Forecasting Time Series with a Mixture of Stationary and Non-stationary Factors

Sium Bodha Hannadige

B.Sc.(Hons), University of Peradeniya, Sri Lanka

A thesis submitted for the degree of Doctor of Philosophy at

Monash University in 2022

Department of Econometrics and Business Statistics

Contents

Copyright notice

List of Tables

List of Figures

Abstract

This thesis contributes to the literature by developing new methods for estimating and forecasting univariate time series, such as *GDP, GDP growth rate,* and *inflation* using a large number of variables in a set of high-dimensional panel data as potential predictors. To this end, we use a factor augmented regression [FAR] model that contains a small number of estimated factors as predictors; the factors are estimated using the aforementioned set of panel data. The validity of this forecasting method has been established when all the variables are stationary, I(0), and when they all non-stationary with unit roots, I(1), but not when they consist of a mixture of I(0) and I(1) variables. The central theme of this thesis is to advance the literature by extending the FAR method to: include a mixture of stationary and non-stationary factors and observed variables as predictors; allow time varying parameters; and allow two-level factors. Since the proposed model relaxes the three assumptions that underlie the FAR method, indications are that the new approaches explored in this thesis have the potential to improve over the corresponding ones in the literature. The approaches are developed with the underlying assumptions being progressively more general, and hence the methodology being progressively more challenging.

First, Chapter 2 in this thesis develops a method for constructing an asymptotically valid prediction interval using an FAR model when the set of predictors includes a mixture of I(0) and I(1) factors; we refer to this as *mixture-FAR* model. This method is important because a set of factors estimated using a large set of panel data, such as FRED-QD, is likely to contain a mixture of stationary and non-stationary variables. Although, in general, the form of non-stationarity may be more general than I(1), this Chapter is restricted to I(1) variables only; despite this restriction, the resulting model is still quite general and is an improvement over the current literature. The methodology has been fully developed with rigorous mathematical proofs for the proposed method. In a simulation study, we observed that the proposed mixture-FAR method performed better than its competitor that requires all the predictors to be I(1); the mean squared error of prediction was at least 33% lower for the mixture-FAR model. As

an empirical illustration, we evaluated the aforementioned methods for forecasting the non-stationary variables, GDP and Industrial Production [IP], using the quarterly panel data FRED-QD on the US economic variables. We observed that the mixture-FAR model proposed in Chapter 2 performed better than its competitors. The proofs of the theorems on the asymptotic validity of the proposed method are provided in this thesis.

Second, for empirical application, a limitation of the mixture-FAR model studied here is that it assumes that the regression parameters are constant throughout the estimation period. If the panel data set spans a long time period, it is possible that the model parameter values may change over time, in which case the constant parameter model studied previously would be misspecified and hence the forecasts based on it are likely to be invalid. Therefore, we generalize the mixture-FAR model by allowing key parameters to be time varying, albeit in a controlled manner. For example, we assume that the factors follow a time-varying vector moving average model of infinite order; this allows the factors to be locally stationary.We refer to the proposed model as a *semi-parametric FAR model.* Chapter 3 in this thesis develops a new method of estimating the semi-parametric FAR model and then using the estimated model for forecasting. Work is in progress to derive the asymptotic properties of the proposed method derived for the case when the factors are observed. In the numerical studies, we explored the latent factors being estimated by two different methods: the conventional Principal Components Analysis (PCA) and a nonparametric local estimation method. In a simulation study, we observed that the factors estimated by a non-parametric method were not very sensitive to different bandwidths, but the estimated coefficients in the FAR model were sensitive to the bandwidth choice. Therefore, we explored cross-validation for choosing suitable bandwidths for factor estimation and parameter estimation. Using the FRED-QD data set, we evaluated the performance of the new method for forecasting the I(1) log(GDP), and two I(0) GDP growth rate and inflation. We observed that the model with nonparametric estimates of factors forecasts inflation better than the semi-parametric FAR model with PCA factors, whereas the model with PCA factors generated log(GDP) and GDP growth rate forecasts better than the model with nonparametric counterparts.

Third, we further extend the semi-parametric FAR model in Chapter 4 by allowing two-level factors, called *global factors* and *group factors*, where global factors are pervasive and group factors are non-pervasive; global factors are related to all the predictor variables and the group factors are related to a group of variables that forms a strict subset of the predictor variables. For example, in the context of FREDQ data, a group factor could be related to a group of price variables. To allow for such two-level factors, we generate global and group factors from the two-level factor panel model, and include them in the FAR model. We refer to the new model as *semi-parametric two-level FAR model*. There is hardly any literature on forecasting using such an FAR model; the presence of two-level factors and time varying parameters result in a model that is complex. Based on the insights gained in the two previous extensions of FAR models, we propose a method for estimating the model. For the estimation of the two-level factor model, we adapt the method introduced by Breitung and Eickmeier [2016], where only the stationary case was studied. The results of our empirical studies show that (a) the semi-parametric one-level FAR model forecasts GDP in level better than its competitors, and (b) the semi-parametric two-level FAR model forecasts GDP growth rate and inflation better than its competitors. These observations provide some evidence that the proposed method is promising and it is likely to be an improvement over the existing ones in the literature.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: ....

Print Name: Sium Bodha Hannadige

Date: February 22, 2022

Acknowledgement

This thesis would have never been a success without the guidance, encouragement, and support of my supervisors, family, and friends. This note is to show my gratitude to them.

First, I would like to sincerely thank my supervisors Professor Jiti Gao and Professor Mervyn Silvapulle for being incredible supervisors and mentors. Their endless enthusiasm for research, productivity, and wealth of knowledge has motivated me throughout my Ph.D. at Monash university. No matter how busy they were, they always kept the doors open to welcome me, whenever I needed their help. I believe that they are the best supervisors I could ever find in my life long research journey. I am also deeply grateful to Professor Param Silvapulle who, although is not my supervisor, gave me invaluable guidance and unreservedly shared her experience and expertise with me, which helped me to achieve my objectives in research. I am deeply thankful for Professor Mervyn Silvapulle and Professor Param Silvapulle for supporting me not only professionally but also at a personal level. Especially, during the COVID-19 lockdown period, they have been constantly motivating me and looking after my health and mental wellbeing. Without their help, I would not have found the beauty of econometrics and research.

I would also like to thank my milestone evaluation panel Professor Gael Martin, Professor Farshid Vahid, Professor Xibin Zhang, Associate Professor Tashushi Oka, and Assistant Professor Yi He for their constructive comments to improve my research findings. A special thanks goes to Dr. Benjamin Wong for helping me with an empirical application. I am deeply grateful to Dr. Indeewara Perera for his guidance and encouragements throughout my candidature. I would like to extend my sincere thanks to Professor Gael Martin and Professor Farshid Vahid for being wonderful Ph.D. directors and providing invaluable support to make my candidature fulfilling. I also thank Professor Donald Poskitt from whom I learned many advanced techniques in econometric theory during my first year of Ph.D..

This research was financially supported by Monash University and the Dean's Excellence

Award of Monash Business School. I am grateful for the financial support that I received. Further, I would like to thank Professor Rob Hyndman, Head of the Department of Econometrics and Business Statistics, for arranging financial support that helped me to comfortably focus on my research at the end of my candidature. I take this opportunity to express my sincere gratitude to all other academic and administrative staff members of the Department of Econometric and Business Statistics at Monash University.

Chapter 2 is being revised for the *Journal of Business and Economic Statistics*. I take this opportunity to thank the two anonymous reviewers for their constructive comments. I presented my research work at seminars in the Department of Econometrics and Business Statistics, Monash-Xiamen University workshop, and at the 40th International Symposium on Forecasting organized by the International Institute of Forecasters. I would like to thank the participants of these seminars, workshops, and the conferences for their valuable comments. I was fortunate to discuss my research with Professor Peter Phillips and Professor Sydney Ludvigson during their visit to Monash University and receive valuable comments and suggestions.

I would like mention many of my fellow students who supported me in many ways. First, I would like to thank Sombut and Jieyang for helping me with methodological ideas, program coding, and encouraging me at times when I felt stressed. I would also like to thank my dearest friends Sayani, Puwasala, and Abby for being such caring friends and supporting me in every way. Further, I would like to thank Dilini, Semini, Nick, Kanchana, Penjie, Vivian, and Madeleine for their enduring friendship throughout my Ph.D. journey. A special thanks go to all my house-mates Sandali, Madeleine Smith, Sandith, Jirarote, Mehek, and Jia-En for their support. Without them, the COVID-19 lockdown would not have been easy to survive.

Last, but by no means least, I would like to thank my family members. I am truly grateful to my father, Bodha Hannadige Chandrasena, and my mother Malani Bopage for their unconditional selfless love, care, and sacrifices that they made to make my life better. Your inspiration and guidance always kept me motivated to perceive better achievements in my life. I am fortunate to be your daughter. I would also like to thank my sisters Kalpani, Harshani, and Nadeeshani, my brother Sajith, and my brothers-in-law, Manoj, Prasanna, and Mayura for their unconditional love, understanding, and moral support given for completing my research.

Chapter 1

Introduction

1.1 Background and motivation

Inflation and GDP growth forecasts are frequently produced and used to improve decision-making at the micro and macro levels. Therefore, generating accurate forecasts of GDP growth and inflation of major countries and regions has been the primary focus of a vast number of studies in the economic and econometric literatures. For example, see Banerjee and Marcellino [2006], Demers and Cheung [2007], Barnett et al. [2014], Banerjee et al. [2005], Lahiri and Sheng [2010], and Abbate and Marcellino [2014]. See also the recent survey papers Eickmeier and Ziegler [2008] for forecasting GDP growth and inflation and Kavtaradze and Mokhtari [2018] for inflation.

Abbate and Marcellino [2014] showed that the main reason for predictive failure is the use of univariate models that by necessity can incorporate only a small subset of the variables. Thus, predictive failure or inaccurate forecast is the result of (1) not taking account of all information in the data, and (2) not taking account of model uncertainty. Since the advent of the factor models by Stock and Watson [2002a, 2007] and Bai and Ng [2002, 2006] and thus the factor augmented-regression (FAR) model, many of the aforementioned studies evaluated the accuracy of forecasts by the univariate FAR model relative to standard time series models and those based on economic theory such as Phillips curve. Evidence in the huge literature on this topic demonstrate that the FAR model captures a large proportion of the information content in the high-dimensional panel data through only a few factors and such FAR models out-perform the time series models and economic theory based models in the out-of-sample predictions of macroeconomic variables.

The main objective of the thesis is to relax three assumptions that underlie the FAR method and to propose improved FAR methods, and develop methodology for estimating and forecasting key macroeconomic variables such as GDP (in level), GDP growth rate and inflation.

The widely studied FAR model includes only I(1) factors as predictors when forecasting I(1) variables, whereas it includes only I(0) factors when forecasting I(0) variables. Moreover, the I(1) factors were generated from the panel of I(1) variables, while I(0) factors were generated from the panel of I(0) variables; see Bai and Ng [2002, 2006] and Choi [2017] for details. However, the panel of FRED-QD data set (Stock and Watson 2002a) consists of I(0) and I(1) variables and thus the factors estimated therein also consist of I(1) and I(0) mixture. The FAR method with a mixture of I(1) and I(0) factors as predictors would improve the statistical efficiency and thus accuracy of forecasting; we call this *mixture-FAR model*. This thesis develops some asymptotic results, conducts a simulation study to evaluate the finite sample properties of the mixture-FAR model parameter estimates and prediction intervals, and illustrate an empirical application of the proposed method for forecasting non-stationary macroeconomic variables using the FRED-QD data set.

The standard FAR models frequently ignore structural changes over time. When the relationship between a macroeconomic variable and its fundamentals changes over time, then the underlying parameters and/or structure might change with time as well. The models with time-varying parameters (TVP) explicitly allow non-linear reactions to the structural changes. These models were found to produce more precise estimates over other econometric models (Abbate and Marcellino 2014 and Kavtaradze and Mokhtari 2018).

Although progress has been made in the estimation and forecasting of univariate time series models with time varying parameters, the literature on the FAR model with TVP for forecasting is underdeveloped. Cai [2007] and Stock and Watson [2009] showed that time series models with TVP can generate forecasts that are robust to structural changes. Furthermore, Cai [2007] studied a time series model with TVP and serially correlated errors, and developed a non-parametric local linear kernel method to estimate the trend and coefficient functions. Li et al. [2011a] proposed a non-parametric method for estimating time-varying parameters in the panel model and obtained faster convergence rates for the time trend and time-varying coefficients. Fan and Huang [2005] introduced a profile least square (PLS) technique to estimate parametric coefficients in a semi-parametric varying model. Chen et al. [2012] extended the semi-parametric time-varying coefficients regression model to the panel data model structure. In this thesis, we adapt these estimation methods to our setting where the FAR model includes generated local stationary factors as predictors and the parameters of these factor being time varying.

Albeit recent, the development of methodology for the estimation of multi-level factor model together with empirical applications is a rapidly growing literature (Wang 2008, Boivin and Ng 2006, Beck et al. 2009, Breitung and Eickmeier 2016, and Rodríguez-Caballero 2021, to name a few). A scenario in which the need for such multi-level factors may arise is the following. The economy could be partitioned into several sectors, and a set of factors that are specific to only one sector and hence has no effects on the variables in the other sectors. We call such factors *level-2* (non-pervasive) factors. In this case, some of the factor loadings would be structurally zero. In such cases, making use of the structural zero restrictions could be expected to improve statistical efficiency. To see the benefit of such factors, consider a group of price variables that constitutes a sector. Generating a factor from this group and including it in the FAR model would improve the out-of-sample forecast of inflation. Another example where two-level factors arise is in international bilateral trade, wherein global factors may impact all the countries in the panel, while level-two factors may impact countries from specific regions. For other examples, see Breitung and Eickmeier [2016].

In this thesis, we progressively relax the three aforementioned assumptions that underpin the FAR model and propose improved FAR methods and develop methodologies to estimate these models. We assess the performance of models in terms of out-of-sample predictions of three key macroeconomic variables GDP, GDP growth rate and inflation, and compare them against its competitors. The improved methods and methodologies are proposed and studied in three main Chapters of the thesis.

1.2 Outline of the thesis

In Chapter 2, we relax the assumption that the FAR model consists of only I(1) factors when forecasting I(1) variables and allow the FAR to include a mixture of I(0) and I(1) factors as predictors. The mixture of I(0) and I(1) factors are estimated from the large panel data such as FRED-QD, which typically contains a mixture of stationary and non-stationary variables and include them in the FAR model. We call this mixture-FAR model. The main methodological contributions of this Chapter are: (a) it establishes the consistency of the estimated factors up to a rotation of the latent factors; (b) it derives the consistency and asymptotic normality of the parameter estimators; and (c) it constructs an asymptotically valid prediction interval. In a simulation study, we evaluate the finite sample properties of the mixture-FAR model and compare its accuracy of out-of-sample forecasts relative to its competitors. In the empirical application, we evaluate the out-of-sample predictability of the proposed method relative to its competitors in the literature in forecasting log(GDP) and industrial production [IP].

In Chapter 3, we relax the assumption that parameters of the mixture FAR model are constant over time and allow the key parameters of the mixture-FAR model to be time varying, albeit in a controlled manner. Also, the factors are allowed to be locally stationary but globally non-stationary. We refer to the proposed model as *semi-parametric FAR model.* Since developing a new methodology for this rigorous model is very challenging, in this Chapter, we develop a methodology under some assumptions. First, we assume that the latent factors are known and follow a time-varying $VMA(\infty)$. So, these factors are locally stationary. Under this assumption, work is in progress to drive the asymptotic properties of the estimated regression coefficients. An extension to the case where the factors are unknown and estimated will be undertaken in the future. Second, we allow the parameters of factors in the FAR model to be time-varying. A profile least square (PLS) technique and non-parametric local estimation method are together used to estimate both constant and time varying coefficients of the FAR model. In the numerical studies, this Chapter explores two different methods for the estimation of the latent factors: the conventional Principal Components Analysis (PCA) and a nonparametric local estimation method. We conduct a simulation study to assess the sensitivity of the factors estimated by a non-parametric method and estimated coefficients in the FAR model to different bandwidths. Using the FRED-QD data set, we evaluate the performance of the new method for forecasting a non-stationary variable, log(GDP), and two stationary variables GDP growth rate and inflation.

In Chapter 4, we relax the assumption that factor model consists of only global (one-level) factors in the panel data model and allow two-level factors, which include level-1 and level-2 factors; see the previous section for a brief discussion. Since the panel data contains I(1) and I(0) economic variables, the two-level factors estimated therein are also a mixture of I(0) and I(1). Furthermore, we allow the parameters of the two-level factors in the FAR model to be time varying. We refer to the proposed model as *semi-parametric two-level FAR model.* For the estimation of the two-level factor model, we adapt the method introduced by Breitung and Eickmeier [2016], where only the stationary panel data was studied, to our panel model setup with mixture of I(0) and I(1) variables. Based on the insights gained in the previous two Chapters, we propose a kernel method for estimating the model. To improve the out-of-sample predictive performance of the method in this Chapter, we included the economic policy uncertainty index (EUI) as a predictor in the two-level FAR model, which is known to reduce the instability in the forecast errors. We apply the new semi-parametric two-level FAR model for forecasting log(GDP), GDP growth rate and inflation. The FRED-QD dataset is categorized into 12 groups, such as employment and unemployment, prices, interest rates, and exchange rates. These groups constitute 12 level-2 factors.

The novelty of the proposed method is that, based on prior knowledge and/or some information criteria such as goodness-of-fit and correlation measures, we can select the number of global and group factors that would be predictors in the FAR model for forecasting the desired macroeconomic variable. Thus, we can specify three separate semi-parametric two-level FAR models for forecasting log(GDP), GDP growth rate and inflation. By contrast, previous studies mostly used the same global factors in the FAR model for forecasting macroeconomic variables. By construction, the proposed model is likely to generate more accurate forecasts than the FAR models used in the literature. The methodological developments for the semi-parametric two-level FAR model is challenging, which will be a topic for the future research.

The final concluding Chapter briefly summarises the main findings of the research topics addressed in the three Chapters and provide some directions for future research.

Chapter 2

Time Series Forecasting using a Mixture of Stationary and Nonstationary Predictors

2.1 Introduction

Construction of valid probability forecasts of key economic variables, such as GDP and Inflation, is central to making reliable economic policy decisions. There is a large body of literature on constructing probability forecasts for a stationary variable using other stationary variables as predictors. By contrast, the literature on making probability forecasts for a nonstationary variable using a mixture of stationary and nonstationary predictors remains underdeveloped. In a method that has attracted considerable attention, a two-step method involving a *factor model* for panel data and a *regression model* for predicting the time series are used jointly (Stock and Watson 2002a). In the first step, the factor model is used for generating a small number of factors to capture most of the information in a set of panel data for a large number of potential predictors. In the second step, the regression model uses the generated factors as predictors, instead of the large number of potential predictors in the panel data. The resulting regression model is known as *factor augmented regression*[FAR] model, which is one of the well-known models for constructing probability forecasts for a time series (Bernanke et al. 2005, Stock and Watson [1998a, 1998b, 2002b]). The large number of economic variables that are potential predictors typically includes a mixture of stationary and nonstationary variables. Consequently, the collection of factors is also typically a mixture of stationary and nonstationary ones (Bai

2004, Eickmeier 2005, Moon and Perron 2007, Smeekes and Wijler 2019). The objective of this chapter is to develop a new method for constructing a valid prediction interval when the predictors in the prediction model include a mixture of stationary and nonstationary factors.

For the main results of this chapter, the only nonstationary variables considered are $I(1)$; therefore, we use the term nonstationary as a synonym for $I(1)$.

Related literature

The validity of the aforementioned general approach for forecasting using an FAR model with estimated factors has been established when all the variables, including the factors, are stationary (Bai 2003, Bai and Ng [2002, 2006], Gonçalves and Perron 2014), and also when they are all nonstationary (Choi 2017), but not when they form a mixture of stationary and nonstationary ones. This chapter builds on the aforementioned literature and develops a method based on FAR models for forecasting, more specifically for constructing an asymptotically valid prediction interval when the chosen set of factors is a mixture of stationary and nonstationary ones.

Suppose that the variables are all nonstationary. Bai [2004] studied the consistency of the estimated factors and proposed a method for estimating an optimal number of factors. The limiting distributions of the estimators of factors and their loadings have also been obtained. Choi [2017] used a method based on generalized principal components for estimating factors, and studied the asymptotic properties of the generated nonstationary factors, their loadings, and forecasts. Under the assumption $T/N \to 0$, Choi [2017] showed that estimators of the parameters in the forecasting model are consistent and asymptotically normal, and that the forecasts converge at the rate T, where T and N are the time and cross-section dimensions respectively.

Since the method in this chapter is based on the large literature for forecasting a stationary variable using an FAR model, a few comments would be helpful. Suppose that we wish to predict a stationary variable, such as inflation, using a method that requires all the predictors in the prediction equation to be stationary, for example the method in Bai [2003] or that in Bai and Ng [2006]. For this scenario, one could either delete all the I(1) variables or use the first differences of the I(1) predictors instead of the original I(1) predictors (Ludvigson and Ng 2007, Stock and Watson 2012, Cheng and Hansen 2015). While this adaptation is methodologically valid, a natural question that arises is whether differencing a nonstationary variable could result in loss of information in the level-data that may be important for forecasting. Similar questions also arise when forecasting a nonstationary variable, the topic of this chapter.

Suppose that the set of generated factors is a mixture of stationary and nonstationary variables, and we wish to predict a nonstationary variable, such as GDP, using a method that requires all the predictors to be nonstationary, for example the method in Choi [2017]. For this scenario, it has been suggested to delete all the predictors that are stationary and apply the method. While this method is valid, deletion of predictors to suit a method is likely to result in loss of information and hence loss of statistical efficiency.

The development of methodology for factor models has contributed to improve time series forecasting, macroeconomic analysis, and monetary policy analysis. Empirical results from several studies indicate that the generated factors often tend to be a mixture of stationary and nonstationary variables. For example, Bai [2004] studied employment fluctuations across 60 industries in the US and found that two nonstationary and one stationary factors explain a large part of the fluctuations in employment. Bernanke et al. [2005] used factor augmented vector auto-regression and found that it contained information to accurately identify the monetary transmission mechanism in the US. Eickmeier [2005] used a large-scale ($N > 300$) dynamic factor model and concluded that the Euro-area economies shared four non-stationary factors and one stationary factor. Eickmeier found that the factors represent mainly the variations in German and French real economic activity as well as of producer prices and financial prices through which they also studied the transmission channels and the impacts of macroeconomic shocks. Moon and Perron [2007] studied the Canadian and US interest rates for different maturities and risk, and found a single nonstationary factor and several stationary ones. The dominant factors were interpreted as level and slope, as in the term structure literature. In a recent study, Smeekes and Wijler [2019] provided an overview of forecasting macroeconomic time series in the presence of unit roots and cointegration. They compared point forecasts of some key economic variables in FRED-MD and FRED-QD data set and nowcasting of unemployment in another data set that was constructed from Google trend using the two methods (a) transforming every series to stationarity, and (b) directly modelling the level data. However, rigorous justification for modelling the level data with unit roots and cointegration in the forecasting model is yet to be provided.

The method in this chapter

In this chapter, we use the methods in the literature (Bai 2004, and Moon and Perron 2007) for generating factors that may be a mixture of stationary and nonstationary variables. Once they have been generated, we use them as predictors in a *factor-augmented regression* [FAR] model for forecasting; we refer to this as a *mixture-FAR* model. We develop new methods for constructing asymptotically valid prediction intervals using the mixture-FAR model. Our results, under the additional assumption that all the variables are stationary, reduce to the corresponding ones in Bai and Ng [2006]. Similarly, our results, under the additional assumption that all the variables are nonstationary, reduce to the corresponding ones in Choi [2017]. In this sense, our results provide a way of combining and extending the existing results on this topic that are limited to the two cases (a) when all the variables are stationary and (b) when all the variables are nonstationary.

To state the asymptotic results, we introduce a diagonal matrix, denoted D_{1T}; its dimension is equal to the number of predictors in the FAR model, and each of its diagonal element is equal to either $\sqrt{T}$ or T according as the corresponding predictor is stationary or nonstationary. The joint limiting distribution of the generated factors is derived under the assumption $\sqrt{N}\left\|D_{1T}^{-2}\right\| \to 0$, where and in what follows $\|A\| = trace(A'A)^{1/2}$. We develop the main part of the asymptotic results under the assumption $T/N \to 0$. We show the consistency and asymptotic normality of estimators of the parameters of the forecasting model. For the case of normally distributed errors in the prediction model with $\sqrt{N}\left\|D_{1T}^{-2}\right\| \to 0$ and $T/N \to 0$, we show that forecast error has an asymptotically normal distribution, and use it to construct an asymptotically valid prediction interval for the dependent variable in the forecast equation. To examine the finite sample properties of the estimates, we conducted a simulation study with data generating processes [DGP] that contain mixtures of stationary and nonstationary variables. In these simulations, we observed that the mixture-FAR method performed overall better than the method that requires all the variables to be nonstationary. As an empirical illustration, we evaluated the aforementioned methods for forecasting the nonstationary variables, GDP and industrial production [IP], using the quarterly panel data on US macroeconomic variables, known as FRED-QD. We observed that the mixture-FAR model performed better than its aforementioned competitors. This observation also corroborates the general observation of our simulation study, namely, the mixture-FAR method performed better than the competing methods.

The rest of this chapter is organized as follows. Section 2.2 introduces the model and the assumptions, and establish the consistency and limiting distributions of the estimators. Section

2.3 reports the results of the simulation study. The empirical example using the FRED-QD data is presented in Section 2.4, and Section 2.5 concludes. The proofs of the theorems and lemmas, and some simulation results are provided in the Appendix section 2.6.

2.2 Methodology

2.2.1 Model and notation

Let $\{Y_t, t = 1, 2, \ldots\}$ denote an observable univariate time series that we wish to predict at a future time $T + h$ $(h \geq 1)$, using the information available up to time T. Let $\{X_{it} \in \mathbb{R} : i = 1, \ldots, N; t = 1, \ldots, T\}$ denote a set of panel data and $\{W_t \in \mathbb{R}^m : t = 1, \ldots, T\}$ denote a set of observable predictors; W_t may contain lagged values of Y_t. The aforementioned *factor augmented regression*[FAR] method for predicting Y_{T+h} uses the following two models:

$$\text{Factor model:} \quad X_{it} = \lambda_i' F_t + e_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \tag{2.1}$$

$$\text{FAR model:} \quad Y_{t+h} = \theta' F_t + \omega' W_t + \epsilon_{t+h} \qquad (t = 1, ..., T), \tag{2.2}$$

where F_t is an $r \times 1$ vector of unobservable factors, $\{e_{it}, \epsilon_t\}$ are idiosyncratic errors, λ_i is an $r \times 1$ vector of factor loadings, and $\theta_{r\times 1}$ and $\omega_{m\times 1}$ are unknown parameters $(i = 1, \ldots, N; t = 1, \ldots, T)$; the number of factors r is assumed known. This was called a "diffusion index forecasting model" by Stock and Watson [2002a].

A point of departure of this chapter from the current literature is that we allow the r factors to be used in the FAR-method, to be a mixture of stationary and nonstationary variables. Further, we assume that Y_t and W_t are nonstationary; as indicated previously, the only nonstationary variables that we consider are $I(1)$. We conjecture that the results in this chapter can be extended to the case when the factors are $I(d)$ $(d = 2, 3, \ldots)$; but, we do not consider such extensions in this chapter.

Remark 1: (a). In model (2.1), the time series $\{X_{it}\}_{t\in\mathbb{N}}$, for a given i $(i = 1, \ldots, N)$, may be stationary or non-stationary. Therefore, the model allows the panel to consist of only stationary variables, or only nonstationary variables, or a mixture of stationary and nonstationary variables. If all the panel variables are stationary, then the estimated factors would also be stationary. This is because each factor is a linear combination of the panel variables; for example, F_{1t} is of the form $\sum_{i=1}^{N} a_i X_{it}$ for some $\{a_1, \ldots, a_N\}$.

(b). One might ask whether it is possible to partition the variables into two groups such that group 1 consists of only stationary variables and group 2 consists of only nonstationary variables, and then estimate the stationary factors using group 1 and nonstationary factors using group 2. Such a method is not valid for the following reasons. A stationary factor may be a function of stationary and nonstationary variables in the panel. Therefore, it is not possible to use only the stationary variables to estimate the stationary factors. Similarly, a nonstationary factor may be a function of stationary and nonstationary variables in the panel. Therefore, it is not possible to use only the nonstationary variables to estimate the nonstationary factors. Apart from these limitations, the probability of misclassification resulting from the multiple tests needed to partition the variables is likely to be very high.
(c). As it is unlikely to estimate the I(0) and I(1) factors separately from two sub-panels, the estimation of the number of I(0) factors using IC based on the sub-panel with stationary X_{it} and estimating the number of I(1) factors using IPC based on the nonstationary X_{it} is not valid.

