 predictors receive less than 0.05 percent of the total posterior weight, the progressively larger models are considered until we reach this lower bound. The only application where this is an issue is finance 2, where small models are essentially never visited.

models is a mixture of Gaussian densities with means $u'_{T+1}\phi^{(j)} + x'_{T+1}\beta(j)$ and variances $\sigma^{2(j)}$, where $\phi^{(j)}, \beta^{(j)}$ and $\sigma^{2(j)}, j = 1, \ldots, M$, are draws from their posterior distribution. We use the mean of the predictive density as the point forecast for the computation of the mean squared forecast error.

- **Lasso (L) and Post-lasso (PL).** As an alternative way to identify good-fitting individual small models, we also consider the popular lasso method (Tibshirani (1996)). We consider the following variants of this methodology. (i) L-5 and L-10: lasso with a fixed number of five and ten predictors; (ii) L-asy: lasso with a penalty parameter based on the asymptotic criterion proposed by Bickel et al. (2009), implemented using the iterative procedure and the tuning constants recommended by Belloni et al. (2011) (notice that this criterion is designed for valid inference, not necessarily best prediction); (iii) L-cv2, L-cv5 and L-cv10: lasso with selection of the number of predictors based on 2-, 5- and 10-fold cross validation. It is well known that constructing the full predictive density implied by lasso is challenging, and there is no agreement in the literature about how to tackle this problem(Hastie et al. (2015)). For this reason, we use two alternative rough approximations of the density of y_{T+1}.

  The first method consists of treating the lasso parameter estimates as known, and assuming Gaussian errors and a flat prior on their variance. Under these assumptions, the density of y_{T+1} is a non-centered Student-t distribution, with mean $u'_{T+1}\hat{\phi}_L + x'_{T+1}\hat{\beta}_L$, scale $\sqrt{\hat{r}_L/(T-2)}$ and degrees of freedom $T-2$, where $\hat{\phi}_L$, $\hat{\beta}_L$ and $\hat{r}_L$ are the lasso estimates of ϕ, β and the sum of squared residuals. As before, we use the mean of the predictive density $\left(u'_{T+1}\hat{\phi}_L + x'_{T+1}\hat{\beta}_L\right)$ as the point forecast for the computation of the mean squared forecast error.

An alternative method to construct the predictive density is based on post-selection inference. It consists of running a simple ordinary least squares regression of the response variable on the regressors selected by lasso (Belloni and Chernozhukov (2013)). This "post-lasso" procedure reduces the bias of the lasso estimator and may better approximate the solution of the best subset selection problem (Beale et al. (1967) and Hocking and Leslie (1967)). With Gaussian errors and a flat prior on the second-stage regression, the implied predictive density of y_{T+1} s a non-centered Student-t distribution, with mean $u'_{T+1}\hat{\phi}_{PL} + x'_{T+1}\hat{\beta}_{PL}$, scale $\sqrt{\left(\left[u'_{T+1}, x'_{T+1}\right]\left([U,X]'[U,X]\right)^{-1}\left[u'_{T+1}, x'_{T+1}\right]' + 1\right)\hat{r}_{PL}/(T-l-n-2)}$ and degrees of freedom $T-l-n-2$, where $\hat{\phi}_{PL}, \hat{\beta}_{PL}$ and $\hat{T}_{PL}$ are the ordinary least squares estimates of ϕ, β and the sum of squared residuals in the second-stage regression, and n is the dimension of the vector $\hat{\beta}_{PL}$. This post-selection approach allows us to incorporate parameter uncertainty in the predictive density, although the parameter estimates in the second stage are of course different from the lasso estimates. It is important to stress that this strategy is appropriate only under the stringent assumptions guaranteeing that model selection does not impact the asymptotic distribution of the parameters estimated in the post-selection step((Bühlmann and Van De Geer, 2011); see also (Leeb and Pötscher, 2005, 2008a,b) for a thorough discussion of the fragility of this approach, and Chernozhukov et al. (2015) for a comprehensive review of these topics). In the figures of the paper, we denote the log-predictive scores implied by this method as PL-5, PL-10, PL-asy, PL-cv2, PL-cv5 and PLcv10, depending on the lasso variant used in the selection stage. For completeness, we also report the mean squared forecast error based on post-lasso, using the mean of the predictive density $\left(u'_{T+1}\hat{\phi}_{PL} + x'_{T+1}\hat{\beta}_{PL}\right)$ as the point forecast.

- **Single best replacement (SBR)**. This class of methods (also known as forward stepwise) is a fast and scalable approximation of the solution of the best subset selection problem, and thus provides yet another way to choose good-fitting sparse individual models. We use the SBR computation algorithm of Soussen et al. (2011) and Polson and Sun (2019b), and consider the following variants of this method. (i) SBR-5 and SBR-10: SBR with a fixed number of five and ten predictors; (ii) SBRcv2, SBR-cv5 and SBR-cv10: SBR with selection of the number of predictors based on 2-, 5- and 10-fold cross validation. The predictive density and point forecast of y_{T+1} implied by these models are constructed as in the post-lasso case.

- **Test-based forward model selection (TBFMS).** As an alternative forward model selection procedure, we also experiment with the test-based method proposed by Kozbur (2020). This class of algorithms selects covariates of progressively larger-scale models and determines model size based on the outcome of statistical hypothesis tests. Following Kozbur (2020), we consider four versions of this method: (i) TBFMS-I, based on hypothesis tests for heteroskedastic disturbances; (ii) TBFMS-II, based on simplified hypothesis tests for heteroskedastic disturbances; (iii) TBFMS-III, based on fit-streamlined hypothesis tests for heteroskedastic disturbances; and (iv) TBFMS-IV, based on hypothesis tests for homoskedastic disturbances. The predictive density and point forecast of y_{T+1} by these models are constructed as in the post-lasso case.