Let $X = \left[X_{it}\right]_{T\times N}$ denote the panel data in matrix form, $F = (F_1, \ldots, F_T)'$ denote the $T\times r$ matrix of unobservable common factors, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ denote the matrix of factor loadings, and $e = \left[e_{it}\right]_{T\times N}$ denote the matrix of error terms from the factor model. Then the factor model (2.1) can also be expressed as $X = F\Lambda' + e$. Since the stationary and nonstationary terms need to be treated differently, let us write $F_t' = (E_t', G_t')'$, where E_t is $p\times 1$ and nonstationary, G_t is $q\times 1$ and stationary; p and q are assumed known. Therefore, $E_t = E_{t-1} + u_t$, where u_t is stationary. Substituting $F_t' = (E_t', G_t')'$, the factor model (2.1) and the FAR model (2.2) take the forms

$$X_{it} = \lambda_i^{(1)'} E_t + \lambda_i^{(2)'} G_t + e_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \tag{2.3}$$

$$Y_{t+h} = \alpha' E_t + \beta' G_t + \omega' W_t + \epsilon_{t+h} \qquad (t = 1, ..., T), \tag{2.4}$$

respectively, where $\lambda_i = (\lambda_i^{(1)'}, \lambda_i^{(2)'})'$ and $\theta = (\alpha', \beta')'$. As expected, estimates of the coefficients α and β of the nonstationary and stationary variables in the FAR model (2.2), converge at the rates T and $T^{1/2}$ respectively. Similarly, since W_t is $I(1)$, we would expect that the estimator $\hat{\omega}$ to converge at the rate T. Since estimators of coefficients corresponding to stationary and nonstationary variable converge at different rates, we introduce the following scaling matrices:

$$D_{1T} = \mathrm{diag}(TI_p,\ T^{1/2}I_q)_{r\times r}, \qquad D_{2T} = TI_m, \qquad D_T = \mathrm{diag}(D_{1T},\ D_{2T}). \tag{2.5}$$

Remark 2: Since the model allows a mixture of stationary and non-stationary factors, it is also reasonable to allow the components of W_t to be a mixture of stationary and non-stationary variables. Although, the methodological details in this thesis are presented only for the case when W_t is nonstationary, the proposed method can be modified to allow for mixture of stationary and non-stationary variables in W_t. The modifications would involve treating the nonstationary components of W_t and nonstationary factors together, and the stationary components of W_t and stationary factors together. Let us write $W_t = (W'_{1t}, W'_{2t})'$, where W_{1t} is an $m_1 \times 1$ vector of a stationary observable variables, and W_{2t} is an $m_2 \times 1$ vector of a nonstationary observable variables. Then the model for Y_{t+h} in (2.4) could be expressed as

$$Y_{t+h} = \alpha' E_t + \beta' G_t + \omega_1' W_{1t} + \omega_2' W_{2t} + \epsilon_{t+h} \qquad (t = 1, \ldots, T). \tag{2.6}$$

The derivations that appear below for (2.4) could be modified to accommodate (2.6). For a given matrix A, let $A > 0$ denote that it is positive definite. For given matrices X and Y, let $X \oplus Y$ denote $diag(X, Y)$. Finally, let $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and in distribution, respectively.

2.2.2 Estimation of the common factors

To estimate the latent factors for a given panel dataset X, we may use either the Gaussian Maximum Likelihood Estimator (MLE) or the method based on Principal Component Analysis [PCA]. In this chapter, we use the latter. To choose an optimal number of factors, r, we use the Integrated Panel Criterion [IPC] and the panel Information Criterion [IC] introduced by Bai [2004] and Bai and Ng [2002], respectively.

For estimating the number of stationary and nonstationary factors in the model, we first estimate the total number, r, of factors in the model. To this end, we first form the differences of the data, and then apply the panel information criterion (IC) in Bai and Ng[2002]; we use differenced data since the IC method in Bai and Ng(2002) is for stationary variables. Since we use stationary panel in this step, the conditions in Bai and Ng[2002] are satisfied and hence the method estimates the optimal number of factors consistently. It is worthy of noting that, since a factor is a linear combination of the variables in the panel, an $I(1)$ factor in the level data would become an $I(0)$ factor for the first differences. Therefore, it is valid to estimate the total number factors, using the first differenced data. After estimating the total number of factors, we estimate the number of I(1) factors using the IPC from Bai[2004] applied to the data without differencing; it is has been shown that this criterion is applicable to a mixture of I(0) and I(1) factors (Moon and Perron[2007]). Therefore, by adapting the IC and IPC criteria, we estimate the numbers of $I(0)$ and $I(1)$ factors. In our simulation study, presented later in section 2.3.3, this method performed well.

Let $\tilde{F}_{T\times r} = (\tilde{F}_1, \ldots, \tilde{F}_T)'$ be defined as equal to D_{1T} times the matrix formed by the r eigenvectors corresponding to the r largest eigenvalues of the matrix XX'. Since we use PCA, $\tilde{F}$ is an estimator of F, the matrix of common factors. For the derivations of the asymptotic results, we assume that the numbers of stationary and nonstationary factors is known. However, in empirical studies, we apply one or more tests to each factor to determine whether it is stationary or nonstationary. Therefore, it is clear that the estimation method has an element of pre-testing. Assuming that a consistent test is applied for classifying a variable as stationary or nonstationary, it follows that the probability of misclassification tends to zero, and hence the asymptotic results would be unaffected by the pre-testing.

In empirical studies, after applying PCA to estimate the factors as a group, a challenge encountered is in determining the appropriate number of $I(0)$ and $I(1)$ factors. It is possible that a rotation of the factors may result in more or fewer $I(1)$ factors. In empirical studies, researchers usually prefer the factors to be interpretable in the context of the study. For statistical efficiency, there may be some merit in ensuring that there is no cointegration. Thus, in view of the number competing qualitative criteria, the choice of the appropriate number of $I(1)$ is not that well defined. In this thesis, we do not investigate these issues. Instead, our broad aim is to adapt the existing methods and develop a new method that is methodologically sound and improves over the current literature. We apply the integrated panel criterion studied in Bai (2004) to identify the number nonstationary factors, and apply unit root tests to determine whether or not each factor is stationary or nonstationary. This method has been used in the literature (see Moon and Perron [2007]). We also conducted a simulation study to evaluate the proposed method. The details of the method and simulation results are discussed in the rest of this chapter.

Once the factors have been estimated, a corresponding estimator of the factor loading matrix Λ is $\tilde{\Lambda} = X'\tilde{F}D_{1T}^{-2}$. Without loss of generality, we assume that the columns of $\tilde{F}$ are arranged such that the first p have been classified as nonstationary and their corresponding eigenvalues are in the decreasing order, and the remaining q columns have been classified as stationary and their corresponding eigenvalues are in the decreasing order. Therefore, without loss of generality, we write $\tilde{F}_t = (\tilde{E}_t', \tilde{G}_t')'$ and $F_t = (E_t', G_t')'$. Let $\tilde{V}_{p,NT}$ denote the diagonal matrix with diagonal elements equal to the largest p eigenvalues of XX' divided by T^2N and each of the corresponding eigenvector has been classified as nonstationary; further, without loss of generality, assume that the diagonal elements appear in the decreasing order. Similarly, let $\tilde{V}_{q,NT}$ denote the diagonal matrix with diagonal elements equal to the largest q eigenvalues of XX' divided by TN and each of the corresponding eigenvector has been classified as stationary; again, without loss of generality, assume that the diagonal elements appear in the decreasing order. Let $\tilde{V}_{NT} = diag(\tilde{V}_{p,NT}, \tilde{V}_{q,NT})$. Therefore, $\tilde{V}_{NT}$ is equal to the diagonal matrix whose diagonal elements are the $r = (p+q)$ largest eigenvalues of the matrix XX' multiplied by D_{1T}^{-2}/N.

We adopt the standard procedure to ensure that the factors are identified up to a rotation. To this end, we assume that $\tilde{F}$ satisfies the normalization $D_{1T}^{-2}\tilde{F}'\tilde{F} = I_r$, $\tilde{\Lambda}'\tilde{\Lambda}$ is diagonal, and define the rotation matrix $H = N^{-1}\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda$. If all the variables are stationary then the foregoing H reduces to the expression in Bai and Ng [2002], and if all the variables are nonstationary then it reduces to the forms in Bai [2004] and Choi [2017].

Let $\hat{L}_t = (\tilde{F}_t', W_t')'$ and $\delta = (\theta' H^{-1}, \omega')'$; then, H and δ are also a functions of the data and unknown population parameters. Then, the FAR model (2.4) can be written as

$$\begin{aligned} Y_{t+h} &= \theta' F_t + \omega' W_t + \epsilon_{t+h} = \theta' H^{-1}(HF_t - \tilde{F}_t + \tilde{F}_t) + \omega' W_t + \epsilon_{t+h} \\ &= \theta' H^{-1}\tilde{F}_t + \omega' W_t + \theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h} \\ &= \delta' \hat{L}_t + \theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h}. \end{aligned} \tag{2.7}$$

Let $(\hat{\alpha}', \hat{\beta}', \hat{\omega}')$ denote the ordinary least squares [OLS] estimator of $(\alpha', \beta', \omega')$ obtained by regressing Y_{t+h} on $\hat{L}_t$ $(t = 1, \ldots, T-h)$. Then

$$\hat{\delta} = (\hat{\alpha}', \hat{\beta}', \hat{\omega}')' = \left(\sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t'\right)^{-1} \sum_{t=1}^{T-h} \hat{L}_t Y_{t+h}. \tag{2.8}$$

Later we will show that $\theta' H^{-1}(HF_t - \tilde{F}_t)$ in (2.7) is asymptotically centered at zero in the limit, and hence $\{\theta' H^{-1}(HF_t - \tilde{F}_t) + \epsilon_{t+h}\}$ could be treated as an error term centered at zero for the purposes of estimating δ. In consequence, it turns out that $\hat{\delta} - \delta$ is asymptotically normal with mean zero, which will be used later for deriving a prediction interval.

Remark 3: While it is not essential for the derivations, the following observation is helpful. Let $H_1 = \tilde{V}_{p,NT}^{-1} \frac{\tilde{E}'E}{T^2} \frac{\Lambda_1'\Lambda_1}{N}$, $H_2 = \tilde{V}_{q,NT}^{-1} \frac{\tilde{G}'G}{T} \frac{\Lambda_2'\Lambda_2}{N}$, and $H_0 = diag(H_1, H_2)$. Then, it may be verified that $(H - H_0)$ converges, in probability, to zero. Consequently, for the asymptotic results, the rotation of the entire factor by H leads to the same asymptotic results as performing the rotations separately for the nonstationary and stationary factors by H_1 and H_2, respectively. In this sense, the two rotations can be performed independently.

Remark 4: In Principal Component (PC) method, we estimate the factors using the eigenvectors corresponding to the largest eigenvalues of the square matrix XX', where X is the panel data set. To estimate the eigenvalues and eigenvectors of the square matrix, we use single value decomposition: $XX' = U\Sigma V^*$, where the diagonal elements of the matrix Σ are corresponding to the eigenvalues (descending order) of XX', and U, V are the matrices of corresponding eigenvectors. Hence, the set of first r eigenvectors of matrix U is the estimated r factors of the panel matrix x (without using the normalization condition). After estimating r eigenvectors, we apply the corresponding normalization condition and estimate the factors.

2.2.3 Distribution theory

In this section, we study the asymptotic distributions of the generated factors and the estimators of the regression parameters. First, we introduce some assumptions; in these assumptions, $M \in \mathbb{R}$ denotes a generic constant, and hence it may be different in its different appearances.

Assumption 2.1 (Factors and factor loadings)**.**
(i) The strictly stationary process u_t in $E_t = E_{t-1} + u_t$, satisfies $\max_{t \geq 1} E\|u_t\|^{4+\delta} \leqslant M$, for some $\delta > 0$.
(ii) $E\|F_1\|^4 \leqslant M$ and $D_{1T}^{-1} \sum_{t=1}^{T} F_t F_t' D_{1T}^{-1} \xrightarrow{d} \Sigma_F$ as $T \to \infty$, where Σ_F is a positive definite random matrix.
(iii) The number of factors r is known and does not depend on N or T; further, factors are not cointegrated.
(iv) The loadings λ_i are either deterministic and $\|\lambda_i\| \leqslant M$ satisfying $\Lambda'\Lambda/N \to \Sigma_\Lambda$ as $N \to \infty$, or they are stochastic and $E\|\lambda_i\|^4 \leqslant M$ satisfying $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$, for some $r \times r$ positive definite non-random matrix Σ_Λ.
(v) The eigenvalues of the matrix $\Sigma_\Lambda \Sigma_F$ are distinct, almost surely.

To estimate the number of factors, we assume that the factors are not cointegrated. If they are cointegrated then the stationary and nonstationary factors cannot be identified because one $I(0)$ factor may represent a combination of cointegrated $I(1)$ factors. By assuming Σ_F and Σ_Λ are positive definite and the eigenvalues of $\Sigma_\Lambda \Sigma_F$ are distinct, we ensure the identifiability of the r factors. If all the factors are nonstationary then Σ_F is distributed as $\int_0^1 B_F(r)B_F'(r)dr$, and if the factors are all stationary then Σ_F converges to the variance-covariance matrix of the factors. To state the next assumption, let us introduce the following notation:

$$\gamma_{st} = E\left(N^{-1}\sum_{i=1}^{N} e_{is}e_{it}\right), \ \tau_{ij,t} = E\left(e_{it}e_{jt}\right), \ \tau_{ij,ts} = E\left(e_{it}e_{js}\right) \ \ (i,j=1,\ldots,N; \ s,t=1,\ldots,T).$$

Assumption 2.2 (Idiosyncratic errors).
(i) $E\left(e_{it}\right)=0$ *and* $E\left|e_{it}\right|^8 \leqslant M$ $(i=1,\ldots,N; t=1,\ldots,T)$.
(ii) $|\gamma_{ss}| \leqslant M$ $(s=1,...,T)$, *and* $T^{-1}\sum_{s=1}^T\sum_{t=1}^T|\gamma_{st}| \leqslant M$.
(iii) $|\tau_{ij,t}| \leqslant |\tau_{ij}|$, *for some* τ_{ij} $(i,j=1,\ldots,N; \ t=1,\ldots,T)$, *and* $N^{-1}\sum_{i=1}^N\sum_{j=1}^N|\tau_{ij}| \leqslant M$.
(iv) $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T|\tau_{ij,ts}| \leqslant M$.
(v) $E\left|N^{-1/2}\sum_{i=1}^N\left[e_{is}e_{it}-E\left(e_{is}e_{it}\right)\right]\right|^4 \leqslant M$ $(t,s=1,\ldots,T)$.

Assumption 2.2 allows the idiosyncratic errors to have weak serial and cross sectional dependence. Heteroskedasticity is also allowed in both the serial and the cross-section dimensions. Since we allow weak correlations among the idiosyncratic errors in (2.1), it is an *approximate factor model*; for simplicity, we refer to it simply as a *factor model.*

Assumption 2.3 (Dependence among $\lambda_i, F_t,$ and e_{it}).
(i) $E\left(\frac{1}{N}\sum_{i=1}^N\left\|D_{1T}^{-1}\sum_{t=1}^T F_t e_{it}\right\|^2\right) \leqslant M$, *and* $E(F_t\, e_{it})=0$ $(i=1,\ldots,N; \ t=1,\ldots,T)$.
(ii) $(1/N)\sum_{i=1}^N\sum_{j=1}^N E\left(\lambda_i\lambda_j' e_{it}e_{jt}\right) \to \Gamma_t$ *as* $N\to\infty$, *for some* Γ_t, *and* $N^{-1/2}\Lambda' e_t \xrightarrow{d} N\left(0,\Gamma_t\right)$ *as* $N\to\infty$, *for each fixed* t $(t=1,\ldots,T)$.
(iii) $E\left\|N^{-1/2}D_{1T}^{-1}\sum_{t=1}^T\Lambda' e_t F_t'\right\|^2 \leqslant M$.

Assumption 2.3 allows the factor loadings $\{\lambda_1,\ldots,\lambda_N\}$, the factors $\{F_1,\ldots,F_T\}$, and the idiosyncratic errors $\{e_{it}, i=1,\ldots,N; t=1,\ldots,T\}$ to have a weak dependence among them.

Consistency of the generated factors

In the literature on FAR models, the consistency of the generated factors has been established for both stationary and nonstationary factors separately. Bai and Ng [2002] and Bai [2004] showed that the time-averaged mean square of factor estimation error [MSE] has $\min\{N, T\}$ and $\min\{N, T^2\}$ convergence rates for $I(0)$ and $I(1)$ factors separately. In our setting, the set of latent factors F contains a mixture of $I(1)$ and $I(0)$ series, and we show that the generated factors are jointly consistent and the convergence rate of MSE is $\min\{N, ||D_{1T}^{-2}||^{-1}\}$. To state the consistency of generated factors in the next lemma, let us recall that the rotation matrix H was defined as $N^{-1}\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda$.

Lemma 2.1. *Suppose that Assumptions 2.1–2.3 are satisfied. Let* $\delta_{NT}^{-1} = \max\left[N^{-1/2}, ||D_{1T}^{-1}||\right]$. *Then,* $T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$.

This lemma states that the time averaged square of factor estimation error converges to zero as $N, T \to \infty$ and the convergence rate is $\min\{N, ||D_{1T}^{-2}||^{-1}\}$. Therefore, we may estimate a rotation of the mixture of latent factors consistently by the method of principal component analysis. For the case when all the factors are stationary, the scaling matrix D_{1T} is $\sqrt{T}I_r$ and the convergence rate is $\min\{N, T\}$; this is consistent with the corresponding result in Bai [2003]. For the case when all the factors are nonstationary, $D_{1T} = TI_r$ and the convergence rate is $\min\{N, T^2\}$; this is consistent with Bai [2004].

Asymptotic distribution of the generated factors

To derive the asymptotic distributions of the estimated factors, we introduce the following additional assumption.

Assumption 2.4 (Weak dependence of idiosyncratic errors)**.**
(i) $\sum_{s=1}^{T}|\gamma_{st}| \leqslant M \ (t = 1, ..., T)$, *and*
(ii) $\sum_{j=1}^{N}|\tau_{ij}| \leqslant M \ (i = 1, ..., N)$, *where* γ_{st} *and* τ_{ij} *are as in Assumption 2.2.*

Lemma 2.2. *Suppose that Assumptions 2.1–2.4 are satisfied. Then* $D_{1T}^{-2}\tilde{F}'F \xrightarrow{d} Q$ *as* $N, T \to \infty$, *where* $Q = V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$ *is a random matrix,* $V = diag(v_1, ..., v_r)$ *with* $\{v_1, \ldots, v_r\}$ *denoting the eigenvalues of* $\Sigma_\Lambda\Sigma_F$, *and* Υ *is the corresponding matrix formed by scaled eigenvectors such that* $\Upsilon'\Upsilon = I_r$.

Lemma 2.3. *Suppose that Assumptions 2.1–2.4 hold. Then, as* $N, T \to \infty$ *with* $\sqrt{N}||D_{1T}^{-2}|| \to 0$, *for each* t, *we have* $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} V^{-1}QN\left(0, \Gamma_t\right) \overset{d}{=} N(0, \Sigma_{\tilde{F}})$, *where* Q *is defined in Lemma 2.2,* Γ_t *is defined in Assumption 2.3, and* Q *is independent of* $N(0, \Gamma_t)$.

This lemma shows that the factor estimation error is asymptotically normal with mean zero; this is important for estimating the parameters of the FAR model consistently, as indicated previously. Later we show that the asymptotic variance of $\left(\tilde{F}_t - HF_t\right)$ can be estimated consistently by $\tilde{V}_{NT}^{-1}\hat{\Gamma}_t\tilde{V}_{NT}^{-1}$, which is used for constructing the prediction interval of h-step ahead forecasts.

Asymptotic distribution of the estimators

To obtain the asymptotic distribution of the OLS estimator $\hat{\delta}$ of δ, we introduce the following additional assumptions.

Assumption 2.5 (Weak dependence between idiosyncratic and regression errors).
(i) $E\left|(TN)^{-1/2}\sum_{s=1}^{T-h}\sum_{i=1}^{N}\{e_{is}e_{it} - E\left(e_{is}e_{it}\right)\}\epsilon_{s+h}\right|^2 \leqslant M \quad (t = 1, \dots, T; h > 0)$.
(ii) $E\left\|(TN)^{-1/2}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\lambda_i e_{it}\epsilon_{t+h}\right\|^2 \leqslant M$, *and* $E\left(\lambda_i e_{it}\epsilon_{t+h}\right) = 0 \quad (i = 1, \dots, N; t = 1, \dots, T)$.

Assumption 2.6 (Moment and CLT for score vector). *Let* $L_t = \left(F_t', W_t'\right)'$. *Then, the following conditions are satisfied.*
(i) $E\left(\epsilon_{t+h}\right) = 0$ *and* $E\left|\epsilon_{t+h}\right|^2 < M \ (t = 1, \dots, T)$.
(ii) $D_T^{-1}\sum_{t=1}^{T}L_tL_t'D_T^{-1} \xrightarrow{d} \Sigma_L$ *as* $N, T \to \infty$, *where* Σ_L *is a positive definite random matrix with probability one.*
(iii) $D_T^{-1}\sum_{t=1}^{T}L_t\varepsilon_{t+h} \xrightarrow{d} \Sigma_{\epsilon L}^{1/2}\ N(0, I)$, *where* $\Sigma_{\epsilon L} > 0$ *with probability one.*

Assumption 2.5 imposes restrictions on the degree of dependence among the idiosyncratic errors over time, and between the idiosyncratic and regression errors. Part (ii) of Assumption 2.5 holds if $\{\lambda_i\}, \{e_{it}\}$, and $\{\epsilon_t\}$ are mutually independent and Assumption 2.2 holds. Assumption 2.6 part (iii) assumes the limiting distribution of $D_T^{-1}\sum_{t=1}^{T}L_t\epsilon_{t+h}$. Under the following four cases we can assume the limiting distribution of $D_T^{-1}\sum_{t=1}^{T}L_t\epsilon_{t+h}$ to follow a normal distribution; see Choi [2017].

Case 1: Suppose $\{F_t\}$ are all I(1) and independent of $\{\epsilon_{t+h}\}$, and $\{W_t, \epsilon_{t+h}\}$ are I(0), then together with $D_{1T} = diag[T, \dots, T]$ and $D_{2T} = diag[\sqrt{T}, \dots, \sqrt{T}]$, we have

$$\Sigma_{\epsilon L} = \left(\int_0^1 B_F(r)B_F(r)'dr\right)\sigma_\epsilon^2 \oplus \lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=1}^{T-h}W_t\epsilon_{t+h}\right)\left(\sum_{t=1}^{T-h}W_t\epsilon_{t+h}\right)',$$

Case 2: If $\{F_t\}$ is a mixture of I(0) and I(1) and independent of $\{\epsilon_{t+h}\}$, and $\{W_t, \epsilon_{t+h}\}$ are I(0),

$$\Sigma_{\epsilon L} = \Sigma_F.\sigma_\epsilon^2 \oplus \lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=1}^{T-h}W_t\epsilon_{t+h}\right)\left(\sum_{t=1}^{T-h}W_t\epsilon_{t+h}\right)',$$

Case 3: If $\{F_t\}$ are all I(1) and independent of $\{\epsilon_{t+h}\}$, and $\{W_t\}$ are I(1) and independent of $\{\epsilon_{t+h}\}$, we have

$$\Sigma_{\epsilon L} = \left(\int_0^1 B_{FW}(r)B'_{FW}(r)dr\right)\sigma_\epsilon^2,$$

Case 4: If $\{F_t\}$ is a mixture of I(0) and I(1) and independent of $\{\epsilon_{t+h}\}$, and $\{W_t\}$ are I(1) and independent of $\{\epsilon_{t+h}\}$, we have

$$\Sigma_{\epsilon L} = \Sigma_F \sigma_\epsilon^2 \left(\int_0^1 B_W(r)B'_W(r)dr\right),$$

where σ_ϵ^2 is the variance of $\{\epsilon_t\}$, $B_{FW}(r) = (B'_F(r), B'_W(r))'$, and $B_W(r)$ is the weak limit of $\frac{1}{\sqrt{T}}W_{[Tr]}$.

Theorem 2.1. *Suppose that Assumptions 2.1–2.6 hold and that $T/N \to 0$. Let δ and the OLS estimator $\hat{\delta}$ be as in (2.7) and (2.8), respectively. Then, as $(N,T) \to \infty$, we have $D_T(\hat{\delta}-\delta) \xrightarrow{d} N\big(0,\Sigma_\delta\big)$ and $H \oplus I \xrightarrow{d} \Psi$, where $\Sigma_\delta = (\Psi')^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}\Sigma_L^{-1}\Psi^{-1}$, and Σ_L and $\Sigma_{\epsilon L}$ are defined in Assumption 2.6.*

The appearance of the scaling matrix $D_T = diag(TI_p, T^{1/2}I_q, TI_m)$ in Theorem 2.1 shows that the estimators $\hat{\alpha}, \hat{\beta}$, and $\hat{\omega}$ converge at the rates $T, T^{1/2}$, and T, respectively. Consequently, the limiting distribution in this theorem reduces to the following known corresponding results in: (a) Bai and Ng [2002] for the FAR model with $I(0)$ variables only, and (b) Choi [2017] for the FAR model with $I(1)$ variables only.

Since the limiting normal distribution in Theorem 2.1 has mean zero, it follows that the use of generated factors, instead of the original unobservable factors in the model, does not affect the consistency of the estimators. To arrive at this result, we used the assumption $T/N \to 0$, which ensures that the effect of the error resulting from factor estimation becomes negligible in the limit. By contrast, if the assumption $T/N \to 0$ is replaced by $T/N \to c$, for some $0 < c < \infty$, then the limiting normal distribution would have a nonzero mean, and hence the estimator would not be consistent. In fact, Gonçalves and Perron [2014] showed, for the case when all the variables are $I(0)$, that if $\sqrt{T}/N \to c$ for some $0 < c < \infty$, then there would be asymptotic bias.

The unknown covariance matrix Σ_δ may be estimated consistently by

$$\hat{\Sigma}_\delta = \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{\epsilon}_{t+h}^2 \hat{L}_t \hat{L}_t' D_T^{-1} \right) \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} . \tag{2.9}$$

This estimator is robust against heteroskedasticity in the regression error. For the special case of homoskedastic errors, a simpler consistent estimator of Σ_δ is

$$\hat{\Sigma}_\delta = \hat{\sigma}_\epsilon^2 \left(D_T^{-1} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' D_T^{-1} \right)^{-1} , \tag{2.10}$$

where $\hat{\sigma}_\epsilon^2 = T^{-1} \sum_{t=1}^{T-h} \hat{\epsilon}_{t+h}^2$ is an estimator of the variance of regression errors.

2.2.4 Prediction interval

Let $Y_{T+h|T}$ denote the conditional mean $E[Y_{T+h} \mid \mathcal{F}_T]$ where $\mathcal{F}_T$ is the information up to time T, and let $(\hat{\delta}, \hat{L}_t)$ be as in (2.7) and (2.8). Then, an estimator of $Y_{T+h|T}$ is $\hat{Y}_{T+h|T} = \hat{\delta}' \hat{L}_T$; similarly, $\hat{Y}_{T+h} = \hat{\delta}' \hat{L}_T$ is also a point forecast of Y_{T+h}. In this section, we obtain a confidence interval for $Y_{T+h|T}$ and a prediction interval for Y_{T+h}. These are obtained by using the next theorem.

Theorem 2.2. *Suppose that Assumptions 2.1–2.6 hold. Further, suppose also that* $\sqrt{N}||D_{1T}^{-2}|| \to 0$ *and* $T/N \to 0$ *as* $N, T \to \infty$, *and that* $(\hat{\Sigma}_\delta, \hat{\Sigma}_{\tilde{F}})$ *is a given consistent estimator of* $(\Sigma_\delta, \Sigma_{\tilde{F}})$. *Then, we have* $\hat{B}_T^{-1/2}\{\hat{Y}_{T+h|T} - Y_{T+h|T}\} \xrightarrow{d} N(0,1)$ *as* $N, T \to \infty$, *where* $\hat{B}_T = \left[\hat{L}_T D_T^{-1} \hat{\Sigma}_\delta D_T^{-1} \hat{L}_T' + N^{-1} \hat{\theta}' \hat{\Sigma}_{\tilde{F}} \hat{\theta}\right]$ *is a consistent estimator of the asymptotic variance, denoted* B_T, *of the conditional forecasting error that appears in the numerator.*

To provide some insight into the foregoing suggested form for $\hat{B}_T$, note that the forecast error can be expressed as

$$\hat{Y}_{T+h|T} - Y_{T+h|T} = (\hat{\delta} - \delta)' \hat{L}_T + \theta' H^{-1} (\tilde{F}_T - H F_T). \tag{2.11}$$

This forecast error is the sum of two components: the first is due to the error in estimating δ and the other is due to estimating the factor. Theorem 2.1 and Lemma 2.3 show that each of these is asymptotically normal with mean zero. It turns out that these two are essentially asymptotically independent and hence the asymptotic variances simply add up.

To use Theorem 2.2 for inference in empirical studies, we need a suitable consistent estimator $(\hat{\Sigma}_\delta, \hat{\Sigma}_{\tilde{F}})$ of $(\Sigma_\delta, \Sigma_{\tilde{F}})$. For, $\hat{\Sigma}_\delta$, we may use the estimators in (2.9) or (2.10) depending on the assumptions. Using Lemmas 2.2 and 2.3, a consistent estimator of $\Sigma_{\tilde{F}_T}$ is

$$\hat{\Sigma}_{\tilde{F}_T} = \tilde{V}_{NT}^{-1} \hat{\Gamma}_T \tilde{V}_{NT}^{-1}, \tag{2.12}$$

where $\hat{\Gamma}_T$ is an estimator of the asymptotic covariance matrix of $\left(N^{-1/2}\Lambda' e_t\right)$, and $\tilde{V}_{NT}$ was defined as a diagonal matrix of the largest r eigenvalues of XX' multiplied by $D_{1T}^{-2}N^{-1}$.

To make use of the form in (2.12), we need a feasible estimator of Γ_T. As suggested by Bai and Ng [2006], depending on the assumptions, $\hat{\Gamma}_T$ may take one of the following forms:

$$(a)\ \hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N} \hat{e}_{iT}^2 \tilde{\lambda}_i \tilde{\lambda}_i', \quad (b)\ \hat{\Gamma}_T = \hat{\sigma}_e^2 \frac{1}{N}\sum_{i=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_i', \quad (c)\ \hat{\Gamma}_T = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_j' \hat{e}_{iT}\hat{e}_{jT}, \tag{2.13}$$

where $\hat{e}_{it} = X_{it} - \tilde{\lambda}_i' \tilde{F}_t$. For cross sectionally uncorrelated idiosyncratic errors, the first two forms of $\hat{\Gamma}_T$ in (2.13) are suitable. If the errors are homoskedastic and $E(e_{it}^2) = \sigma_e^2$, say, then σ_e^2 can be estimated by $\hat{\sigma}_e^2 = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{e}_{it}^2$ and the second form in (2.13) would be suitable. The third form in (2.13) is suitable for estimating the asymptotic variance of generated factors when the idiosyncratic errors have cross sectional correlation. By combining the aforementioned estimators, we obtain a feasible estimator $\hat{B}_T$. Using these, a $100(1-\alpha)\%$ confidence interval for the conditional mean $Y_{T+h|T}$ is

$$\left(\hat{Y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T} \quad , \quad \hat{Y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T}\right), \tag{2.14}$$

where $z_{1-\alpha/2}$ stands for $(1-\alpha/2)th$ quantile of the standard normal distribution.

Next, consider constructing a forecast interval for Y_{T+h}. To this end, first note that the forecast error is

$$\hat{\epsilon}_{T+h} = \hat{Y}_{T+h|T} - Y_{T+h} = \hat{L}_T'\left(\hat{\delta} - \delta\right) + \theta' H^{-1}\left(\tilde{F}_T - HF_T\right) - \epsilon_{T+h}. \tag{2.15}$$

Therefore, the limiting distribution of forecast error also depends on the distribution of the regression error ϵ_{T+h}. Let us suppose that $\epsilon_{T+h} \sim N(0, \sigma_\epsilon^2)$. Then, it follows from Theorem 2.2 that the forecasting error $\hat{\epsilon}_{T+h}$ is also asymptotically normal with mean zero and variance $B_T + var(\epsilon)$. Let $\hat{\sigma}_\epsilon^2$ denote a consistent estimator of σ_ϵ^2; for example, if $\{\varepsilon_t\}$ are iid, then we may choose $\hat{\sigma}_\epsilon^2 = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$. Then, an asymptotic 95% prediction interval for Y_{T+h} is

$$\left(\hat{Y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T + \hat{\sigma}_\epsilon^2} \quad , \quad \hat{Y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T + \hat{\sigma}_\epsilon^2}\right). \tag{2.16}$$

Remark 5: The predictors $\hat{L}_t$ $(= (\tilde{F}_t', , W_t')')$, the OLS estimator $\hat{\delta}$ in (2.8), and the point predictor $\hat{Y}_{t+h}$ do not depend on which components of F_t are identified as $I(0)$ and which ones as $I(1)$. By contrast, the rate matrix D_T, $\hat{\Sigma}_\delta$ in (2.10), the matrix $\hat{B}_T$, and the prediction interval (2.16) based on $\hat{Y}_{t+h|T}$, depends on which components of F_t are identified as $I(0)$ and which ones as $I(1)$.

2.3 Simulation study: finite sample properties

Design of the simulation study

The data generating process [DGP] for the FAR is

$$Y_{t+1} = \alpha F_{1t} + \beta F_{2t} + \omega Y_t + \epsilon_{t+1} \qquad (t = 1, \ldots, T-1) \tag{2.17}$$

$$F_{1t} = F_{1,t-1} + v_t, \qquad (v_t, F_{2,t}) \sim MVN(0, C), \quad C = (1, \rho \mid \rho, 1). \tag{2.18}$$

For ρ in (2.18), we considered the values $0.0, 0.5$, and 0.9. For the error term ϵ_t, we considered both homoskedastic and heteroskedastic cases - see below. The $T \times N$ panel data set was generated by

$$X_{it} = \lambda_i^{(1)} F_{1t} + \lambda_i^{(2)} F_{2t} + e_{it}, \tag{2.19}$$

with the λ_i's drawn from $N(0,1)$ and the error terms $\{e_{it}\}$ as stated below.

Sixteen combinations of $[T, N]$ were considered with $T = 30, 50, 100, 200$ and $N = 30, 50, 100, 200$. The parameter values were set at $\alpha = 0.5, \beta = 1$, and $\omega = 0.5$. We considered the following three different DGPs for each ρ: (1) DGP1: $e_{it} \sim N(0,1)$ and $\epsilon_t \sim N(0,1)$; (2) DGP2: $e_{it} \sim N(0,1)$ and $\epsilon_t \sim N\left(0, 3^{-1}F_{2t}^2\right)$; (3) DGP3: $e_{it} \sim N(0, \sigma_i^2)$ and $\epsilon_t \sim N\left(0, 3^{-1}F_{2t}^2\right)$.

Among the three DGP's, DGP1 is the simplest for which the errors are i.i.d. in both time and cross-section dimensions. In DGP2, $var(\epsilon_t)$ depends on the stationary factor, and hence conditionally heteroskedastic over time. In DGP3, $var(\epsilon_t)$ varies over time and $var(e_{it})$ is distributed uniformly over $[0.5, 1.5]$; therefore, the average variance is the same as that for the homoskedastic case. All simulation estimates are based on 5000 repeated samples. Since the FAR model has a lag term, we adopted a burn-in period of 100 time units; thus, for generating each sample, the first 100 observations were discarded. We used the $\hat{\Sigma}_\delta$ in (2.10) and (2.9) for DGP1 and {DGP2, DGP3}, respectively.

The results are reported in two parts. The first part reports the simulation results for the coverage rates of 95% prediction intervals in (2.16). The second part compares the out-of-sample forecast performance of the method based on the mixture-FAR model developed in this chapter with the methods based on a nonstationary-FAR and the AR(4) models.

2.3.1 Coverage rates of prediction intervals

Table 2.1 reports the coverage rates of 95% prediction intervals for Y_{T+1}. These are based on the assumption that the regression errors are normal. The coverage rates of these intervals range from 88% to 98% with most of them being close to the nominal 95%. Therefore, in terms of coverage rates, the prediction intervals performed quite well.

Table 2.1: Coverage rates (%) of 95% prediction intervals for one-step ahead forecasts.

		$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
	T\N	30	50	100	200	30	50	100	200	30	50	100	200
DGP1	30	93	93	91	92	94	94	92	92	94	93	92	91
	50	91	94	93	90	92	94	92	89	92	94	92	90
	100	90	89	91	88	92	90	91	89	91	90	91	88
	200	92	90	89	90	92	90	89	89	93	89	90	89
DGP2	30	97	96	95	94	95	95	95	94	97	98	97	95
	50	96	96	98	94	97	97	94	94	98	97	96	95
	100	97	95	97	95	96	95	95	95	96	96	95	95
	200	96	96	96	94	96	96	94	94	97	96	95	94
DGP3	30	97	97	96	94	97	96	96	95	97	98	95	96
	50	97	95	96	96	97	95	96	96	98	96	96	95
	100	97	96	95	97	96	96	96	95	96	95	96	96
	200	95	97	95	96	95	95	95	95	96	94	95	96

Note: The assumed error distribution of the forecasting model is normal.

2.3.2 Performance of mixture-FAR relative to non-stationary FAR

In this subsection, we evaluate the performance of the mixture-FAR method relative to nonstationary-FAR method. Recall that the nonstationary-FAR model (see Choi 2017), requires all the variables in the FAR model to be $I(1)$. We evaluate the out-of-sample forecast performance in terms of *out-of-sample R-square*, denoted R^2_{os}, defined as

$$R^2_{os} = 1 - \left(\sum_{t=T_1+1}^{T} (Y_t - \hat{Y}_t)^2 \right) \left(\sum_{t=T_1+1}^{T} (Y_t - \tilde{Y}_t)^2 \right)^{-1}, \tag{2.20}$$

where $\hat{Y}_t$ = prediction using the mixture-FAR model, $\tilde{Y}_t$ = prediction using the competing or reference model, the observations from the first $(T_1 + j)$ time points are used for estimating the model, and the observation at time $T_1 + j + 1$ is used for evaluating the performance of the out-of-sample forecast at time $(T_1 + j + 1)$ $(j = 0, \ldots, T - (T_1 + 1))$. Thus, R^2_{os} is a measure of how well the mixture-FAR performed during the period $[T_1 + 1, T]$, relative to the competing model. As an example, if $R^2_{os} = 0.1$ (respectively, $R^2_{os} = -0.1$) then an estimate of the MSE of prediction for the mixture-FAR model is 10% lower (respectively, higher) than that for the competing model. In this simulation study, we chose the nonstationary-FAR as the competing model. Throughout this chapter, forecast evaluations are based on expanding windows for the estimation period, unless the contrary is made clear.

In this part of the study, we considered the DGP1 with $T = [60, 90, 150, 300]$ and $T_1 = [40, 60, 100, 200]$. First, we consider forecasting a nonstationary series Y_t using the mixture-FAR model and compare it with the corresponding nonstationary-FAR. Table 2.2 provides the results for this comparison. It is evident that the mixture-FAR method performed significantly better than the competing nonstationary-FAR in terms of R^2_{os}. As an example, the entry 0.43 in the cell for $T_1 = 40$ and $N = 30$ shows that the MSE of prediction for the mixture-FAR model is 43% lower than that for the nonstationary-FAR model. The table also shows that, for every case considered in Table 2.2, the MSE of prediction for the proposed mixture-FAR model is at least 33% lower than that for the nonstationary-FAR model. Therefore, in this simulation study, the improvement of the mixture-FAR model compared to the nonstaionary-FAR model is substantial.

In summary, for every case that we studied, the mixture-FAR model performed significantly better than its corresponding competitor, the nonstationary-FAR, for forecasting a nonstationary variable.

Table 2.2: Values of R^2_{os} for the performance of mixture-FAR model relative to the corresponding nonstationary-FAR model.

$T_1 \backslash N$	$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
	30	50	100	200	30	50	100	200	30	50	100	200
40	0.43	0.44	0.45	0.46	0.41	0.42	0.42	0.43	0.33	0.34	0.35	0.35
60	0.45	0.46	0.46	0.47	0.42	0.43	0.44	0.45	0.36	0.36	0.36	0.37
100	0.46	0.47	0.48	0.48	0.44	0.45	0.46	0.46	0.37	0.38	0.38	0.39
200	0.47	0.48	0.48	0.49	0.45	0.46	0.46	0.46	0.38	0.39	0.39	0.39

Note: The values in this table are for one-step ahead out-of-sample forecasts. The forecasting variable Y is nonstationary $[I(1)]$.

2.3.3 Evaluation of the proposed factor estimation method

Simulation design

Considered DGP for the factor model is similar to the one in Bai and Ng [2005],

$$X_{it} = \lambda_i^{(1)'} E_t + \lambda_i^{(2)'} G_t + e_{it}, \tag{2.21}$$

where $\lambda_i^{(1)}, \lambda_i^{(2)} \sim N(0,1), e_{it} \sim N(0,1)$, sets of nonstationary and stationary factors are generated as,

$$E_t = \alpha_1 E_{t-1} + U_{1t}, \tag{2.22}$$

$$G_t = \alpha_2 G_{t-1} + U_{2t}, \tag{2.23}$$

with $\alpha_1 = 1, \alpha_2 = 0$. The errors $(U_{1t}, U_{2t}) \sim MVN(0, C), C = (1, 0|0, 1)$. Nine combinations of $[T, N]$ were considered with $T = [50, 100, 200]$ and $N = [50, 100, 200]$. The total number of factors, r, is considered to be 3 while p changes from 1 to 3. All the simulation results are based on 5000 replications.

Using IC and IPC from Bai and Ng [2002], we calculate the total number of factors, $\hat{r}$, for 5000 replications and estimate the percentage of times $\hat{r} = r$ is obtained. The results are reported in Table 2.3. According to the results, we can see that the information criterion (IC) in Bai and Ng [2002] accurately estimates number of total factors in the panel data set that contains both stationary and nonstationary variables.

Results for estimating the number of nonstationary factors using IPC from Bai [2004] is reported in Table 2.4. According to the results, we can see that when the latent factors is

Table 2.3: Percentage for obtaining $\hat{r} = r$ when panel contains both I(0) and I(1) factors

			IC			PIC		
r	p	T\N	50	100	200	50	100	200
3	3	50	100	100	100	23.4	64.9	95.9
		100	100	100	100	99.4	100	100
		200	100	100	100	100	100	100
	2	50	100	100	100	21.9	59.7	94.2
		100	100	100	100	99.2	100	100
		200	100	100	100	100	100	100
	1	50	100	100	100	17.6	54.1	91.5
		100	100	100	100	99.1	100	100
		200	100	100	100	100	100	100

r is the number of total factors and p is the number of I(1) factors in the model.

a mixture of I(0) and I(1) series, IPC well estimate the number, p, of nonstationary factors. However, performance of this method is low when the set of factors are fully nonstationary.

Table 2.4: Percentage for obtaining $\hat{p} = p$ when panel contains both I(0) and I(1) factors

r	p	T/N	50	100	200
3	3	50	48.8	66.14	73.36
		100	58.72	81.42	92.46
		200	62.24	90.78	98.54
	2	50	86.94	92.82	94.46
		100	89.76	97.12	98.96
		200	91.40	98.70	99.90
	1	50	94.12	100	100
		100	93.02	100	100
		200	94.22	100	100

2.4 Empirical application

In this section we apply the mixture-FAR model for forecasting two key non-stationary macroeconomic variables, namely the GDP and the Industrial Production [IP]. Since we use quarterly data, to capture the cyclic patterns of the response variable, we start with a basic AR(4) model and augment it with factors to construct FAR models. For each model, we compute two sets of prediction intervals, one is based on the asymptotic distribution of the standardized forecast and the other is based on the t-percentile bootstrap; the validity of the bootstrap is yet to be established.

Recall that the forecast of the conditional mean, as shown in Theorem 2.2, is asymptotically normal; this result does not require that the functional form of the error distribution be known. The indications are that a residual based bootstrap method is likely to be valid for constructing confidence interval for the forecast conditional mean and for constructing a prediction interval. With this in mind, we expanded the simulation study in the previous section and evaluated the coverage rates of residual based t-percentile bootstrap prediction intervals when the error distribution is normal and when it is t with 5 degrees of freedom. Since we used residual based bootstrap, it does not assume that the error distribution is known. The results are presented in the two tables at the end of the Appendix to this Chapter; they show that the coverage rates of the bootstrap prediction intervals are close to the nominal level. Therefore, the indications are that it is reasonable to compare the bootstrap intervals with those based on (2.16). We compare and contrast the out-of-sample forecasting performance of the mixture-FAR model with the corresponding non-stationary-FAR and the AR(4) models. To quantify the out-of-sample forecasting performance, we use R^2_{os} defined in (2.20).

2.4.1 Data description

The data were collected from FRED-MD and FRED-QD; these are well-known databases for macroeconomic variables containing monthly and quarterly data, respectively. The latter contains 246 US macroeconomic time series for the period 1959:Q1 to 2018:Q4, with a total of 240 (T=240) observations. We excluded 36 variables because there were missing observations, and used a balanced panel for 210 variables. The variables are categorized into 14 groups; for more details, see the updated appendix for FRED-QD at $https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QDappendix.pdf$. The macroeconomic variables in this balanced panel data set are further categorized into two levels of aggregation, 110 "high-aggregates" and 100 "sub-aggregates". The panel data for $N = 100$ sub-aggregates were used for estimating the factors; to this end, we used principal components analysis [PCA]. These sub-aggregates consist of both stationary and nonstationary time series; for each series, the transformation to $I(0)$ is given in the third row of the data set.

2.4.2 Estimation of factors

We follow a two-step approach by adapting the methods proposed in Bai and Ng [2002] and Bai [2004] for choosing an 'optimal' number of factors. The method proposed by Bai and Ng [2002], which is based on information criteria (IC) for stationary panel, led to the total number of factors being eight for the set of 100 sub-aggregate macroeconomic variables. The corresponding $IC(k)$ for $k = 1, \ldots, 8$ are given in Table 2.5. Then, we applied the method based on *iterated panel criterion* to the original (level) panel dataset proposed by Bai [2004], and concluded that the number of I(1) factors is four. The corresponding $IPC(k)$ for $k = 1, \ldots, 8$ are given in Table 2.6. Thus, we have four I(1) and four I(0) factors in the model.

Finally, we applied the Augmented-Dickey Fuller [ADF] test to each of the factors, and observed that if the factors are ordered according to the magnitude of the eigenvalues, then the factors $\{1, 2, 4, 5\}$ are $I(1)$ and the remaining ones, namely $\{3, 6, 7, 8\}$, are $I(0)$. The overall trends exhibited by the factors in Figure 2.1 are consistent with the aforementioned observation that the factors $\{1, 2, 4, 5\}$ are $I(1)$ and the other four are $I(0)$. Figure 2.1 shows time series plots of the estimated factors. Plots of the two high-aggregate macroeconomic variables, GDP and IP, are presented in Figure 2.2. This figure shows that both variables are nonstationary. Therefore, the mixture-FAR method developed in this chapter is potentially applicable for forecasting GDP and IP.

Table 2.5: Values of $IC(k)$ in the determination of total number of factors

k	1	2	3	4	5	6	7	8
$IC1(k)$	-2.91	-3.65	-4.15	-4.52	-4.74	-5.03	-5.31	**-5.49**
$IC2(k)$	-2.91	-3.64	-4.13	-4.50	-4.71	-5.00	-5.28	**-5.45**
$IC3(k)$	-2.93	-3.68	-4.19	-4.58	-4.81	-5.12	-5.41	**-5.61**

Table 2.6: Values of $IPC(k)$ in the determination of number of nonstationary factors

k	1	2	3	4	5	6	7	8
$IPC1(k)$	1.38	0.89	0.45	**0.21**	0.29	0.23	0.25	0.28
$IPC1(k)$	1.42	0.96	0.57	**0.40**	0.44	0.46	0.52	0.58
$IPC1(k)$	1.43	0.98	0.59	**0.43**	0.47	0.49	0.55	0.62

For the data set in this empirical study, the panel data model and the forecasting model with a mixture of stationary and nonstationary factors take the forms,

$$X_{it} = \lambda_i' F_t + e_{it} = \lambda_i^{(1)\prime} E_t + \lambda_i^{(2)\prime} G_t + e_{it},$$
$$Y_{t+h} = \alpha' \tilde{E}_t + \beta' \tilde{G}_t + \omega_1 Y_t + \omega_2 Y_{t-1} + \omega_3 Y_{t-2} + \omega_4 Y_{t-3} + \epsilon_{t+h} \quad (h > 0),$$

where $\tilde{E}_t$ is the set of four nonstationary generated factors, $\tilde{G}_t$ is the set of four stationary generated factors, and $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ are their coefficients.

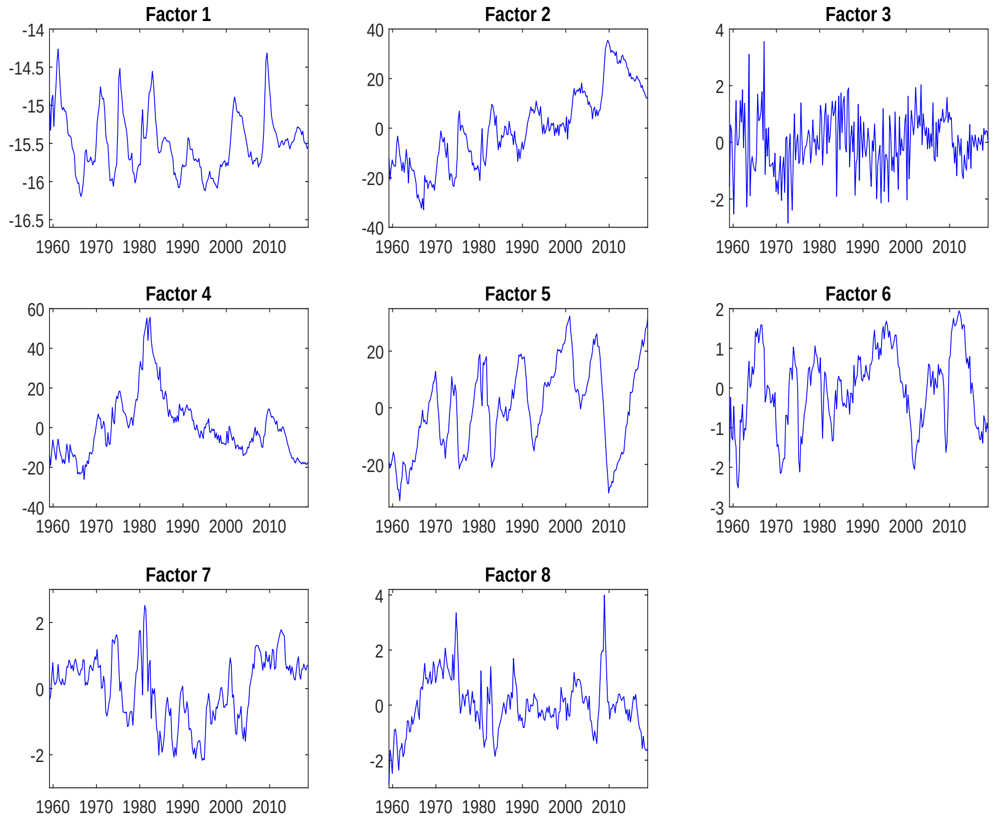


Figure 2.1: A plot of the eight generated PCA factors from the panel data set of 100 variables

Assessing the out-of-sample forecast performance of mixture-FAR method

We considered the following four models; the basic AR(4), and three mixture-FAR models obtained by augmenting the AR(4) with a mixture of the $I(0)$ and the $I(1)$ factors:

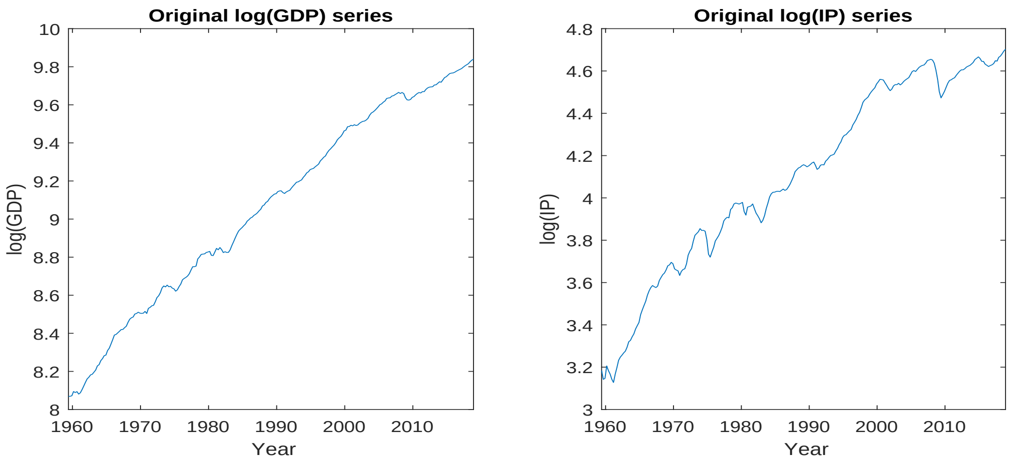


Figure 2.2: Time series plots of $\log(GDP)$ and $\log(IP)$ for 1959:Q1 – 2018:Q4.

$$\begin{aligned}
\text{Model 1}: Y_{t+h} &= \alpha'\tilde{E}_t + \beta'\tilde{G}_t + \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h}, \\
\text{Model 2}: Y_{t+h} &= \alpha_1'\tilde{E}_t + \alpha_2'\tilde{E}_{t-1} + \beta_1'\tilde{G}_t + \beta_2'\tilde{G}_{t-1} + \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h}, \\
\text{Model 3}: Y_{t+h} &= \sum_{i=0}^{3}\alpha_{1+i}'\tilde{E}_{t-i} + \beta_1'\tilde{G}_t + \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h}, \\
\text{Model 4}: Y_{t+h} &= \sum_{i=0}^{3}\omega_{1+i}Y_{t-i} + \epsilon_{t+h}.
\end{aligned}$$

Model 4, the basic AR(4) model, is used as the benchmark for forecast comparison. In the FAR model, we have augmented the estimated factors with AR(4). Therefore, to compare and contrast the contribution from estimated factors in the forecasting, AR(4) is a suitable benchmark. This benchmark model is justified in the literature on forecasting univariate time series using different model specifications such as FAR model, AR models, autoregressive distributed lag models; see Stock and Watson [1998a and 2005]. Model 1 is the AR(4) model augmented with the eight generated factors; this is a mixture-FAR model. Model 2 is Model 1 augmented with one lag of each generated factor. Model 3 is Model 1 augmented with three lags of each nonstationary factor and the stationary factor with no lags. Thus, Model 1 is nested in Models 2 and 3. Much of this section focuses on comparing and contrasting the out-of-sample forecast performance of Models 1 to 4. The plots in Figures 2.3 and 2.4 show the one-step ahead out-of-sample forecasts of $\log(GDP)$ and $\log(IP)$, respectively, for the period 2006:Q1 - 2018:Q4 with the initial estimation period being 1959:Q1 - 2005:Q4. These plots indicate that the out-of-sample predictions of the two I(1) variables, GDP and IP, generally appear to be good.

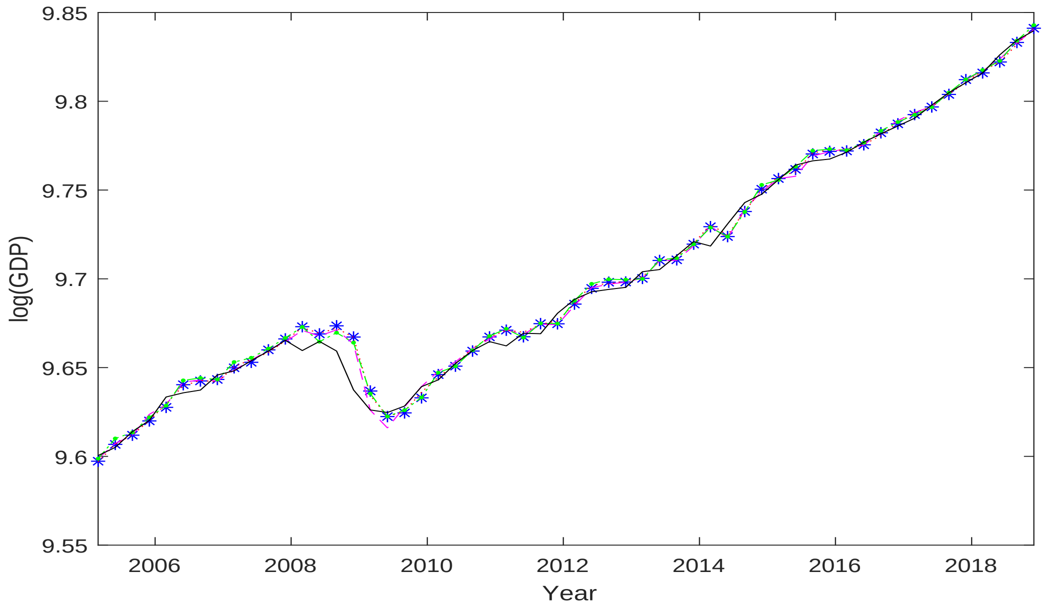


Figure 2.3: The observed $\log(GDP)$, and plots of one-step ahead out-of-sample forecasts of $\log(GDP)$ for 2006:Q1 – 2018:Q4. Blue * : predicted series with Model 1. Magenta dashed line - -: predicted series with Model 2. Red dotted line: predicted series with Model 3. Green - . . : predicted series with AR(4) model. Black solid line: the observed data.

One-step ahead out-of-sample forecast evaluations

We assess the relative predictive performance of the four models in terms of the out-of-sample R^2, denoted R^2_{os}, defined in (2.20). In this subsection, AR(4) is used as the basic benchmark; later we consider a nonstationary-FAR as the benchmark. We considered three different first estimation periods; we also evaluated the forecasts with a rolling window of 40 years for the estimation period, but there were no improvements in the forecast performance, compared to the expanding window. The values of R^2_{os} in Table 2.7 indicate that the mixture-FAR model, Model 2, outperforms the benchmark model, AR(4) for forecasting GDP and IP. Overall, the results presented in Table 2.7 indicate that Model 2 performs better than the other two models as well.

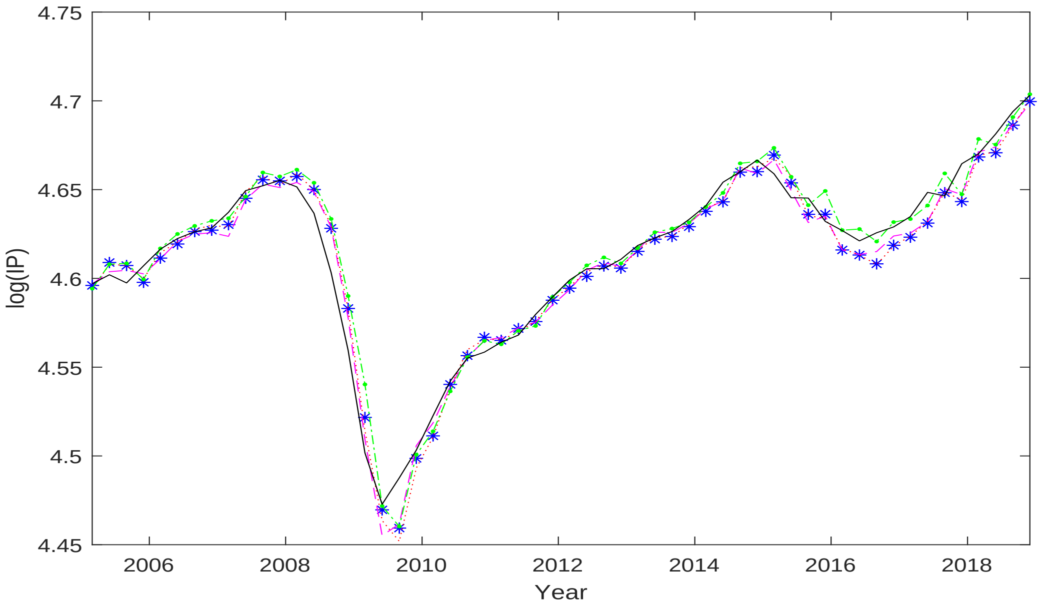


Figure 2.4: The observed $\log(IP)$, and plots of one-step ahead out-of-sample forecasts of $\log(IP)$ for 2006:Q1 – 2018:Q4. Blue * : predicted series with Model 1. Magenta dashed line - -: predicted series with Model 2. Red dotted line: predicted series with Model 3. Green - . . : predicted series with AR(4) model. Black solid line: the observed data.