C.3 Data Appendix

All series cover the period 1993M1 to 2021M8. All series are transformed to be approximately stationary. In particular, if $w_{i,t}$ is the original un-transformed series in levels, when the series is used as a predictor the transformation codes (column T of the table) are: 1 - no transformation (levels), $x_{i,t} = w_{i,t}$; 2 - first difference, $x_{i,t} = w_{i,t} - w_{i,t-1}$; 3- second difference, $x_{i,t} = \Delta w_{i,t} - \Delta w_{i,t-1}$ 4 - logarithm, $x_{i,t} = \log w_{i,t}$; 5 - first difference of logarithm, $x_{i,t} = \log w_{i,t} - \log w_{i,t-1}$; 6 - second difference of logarithm, $x_{i,t} = \Delta \log w_{i,t} - \Delta \log w_{i,t-1}$.

Table C.1: Variable List

No	Variable Name	Transform Code
1	MONEY SUPPLY M1	5
2	MONEY SUPPLY M2	5
3	MONEY SUPPLY M3	5
4	REAL EFFECTIVE EXCHANGE RATES	5

Table C.1 (continued)

5	EER_CPI	5
6	EER_LABOR COSTS	5
7	USD	5
8	YEN	5
9	Crude oil production	2
10	Petroleum production	5
11	Crude oil import	5
12	Import	5
13	Trade balance	2
14	ITS Exports	5
15	Export	5
16	CPI	6
17	CPI exc energy(MoM)	2
18	HICP: SERVICES (%MOM)	2
19	HICP: ENERGY (%MOM)	2
20	HICP: GOODS(%MOM)	2
21	HICP: FOOD & NON-ALCOHOLIC BEVERAGES (%MOM)	2
22	HICP: ALCOHOLIC BEVERAGES, TOBACCO & NARCOTICS (%MOM)	2
23	INDL PROD: INDUSTRY INCL CNSTR	6
24	INDL PROD: MANUFACTURING	6
25	INDUSTRIAL PRODUCTION - ENERGY	6
26	INDUSTRIAL PRODUCTION - INVESTMENT GOODS	6

Table C.1 (continued)

27	INDL. PROD.: MFG. - RUBBER & PLASTIC PRODUCTS	6
28	INDL. PROD.: MFG - WOOD,PAP	6
29	MANUFACTURE OF CHEMICALS AND CHEMICAL PRODUCTS	6
30	INDUSTRIAL PRODUCTION - ELECTRICAL EQUIPMENT	6
31	MANUFACTURE OF BASIC METALS AND METAL PRODUCTS	6
32	INDUSTRIAL PRODUCTION: MACH AND EQUIPMENT	6
33	MANUFACTURE OF TRANSPORT EQUIPMENT(MoM%)	2
34	MFG - PROD: FUTURE TENDENCY	2
35	CONSUMER CONFIDENCE INDEX	2
36	INDUSTRIAL CONFIDENCE INDICATOR	2
37	RETAIL CONFIDENCE INDICATOR	2
38	MFG - EMPLOYMENT: FUTURE TENDENCY	2
39	RETAIL TRADE - EMPLOYMENT: FUTURE TENDENCY	2
40	MFG - SELLING PRICES: FUTURE TENDENCY	2
41	RETAIL TRADE - BUSINESS SITUATION	2
42	CONSUMERS - EXPECTED ECONOMIC SITUATION	2
43	MFG - ORDER BOOKS	2
44	MFG - EXPORT ORDER BOOKS	2
45	RETAIL TRADE - VOLUME OF STOCKS	2
46	MFG - FINISHED GOODS STOCKS	2
47	CNSTR.: OVERALL - PRICE EXPECT	2
48	UNEMPLOYMENT: TOTAL - TOTAL% ACTIVE POP	2
49	NEW PASSENGER CAR REGISTRATIONS	5

Table C.1 (continued)

50	Composite Leading Indicator	2
51	UNEMPLOYMENT RATE: FEMALES	2
52	YIELD 10-YEAR GOVERNMENT BONDS	2
53	SHORT TERM EURO REPO RATE	2
54	INTERBANK RATES	2
55	STOXX EUROPE 50 - PRICE INDEX	5
56	STOXX EUROPE 600 E - PRICE INDEX	5
57	FTSE EUROTOP 100 E - PRICE INDEX	5
58	S&P 500 COMPOSITE - PRICE INDEX	5
59	US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ	5
60	US MONEY SUPPLY M2 CURA	5
61	US UNEMPLOYMENT RATE SADJ	2
62	US INDUSTRIAL PRODUCTION - TOTAL INDEX VOLA	6
63	US TOTAL CIVILIAN EMPLOYMENT VOLA	5
64	US CPI - ALL URBAN: ALL ITEMS SADJ	5
65	US CONSUMER CONFIDENCE INDEX SADJ	2
66	US TOTAL RETAIL TRADE (VOLUME) VOLA	2
67	US CONSTRUCTION EXPENDITURES - TOTAL (AR) CURA	5
68	US INDUSTRIAL PRODUCTION-DURABLE VOLA	6
69	US INDUSTRIAL PRODUCTION - NONDURABLE VOLA	6