2.4.3 Forecast evaluations with long forecast horizons

So far, we considered one-step ahead forecasts. Next, we evaluate and compare Models 1, 2, and 3 in terms of the accuracy of their forecasts over longer forecast horizons, instead of just one-step ahead. We computed the forecasts with the first estimation period being 1959:Q1 - 1999:Q4. We calculated R^2_{os} for different specifications of mixture-FAR relative to AR(4). Figures 2.5 and 2.6 provide plots of R^2_{os} against the forecast horizon h. These figures show that for forecasting GDP and IP over horizons longer than 12 months ($h > 12$), the mixture-FAR models performed better than the AR(4) model. Combining these observations with those in the previous sections, we conclude that the mixture-FAR model performed better than the AR(4) for forecasting GDP and IP over short and long horizons.

2.4.4 Mixture-FAR vs nonstationary-FAR models for forecasting GDP and IP

For forecasting a nonstationary variable such as GDP and IP, a nonstationary-FAR model, wherein all the variables including the factors are nonstationary, has been proposed in the literature (Choi 2017). As in the earlier sections, we refer to this model as a nonstationary-FAR

Table 2.7: Performance of mixture-FAR relative to AR(4), in terms of R^2_{os}.

	$\log(GDP)$			$\log(IP)$		
First estimation period	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1959:Q1 - 1999:Q4	-0.01	0.07	-0.00	-0.14	0.10	-0.21
1959:Q1 - 2005:Q4	-0.08	0.11	-0.01	0.17	0.33	0.11
1959:Q1 - 2008:Q4	0.09	0.16	0.18	0.06	0.28	-0.01

Note: Each entry in the table is the R^2_{os} for a given mixture-FAR model relative to the AR(4) model. These are for one-step ahead forecasts with expanding windows for estimation. The out-of-sample forecast period starts from the end of the estimation period and extends to the end of 2018.

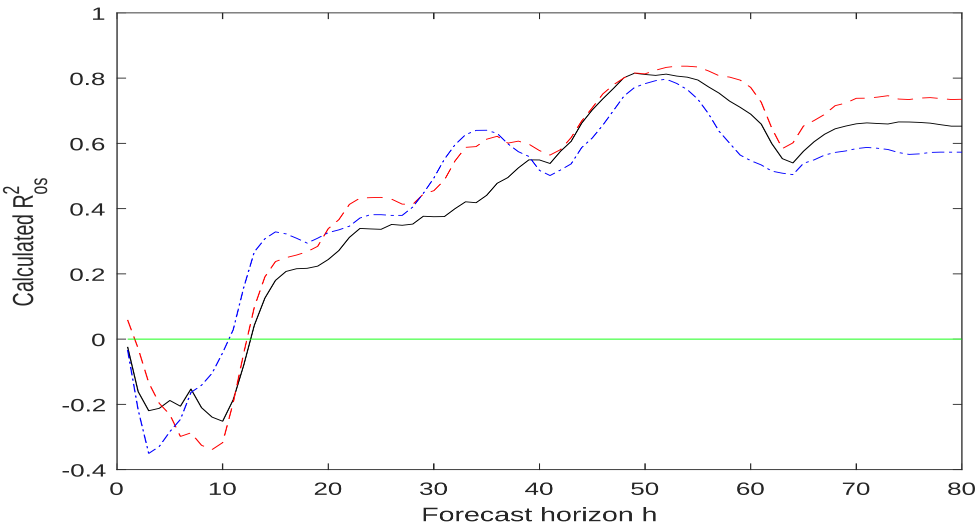


Figure 2.5: Performance of mixture-FAR relative to AR(4) for long-term forecast of $\log(GDP)$. The graph is a plot of R^2_{os} against the forecast horizon h. Black solid line is the R^2_{os} for the model 1 relative to AR(4). Red dash line is the R^2_{os} for model 2 relative to AR(4). Blue dotted dash line is the R^2_{os} for the model 3 relative to AR(4).

model. To implement this method, first we performed principal component analysis on XX', and chose only the nonstationary factors for use as predictors in the prediction model (2.2). We wish to compare the aforementioned nonstationary-FAR method with the mixture-FAR method. To this end, we chose Model 1 as our mixture-FAR model; for the nonstationary-FAR model, we chose

$$\text{Model 5[Nonstationary-FAR]:} \quad Y_{t+h} = \beta' \tilde{E}_t + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+h}.$$

The values of R^2_{os} for these two models are provided in Table 2.8. Consider the entry 0.31 in the column for $\log(IP)$. This says that the *sum of squares of forecast error*[SSFE] for $\log(IP)$ over the period 2009–2018 is 31% lower for mixture-FAR compared to the nonstationary-FAR.

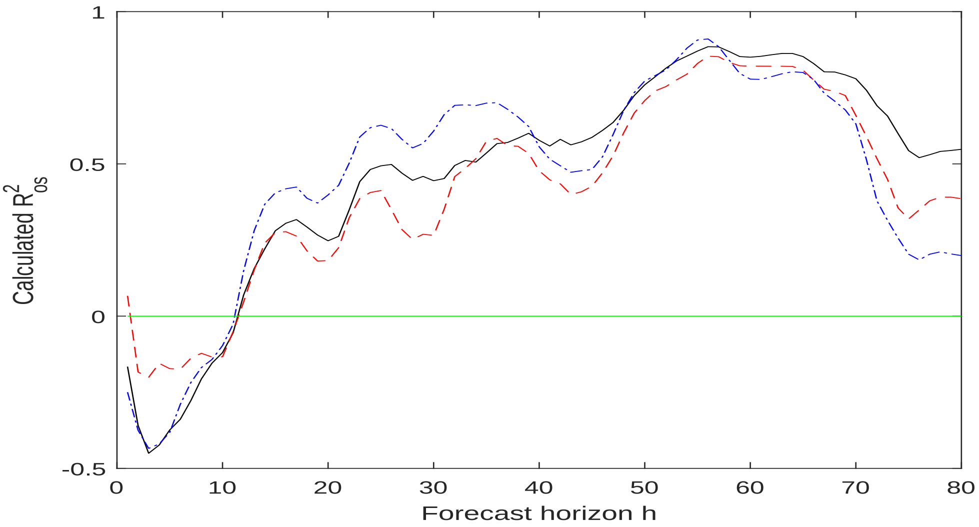


Figure 2.6: Performance of mixture-FAR relative to AR(4) for forecasting $\log(IP)$ over long forecast horizons. Each curve is a plot of R^2_{os} for a given mixture-FAR relative to AR(4), against the forecast horizon h. Black solid line: Model 1. Red dashed line: Model 2. Blue dotted dash line: Model 3.

Table 2.8: Values of R^2_{os} for mixture-FAR compared to the nonstationary-FAR.

First estimation period	$log(GDP)$	$log(IP)$
1959:Q1 - 1999:Q4	0.21	0.17
1959:Q1 - 2005:Q4	0.21	0.39
1959:Q1 - 2008:Q4	0.50	0.31

Note: The mixture-FAR is Model 1 and the nonstationary-FAR is Model 5.

The table also shows that the SSFE for $\log(GDP)$ over the period 2009–2018 is 50% lower for mixture-FAR compared to nonstationary-FAR. In fact, Table 2.8 shows that the mixture-FAR method proposed in this chapter performed significantly better than the method based on a nonstationary-FAR for forecasting GDP and IP.

2.4.5 Prediction intervals

We computed 95% prediction intervals, based on the asymptotic results in Section 2.2 and the bootstrap, for one-step-ahead prediction intervals for $\log(GDP)$ and $\log(IP)$ with expanding window for the estimation period. Assuming that the regression error ϵ_t in (2.2) is normally distributed, we constructed the asymptotic theory-based point-wise 95% prediction intervals for $\log(GDP)$ and $\log(IP)$ for the out-of-sample period 2006:Q1 to 2018:Q4. These intervals are shown in Figures 2.7 and 2.8 together with the observed values of $\log(GDP)$ and $\log(IP)$.

We also estimated the symmetric bootstrap t-percentile prediction intervals using residual bootstrap; to this end, we used 399 bootstrap replications. These are also shown in Figures 2.7 and 2.8. One important difference between the asymptotic theory-based and bootstrap prediction intervals is that the latter does not assume that the error distribution has a known functional form.

Figure 2.7 shows that, except for a very short interval around the crisis period 2009, the observed values of GDP lie within the two prediction intervals. Figure 2.8 shows that every observed value of IP lies within the two prediction intervals. Overall, the bootstrap prediction interval is narrower than the one based on the asymptotic distribution of the forecast, for GDP and IP. The bootstrap prediction interval for IP around the crisis period is large; this may be because the financial and economic crisis introduced large fluctuations in IP. Overall, both prediction intervals have high coverage rates for GDP and IP.

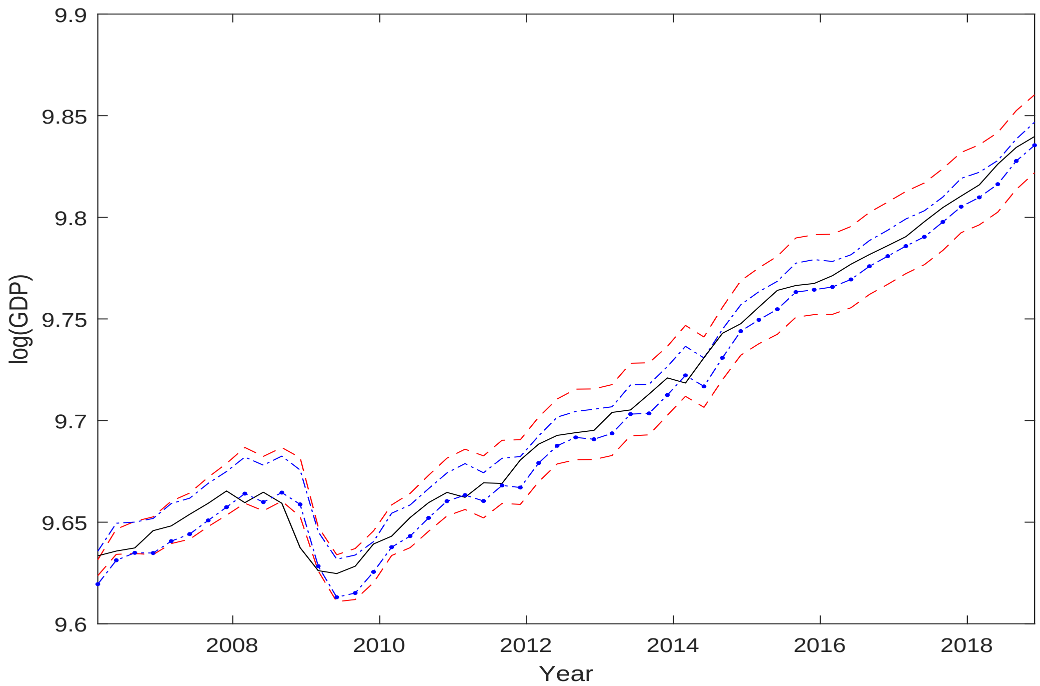


Figure 2.7: One-step ahead point-wise 95% prediction intervals for $\log(GDP)$ using mixture-FAR. The solid black line in the middle is a plot of the observed values of $\log(GDP)$. Red dashed line: the asymptotic theory based prediction interval. Blue dotted and dashed line: the bootstrap prediction interval.

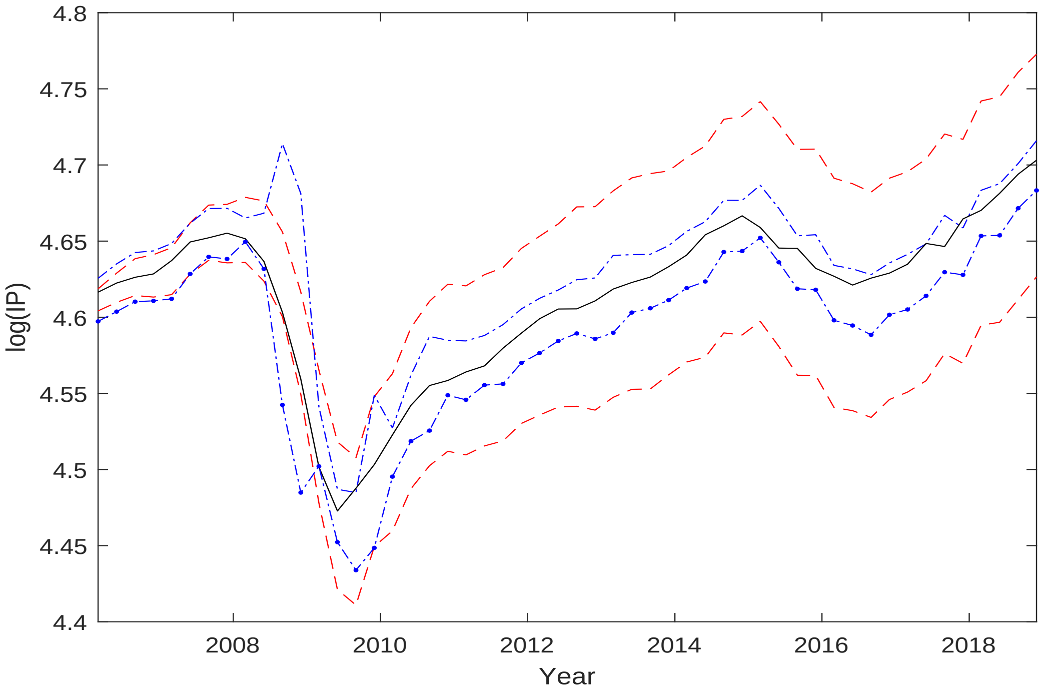


Figure 2.8: One-step ahead point-wise 95% prediction intervals for IP using mixture-FAR. The solid black line in the middle is a plot of the observed values of IP. Red dashed line: the asymptotic theory based prediction interval. Blue dotted and dashed line: the bootstrap prediction interval.

Remark 7: It is not possible to specify a guide on how a researcher may specify an FAR model in empirical studies. Nevertheless, it is worth noting that the reason for studying mixture FAR model in this thesis is to improve the current literature by providing a method to combine the currently available two methods, one for FAR with stationary factors only and the other for FAR with nonstationary factors only. Since mixture FAR models are more challenging to use, a desirable approach to specifying a suitable mixture model is to start with a suitable FAR model with only stationary factors, and another FAR model with only nonstationary factors. Then, the mixture FAR approach could use the aforementioned preliminary analyses as a starting points for mixture FAR approach.

2.5 Conclusion

This chapter developed methodology for forecasting nonstationary macroeconomic variables, such as GDP and industrial production[IP], when a set of panel data is available for a large number of potential predictors. We propose to estimate a small number of factors using the panel data, and use them as predictors for forecasting. The factors are chosen such that they contain a large proportion of the information in the large number of potential predictors. The validity of this method for forecasting has been established in the literature when all the variables are stationary (Bai and Ng 2006), and also when they are all nonstationary (Choi 2017), but not when they consist of a mixture of stationary and nonstationary variables. Typically, a set of panel data on such a large number of macroeconomic variables would contain mixture of stationary and nonstationay variables, which turned out be also the case in our empirical example. Therefore, the method developed in this chapter is of practical significance. To use the estimated mixture of stationary and nonstationary factors as predictors, and construct an asymptotically valid prediction interval, this chapter developed the methodology. In our simulation study, the mixture-FAR method developed in this chapter performed significantly better than the one that uses only nonstationary variables. We applied the proposed method for forecasting GDP and IP. We assessed the out-of-sample forecast performance of the mixture-FAR relative to the corresponding nonstationary-FAR and the AR(4) models. We observed that the mixture-FAR model performed significantly better than the aforementioned two competing methods. In summary, this chapter provides an improved method of forecasting a nonstationary variable using information from stationary and nonstationary variables.

2.6 Appendices

Appendix A defines mathematical symbols. Appendix B gives the proofs of Lemmas 1–3 listed in the main paper and then contains the proofs of the auxiliary lemmas used to prove the main results in Appendix C. Simulation results for bootstrap are presented in Section D.

2.6.1 Appendix A: Mathematical Symbols

First, we introduce the following notation and mathematical symbols:

$$\gamma_{st} = E\left(N^{-1}\textstyle\sum_{i=1}^{N} e_{is}e_{it}\right), \quad \zeta_{st} = N^{-1}\sum_{i=1}^{N}\left(e_{is}e_{it} - E(e_{is}e_{it})\right), \tag{A1.1}$$

$$\eta_{st} = N^{-1}F_s'\Lambda' e_t, \qquad \xi_{st} = N^{-1}F_t'\Lambda' e_s \tag{A1.2}$$

$$\tau_{ij,ts} = E(e_{it}e_{js}), \qquad \tau_{ij,t} = \tau_{ij,tt}; \tag{A1.3}$$

for $i,j = 1,\dots N; s,t = 1,\dots,T$, and let us note that $\eta_{st} = \xi_{ts}$.

Variables:

$Y = (Y_t)_{T\times 1} = (Y_1,\dots,Y_T)'$, uni-variate dependent variable,

$W = (W_{it})_{T\times m} = \begin{pmatrix} W_{11} & \dots & W_{m1} \\ \vdots & \ddots & \vdots \\ W_{1T} & \dots & W_{mT} \end{pmatrix} = (W_1',\dots,W_T')'$; m is the number of observable (nonstationary) regressors,

$F = (F_{it})_{T\times r} = \begin{pmatrix} F_{11} & \dots & F_{r1} \\ \vdots & \ddots & \vdots \\ F_{1T} & \dots & F_{rT} \end{pmatrix} = \begin{pmatrix} F_1' \\ \vdots \\ F_T' \end{pmatrix}$; $F' = (F_1,\dots,F_T)$,

$\Lambda = (\lambda_{ij})_{N\times r} = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{r1} \\ \vdots & \ddots & \vdots \\ \lambda_{1N} & \dots & \lambda_{rN} \end{pmatrix} = \begin{pmatrix} \lambda_1' \\ \vdots \\ \lambda_N' \end{pmatrix}$; $\Lambda' = (\lambda_1,\dots,\lambda_N)$,

$e = (e_{it})_{T\times N} = \begin{pmatrix} e_{11} & \dots & e_{N1} \\ \vdots & \ddots & \vdots \\ e_{1T} & \dots & e_{NT} \end{pmatrix} = \begin{pmatrix} e_1' \\ \vdots \\ e_T' \end{pmatrix}$; $e' = (e_1,\dots,e_T)$.

Scaling matrices:

$D_{1T} = \begin{pmatrix} TI_p & 0 \\ 0 & T^{1/2}I_q \end{pmatrix}$; p, q are the number of nonstationary and stationary factors respectively.

$$D_T = \begin{pmatrix} TI_p & 0 & 0 \\ 0 & T^{1/2}I_q & 0 \\ 0 & 0 & TI_m \end{pmatrix}.$$

$$\|D_{1T}\| = O(T), \quad \|D_{1T}\|^{-1} = O(T^{-1}), \quad \|D_{1T}^{-1}\| = O(T^{-1/2}), \quad \|D_{1T}^{-1}\|^{-1} = O(T^{1/2}).$$

Matrices from Assumption 2.1:

$D_{1T}^{-1} \sum_{t=1}^{T} F_t F_t' D_{1T}^{-1} \xrightarrow{d} \Sigma_{F_{(r\times r)}}$; a positive definite random matrix,

$N^{-1}\Lambda'\Lambda \to \Sigma_{\Lambda_{(r\times r)}}$ or $N^{-1}\Lambda'\Lambda \xrightarrow{p} \Sigma_{\Lambda_{(r\times r)}}$; a positive definite non-random matrix.

Matrices with eigenvalues:

$\tilde{V}_{NT} = \text{diag}(\tilde{Z}) \times N^{-1}D_{1T}^{-2}; \tilde{Z} = (\tilde{v}_1, \ldots, \tilde{v}_r)$; the r largest eigenvalues of XX',

$\tilde{V}_{NT}^{*} = \text{diag}(\tilde{Z}^{*}) \times D_{1T}^{-2}; \tilde{Z}^{*} = (v_1^{*}, \ldots, v_r^{*})$; the r largest eigenvalues of $N^{-1}F\Lambda'\Lambda F'$,

$V = \text{diag}(Z); Z = (v_1, \ldots, v_r)$; the eigenvalues of $\Sigma_\Lambda \Sigma_F$.

Rotation matrix:

$H = N^{-1}\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda$; $r \times r$ matrix.

Parameters:

$\delta = \begin{pmatrix} \alpha' & \beta' & \omega' \end{pmatrix}' = \begin{pmatrix} \theta' & \omega' \end{pmatrix}'; \theta' = \begin{pmatrix} \alpha' & \beta' \end{pmatrix}$, where

$\alpha : p \times 1$ column vector of parameters corresponding to nonstationary factors,

$\beta : q \times 1$ column vector of parameters corresponding to stationary factors,

$\omega : m \times 1$ column vector of parameters corresponding to observable regressors.

$\hat{\delta}$ denotes the OLS estimated coefficients.

Information criteria – Bai and Ng [2002]:

To estimate the total number of factors, r, we minimized the following criteria functions proposed by Bai and Ng [2002]:

$$IC_{P1}(k) = ln(V_1(k, \tilde{F}^k)) + k\left(\frac{N+T}{NT}\right)ln\left(\frac{NT}{N+T}\right), \tag{A1.4}$$

$$IC_{P2}(k) = ln(V_1(k, \tilde{F}^k)) + k\left(\frac{N+T}{NT}\right)lnC, \tag{A1.5}$$

$$IC_{P3}(k) = ln(V_1(k, \tilde{F}^k)) + k\left(\frac{N+T}{NT}\right)\left(\frac{lnC}{C}\right), \tag{A1.6}$$

where $C = min[T, N]$, $\hat{\sigma}^2$ be a consistent estimate of $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E(e_{it})^2$, k be the number of factors that we estimated by the principal components and $\tilde{F}^k$ be the matrix of all

k factors. The sum of squared residuals of the stationary data matrix ΔX on k factors is given as,

$$V_1(k, \tilde{F}^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\Delta X_{it} - \tilde{\lambda}_i'^k \tilde{F}_t^k\right)^2, \tag{A1.7}$$

and $lnV_1(k, \tilde{F}^k) + kg(N, T)$ be the loss function where $g(N, T)$ is the penalty for over-fitting. The optimal total number of factors is the inter that minimized the information criteria.

Integrated Panel Criteria – Bai [2004]:

To estimate the number of I(1) factors, we use the integrated panel information criteria proposed by Bai [2004] to the estimated factors using level data.

$$IPC_{P1}(k) = V_2(k, \tilde{F}^k) + k\hat{\sigma}^2 \alpha_T \left(\frac{N+T}{NT}\right) ln\left(\frac{NT}{N+T}\right), \tag{A1.8}$$

$$IPC_{P2}(k) = V_2(k, \tilde{F}^k) + k\hat{\sigma}^2 \alpha_T \left(\frac{N+T}{NT}\right) lnC, \tag{A1.9}$$

$$IPC_{P3}(k) = V_2(k, \tilde{F}^k) + k\hat{\sigma}^2 \alpha_T \left(\frac{N+T-k}{NT}\right) lnNT, \tag{A1.10}$$

where $\alpha_T = T/[4lnln(T)]$ and

$$V_2(k, \tilde{F}^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \tilde{\lambda}_i'^k \tilde{F}_t^k\right)^2. \tag{A1.11}$$

The optimal number of nonstationary factors in the model is $\hat{p} = \arg\min_{0 \leqslant k \leqslant Kmax} IPC(k)$. Thus, we can obtain the optimal number of I(0) factors, $\hat{q} = \hat{r} - \hat{p}$.

2.6.2 Appendix B: Proofs of the Auxiliary Lemmas

Before we provide the proofs of Theorems 1 and 2 in Appendix C below, we introduce a series of auxiliary lemmas.

Lemma A. 1. *Suppose that Assumptions 2.1–2.3 are satisfied. Then we have*

(i) $T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \gamma_{st}^2 \leqslant M$,

(ii) $T^{-1} N^{-1} \sum_{t=1}^{T} \|\Lambda' e_t\|^2 = O_P(1)$.

Proof. Part (i) Let $\rho_{st} = \gamma_{st}(\gamma_{ss}\gamma_{tt})^{-1/2}$ $(s = 1, \ldots, T; t = 1, \ldots, T)$. Then $|\rho_{st}| \leq 1$, and by Assumption 2.2(ii), we have $|\gamma_{ss}| \leqslant M$.

$$\frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \gamma_{st}^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma_{ss}\gamma_{tt}\rho_{st}^2 \leq M \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |\gamma_{ss}\gamma_{tt}|^{1/2} |\rho_{st}| = M \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |\gamma_{st}| \leq M^2,$$

since $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_{st}| \leqslant M$ by Assumption 2.2(ii).

Part (ii) Using the Assumptions 2.1(iv) and 2.2(iii), we obtain

$$\frac{1}{TN}\sum_{t=1}^{T}\left\|\Lambda' e_t\right\|^2 = \frac{1}{TN}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}\lambda_i e_{it}\right)'\left(\sum_{j=1}^{N}\lambda_j e_{jt}\right) = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i'\lambda_j e_{it}e_{jt},$$

$$E\left(\frac{1}{TN}\sum_{t=1}^{T}\left\|\Lambda' e_t\right\|^2\right) = E\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i'\lambda_j e_{it}e_{jt}\right) \leqslant \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\lambda_i'\lambda_j\right|\left|E(e_{it}e_{jt})\right|$$

$$= \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\lambda_i'\lambda_j\right|\left|\tau_{ij,t}\right| \leqslant M^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| \leq M^3.$$

Therefore, $T^{-1}N^{-1}\sum_{t=1}^{T}\|\Lambda' e_t\|^2 = T^{-1}\sum_{t=1}^{T}\left\|N^{-1/2}\sum_{i=1}^{N}e_{it}\lambda_i\right\|^2 = O_P(1)$. ■

Let $\tilde{V}_{NT}^{*} = \operatorname{diag}(v_1^*,\ldots,v_r^*)D_{1T}^{-2}$, where $v_1^*,\ldots,v_r^*$ are the r non-zero eigenvalues of $N^{-1}F\left(\Lambda'\Lambda\right)F'$. Similarly, let $V = \operatorname{diag}(v_1,...,v_r)$ where $\{v_1,\ldots,v_r\}$ are the eigenvalues of $\Sigma_\Lambda\Sigma_F$.

Lemma A. 2. *Suppose that Assumptions 2.1–2.3 are satisfied. Then, as $N,T \to \infty$ with $T/N \to 0$,*

(i) $\left\|D_{1T}^{-2}\tilde{F}'\left(\frac{XX'}{N}\right)\tilde{F}D_{1T}^{-2} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| = o_P(1)$,

(ii) $N^{-1}D_{1T}^{-2}\tilde{F}'XX'\tilde{F}D_{1T}^{-2} = \tilde{V}_{NT} \xrightarrow{d} V$,

(iii) $\|H\| = \left\|\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\frac{\Lambda'\Lambda}{N}\right\| = O_P(1)$.

Proof. ***Part (i)*** Let $W = D_{1T}^{-2}\tilde{F}'\left(XX'N^{-1}\right)\tilde{F}D_{1T}^{-2} - D_{1T}^{-2}\tilde{F}'F\left(\Lambda'\Lambda N^{-1}\right)F'\tilde{F}D_{1T}^{-2}$. By substituting $X = F\Lambda' + e$, we obtain,

$$\begin{aligned} W &= N^{-1}D_{1T}^{-2}\tilde{F}'\left(F\Lambda' + e\right)\left(F\Lambda' + e\right)'\tilde{F}D_{1T}^{-2} - N^{-1}D_{1T}^{-2}\tilde{F}'F\Lambda'\Lambda F'\tilde{F}D_{1T}^{-2} \\ &= N^{-1}D_{1T}^{-2}\tilde{F}'\left\{F\Lambda'\Lambda F' + F\Lambda' e' + e\Lambda F' + ee' - F\Lambda'\Lambda F'\right\}\tilde{F}D_{1T}^{-2} \\ &= N^{-1}D_{1T}^{-2}\tilde{F}'\left\{F\Lambda' e' + e\Lambda F' + ee'\right\}\tilde{F}D_{1T}^{-2} = D_{1T}^{-2}\tilde{F}'W^*, \end{aligned}$$

where $W^* = N^{-1}\left\{F\Lambda' e' + e\Lambda F' + ee'\right\}\tilde{F}D_{1T}^{-2}$. Using Cauchy Schwarz inequality, we have $\|W\|^2 \leqslant \left\|D_{1T}^{-2}\tilde{F}\right\|^2\|W^*\|^2$. Consider W^*. Let $W_t^* = N^{-1}D_{1T}^{-2}\tilde{F}'\left\{ee_t + F\Lambda' e_t + e\Lambda F_t\right\}$ for $t = 1,\ldots,T$. Then,

$$\begin{aligned} \|W^*\|^2 &= \left\|(W^*)'\right\|^2 = \left\|\frac{1}{N}D_{1T}^{-2}\tilde{F}'\left(e\Lambda F' + F\Lambda' e' + ee'\right)\right\|^2 \\ &= \left\|\frac{1}{N}D_{1T}^{-2}\tilde{F}'\left(e\Lambda\left(F_1,\ldots,F_T\right) + F\Lambda'\left(e_1,\ldots,e_T\right) + e\left(e_1,\ldots,e_T\right)\right)\right\|^2 \\ &= \left\|(W_1^*,\ldots,W_T^*)\right\|^2 = \sum_{t=1}^{T}\|W_t^*\|^2, \end{aligned}$$

where $W_t^* = N^{-1}D_{1T}^{-2}\tilde{F}'(e\Lambda F_t + F\Lambda' e_t + ee_t)$.
From the fact that $(a+b+c+d)^2 \leqslant 4\left(a^2+b^2+c^2+d^2\right)$, we obtain,

$$
\begin{aligned}
\|W^*\|^2 &= \sum_{t=1}^{T}\left\|N^{-1}D_{1T}^{-2}\tilde{F}'\left\{ee_t + F\Lambda' e_t + e\Lambda F_t\right\}\right\|^2 \\
&= \sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2 \\
&\leqslant 4\sum_{t=1}^{T}\left(\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2 + \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2 + \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2 + \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2\right) \\
&= 4\left(\sum_{t=1}^{T}a_t + \sum_{t=1}^{T}b_t + \sum_{t=1}^{T}c_t + \sum_{t=1}^{T}d_t\right) \qquad (say),
\end{aligned}
$$

where

$$
a_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2, b_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2, c_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2, d_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2. \tag{B.1}
$$

Now, consider each term separately. By Cauchy Schwarz inequality,

$$
\frac{1}{T}\sum_{t=1}^{T}a_t = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2 \leqslant \|D_{1T}^{-1}\|^2\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2
$$

From the normalization condition $D_{1T}^{-2}\tilde{F}'\tilde{F} = I_r$, we have $\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2 = O_P(1)$. Using Lemma (B.1), we have $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{st}^2 = O(1)$. Therefore, we have $T^{-1}\sum_{s=1}^{T}a_t = O_P\left(\|D_{1T}^{-1}\|^2\right)$.
Since $\left\|D_{1T}^{-1}\right\|^2 = O(T^{-1})$, we obtain

$$
\sum_{t=1}^{T}a_t = O_P\left(T\|D_{1T}^{-1}\|^2\right) = O_P(1). \tag{B.2}
$$

Now, consider the second term. Using the Cauchy Schwarz inequality, we have

$$
\frac{1}{T}\sum_{t=1}^{T}b_t = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2 \leqslant \|D_{1T}^{-1}\|^2\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\underbrace{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\zeta_{st}|^2}_{K_1\ \text{(say)}}.
$$

Using Assumption 2.2(v), we have $E\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{is}e_{it} - E\left(e_{is}e_{it}\right)\right]\right|^4 = N^2E\left|\zeta_{st}\right|^4 \leqslant M < \infty$. Therefore,

$$
E\left|K_1\right| \leqslant \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\zeta_{st}\right|^2 \leqslant \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{M_1}{N} = O\left(\frac{T}{N}\right),
$$

where M_1 is a finite constant. Thus, together with the fact that, $T||D_{1T}^{-2}|| = O(1)$, we obtain

$$\frac{1}{T}\sum_{t=1}^{T} b_t = O_P\left(\|D_{1T}^{-1}\|^2\right) O_P\left(\frac{T}{N}\right) = O_P\left(\frac{1}{N}\right).$$

Hence, we have

$$\sum_{t=1}^{T} b_t = O_P\left(\frac{T}{N}\right). \tag{B.3}$$

Now, consider the third term. Using Cauchy Schwartz inequality, Assumption 2.1, normalization condition on factors, and Lemma (B.1), we obtain,

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} c_t &= \frac{1}{T}\textstyle\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\frac{F_s'\Lambda' e_t}{N}\right\|^2 \\ &\leqslant \frac{1}{N}\underbrace{\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\Lambda' e_t}{\sqrt{N}}\right\|^2}_{O_P(1)}\underbrace{\left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s F_s'\right\|^2}_{K_2 \text{ (say)}}, \end{aligned} \tag{B.4}$$

where we have

$$K_2 = O_P\left(\textstyle\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\right) O_P\left(\sum_{s=1}^{T}\left\|D_{1T}^{-1}F_s\right\|^2\right) = O_P(1)$$

Therefore, $T^{-1}\sum_{t=1}^{T} c_t = O_P\left(N^{-1}\right)$. Hence we have,

$$\sum_{t=1}^{T} c_t = O_P\left(\frac{T}{N}\right). \tag{B.5}$$

Similarly, we can show that $\sum_{t=1}^{T} d_t = O_P\left(TN^{-1}\right)$. Using Cauchy Schwarz inequality,

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} d_t = \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\xi_{st}\tilde{F}_s\right\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\frac{F_t'\Lambda' e_s}{N}\tilde{F}_s\right\|^2 \\ \leqslant \frac{1}{TN}\left(\sum_{t=1}^{T}\left\|D_{1T}^{-1}F_t\right\|^2\right)\left(\frac{1}{N}\sum_{s=1}^{T}\|\Lambda' e_s\|^2\right)&\left(\sum_{s=1}^{T}\left\|D_{1T}^{-1}\tilde{F}_s\right\|^2\right) = O_P\left(\frac{1}{N}\right). \end{aligned}$$

Thus,

$$\sum_{t=1}^{T} d_t = O_P\left(\frac{T}{N}\right). \tag{B.6}$$

From equations (B.2)–(B.6), we have

$$\|W^*\|^2 \leqslant 4\left(\sum_{t=1}^{T} a_t + \sum_{t=1}^{T} b_t + \sum_{t=1}^{T} c_t + \sum_{t=1}^{T} d_t\right) = O_P(1) + O_P\left(\frac{T}{N}\right).$$

Thus,

$$\|W\| \leqslant \left\|D_{1T}^{-1}\right\| \left\|D_{1T}^{-1}\tilde{F}'\right\| \|W^*\| = O\left(\left\|D_{1T}^{-1}\right\|\right) O_P(1) \left\{O_P(1) + O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right)\right\} = o_P(1),$$

as $N, T \to \infty$.

Part (ii) By the definition of eigenvalues and eigenvectors, $N^{-1}XX'\tilde{F}D_{1T}^{-2} = \tilde{F}\tilde{V}_{NT}$, and the normalization condition, $D_{1T}^{-2}\tilde{F}'\tilde{F} = I_r$, we have, $N^{-1}D_{1T}^{-2}\tilde{F}'XX'\tilde{F}D_{1T}^{-2} = \tilde{V}_{NT}$. Now, together with part (i), we can write

$$\left\|\tilde{V}_{NT} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| = o_P(1). \tag{B.7}$$

Furthermore, recall that $\tilde{V}_{NT}^*$ is the diagonal matrix with largest r eigenvalues of $F\left(\Lambda'\Lambda N^{-1}\right)F'$ multiplied by D_{1T}^{-2} and $\tilde{F}^*$ be the corresponding eigenvector matrix such that $D_{1T}^{-2}\tilde{F}^{*\prime}\tilde{F}^* = I_r$. Then, using similar arguments as in *part (i)*, we have

$$\left\|D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2} - D_{1T}^{-2}\tilde{F}^{*\prime}F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}^*D_{1T}^{-2}\right\| = o_P(1). \tag{B.8}$$

Again, by the definition of eigenvalues and eigenvectors, and the normailzation condition, we may write $N^{-1}D_{1T}^{-2}\tilde{F}^{*\prime}(F\Lambda'\Lambda F')\tilde{F}^*D_{1T}^{-2} = \tilde{V}_{NT}^*$. Then, equation (B.8) gives

$$\left\|\tilde{V}_{NT}^* - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| = o_P(1). \tag{B.9}$$

Therefore, we have,

$$\begin{aligned}\left\|\tilde{V}_{NT} - \tilde{V}_{NT}^*\right\| &= \left\|\tilde{V}_{NT} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2} - \left(\tilde{V}_{NT}^* - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right)\right\| \\ &\leqslant \left\|\tilde{V}_{NT} - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| + \left\|\tilde{V}_{NT}^* - D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)F'\tilde{F}D_{1T}^{-2}\right\| \\ &= o_P(1).\end{aligned}$$

Thus, we have $\tilde{V}_{NT} = \tilde{V}_{NT}^* + o_P(1)$.

Since the r largest eigenvalues of $F\left(\Lambda'\Lambda N^{-1}\right)F'$ are the same as those of $\left(\Lambda'\Lambda N^{-1}\right)F'F$, elements of $\tilde{V}_{NT}^*$ are equal to the eigenvalues of $(\Lambda'\Lambda N^{-1})F'F$ multiplied by D_{1T}^{-2}. Then, under the Assumption 2.1, we have that $\tilde{V}_{NT}^*$ converges, in distribution, to V, positive definite diagonal matrix with eigenvalues of $\Sigma_\Lambda\Sigma_F$. Therefore, $\tilde{V}_{NT} \xrightarrow{d} V$ and $\tilde{V}_{NT}^{-1} = O_P(1)$.

Part (iii) This part directly holds form the *Part (ii)*, and Assumption 2.1.

$$\|H\| = \left\|\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\frac{\Lambda'\Lambda}{N}\right\| \leq \left\|\tilde{V}_{NT}^{-1}\right\| \left\|D_{1T}^{-2}\tilde{F}'F\right\| \left\|\frac{\Lambda'\Lambda}{N}\right\| = O_P(1).$$

■

The next lemma proves the consistency of factors, and is stated in the main paper.

Lemma A. 3. *Suppose that Assumptions 2.1–2.3 are satisfied. Let* $\delta_{NT}^{-1} = \max\left[N^{-1/2}, ||D_{1T}^{-1}||\right]$. *Then, there exists an* $(r \times r)$ *non-singular matrix* H*, called a rotation matrix, such that*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 = O_P\left(\delta_{NT}^{-2}\right).$$

Proof. By the identity $\tilde{F} = N^{-1}X\tilde{\Lambda}\tilde{V}_{NT}^{-1}$, as stated in Bai and Ng(2002), and $\tilde{\Lambda} = X'\tilde{F}D_{1T}^{-2}$, we have $\tilde{F} = N^{-1}X(X'\tilde{F}D_{1T}^{-2})\tilde{V}_{NT}^{-1}$. Using $H = \tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\left\{N^{-1}\Lambda'\Lambda\right\}$ and expanding XX', we obtain

$$\begin{aligned}
\tilde{F} - FH' &= \frac{1}{N}XX'\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} - F\left(\tilde{V}_{NT}^{-1}D_{1T}^{-2}\tilde{F}'F\frac{\Lambda'\Lambda}{N}\right)' \\
&= \frac{1}{N}XX'\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} - F\frac{\Lambda'\Lambda}{N}F'\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} \\
&= \left\{\frac{1}{N}XX' - \frac{F\Lambda'\Lambda F'}{N}\right\}\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} \\
&= \frac{1}{N}\left\{F\Lambda'e' + e\Lambda F' + ee'\right\}\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1}.
\end{aligned}$$

Since $\tilde{F} = \left(\tilde{F}_1, \ldots, \tilde{F}_T\right)'$, and $F = (F_1, \ldots, F_T)'$, in vector notation, the above equation becomes

$$\begin{aligned}
\tilde{F}_t - HF_t &= \tilde{V}_{NT}^{-1}D_{1T}^{-2}\frac{1}{N}\tilde{F}'\left\{ee_t + F\Lambda'e_t + e\Lambda F_t\right\} \\
&= \tilde{V}_{NT}^{-1}D_{1T}^{-2}\left\{\frac{1}{N}\sum_{s=1}^{T}\tilde{F}_s e_s' e_t + \frac{1}{N}\sum_{s=1}^{T}\tilde{F}_s F_s'\Lambda' e_t + \frac{1}{N}\sum_{s=1}^{T}\tilde{F}_s e_s'\Lambda F_t\right\}.
\end{aligned}$$

Then we have,

$$\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1}\left\{D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\}, \quad \text{(B.10)}$$

where $\gamma_{st}, \zeta_{st}, \eta_{st}$ and ξ_{st} as defined before.

Since $\tilde{V}_{NT}^{-1} = O_P(1)$ by Lemma B.2, and the fact that $(a+b+c+d)^2 \leqslant 4\left(a^2+b^2+c^2+d^2\right)$, we obtain $\left\|\tilde{F}_t - HF_t\right\|^2 \leq 4\left(a_t + b_t + c_t + d_t\right)$ where,

$$a_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right\|^2, b_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right\|^2, c_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right\|^2, d_t = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\|^2.$$

Then, $T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 \leqslant T^{-1}\sum_{t=1}^{T} 4\left(a_t + b_t + c_t + d_t\right)$. In the proof of Lemma (B.2) (equations (B.2)–(B.6)), we have

$$\frac{1}{T}\sum_{t=1}^{T} a_t = O_P\left(\left\|D_{1T}^{-2}\right\|\right), \quad \frac{1}{T}\sum_{t=1}^{T} b_t = O_P\left(\frac{1}{N}\right), \tag{B.11}$$

$$\frac{1}{T}\sum_{t=1}^{T} c_t = O_P\left(\frac{1}{N}\right), \quad \frac{1}{T}\sum_{t=1}^{T} d_t = O_P\left(\frac{1}{N}\right). \tag{B.12}$$

Thus, we obtain

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2 = \left\{O_P\left(||D_{1T}^{-2}||\right) + O_P\left(N^{-1}\right)\right\} = \max[O_P\left(||D_{1T}^{-2}||\right), O_P\left(N^{-1}\right)] = O_P\left(\delta_{NT}^{-2}\right),$$

where $\delta_{NT}^{-1} = \max\left[N^{-1/2}, ||D_{1T}^{-1}||\right]$. ■

Lemma A. 4. *Suppose that Assumptions 2.1–2.4 satisfied. Then, as $N, T \to \infty$, we have*

(i) $A_{1t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right)$

(ii) $A_{2t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)$,

(iii) $A_{3t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} = O_P\left(N^{-1/2}\right)$,

(iv) $A_{4t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)$.

Proof. ***Part (i)*** Let us write

$$A_{1t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} = D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T} HF_s\gamma_{st} = A_{11t} + HA_{12t}, \tag{B.13}$$

and consider each part separately. Using Cauchy Schwartz inequality, Assumption 2.4, Lemma (B.1), and Lemma (2.1), and the fact that $\left\|D_{1T}^{-2}\right\| = O(T^{-1})$, we have,

$$\begin{aligned}
\|A_{11t}\| &= \left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\right\| \\
&\leqslant \left\|D_{1T}^{-2}\right\|\underbrace{\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}}_{O_P\left(\sqrt{T}\delta_{NT}^{-1}\right)}\underbrace{\left(\sum_{s=1}^{T}|\gamma_{st}|^2\right)^{1/2}}_{O(1)} = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right),
\end{aligned}$$

$$\begin{aligned}
E\|A_{12t}\| &= E\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s\gamma_{st}\right\| \leqslant \left\|D_{1T}^{-2}\right\|\sum_{s=1}^{T}E\left\|F_s\gamma_{st}\right\| \leqslant ||D_{1T}^{-2}||\sum_{s=1}^{T}\left(E||F_s||^2\right)^{1/2}\left(E\left|\gamma_{st}\right|^2\right)^{1/2} \\
&\leqslant M||D_{1T}^{-2}||\sum_{s=1}^{T}|\gamma_{st}| = O(||D_{1T}^{-2}||),
\end{aligned}$$

since we assume that $\left(E||F_t||^4\right) \leqslant M$ and $\sum_{s=1}^{T}|\gamma_{st}| \leqslant M$ for some finite constant M in the Assumptions 2.1 and 2.4(i). Then, together with $\|H\| = O_P(1)$ from Lemma (B.2), and the fact that $\delta_{NT}^{-1} = \max[N^{-1/2}, ||D_{1T}^{-1}||]$, we obtain

$$A_{1t} = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right) + O_P(1)O_P\left(||D_{1T}^{-2}||\right) = O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right).$$

Part (ii) Similar to Part (i), we can decompose A_{2t} as follows:

$$A_{2t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} = D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}HF_s\zeta_{st} = A_{21t} + HA_{22t}. \tag{B.14}$$

By Cauchy Schwarz inequality and Assumption 2.2(v), we have

$$\|A_{21t}\| = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right\| \leqslant \|D_{1T}^{-2}\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\underbrace{\left(\sum_{s=1}^{T}\|\zeta_{st}\|^2\right)^{1/2}}_{A_{23t}},$$

where we have

$$\begin{aligned} E(A_{23t}^2) &= E\left(\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(e_{is}e_{it} - E\left(e_{is}e_{it}\right)\right)\right\|^2\right) \\ &\leq \frac{1}{N}\left(\sum_{s=1}^{T}E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{is}e_{it} - E\left(e_{is}e_{it}\right)\right)\right|^2\right) = O\left(\frac{T}{N}\right). \end{aligned}$$

Then, together with $T||D_{1T}^{-2}|| = O(1)$, we get

$$A_{21t} = O\left(||D_{1T}^{-2}||\right)O_P\left(T^{1/2}\delta_{NT}^{-1}\right)O_P\left(T^{1/2}N^{-1/2}\right) = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$$

Consider the second term of A_{2t}. Again by Cauchy Schwarz inequality and $T||D_{1T}^{-2}|| = O(1)$,

$$\begin{aligned} E\left\|A_{22t}\right\| &= E\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s\zeta_{st}\right\| \leqslant \left\|D_{1T}^{-2}\right\|\sum_{s=1}^{T}\left(E\left\|F_s\right\|^2\right)^{1/2}\left(E\left|\zeta_{st}\right|^2\right)^{1/2} \\ &\leqslant M\left\|D_{1T}^{-2}\right\|\sum_{s=1}^{T}\left(E\left|\zeta_{st}\right|^2\right)^{1/2} = O\left(\frac{||D_{1T}^{-1}||}{\sqrt{N}}\right), \end{aligned}$$

thus we have

$$A_{22t} = O_P\left(\frac{||D_{1T}^{-1}||}{\sqrt{N}}\right),$$

where the second inequality holds as we assume $E\left\|F_s\right\|^4 \leqslant M < \infty$ in Assumption 2.1(ii), and the second equality holds by following the Assumption 2.1(v).

Then, together with Lemma (B.2), $\|H\| = O_P(1)$, we obtain

$$A_{2t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right) = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$$

Part (iii) Again, as in the Part (i), let us write,

$$A_{3t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} = D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st} + HD_{1T}^{-2}\sum_{s=1}^{T}F_s\eta_{st} = A_{31t} + HA_{32t}.$$

By Cauchy Schwarz inequality, Lemma (2.1), Lemma (B.1), and Assumptions 2.1 and 2.3, we have,

$$\|A_{31t}\| = \left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right\| \leqslant \left\|D_{1T}^{-2}\right\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\underbrace{\sum_{s=1}^{T}\|\eta_{st}\|^2}_{A_{33t}}\right)^{1/2},$$

where we have

$$\begin{aligned}E\left(A_{33t}\right) &= E\left(\sum_{s=1}^{T}\left\|\frac{F_s'\Lambda' e_t}{N}\right\|^2\right) \leqslant \sum_{s=1}^{T}E\left\|\frac{F_s'\Lambda' e_t}{N}\right\|^2 \\ &\leqslant \sum_{s=1}^{T}E\left\|F_s'\right\|^2 E\left\|\frac{\Lambda' e_t}{N}\right\|^2 = E\left\|\frac{\Lambda' e_t}{N}\right\|^2\sum_{s=1}^{T}E\left\|F_s'\right\|^2 = O\left(\frac{T}{N}\right).\end{aligned}$$

Thus, we have $A_{31t} = O\left(\left\|D_{1T}^{-2}\right\|\right)O_P\left(\sqrt{T}\delta_{NT}^{-1}\right)O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right) = O_P\left(\frac{1}{\sqrt{N}\delta_{NT}}\right)$.

Consider the second part of A_{3t},

$$\begin{aligned}\|A_{32t}\| &= \left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s\eta_{st}\right\| = \left\|D_{1T}^{-2}\sum_{s=1}^{T}F_sF_s'\frac{\Lambda' e_t}{N}\right\| \\ &\leqslant \left\|\frac{\Lambda' e_t}{N}\right\|\left\|\sum_{s=1}^{T}D_{1T}^{-1}F_sF_s'D_{1T}^{-1}\right\| \leqslant \left\|\frac{\Lambda' e_t}{N}\right\|\left(\sum_{s=1}^{T}\left\|D_{1T}^{-1}F_s\right\|^2\right) = O_P\left(N^{-1/2}\right).\end{aligned}$$

Therefore, we have $A_{3t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right) + O_P(1)O_P\left(N^{-1/2}\right) = O_P\left(N^{-1/2}\right)$.

Part (iv) We can similarly decompose A_{4t} as follows:

$$A_{4t} = D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st} = D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\xi_{st} + D_{1T}^{-2}\sum_{s=1}^{T}HF_s\xi_{st} = A_{41t} + HA_{42t}.$$

Using Cauchy Schwartz inequality, Lemma (2.1), Lemma (B.1), and Assumptions 2.1 and 2.3, we obtain,

$$\|A_{41t}\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \xi_{st} \right\| = \left\| D_{1T}^{-2} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) F_t' \frac{\Lambda' e_s}{N} \right\|$$
$$\leqslant \|D_{1T}^{-2}\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\sum_{s=1}^{T} \left\| F_t' \frac{\Lambda' e_s}{N} \right\|^2}_{A_{43t}} \right)^{1/2},$$

where we have

$$E(A_{43t}) = E\left(\sum_{s=1}^{T} \left\| F_t' \frac{\Lambda' e_s}{N} \right\|^2 \right) \leqslant \sum_{s=1}^{T} E \left\| F_t' \frac{\Lambda' e_s}{N} \right\|^2 \leqslant \sum_{s=1}^{T} E \|F_t\|^2 E \left\| \frac{\Lambda' e_s}{N} \right\|^2$$
$$= E \|F_t\|^2 \sum_{s=1}^{T} E \left\| \frac{\Lambda' e_s}{N} \right\|^2 = O\left(\frac{T}{N} \right),$$

and thus

$$\|A_{41t}\| = \left\| D_{1T}^{-2} \right\| O_P\left(\frac{\sqrt{T}}{\delta_{NT}} \right) O_P\left(\frac{\sqrt{T}}{\sqrt{N}} \right) = O_P\left(\frac{1}{\sqrt{N}\delta_{NT}} \right).$$

Consider the second part of A_{4t}. Using Cauchy Schwartz inequality, Assumptions 2.1 and 2.3(iii), we have,

$$E\|A_{42t}\| = E \left\| D_{1T}^{-2} \sum_{s=1}^{T} \frac{F_t' \Lambda' e_s F_s'}{N} \right\| \leqslant \left\| D_{1T}^{-1} \right\| E \left\| D_{1T}^{-1} F_t' \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right\|$$
$$\leqslant \left\| D_{1T}^{-1} \right\| \left(E \|F_t\|^2 \right)^{1/2} \left(E \left\| D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right\|^2 \right)^{1/2} = O\left(\frac{\left\| D_{1T}^{-1} \right\|}{\sqrt{N}} \right).$$

Thus, we have $||A_{4t}|| = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(||D_{1T}^{-1}||N^{-1/2}\right) = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$ ■

Lemma B. 1. *Suppose that Assumptions 2.1–2.4 are satisfied. Then,* $D_{1T}^{-2}\tilde{F}'F \xrightarrow{d} Q$ *as* $N, T \to \infty$, *where* $Q = V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$ *is a random matrix,* $V = \mathrm{diag}(v_1, ..., v_r)$ *with* $\{v_1, \ldots, v_r\}$ *denoting the eigenvalues of* $\Sigma_\Lambda\Sigma_F$, *and* Υ *is the corresponding matrix formed by scaled eigenvectors such that* $\Upsilon'\Upsilon = I_r$.

Proof. Since $\tilde{V}_{NT}$ is the diagonal matrix of r largest eigenvalues of XX' multiplied by $N^{-1}D_{1T}^{-2}$, we have $XX'\tilde{F}N^{-1}D_{1T}^{-2} = \tilde{F}\tilde{V}_{NT}$. Now, multiplying both side by $\left(\Lambda'\Lambda N^{-1}\right)^{1/2}\left(D_{1T}^{-2}F'\right)$, we have

$$\left(\frac{\Lambda'\Lambda}{N} \right)^{1/2} \left(D_{1T}^{-2} F' \right) XX' \tilde{F} \frac{1}{N} D_{1T}^{-2} = \left(\frac{\Lambda'\Lambda}{N} \right)^{1/2} \left(D_{1T}^{-2} F' \right) \tilde{F} \tilde{V}_{NT}. \qquad \text{(B.15)}$$

Since $X = F\Lambda' + e$, we can expand XX', and we obtain,

$$XX'\tilde{F}\frac{1}{N}D_{1T}^{-2} = \frac{F\Lambda'\Lambda F'}{N}\tilde{F}D_{1T}^{-2} + \left(\frac{F\Lambda'e' + e\Lambda F' + ee'}{N}\right)\tilde{F}D_{1T}^{-2}.$$

Then, we may rewrite equation (B.15) as

$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\left(\frac{F\Lambda'\Lambda F'}{N}\tilde{F}D_{1T}^{-2} + \left(\frac{F\Lambda'e' + e\Lambda F' + ee'}{N}\right)\tilde{F}D_{1T}^{-2}\right) = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\tilde{F}\tilde{V}_{NT}$$
$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'F\left(\frac{\Lambda'\Lambda}{N}\right)\left(F'\tilde{F}D_{1T}^{-2}\right) + A_{NT} = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\tilde{F}\tilde{V}_{NT},$$

where $A_{NT} = \left(N^{-1}\Lambda'\Lambda\right)^{1/2} D_{1T}^{-2}F'\left(F\Lambda'e' + e\Lambda F' + ee'\right)\tilde{F}N^{-1}D_{1T}^{-2}$. We may show that $A_{NT} = o_P(1)$.

Using the proof of Lemma (2.1), we have $N^{-1}\left(F\Lambda'e' + e\Lambda F' + ee'\right)\tilde{F}D_{1T}^{-2}\tilde{V}_{NT}^{-1} = \tilde{F} - FH'$. Therefore, we can write

$$\begin{aligned}
A_{NT} &= \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2} D_{1T}^{-2}F'\left(F\Lambda'e' + e\Lambda F' + ee'\right)\tilde{F}\frac{1}{N}D_{1T}^{-2} \\
&= \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}F'\left(\tilde{F} - FH'\right)\right)\tilde{V}_{NT} \\
&= \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}\sum_{t=1}^{T}F_t\left(\tilde{F}_t - HF_t\right)'\right)\tilde{V}_{NT}.
\end{aligned}$$

Using Cauchy-Schwarz inequality and Lemma (B.2), $\tilde{V}_{NT} = O_P(1)$, we have

$$\begin{aligned}
\|A_{NT}\|^2 &= \left\|\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}\sum_{t=1}^{T}F_t\left(\tilde{F} - HF_t\right)'\right)\tilde{V}_{NT}\right\|^2 \\
&\leqslant \left\|\frac{\Lambda'\Lambda}{N}\right\|\left\|\sum_{t=1}^{T}D_{1T}^{-2}F_t\left(\tilde{F}_t - HF_t\right)'\right\|^2\left\|\tilde{V}_{NT}\right\|^2 \\
&\leqslant \left\|\frac{\Lambda'\Lambda}{N}\right\|\|D_{1T}^{-1}\|^2\left(\sum_{t=1}^{T}\|D_{1T}^{-1}F_t\|^2\right)\left(\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2\right)
\end{aligned}$$

where Lemma (B.2) has been used for the last inequality.

Using Assumption 2.1, we have $\left\|N^{-1}\Lambda'\Lambda\right\| = O_P(1)$ and $\left(\sum_{t=1}^{T}\left\|D_{1T}^{-1}F_t\right\|^2\right) = O_P(1)$. Also, by Lemma (2.1), $\left(T^{-1}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2\right) = O_P(\delta_{NT}^{-2})$. Then, as $T\left\|D_{1T}^{-2}\right\| = O(1)$, we obtain $A_{NT} = O_P\left(\delta_{NT}^{-1}\right) = o_P(1)$.

Therefore, we have

$$\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}F'F\right)\left(\frac{\Lambda'\Lambda}{N}\right)\left(F'\tilde{F}D_{1T}^{-2}\right) + o_P(1) = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(D_{1T}^{-2}F'\tilde{F}\right)\tilde{V}_{NT}. \qquad \text{(B.16)}$$

Let $B_{NT} = \left(N^{-1}\Lambda'\Lambda\right)^{1/2}\left(D_{1T}^{-2}F'F\right)\left(N^{-1}\Lambda'\Lambda\right)^{1/2}$ and

$$C_{NT} = \left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(F'\tilde{F}D_{1T}^{-2}\right). \tag{B.17}$$

We may show that, $D_{1T}^{-2}F'\tilde{F} = F'\tilde{F}D_{1T}^{-2}$, asymptotically. Consider,

$$D_{1T}^{-2}F'\tilde{F} = D_{1T}^{-2}F'\left(\tilde{F} - FH' + FH'\right) = D_{1T}^{-2}F'FH' + D_{1T}^{-2}F'\left(\tilde{F} - FH'\right).$$

Using Cauchy Schwartz inequality, Lemma (2.1) and Assumption 2.1, we have,

$$\left\|D_{1T}^{-2}F'\left(\tilde{F} - FH'\right)\right\| \leqslant \left\|D_{1T}^{-1}\right\|\left\|D_{1T}^{-1}F'\right\|\left\|\tilde{F} - FH'\right\| = o_P(1).$$

Thus, we have $D_{1T}^{-2}F'\tilde{F} = D_{1T}^{-2}(F'F)H' + o_P(1)$.

Since we assume that the factors are not cointegrated, and define the rotation matrix, H, to be asymptotically diagonal (with ±1), we have $D_{1T}^{-2}F'FH'$ as a block diagonal matrix. This implies that $F'\tilde{F}$ is asymptotically block diagonal. Therefore, we can write $D_{1T}^{-2}F'\tilde{F} = F'\tilde{F}D_{1T}^{-2}$ asymptotically.

Then, we can rewrite the equation (B.16) as

$$\left(B_{NT}C_{NT} + A_{NT}\right) = C_{NT}\tilde{V}_{NT} \quad \text{or} \quad \left(B_{NT} + A_{NT}C_{NT}^{-1}\right)C_{NT} = C_{NT}\tilde{V}_{NT}.$$

Since $\tilde{V}_{NT}$ is diagonal, it follows that the columns of C_{NT} are eigenvectors of the matrix $B_{NT} + A_{NT}C_{NT}^{-1}$. However, this C_{NT} is not of unit length. Let $\Upsilon_{NT} = C_{NT}\tilde{V}_{NT}^{\dagger-1/2}$, where $\tilde{V}_{NT}^{\dagger}$ is a diagonal matrix with $\operatorname{diag}\left(\tilde{V}_{NT}^{\dagger}\right) = \operatorname{diag}\left(C'_{NT}C_{NT}\right)$, the r largest eigenvalues of $FN^{-1}\Lambda'\Lambda F'$. Then, we have $\left(B_{NT} + A_{NT}C_{NT}^{-1}\right)\Upsilon_{NT} = \Upsilon_{NT}\tilde{V}_{NT}$, where Υ_{NT} is the collection of unit length eigenvectors of the matrix $B_{NT} + A_{NT}C_{NT}^{-1}$. By the Assumption 2.1, we have $\left(N^{-1}\Lambda'\Lambda\right) \xrightarrow{p} \Sigma_\Lambda$ and $\left(D_{1T}^{-1}F'F\right) \xrightarrow{d} \Sigma_F$. Also, we have $A_{NT} = o_P(1)$. Furthermore, we may show that $C_{NT}^{-1} = O_P(1)$.

Since $F'\tilde{F}D_{1T}^{-2}$ is asymptotically block diagonal, we have C_{NT} as an asymptotically diagonal matrix. We may show that $\lim_{T,N\to\infty}C_{NT,ii} \neq 0$, in probability, for $i = 1, \ldots r$, where $C_{NT,ii}$ is the ii^{th} term of the matrix C_{NT}. Consider $\left(C_{NT}\right)'\left(C_{NT}\right)$.

$$\begin{aligned} C'_{NT}C_{NT} &= D_{1T}^{-2}\tilde{F}'F\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}\left(\frac{\Lambda'\Lambda}{N}\right)^{1/2}F'\tilde{F}D_{1T}^{-2} = D_{1T}^{-2}\tilde{F}'\left(\frac{XX'}{N}\right)\tilde{F}D_{1T}^{-2} + o_P(1) \\ &= \tilde{V}_{NT} + o_P(1), \end{aligned}$$

where the second and third equalities followed by the Lemma (B.2) (i) and (ii). Using Lemma (B.2), we have $\tilde{V}_{NT} = O_P(1)$ and $\tilde{V}_{NT}^{-1} = O_P(1)$, which implies that $\lim_{N,T\to\infty}\tilde{V}_{NT,ii} \neq 0$, in probability. Then, $\lim_{N,T\to\infty}C_{NT,ii}^2 \neq 0$, in probability for $i = 1, \ldots r$. Thus, we have C_{NT}

is bounded away from zero, $C_{NT}^{-1} = O_P(1)$. Hence, together with $A_{NT} = o_P(1)$, we obtain $B_{NT} + A_{NT}C_{NT}^{-1}$ converges, in distribution, to $B = \Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$.

Furthermore, by Assumption 2.1(v), we assume that the eigenvalues of $\Sigma_\Lambda\Sigma_F$ are distinct. Then, the eigenvalues of $B_{NT} + A_{NT}C_{NT}^{-1}$ are also distinct for large N and T. This implies that the eigenvectors of the $B_{NT} + A_{NT}C_{NT}^{-1}$ are unique except for the fact that these eigenvectors can be replaced by their negative (other sign) of themselves. Since C_{NT} and $\Upsilon_{NT} = C_{NT}\tilde{V}_{NT}^{\dagger -1/2}$ are functions of $\tilde{F}$, given the column sign of $\tilde{F}$, Υ_{NT} is uniquely determined. Using the eigenvalue perturbation theory, we have a unique eigenvector matrix Υ of B such that $\Upsilon_{NT} \xrightarrow{d} \Upsilon$. Since $\left(N^{-1}\Lambda'\Lambda\right)$ is positive definite, we can rewrite equation (B.17) as $D_{1T}^{-2}\tilde{F}'F = C_{NT}'\left(N^{-1}\Lambda'\Lambda\right)^{-1/2} = \tilde{V}_{NT}^{\dagger 1/2}\Upsilon_{NT}'\left(N^{-1}\Lambda'\Lambda\right)^{-1/2}$. Thus, together with $\tilde{V}_{NT}^{\dagger} \xrightarrow{d} V$ in Lemma (B.2), we have

$$D_{1T}^{-2}\tilde{F}'F \xrightarrow{d} V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2} := Q.$$

$\blacksquare$

Lemma B. 2. *Suppose that Assumptions 2.1–2.4 hold. Then, as $N, T \to \infty$ with $\sqrt{N}||D_{1T}^{-2}|| \to 0$ for each given t, we have $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} V^{-1}QN\left(0, \Gamma_t\right) \overset{d}{=} N(0, \Sigma_{\tilde{F}})$, where Q is defined in Lemma (2.2), Γ_t is defined in Assumption 2.3, and Q is independent of $N(0, \Gamma_t)$.*

Proof of Lemma 2.3. Using the identity for $\tilde{F}, \tilde{\Lambda}$ and H as in the proof of Lemma (2.1), we obtain

$$\begin{aligned}\tilde{F}_t - HF_t &= \tilde{V}_{NT}^{-1}\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right) \\ &= \tilde{V}_{NT}^{-1}\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right).\end{aligned}$$

Using Lemma (B.2), we have $\tilde{V}_{NT}^{-1} = O_P(1)$. Furthermore, in Lemma (B.4), we have shown that $A_{1t} = O_P\left(\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right), A_{2t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right), A_{3t} = O_P\left(N^{-1/2}\right)$ and $A_{4t} = O_P\left(N^{-1/2}\delta_{NT}^{-1}\right)$. Then, we have,

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = O_P\left(\sqrt{N}\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right) + O_P\left(\delta_{NT}^{-1}\right) + O_P(1).$$

Since $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$, we consider the following two cases:

Case 1. *If $O\left(N^{-1/2}\right) > O(\left\|D_{1T}^{-1}\right\|)$, we have $O\left(\delta_{NT}^{-1}\right) = O\left(N^{-1/2}\right)$, then,*

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = O_P(\left\|D_{1T}^{-1}\right\|) + O_P\left(N^{-1/2}\right) + O_P(1).$$

Case 2. *If $O(\left\|D_{1T}^{-1}\right\|) > O\left(N^{-1/2}\right)$, we have $O\left(\delta_{NT}^{-1}\right) = O\left(\left\|D_{1T}^{-1}\right\|\right)$, then,*

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = O_P(\sqrt{N}\left\|D_{1T}^{-2}\right\|) + O_P\left(\left\|D_{1T}^{-1}\right\|\right) + O_P(1).$$

Thus, for both cases, limiting distribution of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$ is determined by the third term under the condition $\lim_{N,T\to\infty}(N^{1/2}\left\|D_{1T}^{-2}\right\|) \to 0$.

Therefore, we have

$$\sqrt{N}\left(\tilde{F}_t - HF_t\right) = \tilde{V}_{NT}^{-1}N^{1/2}D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\frac{F_s'\Lambda' e_t}{N} + o_P(1) = \tilde{V}_{NT}^{-1}D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_sF_s'\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_P(1).$$

By Assumption 2.3(ii), $N^{-1/2}\sum_{i=1}^{N}\lambda_i e_{it} \xrightarrow{d} N\left(0,\Gamma_t\right)$. Hence, together with Lemma (2.2) and Lemma (B.2), we have, $\sqrt{N}\left(\tilde{F}_t - HF_t\right) \xrightarrow{d} V^{-1}QN(0,\Gamma_t)$,
where $\Gamma_t = \lim_{N,T\to\infty} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(\lambda_i\lambda_j' e_{it}e_{jt}\right)$. Since Q, the limiting distribution of $D_{1T}^{-2}\tilde{F}'F$, is determined only by the common factors, Q is independent of $N(0,\Gamma_t)$. ■

Lemma A. 5. *Suppose that Assumptions 2.1–2.6 are satisfied. Then,* $D_{1T}^{-2}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$ *where* $\delta_{NT}^{-1} = max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$.

Proof. First, recall the identity (B.10),

$$\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1}\left\{D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\}.$$

Using the above identity, we have

$$D_{1T}^{-2}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} = \tilde{V}_{NT}^{-1}\left(I + II + III + IV\right), \tag{B.18}$$

where $I = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\epsilon_{t+h}$, $II = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\epsilon_{t+h}$, $III = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\epsilon_{t+h}$ and $IV = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\epsilon_{t+h}$.

According to Lemma (B.2), we have $\tilde{V}_{NT}^{-1} = O_P(1)$, and we may show that

$$I = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right), \quad II = O_P\left(N^{-1/2}T^{-1/2}\right), \quad III = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right), \quad IV = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right).$$

Consider each term of equation (B.18) separately.

$$I = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\epsilon_{t+h} = D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\epsilon_{t+h} + D_{1T}^{-4}\sum_{t=1}^{T-h}\sum_{s=1}^{T}HF_s\gamma_{st}\epsilon_{t+h} = I_1 + HI_2 \quad \text{(say)}.$$

Using Cauchy Schwarz inequality,

$$\begin{aligned}
||I_1|| &= \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \gamma_{st} \epsilon_{t+h} \right\| = \left\| D_{1T}^{-4} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \left(\sum_{t=1}^{T-h} \gamma_{st} \epsilon_{t+h} \right) \right\| \\
&\leqslant \left\| D_{1T}^{-4} \right\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\sum_{s=1}^{T} \left| \sum_{t=1}^{T-h} \gamma_{st} \epsilon_{t+h} \right|^2 \right)^{1/2} \\
&\leqslant T^2 \left\| D_{1T}^{-4} \right\| \frac{1}{\sqrt{T}} \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T-h} |\gamma_{st}|^2 \frac{1}{T} \sum_{t=1}^{T-h} |\epsilon_{t+h}|^2 \right)^{1/2} \\
&= T^2 \left\| D_{1T}^{-4} \right\| \frac{1}{\sqrt{T}} O_P \left(\delta_{NT}^{-1} \right),
\end{aligned}$$

since $\left(T^{-1} \sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right) = O_P(\delta_{NT}^{-2})$ from Lemma (2.1), $T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T-h} |\gamma_{st}|^2 \leqslant M$ from Lemma (B.1), and $T^{-1} \sum_{t=1}^{T-h} E |\epsilon_{t+h}|^2 = O(1)$ by Assumption 2.6. Therefore, together with the fact that $T||D_{1T}^{-2}|| = O(1)$, we have $\|I_1\| = O_P \left(T^{-1/2} \delta_{NT}^{-1} \right)$.

Now, consider the second part of I. Using Cauchy-Schwarz inequality,

$$\begin{aligned}
E \|I_2\| = E \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_s \gamma_{st} \epsilon_{t+h} \right\| &\leqslant \left\| D_{1T}^{-4} \right\| \sum_{t=1}^{T-h} \sum_{s=1}^{T} E \left\| F_s \gamma_{st} \epsilon_{t+h} \right\| \\
&\leqslant \left\| D_{1T}^{-4} \right\| \sum_{t=1}^{T-h} \sum_{s=1}^{T} |\gamma_{st}| \left(E \|F_s\|^2 \right)^{1/2} \left(E |\epsilon_{t+h}|^2 \right)^{1/2} = O \left(\left\| D_{1T}^{-2} \right\| \right),
\end{aligned}$$

as we have $E \|F_s\|^2 \leqslant M$ and $E |\epsilon_{t+h}|^2 \leqslant M$, and $T^{-1} \sum_{t=1}^{T-h} \sum_{s=1}^{T} |\gamma_{st}| \leqslant M$ for some finite constant M by the Assumptions. Hence, $I = O_P \left(T^{-1/2} \delta_{NT}^{-1} \right) + O_P \left(\left\| D_{1T}^{-2} \right\| \right) = O_P \left(T^{-1/2} \delta_{NT}^{-1} \right)$ as $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\| D_{1T}^{-1} \right\|]$.

Now, we may show that $II = O_P \left(N^{-1/2} T^{-1/2} \right)$.

$$II = D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st} \epsilon_{t+h} = D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \zeta_{st} \epsilon_{t+h} + D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} HF_s \zeta_{st} \epsilon_{t+h} = II_1 + HII_2.$$

Using Cauchy Schwarz inequality and Assumption 2.5(i), we have

$$
\begin{aligned}
||II_1|| &= \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \zeta_{st}\epsilon_{t+h} \right\| \leqslant \|D_{1T}^{-4}\| \left\| \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \sum_{t=1}^{T-h} \zeta_{st}\epsilon_{t+h} \right\| \\
&\leqslant \|D_{1T}^{-4}\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \zeta_{st}\epsilon_{t+h} \right\|^2 \right)^{1/2} \\
&= T^2 \|D_{1T}^{-4}\| \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\frac{1}{T} \sum_{s=1}^{T} \left| \frac{1}{T} \sum_{t=1}^{T-h} \zeta_{st}\epsilon_{t+h} \right|^2}_{II_3(say)} \right)^{1/2} \\
&= T^2 \|D_{1T}^{-4}\| \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} (II_3)^{1/2},
\end{aligned}
$$

where we have

$$
E(II_3) \leqslant \frac{1}{T} \sum_{s=1}^{T} E \left| \frac{1}{T} \sum_{t=1}^{T-h} \zeta_{st}\epsilon_{t+h} \right|^2 = \frac{1}{T} \sum_{s=1}^{T} E \left| \frac{1}{T} \sum_{t=1}^{T-h} \frac{1}{N} \sum_{i=1}^{N} \left(e_{is}e_{it} - E(e_{is}e_{it})\right) \epsilon_{t+h} \right|^2 = O\left(\frac{1}{NT}\right).
$$

Then, using Lemma (2.1), we obtain, $||II_1|| = O(1)O_P(\delta_{NT}^{-1})O_P\left(N^{-1/2}T^{-1/2}\right) = O_P\left(N^{-1/2}T^{-1/2}\delta_{NT}^{-1}\right)$.

By Cauchy Schwarz inequality, Assumption 2.1, and the fact that $E(II_3) = O\left(N^{-1}T^{-1}\right)$, we have,

$$
\begin{aligned}
\|II_2\| &= \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_s\zeta_{st}\epsilon_{t+h} \right\| \leqslant \|D_{1T}^{-3}\| \left\| \sum_{t=1}^{T-h} \sum_{s=1}^{T} D_{1T}^{-1} F_s\zeta_{st}\epsilon_{t+h} \right\| \\
&= \|D_{1T}^{-3}\| \left\| \sum_{s=1}^{T} D_{1T}^{-1} F_s \sum_{t=1}^{T-h} \zeta_{st}\epsilon_{t+h} \right\| \leqslant \|D_{1T}^{-3}\| \left(\sum_{s=1}^{T} \|D_{1T}^{-1}F_s\|^2 \right)^{1/2} \left(\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \zeta_{st}\epsilon_{t+h} \right\|^2 \right)^{1/2} \\
&= O\left(\|D_{1T}^{-3}\|\right) O_P(1) O_P\left(\frac{T^{3/2}}{\sqrt{NT}}\right) = O_P\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}
$$

Thus, together with Lemma (B.2) and $H = O_P(1)$, we obtain

$$
II = II_1 + HII_2 = O_P\left(N^{-1/2}T^{-1/2}\delta_{NT}^{-1}\right) + O_P(1)O_P\left(N^{-1/2}T^{-1/2}\right) = O_P\left(N^{-1/2}T^{-1/2}\right).
$$

Similar to the first and second terms, we can decompose the third term in equation (B.18) as follows:

$$
III = D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \tilde{F}_s\eta_{st}\epsilon_{t+h} = D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \eta_{st}\epsilon_{t+h} + D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} HF_s\eta_{st}\epsilon_{t+h} = III_1 + HIII_2.
$$

Using Cauchy Schwarz inequality,

$$\|III_1\| = \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \eta_{st} \epsilon_{t+h} \right\| \leqslant \|D_{1T}^{-4}\| \left\| \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \sum_{t=1}^{T-h} \eta_{st} \epsilon_{t+h} \right\|$$

$$\leqslant \|D_{1T}^{-4}\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \eta_{st} \epsilon_{t+h} \right\|^2}_{III_3(say)} \right)^{1/2},$$

where we have

$$E(III_3) \leqslant \sum_{s=1}^{T} \left(E \left\| \sum_{t=1}^{T-h} \eta_{st} \epsilon_{t+h} \right\|^2 \right) = \sum_{s=1}^{T} E \left\| \sum_{t=1}^{T-h} \frac{F_s' \Lambda' e_t}{N} \epsilon_{t+h} \right\|^2$$

$$= \sum_{s=1}^{T} E \left\| F_s' \sum_{t=1}^{T-h} \frac{\Lambda' e_t \epsilon_{t+h}}{N} \right\|^2 \leqslant \sum_{s=1}^{T} E \|F_s\|^2 \left(E \left\| \sum_{t=1}^{T-h} \frac{\Lambda' e_t \epsilon_{t+h}}{N} \right\|^2 \right) = O\left(\frac{T^2}{N} \right),$$

since $E\|F_s\|^2 = O(1)$ and $E \left\| N^{-1/2} T^{-1/2} \sum_{t=1}^{T-h} \sum_{i=1}^{N} \lambda_i e_{it} \epsilon_{t+h} \right\|^2 = O(1)$ by Assumptions 2.1 and 2.5.

Therefore, using Lemma (2.1) and the fact that $T \left\| D_{1T}^{-2} \right\| = O(1)$, we have,

$$\|III_1\| = O_P\left(||D_{1T}^{-4}|| \sqrt{T} \delta_{NT}^{-1} T N^{-1/2} \right) = O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \delta_{NT}^{-1} \right).$$

Again, using Cauchy Schwartz inequality, $T \left\| D_{1T}^{-2} \right\| = O(1)$, and Assumptions 2.1 and 2.5, we have,

$$III_2 = \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_s \eta_{st} \epsilon_{t+h} \right\| = \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_s F_s' \frac{\Lambda' e_t}{N} \epsilon_{t+h} \right\|$$

$$\leqslant \|D_{1T}^{-2}\| \left\| \sum_{t=1}^{T-h} \sum_{s=1}^{T} D_{1T}^{-2} F_s F_s' \frac{\Lambda' e_t \epsilon_{t+h}}{N} \right\| = \|D_{1T}^{-2}\| \left\| \left(\sum_{s=1}^{T} D_{1T}^{-2} F_s F_s' \right) \left(\sum_{t=1}^{T-h} \frac{\Lambda' e_t \epsilon_{t+h}}{N} \right) \right\|$$

$$\leqslant \|D_{1T}^{-2}\| \left\| D_{1T}^{-2} \sum_{s=1}^{T} F_s F_s' \right\| \left\| \sum_{t=1}^{T-h} \frac{\Lambda' e_t \epsilon_{t+h}}{N} \right\| = \left\| D_{1T}^{-2} \right\| O_P(1) O_P\left(\frac{\sqrt{T}}{\sqrt{N}} \right) = O_P\left(\frac{\left\| D_{1T}^{-1} \right\|}{\sqrt{N}} \right).$$

Therefore, we have, $III = O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \delta_{NT}^{-1} \right) + O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \right) = O_P\left(N^{-1/2} \left\| D_{1T}^{-1} \right\| \right).$

Again from the decomposition, we have

$$IV = D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \tilde{F}_s \xi_{st} \epsilon_{t+h} = D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s \right) \xi_{st} \epsilon_{t+h} + D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} HF_s \xi_{st} \epsilon_{t+h} = IV_1 + HIV_2.$$

Consider each term separately. Using Cauchy-Schwartz inequality,

$$
\begin{aligned}
\|IV_1\| &= \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \xi_{st}\epsilon_{t+h} \right\| \leqslant \left\| D_{1T}^{-4} \right\| \left\| \sum_{s=1}^{T} \left(\tilde{F}_s - HF_s\right) \sum_{t=1}^{T-h} \xi_{st}\epsilon_{t+h} \right\| \\
&\leqslant \left\| D_{1T}^{-4} \right\| \left(\sum_{s=1}^{T} \left\| \tilde{F}_s - HF_s \right\|^2 \right)^{1/2} \left(\underbrace{\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \xi_{st}\epsilon_{t+h} \right\|^2}_{IV_3 \text{ (say)}} \right)^{1/2} .
\end{aligned}
$$

From Lemma (B.1), and Assumptions 2.1 and 2.6, we obtain

$$
\begin{aligned}
E(IV_3) &= E\left(\sum_{s=1}^{T} \left\| \sum_{t=1}^{T-h} \frac{F_t'\Lambda' e_s}{N} \epsilon_{t+h} \right\|^2 \right) = E\left(\sum_{s=1}^{T} \left\| \left(\frac{\Lambda' e_s}{N}\right)' \sum_{t=1}^{T-h} F_t\epsilon_{t+h} \right\|^2 \right) \\
&\leqslant \sum_{s=1}^{T} E\left\| \frac{\Lambda' e_s}{N} \right\|^2 E\left\| \sum_{t=1}^{T-h} F_t\epsilon_{t+h} \right\|^2 \leqslant \sum_{s=1}^{T} E\left\| \frac{\Lambda' e_s}{N} \right\|^2 \sum_{t=1}^{T-h} E\left\| F_t\epsilon_{t+h} \right\|^2 \\
&\leqslant \sum_{s=1}^{T} E\left\| \frac{\Lambda' e_s}{N} \right\|^2 \sum_{t=1}^{T-h} E\left\| F_t \right\|^2 E\left\| \epsilon_{t+h} \right\|^2 = O\left(\frac{T^2}{N}\right).
\end{aligned}
$$

Therefore, together with the fact that $T\left\|D_{1T}^{-2}\right\| = O(1)$ and Lemma (2.1), we have,

$$
\|IV_1\| = O\left(\left\|D_{1T}^{-4}\right\|\right) O_P\left(T^{1/2}\delta_{NT}^{-1}\right) O_P\left(TN^{-1/2}\right) = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right).
$$

Again, using Cauchy-Schwartz inequality and Assumptions 2.1, 2.3 and 2.6, we have

$$
\begin{aligned}
\|IV_2\| &= \left\| D_{1T}^{-4} \sum_{t=1}^{T-h} \sum_{s=1}^{T} F_s\xi_{st}\epsilon_{t+h} \right\| = \left\| D_{1T}^{-2} \left(D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right) \left(\sum_{t=1}^{T-h} D_{1T}^{-1} F_t\epsilon_{t+h} \right) \right\| \\
&\leqslant \left\| D_{1T}^{-2} \right\| \left\| D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right\| \left\| \sum_{t=1}^{T-h} D_{1T}^{-1} F_t\epsilon_{t+h} \right\| \\
&\leqslant \left\| D_{1T}^{-2} \right\| \left\| D_{1T}^{-1} \sum_{s=1}^{T} \frac{\Lambda' e_s F_s'}{N} \right\| \left(\sum_{t=1}^{T-h} \left\| D_{1T}^{-1} F_t \right\|^2 \right)^{1/2} \left(\sum_{t=1}^{T-h} \left| \epsilon_{t+h} \right|^2 \right)^{1/2} \\
&= O_P\left(\left\| D_{1T}^{-2} \right\| \frac{\sqrt{T}}{\sqrt{N}} \right) = O_P\left(\frac{\left\| D_{1T}^{-1} \right\|}{\sqrt{N}} \right).
\end{aligned}
$$

Thus, we have $IV = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-1}\right) + O_P\left(\left\|D_{1T}^{-1}\right\| N^{-1/2}\right) = O_P\left(\left\|D_{1T}^{-1}\right\| N^{-1/2}\right)$.

Since $\tilde{V}_{NT}^{-1} = O_P(1)$, Lemma (B.2), we have,

$$
\begin{aligned}
&D_{1T}^{-2} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} = \tilde{V}_{NT}^{-1} \left\{ O_P\left(T^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}T^{-1/2}\right) + O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|\right) \right\} \\
&= O_P\left(T^{-1/2}\delta_{NT}^{-1}\right).
\end{aligned}
$$

Therefore, we have shown $D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} \xrightarrow{p} 0$ as $T, N \to \infty$. ■

Lemma A. 6. *Let Assumptions 2.1–2.6 hold. If, in addition, $T/N \to 0$, then as $T, N \to \infty$, we have* $D_{1T}^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t - HF_t\right)'(H^{-1})'\theta \xrightarrow{p} 0$.

Proof. By replacing $\hat{L}_t = \begin{pmatrix} \tilde{F}_t' & W_t' \end{pmatrix}'$, we obtain,

$$
\begin{aligned}
&D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta = D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t W_t \end{pmatrix}\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\
&= D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t - HF_t + HF_t \\ W_t \end{pmatrix}\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\
&= \begin{pmatrix} D_{1T}^{-1}\sum_{t=1}^{T-h}(\tilde{F}_t - HF_t + HF_t)\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\ \\ \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \end{pmatrix} \\
&= \begin{pmatrix} D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\ \\ 0 \end{pmatrix} \\
&+ \begin{pmatrix} D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \\ \\ \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\theta \end{pmatrix} = \begin{pmatrix} A_1 + A_2 \\ A_3 \end{pmatrix}\left(H^{-1}\right)'\theta \qquad \text{(say)},
\end{aligned}
$$

where $A_1 = D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)'$, $A_2 = D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'$, and $A_3 = T^{-1}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'$.

By considering each term separately, we shall show that A_1, A_2, and A_3 converge to 0, in probability, as $N, T \to \infty$ with $T/N \to 0$.

First, recall the equation (B.10).

$$
\begin{aligned}
\tilde{F}_t - HF_t &= \tilde{V}_{NT}^{-1}\left\{D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right\}, \\
&= \tilde{V}_{NT}^{-1}\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right),
\end{aligned}
$$

where $A_{1t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \gamma_{st}$, $A_{2t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \zeta_{st}$, $A_{3t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \eta_{st}$ and $A_{4t} = D_{1T}^{-2} \sum_{s=1}^{T} \tilde{F}_s \xi_{st}$. Then, we may write A_1 as follows:

$$\begin{aligned}
A_1 &= D_{1T}^{-1} \sum_{t=1}^{T-h} \left(\tilde{F}_t - HF_t\right) \left(\tilde{F}_t - HF_t\right)' \\
&= D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{t=1}^{T-h} \left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right) \left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)' \tilde{V}_{NT}^{-1} \\
&= D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} + D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1}.
\end{aligned}$$

Using the triangle inequality,

$$\begin{aligned}
\|A_1\| &= \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} + D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1} \right\| \\
&\leqslant \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} \right\| + \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1} \right\| \\
&= a_1 + a_2 \quad (say).
\end{aligned}$$

Consider the two terms separately. Using Cauchy Schwartz inequality and Lemma (B.2), $\left\|\tilde{V}_{NT}^{-1}\right\| = O_P(1)$, we have

$$\begin{aligned}
a_1 &= \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{it}' \tilde{V}_{NT}^{-1} \right\| \leqslant T \left\|D_{1T}^{-1}\right\| \left\|\tilde{V}_{NT}^{-1}\right\| \left\| \sum_{i=1}^{4} \frac{1}{T} \sum_{t=1}^{T-h} A_{it} A_{it}' \right\| \left\|\tilde{V}_{NT}^{-1}\right\| \\
&\leqslant T \left\|D_{1T}^{-1}\right\| \sum_{i=1}^{4} \frac{1}{T} \sum_{t=1}^{T-h} \|A_{it}\|^2.
\end{aligned}$$

In the proof of Lemma (2.1), we have shown that

$$\begin{aligned}
T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{1t}\|^2 &= O_P\left(\left\|D_{1T}^{-2}\right\|\right), & T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{2t}\|^2 &= O_P\left(N^{-1}\right), \\
T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{3t}\|^2 &= O_P\left(N^{-1}\right), & T^{-1} \textstyle\sum_{t=1}^{T-h} \|A_{4t}\|^2 &= O_P\left(N^{-1}\right).
\end{aligned}$$

Therefore, $a_1 = O\left(T\left\|D_{1T}^{-1}\right\|\right)\left\{O_P\left(\left\|D_{1T}^{-2}\right\|\right) + O_P\left(N^{-1}\right)\right\} = O_P\left(T\left\|D_{1T}^{-1}\right\| \delta_{NT}^{-2}\right)$.

Then, consider the cross terms such as $A_{it}A_{jt}'; i, j = 1, \ldots, 4$ for $i \neq j$. We may prove that the following cross term, a_2, is of order $o_P(1)$ under the condition $T/N \to 0$.

$$\begin{aligned}
a_2 &= \left\| D_{1T}^{-1} \tilde{V}_{NT}^{-1} \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \sum_{t=1}^{T-h} A_{it} A_{jt}' \tilde{V}_{NT}^{-1} \right\| \leqslant T \left\|D_{1T}^{-1}\right\| \left\|\tilde{V}_{NT}^{-1}\right\| \left\| \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \frac{1}{T} \sum_{t=1}^{T-h} A_{it} A_{jt}' \right\| \left\|\tilde{V}_{NT}^{-1}\right\| \\
&\leqslant T \left\|D_{1T}^{-1}\right\| \sum_{i=1}^{4} \sum_{j \neq i=1}^{4} \left\| \frac{1}{T} \sum_{t=1}^{T-h} A_{it} A_{jt}' \right\|.
\end{aligned}$$

Using Cauchy Schwarz inequality and the proof of Lemma (2.1), we have

$$\begin{aligned}
\left\|T^{-1}\sum_{t=1}^{T-h}A_{1t}A_{2t}'\right\| &\leqslant T^{-1}\left(\sum_{t=1}^{T-h}\|A_{1t}\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\|A_{2t}\|^2\right)^{1/2}\\
&= O\left(T^{-1}\right)O_P\left(\sqrt{T}\left\|D_{1T}^{-1}\right\|\right)O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right),\\
\left\|T^{-1}\sum_{t=1}^{T-h}A_{1t}A_{3t}'\right\| &\leqslant T^{-1}\left(\sum_{t=1}^{T-h}\|A_{1t}\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\|A_{3t}\|^2\right)^{1/2}\\
&= O\left(T^{-1}\right)O_P\left(\sqrt{T}\left\|D_{1T}^{-1}\right\|\right)O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right),
\end{aligned}$$

$$\begin{aligned}
\left\|T^{-1}\sum_{t=1}^{T-h}A_{1t}A_{4t}'\right\| &\leqslant T^{-1}\left(\sum_{t=1}^{T-h}\|A_{1t}\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\|A_{4t}\|^2\right)^{1/2}\\
&= O\left(T^{-1}\right)O_P\left(\sqrt{T}\left\|D_{1T}^{-1}\right\|\right)O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right),\\
\left\|T^{-1}\sum_{t=1}^{T-h}A_{2t}A_{3t}'\right\| &\leqslant T^{-1}\left(\sum_{t=1}^{T-h}\|A_{2t}\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\|A_{3t}\|^2\right)^{1/2}\\
&= T^{-1}O_P\left(\sqrt{\frac{T}{N}}\right)O_P\left(\sqrt{\frac{T}{N}}\right) = O_P\left(\frac{1}{N}\right).
\end{aligned}$$

Similarly, we can show that the other cross terms $T^{-1}\sum_{t=1}^{T-h}A_{2t}A_{4t}'$ and $T^{-1}\sum_{t=1}^{T-h}A_{3t}A_{4t}'$ are also $O_P\left(N^{-1}\right)$. Then, as $\sqrt{T}/N\to 0$, we have,

$$a_2 = T\left\|D_{1T}^{-1}\right\|\left\{O_P\left(\frac{\left\|D_{1T}^{-1}\right\|}{\sqrt{N}}\right)+O_P\left(\frac{1}{N}\right)\right\} = O_P\left(\frac{1}{\sqrt{N}}\right)+O_P\left(\frac{\sqrt{T}}{N}\right) = o_P(1).$$

Together with the fact that $T\left\|D_{1T}^{-2}\right\| = O(1)$ and $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$,

$$A_1 = a_1 + a_2 = O_P\left(T\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-2}\right)+O_P\left(\frac{1}{\sqrt{N}}\right)+O_P\left(\frac{\sqrt{T}}{N}\right) = o_P(1).$$

Then, consider the second term $A_2 = D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'$. By Cauchy Schwarz inequality, Assumption 2.1, Lemma (A.2) and Lemma (2.1), we obtain

$$\begin{aligned}
\|A_2\| = \left\|D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'\right\| &\leq ||H||\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|\tilde{F}_t - HF_t\right\|^2\right)^{1/2}\\
= O_P\left(T^{1/2}\delta_{NT}^{-1}\right).
\end{aligned}$$

Since $\delta_{NT}^{-1} = \max[N^{-1/2}, ||D_{1T}^{-1}||]$, and $T\left\|D_{1T}^{-2}\right\| = O(1)$, we have $\delta_{NT}^{-1} = O\left(T^{-1/2}\right)$ as $T/N \to 0$. This implies, $\|A_2\| = O_P(1)$.

Thus, we cannot use this method to show $A_2 \xrightarrow{p} 0$ as $N, T \to \infty$ with $T/N \to 0$. Therefore, using $\tilde{F}_t - HF_t = \tilde{V}_{NT}^{-1}\{A_{1t} + A_{2t} + A_{3t} + A_{4t}\}$, we may rewrite A_2 as follows:

$$D_{1T}^{-1}\sum_{t=1}^{T-h} HF_t\left(\tilde{F}_t - HF_t\right)' = HD_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)'\tilde{V}_{NT}^{-1}, \tag{B.19}$$

$$= H\left(B_1 + B_2 + B_3 + B_4\right)\tilde{V}_{NT}^{-1}, \tag{B.20}$$

where $B_1 = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{1t}'$, $B_2 = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{2t}'$, $B_3 = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{3t}'$, and $B_4 = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{4t}'$. We may consider each term separately.

$$\begin{aligned} B_1 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{1t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right)' \\ &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\gamma_{st}'\left(\tilde{F}_s - HF_s\right)' + D_{1T}^{-2}\sum_{s=1}^{T}\gamma_{st}'F_s'H'\right) = B_{11} + B_{12}H' \qquad \text{(say)}\ . \end{aligned}$$

Using Cauchy Schwarz inequality, Lemma (B.1), and Lemma (2.1) we have,

$$\begin{aligned} \|B_{11}\| &= \left\|D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\gamma_{st}'\left(\tilde{F}_s - HF_s\right)'\right)\right\| \\ &\leqslant \left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\underbrace{\left\|D_{1T}^{-2}\sum_{s=1}^{T}\gamma_{st}'\left(\tilde{F}_s - HF_s\right)'\right\|^2}_{b_{11t}}\right)^{1/2}, \end{aligned}$$

where we have,

$$b_{11t} \leqslant \left\|D_{1T}^{-2}\right\|^2\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)\left(\sum_{s=1}^{T}|\gamma_{st}|^2\right) = O_p\left(\left\|D_{1T}^{-2}\right\|\delta_{NT}^{-2}\right)$$

Hence, we obtain, $\|B_{11}\| = O_P\left(\delta_{NT}^{-1}\right)$.

Consider $B_{12} = D_{1T}^{-1}\left(D_{1T}^{-2}\sum_{t=1}^{T-h}\sum_{s=1}^{T} F_t\gamma_{st}'F_s'\right)$. By Cauchy Schwarz inequality and Assumption 2.1(i), we obtain,

$$\begin{aligned} E\left\|D_{1T}^{-2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_t\gamma_{st}'F_s'\right\| &\leqslant \left\|D_{1T}^{-2}\right\|\sum_{t=1}^{T-h}\sum_{s=1}^{T}E\left\|F_t\gamma_{st}'F_s'\right\| \\ &\leqslant \left\|D_{1T}^{-2}\right\|\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left(E\left\|F_t\right\|^2\right)^{1/2}\left(E\left\|F_s\right\|^2\right)^{1/2}\left(E|\gamma_{st}|^2\right)^{1/2} = O(1), \end{aligned}$$

since $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}| \leqslant M$ by Assumption 2.2, and $E\|F_t\|^2 \leqslant M$ by Assumption 2.1. Therefore, $B_{12} = O_p\left(\|D_{1T}^{-1}\|\right)$. Thus, we have

$$B_1 = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{1t}' = O_P\left(\delta_{NT}^{-1}\right) + O_P\left(\|D_{1T}^{-1}\|\right) = O_P\left(\delta_{NT}^{-1}\right). \tag{B.21}$$

Then, consider the second term in equation (B.19),

$$\begin{aligned}
B_2 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{2t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right)' \\
&= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right)' + D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T} HF_s\zeta_{st}\right)' \\
&= B_{21} + B_{22}H'.
\end{aligned}$$

Using Cauchy Schwartz inequality, we have,

$$E\|B_{22}\| = E\left\|D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T} F_s'\zeta_{st}\right)\right\| \leqslant \|D_{1T}^{-3}\|\sum_{t=1}^{T-h}\left(E\|F_t\|^2\right)^{1/2}\left(E\left\|\sum_{s=1}^{T} F_s'\zeta_{st}\right\|^2\right)^{1/2}.$$

Using Assumptions 2.1 and 2.2, we obtain, $E\left\|\sum_{s=1}^{T} F_s'\zeta_{st}\right\|^2 = O\left(TN^{-1}\right)$. Therefore, together with the fact that $T\left\|D_{1T}^{-2}\right\| = O(1)$, we have, $E\|B_{22}\| = O\left(N^{-1/2}\right)$. Again, using the Cauchy Schwartz inequality and Lemma (2.1), we have,

$$\begin{aligned}
\|B_{21}\| &= \left\|D_{1T}^{-1}\sum_{t=1}^{T-h} F_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right)'\right\| \\
&\leqslant \left\|D_{1T}^{-2}\right\|\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right\|^2\right)^{1/2} \\
&\leqslant \left\|D_{1T}^{-2}\right\|\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\zeta_{st}|^2\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2} \\
&= O_P\left(\left\|D_{1T}^{-2}\right\|\right)O_P\left(\frac{T}{\sqrt{N}}\sqrt{T}\delta_{NT}^{-1}\right) = O_P\left(\frac{\sqrt{T}}{\delta_{NT}\sqrt{N}}\right),
\end{aligned}$$

since $E\left|N^{-1/2}\sum_{i=1}^{N}\left(e_{is}e_{it} - E(e_{is}e_{it})\right)\right|^4 \leqslant M$.

Together with Lemma (B.2), $\|H\| = O_P(1)$, we have

$$B_2 = O_P\left(\frac{\sqrt{T}}{\delta_{NT}\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right). \tag{B.22}$$

Similar to B_1 and B_2, we can rewrite B_3 as follows:

$$\begin{aligned}B_3 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{3t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right)'\\ &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)' + D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T} HF_s\eta_{st}\right)' = B_{31} + B_{32}H'.\end{aligned}$$

Again, using Cauchy Schwartz inequality, Assumptions 2.1 and 2.3(iii), and Lemma (2.1), we obtain,

$$\begin{aligned}\|B_{31}\| &= \left\|D_{1T}^{-1}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\left(D_{1T}^{-2}\sum_{t=1}^{T-h} F_t\eta_{st}\right)\right\|\\ &\leqslant \left\|D_{1T}^{-1}\right\|\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{T}\left\|D_{1T}^{-2}\sum_{t=1}^{T-h} F_t\eta_{st}\right\|^2\right)^{1/2}\\ &= O_p\left(T^{1/2}\delta_{NT}^{-1}N^{-1/2}\right).\end{aligned}$$

Again, using a similar argument, we have,

$$\|B_{32}\| = \left\|D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \sum_{s=1}^{T} D_{1T}^{-2}F_s'\eta_{st}\right\| \leqslant \left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T} F_s'\eta_{st}\right\|^2\right)^{1/2} = O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right)$$

Since $H = O_P(1)$, Lemma (B.2), we have,

$$B_3 = B_{31} + B_{32}H' = O_P\left(\frac{\sqrt{T}}{\sqrt{N}\delta_{NT}}\right) + O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right) = O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right). \tag{B.23}$$

Similarly, we can show that $B_4 = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right)$.

$$\begin{aligned}B_4 &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t A_{4t}' = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right)'\\ &= D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\xi_{st}\right)' + D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T} HF_s\xi_{st}\right)' = B_{41} + B_{42}H'.\end{aligned}$$

Using Cauchy Schwartz inequality, Assumptions 2.1 and 2.3, Lemma (B.1), and the Lemma (2.1), we obtain,

$$B_{41} = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\frac{\Lambda' e_s F_t}{N}\right)' = \left(\sum_{t=1}^{T-h} D_{1T}^{-1}F_tF_t'D_{1T}^{-1}\right)\underbrace{\left(D_{1T}^{-1}\sum_{s=1}^{T} N^{-1}\Lambda' e_s\left(\tilde{F}_s - HF_s\right)\right)}_{B_{43}},$$

where we have,

$$\|B_{43}\| = \left\|D_{1T}^{-1}\right\| \left(\frac{1}{N^2}\sum_{s=1}^{T}\left\|\Lambda' e_s\right\|^2\right)^{1/2} \left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2} = \left\|D_{1T}^{-1}\right\| O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right)\left(\sqrt{T}\delta_{NT}^{-1}\right).$$

Hence, $B_{41} = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right)$.
Given Assumptions 2.1 and 2.3(iii), we obtain,

$$B_{42} = D_{1T}^{-1}\sum_{t=1}^{T-h} F_t \left(D_{1T}^{-2}\sum_{s=1}^{T} F_s \frac{\Lambda' e_s F_t}{N}\right)' = \sum_{t=1}^{T-h} D_{1T}^{-1} F_t F_t' D_{1T}^{-1}\left(D_{1T}^{-1}\sum_{s=1}^{T}\frac{\Lambda' e_s F_s'}{N}\right) = O_P\left(\frac{1}{\sqrt{N}}\right).$$

Therefore, we have

$$B_4 = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right). \tag{B.24}$$

Thus, together with $\|H\| = O_P(1)$ and $\left\|\tilde{V}_{NT}^{-1}\right\| = O_P(1)$, from Lemma (B.2), and equations (B.21)-(B.24) we have

$$\begin{aligned} D_{1T}^{-1}\sum_{t=1}^{T-h} HF_t\left(\tilde{F}_t - HF_t\right)' &= H\left(B_1 + B_2 + B_3 + B_4\right)\tilde{V}_{NT}^{-1} \\ &= O_P\left(\frac{1}{\delta_{NT}}\right) + \left(O_P\left(\frac{\sqrt{T}}{\sqrt{N}\delta_{NT}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)\right) + O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right). \end{aligned}$$

Since $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|]$ and $T\left\|D_{1T}^{-2}\right\| = O(1)$, as $T, N \to \infty$ with $T/N \to 0$, we have $\delta_{NT}^{-1} = \max[N^{-1/2}, \left\|D_{1T}^{-1}\right\|] = \max[N^{-1/2}, T^{-1/2}] = T^{-1/2}$. Hence, as $T, N \to \infty$ with $T/N \to 0$, we obtain,

$$D_{1T}^{-1}\sum_{t=1}^{T-h} HF_t\left(\tilde{F}_t - HF_t\right)' = O_P\left(T^{-1/2}\right) + O_P\left(N^{-1/2}\right) + O_P\left(T^{1/2}N^{-1/2}\right) = o_P(1).$$

Now, consider the third term $A_3 = T^{-1}\sum_{t=1}^{T-h} W_t\left(\tilde{F}_t - HF_t\right)'$. Using Cauchy Schwartz inequality and Lemma (2.1), we have

$$\|A_3\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h} W_t\left(\tilde{F}_t - HF_t\right)'\right\| \le \left(\frac{1}{T}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|\tilde{F}_t - HF_t\right\|^2\right)^{1/2} = O_P\left(\frac{\sqrt{T}}{\delta_{NT}}\right).$$

Thus, we cannot use this method to prove that $A_3 \xrightarrow{p} 0$ as $N, T \to \infty$ with $T/N \to 0$. Therefore, we consider the following method. Rewrite A_3 using A_{1t}, A_{2t}, A_{3t} and A_{4t} defined in Lemma (B.4),

$$A_3 = \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(\tilde{F}_t - HF_t\right)' = \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)'\tilde{V}_{NT}^{-1} = \left(d_1 + d_2 + d_3 + d_4\right)'\tilde{V}_{NT}^{-1}, \tag{B.25}$$

where $d_1 = T^{-1}\sum_{t=1}^{T-h} W_t A_{1t}'$, $d_2 = T^{-1}\sum_{t=1}^{T-h} W_t A_{2t}'$, $d_3 = T^{-1}\sum_{t=1}^{T-h} W_t A_{3t}'$, and $d_4 = T^{-1}\sum_{t=1}^{T-h} W_t A_{4t}'$. Replacing A_{1t} by its definitions, we have

$$d_1 = \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T} F_s'\gamma_{st}' H'\right)$$
$$= d_{11} + d_{12}H'.$$

By Cauchy- Schwartz inequality and Lemma (2.1), we have,

$$||d_{11}|| \leqslant \left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\gamma_{st}'\right\|^2\right)^{1/2} \tag{B.26}$$
$$= O_P\left(\sqrt{T}\right)O_P\left(||D_{1T}^{-1}||\delta_{NT}^{-1}\right) = O_P(\delta_{NT}^{-1}),$$

since we assume that the observable series W_t are $I(1)$, and we have shown that $\left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\gamma_{st}'\right\|^2 = O_P\left(\left\|D_{1T}^{-2}\right\|\delta_{NT}^{-2}\right)$. Using Cauchy Schwartz inequality and Assumption 2.1, we have,

$$\|d_{12}\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right)\right\| \leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right\|^2\right)^{1/2} = O_P\left(\left\|D_{1T}^{-1}\right\|\right).$$

Note that using Cauchy Schwartz inequality and Lemma (B.1), we have,

$$E\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right\|^2\right) \leqslant \sum_{t=1}^{T-h}E\left(\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\gamma_{st}'\right\|^2\right) \leqslant \left\|D_{1T}^{-2}\right\|^2\sum_{t=1}^{T-h}\left(\sum_{s=1}^{T}E\left\|F_s\right\|^2\right)\sum_{s=1}^{T}|\gamma_{st}|^2$$
$$= O\left(\left\|D_{1T}^{-2}\right\|\right).$$

Thus,

$$d_1 = O_P(\delta_{NT}^{-1}) + O_P(||D_{1T}^{-1}||) = O_P(\delta_{NT}^{-1}). \tag{B.27}$$

Now, consider the second term d_2.

$$d_2 = \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h} W_t \left(D_{1T}^{-2}\sum_{s=1}^{T} F_s'\zeta_{st}' H'\right)$$
$$= d_{21} + d_{22}H'.$$

From Cauchy Schwartz inequality, Assumption 2.2(v), and Lemma (2.1), we obtain,

$$
\begin{aligned}
||d_{21}||^2 &\leqslant ||D_{1T}^{-2}||^2 \left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|\sum_{s=1}^{T}\zeta_{st}'\left(\tilde{F}_s - HF_s\right)\right\|^2\right)\\
&= ||D_{1T}^{-2}||^2 \left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\zeta_{st}|^2\right)\left(\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)\\
&= O_P\left(T||D_{1T}^{-2}||^2\right) O_P\left(\frac{T}{N}\right) O_P\left(T\delta_{NT}^{-2}\right) = O_P\left(\frac{T}{N\delta_{NT}^2}\right).\\
||d_{22}|| &= \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t D_{1T}^{-2}\sum_{s=1}^{T}F_s'\zeta_{st}'\right\| \leqslant \left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s'\zeta_{st}\right\|^2\right)^{1/2} = O_P\left(\frac{\sqrt{T}}{\sqrt{N}}\right).
\end{aligned}
$$

Then, we have

$$
d_2 = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(T^{1/2}N^{-1/2}\right) = O_P\left(T^{1/2}N^{-1/2}\right). \tag{B.28}
$$

Now, consider the third term in equation (B.25). Similar with d_2, we can rewrite d_3 as follows:

$$
\begin{aligned}
d_3 &= \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s'\eta_{st}'H'\right)\\
&= d_{31} + d_{32}H'.
\end{aligned}
$$

Using Cauchy Schwartz inequality, we have,

$$
\begin{aligned}
\|d_{31}\| &= \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)'\right\|\\
&\leqslant \left\|D_{1T}^{-2}\right\|\left(\frac{1}{T}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\eta_{st}|^2\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2}\\
&= O\left(\left\|D_{1T}^{-2}\right\|\right) O_P\left(\sqrt{T}\right) O_P\left(\frac{\left\|D_{1T}^{-1}\right\|^{-1}}{\sqrt{N}}\right) O_P\left(\sqrt{T}\delta_{NT}^{-1}\right) = O_P\left(\frac{\left\|D_{1T}^{-1}\right\|^{-1}}{\sqrt{N}\delta_{NT}}\right).
\end{aligned}
$$

Note that from Lemma (B.1) and Assumption 2.1, we have,

$$
\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\|\eta_{st}\|^2 = \frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left\|\frac{F_s'\Lambda' e_t}{N}\right\|^2 \leqslant \frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{\Lambda' e_t}{N}\right\|^2\sum_{s=1}^{T}\|F_s\|^2 = \frac{\left\|D_{1T}^{-2}\right\|^{-1}}{N}.
$$

Since we assume that $W_t \sim I(1)$, and $\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}F_s\eta_{st}\right\|^2 = O_P\left(N^{-1}\right)$, we have,

$$
\|d_{32}\| = \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s\eta_{st}\right)\right\| \leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=}^{T}F_s\eta_{st}\right\|^2\right)^{1/2} = O_P\left(\frac{1}{\sqrt{N}}\right)
$$

Thus,

$$d_3 = O_P\left(N^{-1/2}\left\|D_{1T}^{-1}\right\|^{-1}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right) = O_P\left(N^{-1/2}\right) \tag{B.29}$$

as $T, N \to \infty$ with $T/N \to 0$.

Similarly, we may show that $d_4 = O_P\left(N^{-1/2}\right)$.

$$d_4 = \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right)' = \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\xi_{st}\right)' + \frac{1}{T}\sum_{t=1}^{T-h} W_t\left(D_{1T}^{-2}\sum_{s=1}^{T}F_s\xi_{st}\right)' H'$$
$$= d_{41} + d_{42}H'$$

Using Cauchy Schwartz inequality and Lemma (2.1), we have,

$$\|d_{41}\| \leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}||W_t||^2\right)^{1/2}\left(\sum_{t=1}^{T-h}\left\|D_{1T}^{-2}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\xi_{st}\right\|^2\right)^{1/2}$$
$$\leqslant \left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)^{1/2}\left(\left\|D_{1T}^{-1}\right\|^2\underbrace{\sum_{t=1}^{T-h}\sum_{s=1}^{T}\left\|D_{1T}^{-1}\xi_{st}\right\|^2}_{d_{43}}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)^{1/2},$$

where we have,

$$d_{43} = \sum_{t=1}^{T-h}\sum_{s=1}^{T}\left\|D_{1T}^{-1}\frac{F_t'\Lambda' e_s}{N}\right\|^2 \leqslant \sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\sum_{s=1}^{T}\left\|\frac{\Lambda' e_s}{N}\right\|^2 = O_P\left(\frac{T}{N}\right).$$

We used Lemma (B.1) to bound d_{43}.

Thus, we have

$$d_{41} = O_P(1)\, O_P\left(\frac{\left\|D_{1T}^{-1}\right\|\sqrt{T}}{\sqrt{N}}\right) O_P\left(\sqrt{T}\delta_{NT}^{-1}\right) = O_P\left(\frac{\sqrt{T}}{\sqrt{N}\delta_{NT}}\right).$$

Again, by Cauchy Schwartz inequality and Assumptions 2.1 and 2.3(iii), we obtain,

$$\|d_{42}\|^2 = \left\|\frac{1}{T}\sum_{t=1}^{T-h}W_t D_{1T}^{-1}\left(\sum_{s=1}^{T}\frac{D_{1T}^{-1}F_t'\Lambda' e_s F_s}{N}\right)\right\|^2 \leqslant \left\|D_{1T}^{-1}\right\|^2\left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)\left(\sum_{t=1}^{T-h}\left\|\sum_{s=1}^{T}\frac{D_{1T}^{-1}F_t'\Lambda' e_s F_s'}{N}\right\|^2\right)$$
$$\leqslant \left\|D_{1T}^{-1}\right\|^2\left(\frac{1}{T^2}\sum_{t=1}^{T-h}\|W_t\|^2\right)\sum_{t=1}^{T-h}\left\|D_{1T}^{-1}F_t\right\|^2\left\|\sum_{s=1}^{T}\frac{\Lambda' e_s F_s}{N}\right\|^2$$
$$= O_P\left(\left\|D_{1T}^{-1}\right\|^2\right) O_P(1) O_P\left(\frac{\left\|D_{1T}^{-1}\right\|^{-2}}{N}\right) = O_P\left(\frac{1}{N}\right).$$

Hence,

$$||d_4|| = O_P\left(T^{1/2}N^{-1/2}\delta_{NT}^{-1}\right) + O_P\left(N^{-1/2}\right) = O_P\left(N^{-1/2}\right) \tag{B.30}$$

as $T, N \to \infty$ with $T/N \to 0$.

Therefore, together with Lemma (B.2) and equations (B.27)-(B.30), as $T/N \to 0$ for $T, N \to \infty$, we have,

$$A_3 = (d_1 + d_2 + d_3 + d_4)\,\tilde{V}_{NT}^{-1} = O_P(\delta_{NT}^{-1}) + O_P\left(T^{1/2}N^{-1/2}\right) + O_P\left(N^{-1/2}\right) = o_P(1).$$

Hence, we have shown that

$$\begin{aligned} D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(\tilde{F}_t - HF_t\right)'(H^{-1})'\theta &= \begin{pmatrix} A_1 + A_2 \\ A_3 \end{pmatrix}(H^{-1})'\theta = \begin{pmatrix} o_P(1) + o_P(1) \\ o_P(1) \end{pmatrix} O_P(1)\text{Plim}(\hat{\theta}) \\ &\xrightarrow{p} 0 \qquad \text{as } N, T \to \infty \text{ with } T/N \to 0. \end{aligned}$$

■

2.6.3 Appendix C: Proofs of the Main Results

With the necessary lemmas listed in Appendix B, we are ready to prove the main theorems.

Theorem 1. *Suppose that Assumptions 2.1–2.6 hold and that $T/N \to 0$. Let δ and the OLS estimator $\hat{\delta}$ be as in equation (9). Then, as $(N,T) \to \infty$, we have $D_T(\hat{\delta} - \delta) \xrightarrow{d} N\big(0, \Sigma_\delta\big)$, where $\Sigma_\delta = (\Psi')^{-1}\Sigma_L^{-1}\Sigma_{\epsilon L}\Sigma_L^{-1}\Psi^{-1}$, Σ_L and $\Sigma_{\epsilon L}$ are defined in Assumptions 2.1–2.6, and $H \oplus I \xrightarrow{d} \Psi$.*

Proof of Theorem 1. Let $D_T = [D_{1T} \oplus D_{2T}]$ where D_{1T} defined as $D_{1T} = diag[TI_p \oplus \sqrt{T}I_q]_{r\times r}$ and $D_{2T} = diag[T, ..., T]_m$. Then, the OLS estimator $\hat{\delta}$ of the forecasting model

$$Y_{t+h} = \hat{L}_t{}'\delta + \theta' H^{-1}\left(HF_t - \tilde{F}_t\right) + \epsilon_{t+h}$$

can be written as

$$\begin{aligned} D_T(\hat{\delta} - \delta) &= \left(D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1}\right)^{-1}\left(D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\epsilon_{t+h} + D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(HF_t - \tilde{F}\right)'H^{-1}\theta\right) \\ &= \left(\underbrace{D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1}}_{Z}\right)^{-1}\left\{D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} HF_t \\ W_t \end{pmatrix}\epsilon_{t+h} + D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t - HF \\ 0 \end{pmatrix}\epsilon_{t+h}\right. \\ &\qquad\left. + D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t(HF_t - \tilde{F})'(H^{-1})'\theta\right\} \\ &= (Z)^{-1}(A + B + C) \qquad \text{(say)} \end{aligned}$$

Then, we may consider each term separately.

$$
\begin{aligned}
A &= D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} HF_t \\ W_t \end{pmatrix}\epsilon_{t+h} = D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} H & 0 \\ 0 & I_m \end{pmatrix}\begin{pmatrix} F_t \\ W_t \end{pmatrix}\epsilon_{t+h} \\
&= D_T^{-1}\sum_{t=1}^{T-h}\begin{pmatrix} H_1 & 0 & 0 \\ 0 & H_2 & 0 \\ 0 & 0 & I_m \end{pmatrix}\begin{pmatrix} E_t \\ G_t \\ W_t \end{pmatrix}\epsilon_{t+h} = (H_1 \oplus H_2 \oplus I_m)\, D_T^{-1}\sum_{t=1}^{T-h} L_t\epsilon_{t+h} \\
&\xrightarrow{d} \Psi\Sigma_{\epsilon L}^{1/2}N(0,I),
\end{aligned}
$$

since $D_T^{-1}\sum_{t=1}^{T} L_t\epsilon_{t+h} \xrightarrow{d} \Sigma_{\epsilon L}^{1/2}\times N(0,I)$, using Assumption 2.6(iii) and $(H_1\oplus H_2\oplus I_m) = \Psi_0 \to \Psi$ as $N,T\to\infty$.
By Lemma (B.5), we have $B = D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\epsilon_{t+h} \xrightarrow{p} 0$ as $N,T\to\infty$, and by Lemma (B.6), if $T/N \to 0$ as $T,N\to\infty$, we have $C = D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_T(HF_t - \tilde{F}_t)'(H^{-1})'\theta \xrightarrow{p} 0.$ [1]

Then, we may consider $Z = D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1}$. By writing $\hat{L}_t = \Psi_0 L_t + \hat{L}_t - \Psi_0 L_t$, we obtain,

$$
\begin{aligned}
&D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1} = D_T^{-1}\left(\sum_{t=1}^{T-h}(\Psi_0 L_t + \hat{L}_t - \Psi_0 L_t)(\Psi_0 L_t + \hat{L}_t - \Psi_0 L_t)'\right)D_T^{-1} \\
&= D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\Psi_0 L_t L_t'\Psi_0' + \Psi_0 L_t\left(\hat{L}_t - \Psi_0 L_t\right)'\right)\right)D_T^{-1} \\
&+ D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\left(\hat{L}_t - \Psi_0 L_t\right)(\Psi_0 L_t)' + \left(\hat{L}_t - \Psi_0 L_t\right)\left(\hat{L}_t - \Psi_0 L_t\right)'\right)\right)D_T^{-1} \\
&= \Psi_0 D_T^{-1}\left(\sum_{t=1}^{T-h}L_t L_t'\right)D_T^{-1}\Psi_0' + D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t - \Psi_0 L_t\right)L_t'\right)D_T^{-1}\Psi_0' \\
&+ \Psi_0 D_T^{-1}\left(\sum_{t=1}^{T-h}L_t\left(\hat{L}_t - \Psi_0 L_t\right)'\right)D_T^{-1} + D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t - \Psi_0 L_t\right)\left(\hat{L}_t - \Psi_0 L_t\right)'\right)D_T^{-1} \\
&\equiv z_1 + z_2 + z_3 + z_4 \qquad \text{(say)}.
\end{aligned}
$$

[1] If $T/N \to a(>0)$, then $C \nrightarrow 0$ in probability as $N,T\to\infty$. Then there is a bias term as discussed in Goncalves (2014).

Then, we may show that $z_2 + z_3 + z_4 = o_P(1)$. First, consider z_2.

$$
\begin{aligned}
z_2 &= D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t - \Psi_0 L_t\right)L_t'\right)D_T^{-1}\Psi_0' \\
&= \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1}\left(\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t - HF_t \\ 0 \end{pmatrix}\begin{pmatrix} F_t' & W_t' \end{pmatrix}\right)\begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1}\begin{pmatrix} H & 0 \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1}\begin{pmatrix} \sum_{t=1}^{T-h}(\tilde{F}_t - HF_t)F_t' & \sum_{t=1}^{T-h}(\tilde{F}_t - HF_t)W_t' \\ 0 & 0 \end{pmatrix}\begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1}\begin{pmatrix} H & 0 \\ 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} D_{1T} & 0 \\ 0 & TI_m \end{pmatrix}^{-1}\begin{pmatrix} \left(\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)F_t'HD_{1T}^{-1}\right) & \left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)W_t'\right) \\ 0 & 0 \end{pmatrix}
\end{aligned}
$$

In the proof of Lemma (B.6), we have shown that $A_2 = \left(D_{1T}^{-1}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)'\right) \xrightarrow{p} 0$, and $A_3 = \left(T^{-1}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\right) \xrightarrow{p} 0$ for $T, N \to \infty$ with $T/N \to 0$. Thus, $z_2 = o_P(1)$. Also, $z_3 = z_2' = o_P(1)$. Now, consider the last term of Z.

$$
\begin{aligned}
z_4 &= D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t - \Psi_0 L_t\right)\left(\hat{L}_t - \Psi_0 L_t\right)'\right)D_T^{-1} = D_T^{-1}\left[\sum_{t=1}^{T-h}\begin{pmatrix} \tilde{F}_t - HF_t \\ 0 \end{pmatrix}\begin{pmatrix} (\tilde{F}_t - HF_t)' & 0 \end{pmatrix}\right]D_T^{-1} \\
&= D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)'D_{1T}^{-1}.
\end{aligned}
$$

<u>Method 1</u>

By the proof of Lemma (B.6), as $N, T \to \infty$ with $T/N \to 0$, we have

$$z_4 = A_1 D_{1T}^{-1} = O_P\left(T\left\|D_{1T}^{-1}\right\|\delta_{NT}^{-2}\right)O_P\left(\left\|D_{1T}^{-1}\right\|\right) = O_P\left(\delta_{NT}^{-2}\right) \xrightarrow{p} 0.$$

<u>Method 2</u>: Similar arguments to the proof of Lemma B.7 of Choi (2017)

Using Lemma (2.3), asymptotic distribution of estimated factors, for $N, T \to \infty$ with $\sqrt{N}||D_{1T}^{-2}|| \to 0$, we have $(\tilde{F}_t - HF_t) = O_P\left(N^{-1/2}\right)$. Hence, $\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)' = O_P\left(N^{-1}\right)$. As $T\left\|D_{1T}^{-2}\right\| = O(1)$, we have, $D_{1T}^{-1}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)'D_{1T}^{-1} = O_P\left(\left\|D_{1T}^{-2}\right\|TN^{-1}\right) = o_P(1)$. Therefore, we have $z_4 = o_P(1)$. However, to follow this method, we need to have the condition $\sqrt{N}\left\|D_{1T}^{-2}\right\| \to 0$ also.

Thus, we obtain

$$Z = z_1 + o_P(1) = \Psi_0 D_T^{-1}\left(\sum_{t=1}^{T-h}L_t L_t'\right)D_T^{-1}\Psi_0' + o_P(1) \xrightarrow{d} \Psi\Sigma_L\Psi',$$

by Assumption 2.6(ii) and $\Psi_0 \xrightarrow{p} \Psi$ where Σ_L is a random matrix defined as in Assumptions. Then, together with the previous results for A, B, and C, we obtain

$$\begin{aligned} D_T(\hat{\delta} - \delta) &\xrightarrow{d} \left(\Psi \Sigma_L \Psi'\right)^{-1} \left\{\Psi \Sigma_{\epsilon L}^{1/2} N(0, I)\right\} \\ &\xrightarrow{d} \left(\Psi'\right)^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L}^{1/2} N(0, I). \end{aligned}$$

Hence, $D_T(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta)$ where $\Sigma_\delta = (\Psi')^{-1} \Sigma_L^{-1} \Sigma_{\epsilon L} \Sigma_L^{-1} \Psi^{-1}$.

■

Theorem 2. *Let Assumptions 2.1–2.6 hold. Furthermore, suppose that* $\sqrt{N}||D_{1T}^{-2}|| \to 0$ *and* $T/N \to 0$ *as* $N, T \to \infty$*, and that* $(\hat{\Sigma}_\delta, \hat{\Sigma}_{\tilde{F}})$ *is a given consistent estimator of* $(\Sigma_\delta, \Sigma_{\tilde{F}})$*. Then, we have*

$$\frac{\hat{Y}_{T+h|T} - Y_{T+h|T}}{\sqrt{\hat{B}_T}} \xrightarrow{d} N(0, 1) \quad \text{as } N, T \to \infty$$

where $\hat{B}_T = \left[\hat{L}_T D_T^{-1} \hat{\Sigma}_\delta D_T^{-1} \hat{L}_T' + N^{-1} \hat{\theta}' \hat{\Sigma}_{\tilde{F}} \hat{\theta}\right]$ *is a consistent estimator of the asymptotic variance, denoted* B_T*, of the conditional forecasting error that appears in the numerator.*

Proof of Theorem 2. Since an estimator of $Y_{T+h|T}$ is $\hat{Y}_{T+h|T} = \hat{\delta}' \hat{L}_T$, and

$$Y_{T+h|T} = \delta' \hat{L}_T + \theta' H^{-1} \left(H F_T - \tilde{F}_T\right),$$

we have,

$$\begin{aligned} \hat{Y}_{T+h|T} - Y_{T+h|T} &= \left(\hat{\delta} - \delta\right)' \hat{L}_T + \theta' H^{-1} \left(\tilde{F}_T - H F_T\right) \\ &= \left(\hat{\delta} - \delta\right)' D_T D_T^{-1} \hat{L}_T + \frac{1}{\sqrt{N}} \theta' H^{-1} \sqrt{N} \left(\tilde{F}_T - H F_T\right) \\ &= \left(D_T \left(\hat{\delta} - \delta\right)\right)' D_T^{-1} \hat{L}_T + \frac{1}{\sqrt{N}} \theta' H^{-1} \sqrt{N} \left(\tilde{F}_T - H F_T\right). \end{aligned}$$

Using Theorem (1), the limiting distribution of the estimated parameters, along with $T/N \to 0$, we have $D_T \left(\hat{\delta} - \delta\right) \xrightarrow{d} N(0, \Sigma_\delta)$.

By Lemma (2), the limiting distribution of the estimated factors for $\sqrt{N}||D_{1T}^{-2}|| \to 0$, we have

$$\sqrt{N} \left(\tilde{F}_t - H F_t\right) \xrightarrow{d} N(0, \Sigma_{\tilde{F}_t}).$$

Moreover, $D_T \left(\hat{\delta} - \delta\right)$ and $\sqrt{N} \left(\tilde{F}_t - H F_t\right)$ are asymptotically independent as the limit of the first term determined by the regression errors and the limit of the second term determined by the idiosyncratic errors. Hence, the limiting distribution of the forecast error conditional on $\{L_t\}_{t=1}^T$ is

$$\hat{Y}_{T+h|T} - Y_{T+h|T} \xrightarrow{d} N\left(0, B_T\right),$$

where $B_T = \hat{L}_T' D_T^{-1} \Sigma_\delta D_T^{-1} \hat{L}_T + N^{-1}\theta' \Sigma_{\tilde{F}_T} \theta$.

Furthermore, $\hat{B}_T = \hat{L}_T' D_T^{-1} \hat{\Sigma}_\delta D_T^{-1} \hat{L}_T + N^{-1}\hat{\theta}' \hat{\Sigma}_{\tilde{F}_T} \hat{\theta}$ is the consistent estimator of the asymptotic variance of the forecast error where $\hat{\Sigma}_\delta$ and $\hat{\Sigma}_{\tilde{F}_T}$ defined as in the main paper. Therefore,

$$\frac{\hat{Y}_{T+h|T} - Y_{T+h|T}}{\sqrt{\hat{B}_T}} \xrightarrow{d} N(0,1) \text{ as } N, T \to \infty.$$

■

2.6.4 Appendix D: Simulation Results

In this section we present the simulation results on the coverage rates of residual based t-percentile bootstrap prediction intervals when the error distribution is normal and when it is t with 5 degrees of freedom.

Results when the error distribution is normal

The design of the simulation for this part is the same as that in section 3 of the main paper. Thus, the two DGPs are:

DGP1 : $\epsilon_t \sim N(0, 1)$

DGP2 : $\epsilon_t \sim N(0, 3^{-1}F_{2t}^2)$.

Table 2.9: Coverage rates (%) of residual based 95% bootstrap (t-percentile) prediction intervals for one-step ahead forecasts when the error distribution is normal

		$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
	T\N	30	50	100	200	30	50	100	200	30	50	100	200
DGP1	30	82	84	87	90	82	82	88	89	83	82	86	91
	50	83	86	88	87	82	87	89	86	82	87	90	86
	100	78	82	88	86	80	82	89	87	81	82	89	87
	200	85	84	84	88	85	84	85	89	86	84	85	88
DGP2	30	88	90	91	92	90	88	91	90	90	91	91	92
	50	90	92	93	93	89	93	93	94	91	94	94	95
	100	89	88	91	92	88	89	90	90	90	91	94	93
	200	86	87	93	90	86	86	93	89	89	88	94	92

Results when the error distribution is t_5

For this part of the simulation study, we used DGP1 as for the previous table, except that the errors $\{\epsilon_t\}$ were generated from t_5 distribution, instead of the normal distribution. Top panel of Table 2.10 provides the coverage rates of the 95% asymptotic prediction interval obtained assuming that the errors are normal. Bottom panel of Table 2.10 provides the coverage rates for the residual based 95% bootstrap t-percentile prediction interval.

Table 2.10: Coverage rates (%) of 95% prediction intervals for one-step ahead forecasts when the error distribution is t_5.

	$\rho = 0.0$				$\rho = 0.5$				$\rho = 0.9$			
T\N	30	50	100	200	30	50	100	200	30	50	100	200
Forecast interval assuming that the errors are normally distributed												
30	96	95	97	95	97	95	97	96	97	95	96	95
50	94	93	94	96	94	93	94	96	93	94	93	97
100	95	93	96	97	95	93	96	97	95	93	96	97
200	94	96	94	95	94	96	94	95	94	97	94	96
Residual based bootstrap prediction interval												
30	84	84	87	88	84	84	86	87	84	84	85	87
50	82	84	87	90	84	84	88	90	85	85	87	89
100	83	83	88	91	84	84	89	91	84	84	88	92
200	83	88	83	86	84	89	84	86	85	90	84	87

Chapter 3

Forecasting Univariate Time Series using a Semiparametric FAR model

3.1 Introduction

In macroeconomics, methods for choosing a good forecasting model, estimation of its parameters, and construction of reliable forecasts based on the estimated model have been active areas of research.

Within this broad framework, a two-step factor augmented regression [FAR] method has attracted considerable attention in the recent literature. Chapter 2 discussed validity of the FAR model under the following three scenarios:
(a) All the panel variables and predictors are stationary (Bai and Ng [2002, 2006], Bai 2003, Gonçalves and Perron 2014).
(b) All the variables follow unit root nonstationary processes (Bai 2004, Maciejowska 2010, Choi 2017).
(c) The set of predictors contains I(0) and I(1) processes (Smeekes and Wijler 2019, Hannadige et al. 2021).
The parameters in the models appearing in the aforementioned studies were assumed to be time invariant; we refer to them as the *parametric FAR model.* The objective of this chapter is to propose, evaluate, and illustrate a method for forecasting a univariate time series in a more general setting than those in the aforementioned studies. More specifically, we allow the coefficients of the factors in the FAR model to be time varying, and the factors themselves to be locally stationary but possibly globally nonstationary.

In the literature, time series analysis involving $I(1)$ processes have been extended to include other nonstationary processes. One important direction allows locally stationary processes (see Dette and Wu 2022). They have been used in empirical studies in economics and finance (see Hansen 2001, Stock and Watson [2016], and Yan et al. [2021]). In this chapter we extend the method developed in the previous chapter to include locally stationary processes. Locally stationary process has a more general structure than the usual stationary process. They need not be stationary, but estimation methods for stationary processes can be combined with kernel smoothing to estimate locally stationary processes.

In a setting such as that in FRED-QD, where the panel data set is observed over a long period of time, it is likely that the values of some of the parameters might have changed over time. In such cases, estimates based on a model that assumes that the parameters are invariant over time would be inconsistent. To address this issue, this Chapter allows the parameters to be time varying.

Related literature

Forecasting of univariate time series using stationary factor models with possible structural breaks has a long history (Banerjee et al. 2008, Stock and Watson 2009). Banerjee et al. [2008] reported the results of a comprehensive simulation study on the forecasting performance of factor models when there are structural instabilities. They observed that although discrete changes in the factor loadings affect the performance of factor models, continuous changes do not affect the performance significantly. Stock and Watson [2009] also arrived at a similar conclusion about the effect of structural instabilities in factor models. To make the forecasts more robust to structural breaks, Stock and Watson also recommended using the full sample for estimating the latent factors and sub-samples or time-varying coefficients for forecasting. This is because, even if the factor loadings are structurally unstable, factors can be well estimated if the instability is independent across the observable variables. Several authors have developed methods of forecasting, using sub-samples or time varying coefficients, when there are structural instabilities (Li et al. 2016, Li et al. 2011a). In this chapter, we develop a related method for forecasting that involves time varying parameters in the forecasting model.

Since structural instabilities affect the performance of factors models, methods have been developed for testing the presence of such instabilities, some with known break points and others with unknown break points (Breitung and Eickmeier 2011, Corradi and Swanson 2014, Su and Wang 2017, Bates et al. 2013, Li et al. 2016, Chen and Hong 2012). Corradi and Swanson [2014] emphasized that not only the factor loading instabilities but also the coefficient instabilities may lead to unreliable forecasts. A local version of principal components analysis [PCA] and a penalized PCA have been proposed to accommodate structural breaks in factor models (see Li et al. [2016], Su and Wang [2017], Motta et al. [2011]).

Statistical inference based on time series models with time varying parameters has been studied extensively in the literature (Gao and Hawthorne 2006, Zhou and Wu 2010, Li et al. 2020). Some of the basic ideas underlying such time varying parameter models have also been extended to panel data models, which include the diffusion index model (Wei and Zhang 2020, Li et al. 2011a). Wei and Zhang [2020] used local PCA to estimate the factor loadings in the factor model and the local constant method to estimate the time-varying coefficients in the forecasting model; the asymptotic validity of the method in Wei and Zhang [2020] is yet to be established. These developments are for time series and panel data models with time varying parameters.

The foregoing ideas have also been extended to semiparametric models, more specifically models with a mixture of constant and varying coefficients. A standard method of estimation in such models is *profile least squares*[PLS] (Fan and Huang 2005), which has also been extended to time series models (Li et al. 2011b, Zhang et al. 2012). By replacing the partially varying coefficients in Fan and Huang [2005] with time-varying coefficients, Chen et al. [2012] extended the nonparametric idea to semi-parametric trending panel data models by proposing a pooled profile likelihood estimation method. The aforementioned methods have also been illustrated in different empirical studies (Chen et al. 2018, Silvapulle et al. 2017, Silvapulle and Jayasuriya 2018).

Contribution of this chapter

The first main point of departure of this Chapter is that the panel data model contains locally stationary factors. Once the factors have been estimated, we use them as regressors in the forecasting model. The coefficients corresponding to the factors in the forecasting regression model are assumed to be smooth time-varying functions. By contrast, the coefficients corresponding to the observable predictors are assumed to be time invariant. We refer to the FAR model discussed in this chapter as *semi-parametric FAR model.* Both the time-varying and the constant coefficients are estimated using the PLS approach and the local constant method. Using simulation studies and an empirical example, we evaluated both nonparametric kernel and orthogonal polynomial based methods for estimating the time varying component of the FAR model. In these evaluations, the nonparametric smoothing kernel method performed better for forecasting. Therefore, we proceeded with only the kernel estimation method. The details about the orthogonal polynomial method and the results to support the kernel method are provided in an appendix to the chapter. The main conclusion of this Chapter, supported by simulation studies and an empirical application, is that the method proposed in this Chapter for the intended scenarios, is an improvements over the one in the previous Chapter.

Outline of the Chapter

The rest of the Chapter is organized as follows. Section 3.2 introduces the semi-parametric FAR model. Section 3.3 reports the results of a comprehensive simulation study. Section 3.4 presents an empirical application that uses the FRED-QD data set. Section 3.5 concludes. Appendix A provides the results for comparison the kernel and the sieve method based on two different orthonormal bases. Numerical studies on the choice of bandwidths are presented in Appendix B. Some additional analyses evaluating the use of rolling window in the empirical study is provided in Appendix C.

3.2 Methodology

3.2.1 Model and notations

Let $\{Y_t : t = 1, 2, \ldots\}$ denote an observable univariate time series that we wish to forecast one-step ahead using information available at time T. Let $\{X_{it} \in \mathbb{R} : i = 1, \ldots, N; t = 1, \ldots, T\}$ denote a set of panel data, and $\{V_t \in \mathbb{R}^m : t = 1, \ldots, T\}$ denote a set of observable predictors. The *factor augmented regression*[FAR] model studied in this chapter is

$$Y_{t+1} = \alpha_t' F_t + \beta' V_t + \eta_{t+1} \qquad (t = 1, \ldots, T), \tag{3.1}$$

where $\{\eta_t \in \mathbb{R} : t \in \mathbb{N}\}$ is a sequence of martingale difference sequence, β is an $m \times 1$ vector of time-invariant parameters, α_t is an $r \times 1$ vector of time-varying parameters, and F_t is an $r \times 1$ vector of unobservable factors. The factors $\{F_t\}$ are to be estimated using the set of panel data $\{X_{it} \in \mathbb{R} : i = 1, \ldots, N; t = 1, \ldots, T\}$ as described later in this Chapter; the resulting estimator $\tilde{F}_t$ will be substituted for F_t $(t = 1, \ldots, T)$ in (3.1) for estimating the unknown parameters in that model. Then the estimated model is to be used for forecasting Y_{T+1}.

Let $\tau_t = t/T$ $(t = 1, \ldots, T)$. In contrast to the previous chapter, in this chapter we assume that $\{F_t\}_{t \in \mathbb{N}}$ and $\{V_t\}_{t \in \mathbb{N}}$ are *vector moving average* $[VMA(\infty)]$ processes of infinite order taking the forms

$$F_t = \mu(\tau_t) + \sum_{j=0}^{\infty} B_j(\tau_t)\epsilon_{t-j} \quad (t = 1, \ldots, T), \tag{3.2}$$

$$V_t = \gamma + \sum_{j=0}^{\infty} A_j \zeta_{t-j} \quad (t = 1, \ldots, T), \tag{3.3}$$

where $\mu(\tau_t)$ is an $r \times 1$ vector of unknown trending functions, γ is an $m \times 1$ vector of unknown constant (i.e. time invariant) parameters, $\{\epsilon_t\}_{t\in\mathbb{N}}$ and $\{\zeta_t\}_{t\in\mathbb{N}}$ are two martingale difference sequences of dimension $r \times 1$ and $m \times 1$ respectively, and $\{B_j(\cdot) \in \mathbb{R}^{r\times r}, A_j \in \mathbb{R}^{m\times m} : j = 0, 1, \ldots\}$ are deterministic coefficients. The set of factors $\{F_t\}_{t\in\mathbb{N}}$ is $VMA(\infty)$ with time-varying coefficients, behave almost as a stationary process locally. The set of observable variables $\{V_t\}_{t\in\mathbb{N}}$ is stationary.

The term *locally stationary process* has appeared in the literature. The following definition of local stationarity is consistent with that in Vogt [2012]. Let $S_t \in \mathbb{R}$ denote a given stochastic process. The process S_t is *locally stationary* if for each re-scaled time point $u \in [0, 1]$ there exists an associated process $Z_t(u)$ with the following two properties:
(i). $Z_t(u)$ is strictly stationary with some probability density function,
(ii).

$$|S_t - Z_t(u)| \leq \big(|\frac{t}{T} - u| + \frac{1}{T}\big)U_t(u) \qquad (a.s)$$

where $U_t(u) > 0$ satisfies $E[U(t(u))^a] < C$ for some $a > 0$ and $C < \infty$ independent of $\{u, t, T\}$. Since the moments of $U_t(u)$ are uniformly bounded, we have

$$|S_t - Z_t(u)| = O_p\big(|(t/T) - u| + (1/T)\big).$$

In this sense, S_t is approximated by the stationary process $Z_t(u)$ in the local neighbourhoods of the re-scaled time point $u \in [0, 1]$. A locally stationary process is typically nonstationary, although a stationary process is also locally stationary, and it is not a mixture of $I(0)$ and $I(1)$ processes. In the context of this thesis, an advantage of incorporating locally stationary processes is the combined effect of the following: They allow us to use certain nonstationary processes that were not allowed in the previous Chapter, and they allow us to develop methods of estimation by adapting methods for estimating stationary processes and local kernel smoothing techniques. These qualitative notions do need rigorous developments, nevertheless they do highlight the potential benefits of locally stationary processes.

The method of estimation and forecasting developed in this Chapter do not require the VMA(∞) structure, but it is required for the informal arguments leading to the asymptotic distributions of the estimators and forecasts. These informal arguments not presented in the thesis.

We assume that the α_t in (3.1) is of the form

$$\alpha_t = \alpha(\tau_t) \qquad (t = 1, \ldots, T), \tag{3.4}$$

where $\alpha(.)$ is an unknown smooth function. Let us introduce the notation

$$\mathbb{B}_t(L) = \sum_{j=0}^{\infty} B_j(\tau_t) L^j, \qquad \mathbb{A}(L) = \sum_{j=0}^{\infty} A_j L^j,$$

where L is the lag operator, and $B_j(\tau_t)$ and A_j are the deterministic coefficients in (3.2) and (3.3), respectively. Then we have

$$F_t = \mu(\tau_t) + \sum_{j=0}^{\infty} B_j(\tau_t)\epsilon_{t-j} = \mu(\tau_t) + \mathbb{B}_t(L)\epsilon_t \quad (t = 1, \ldots, T), \tag{3.5}$$

$$V_t = \gamma + \sum_{j=0}^{\infty} A_j \zeta_{t-j} = \gamma + \mathbb{A}(L)\zeta_t \quad (t = 1, \ldots, T). \tag{3.6}$$

When $\mu(\tau_t)$ and $B_j(\tau_t)$ are time invariant, $\mu(\tau_t) = \mu, B_j(\tau_t) = B_j$, the factor structure reduces to a stationary linear process. Therefore, our factor structure can incorporate certain stationary processes as well. More importantly, these models are obtained by smooth but small variations of stationary processes.

Let $\hat{F}_t$ denote an estimator of F_t $(t = 1, \ldots, T)$, where the estimation is performed using the set of panel data $\{X_{it}\}$; the precise form of $\hat{F}_t$ is discussed later in the Chapter. The ultimate goal of the foregoing formulation is to estimate

$$Y_{t+1} = \alpha_t' \hat{F}_t + \beta' V_t + \text{error} \qquad (t = 1, \ldots, T) \tag{3.7}$$

as an approximation of (3.1), and using the estimated model for forecasting Y_{T+1}.

In the simulation and empirical studies presented later in this Chapter, we estimate F_t by nonparametric estimation. It will be seen that we do not estimate the coefficients $\{A_j, B_j(\tau_t) : t = 1, \ldots, T; j = 0, 1, \ldots\}$; this was expected since the method in this Chapter does not require the predictors to be VMA(∞). The assumption that $\{F_t\}$ and $\{V_t\}$ are $VMA(\infty)$ processes is used in the simulations study. These observations emphasize the fact that the method is applicable when $\{F_t\}$ and $\{V_t\}$ are more general than $VMA(\infty)$. In the next subsection, we propose a method of estimating the FAR model (3.1) when the values of the factors are known; then in the following subsection, we propose a method of estimating the factors and then estimating the approximate version (3.7) with estimated factors in place of the unobservable factors.

3.2.2 Estimation of the FAR model when the factors are known

We apply the *profile least squares*[PLS] method in Fan and Huang [2005], with local constant, for estimating the constant parameters and time-varying coefficients. Let $K(.)$ denote a given kernel smoothing function that satisfies $\int K(u)du = 1$, $\int uK(u)du = 0$, and $\int u^2K(u)du < \infty$. Ordinary least squares estimation of the time-varying parameter $\alpha(\tau)$, at any fixed τ in the range $[\tau_1, \tau_T]$, is achieved by minimizing

$$\sum_{s=1}^{T-1} \left(Y_{s+1} - \beta' V_s - \alpha(\tau)' F_s\right)^2 K\left(\frac{\tau_s - \tau}{h}\right), \tag{3.8}$$

where h is a given bandwidth. For *any* β, the expression in (3.8) is minimized when $\alpha(\tau) = \tilde{\alpha}(\tau;\beta)$, where

$$\tilde{\alpha}(\tau;\beta) = \left(\sum_{s=1}^{T-1} F_s K\left(\frac{\tau_s - \tau}{h}\right) F_s'\right)^{-1} \sum_{s=1}^{T-1} F_s K\left(\frac{\tau_s - \tau}{h}\right)(Y_{s+1} - \beta' V_s). \tag{3.9}$$

Therefore, we concentrate out α_t in (3.1) by substituting $\tilde{\alpha}(\tau_t;\beta)$ for α_t to obtain

$$\begin{aligned} Y_{t+1} &= \tilde{\alpha}(\tau_t;\beta)' F_t + \beta' V_t + \text{error} \\ &= F_t' \left(\sum_{s=1}^{T-1} F_s K\left(\frac{\tau_s - \tau_t}{h}\right) F_s'\right)^{-1} \sum_{s=1}^{T-1} F_s K\left(\frac{\tau_s - \tau_t}{h}\right)\left(Y_{s+1} - \beta' V_s\right) + \beta' V_t + \text{error}. \end{aligned} \tag{3.10}$$

Let

$$\tilde{Y}_{t+1} = Y_{t+1} - F_t' \left(\sum_{s=1}^{T-1} F_s K\left(\frac{\tau_s - \tau_t}{h}\right) F_s'\right)^{-1} \sum_{s=1}^{T-1} F_s K\left(\frac{\tau_s - \tau_t}{h}\right) Y_{s+1}, \tag{3.11}$$

$$\tilde{V}_t = V_t - F_t' \left(\sum_{t=1}^{T-1} F_s K\left(\frac{\tau_s - \tau_t}{h}\right) F_t'\right)^{-1} \sum_{t=1}^{T-1} F_s K\left(\frac{\tau_s - \tau_t}{h}\right) V_s. \tag{3.12}$$

Then (3.10) takes the form

$$\tilde{Y}_{t+1} = \beta' \tilde{V}_t + \text{error} \quad (t = 1, \ldots, T). \tag{3.13}$$

Estimation of β in (3.13) by ordinary least squares leads to

$$\hat{\beta} = \left(\sum_{t=1}^{T-1} \tilde{V}_t \tilde{V}_t' \right)^{-1} \left(\sum_{t=1}^{T-1} \tilde{V}_t \tilde{Y}_{t+1} \right), \tag{3.14}$$

which is a function of the data only, except for the bandwidth h. Next, substitute the $\hat{\beta}$ in (3.14) for β in (3.1) to obtain

$$Y_{t+1} - \hat{\beta}' V_t = \alpha(\tau_t)' F_t + \text{error} \qquad (t = 1, \ldots, T). \tag{3.15}$$

Now, estimate $\alpha(\tau)$, at any fixed $\tau \in [\tau_1, \tau_T]$ with $\hat{\beta}$ held fixed, by kernel based least squares. Therefore, the resulting estimator of $\alpha(\tau)$ is

$$\hat{\alpha}(\tau) = \left(\sum_{t=1}^{T-1} F_t K\left(\frac{\tau_t - \tau}{h} \right) F_t' \right)^{-1} \sum_{t=1}^{T-1} F_t K\left(\frac{\tau_t - \tau}{h} \right) \left(Y_{t+1} - \hat{\beta}' V_t \right); \tag{3.16}$$

this process can be repeated for $\tau = \tau_1, \ldots, \tau_T$. Finally, $(\hat{\alpha}(\tau_1), \ldots, \hat{\alpha}(\tau_T), \hat{\beta})$ with $\hat{\alpha}(\tau)$ as in (3.16) and $\hat{\beta}$ in (3.14) provide the PLS estimates of $(\alpha(\tau_1), \ldots, \alpha(\tau_T), \beta)$. We implement the foregoing derivations in the following algorithm for computing the estimates of the coefficients in the model (3.1), for a given bandwidth h:

Algorithm for estimating the FAR model when $\{F_t\}$ are known:

Step 1: Compute $(\tilde{Y}_{t+1}, \tilde{V}_t)$, defined in (3.11) and (3.12) for $t = 1, \ldots, T$.

Step 2: Compute $\hat{\beta}$ in (3.14).

Step 3: Compute $\hat{\alpha}(\tau_T)$ using (3.16).

Once $\hat{\alpha}(\tau_T)$ and $\hat{\beta}$ have been computed, a point forecast of Y_{T+1} is

$$\hat{Y}_{T+1} = \hat{\alpha}(\tau_T)' F_T + \hat{\beta}' V_T,$$

where we used F_T, instead of $\hat{F}_T$ because in this subsection we assumed that the factors are known. Although $\hat{\alpha}(\tau_1), \ldots, \hat{\alpha}(\tau_{T-1})$ do not appear explicitly in $\hat{Y}_{T+1}$, their values enter the calculation of $\hat{Y}_{T+1}$, as is clear from the foregoing derivations.

Informal arguments suggest that (a) $\sqrt{T}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, \Sigma_\beta)$ as $T \to \infty$, for some Σ_β, and (b) $\sqrt{Th}\left(\hat{\alpha}(\tau) - \alpha(\tau) - C(\tau, h)\right) \xrightarrow{d} N\left(0, \Sigma_\alpha\right)$, for some Σ_α and $C(\tau, h)$. Although we have a mixture of locally stationary and stationary regressors $\{F_t, V_t\}$ in the forecasting model, convergence rate for $\hat{\beta}$ did not change compared to the convergence rates of coefficients in semi-parametric stationary regression models (Li et al. [2011b]). Formal rigorous proofs of these two results are not provided in this thesis. These two results are indicative of the asymptotic behaviour of the estimators; these results are not assumed or used in the mathematical arguments in the rest of this thesis. Nevertheless, they play a part informally.

3.2.3 Nonparametric estimation of factors when the factors are unknown

In this subsection, we consider the case when the factors are unobservable and hence estimated factors are used for estimating the FAR model. To this end, we propose a method of estimating the factors; once they have been estimated, we propose to use the estimated factors in the method described in the previous subsection with the known factors therein replaced by the estimated factors. Throughout, we assume that a set of panel data $\{X_{it} : i = 1, \ldots, N; t = 1, \ldots, T\}$ is available for estimating the factors, and that the factors satisfy a standard factor model for the set of panel data. More specifically, we assume that the factor model for the panel data takes the form,

$$X_{it} = \lambda_i' F(\tau_t) + e_{it} \qquad (i = 1, \ldots, N; t = 1, \ldots, T), \tag{3.17}$$

where $F(\tau_t)$ is an r-dimensional vector of unknown factors at time t, λ_i is the factor loading for ith cross-sectional unit, and e_{it} is the idiosyncratic error term.

First, we propose a method of estimating the factor model.

Step 1. Estimate the factor model (3.17) by the standard principal component analysis [PCA] method. Let the resulting estimates of $\{F(\tau_t), \lambda_i\}$ be denoted by $\{\tilde{F}(\tau_t), \tilde{\lambda}_i\}$ ($i = 1, \ldots, N; t = 1, \ldots, T$). As usual, we assume that the factors are standardized by $T^{-1}\tilde{F}'\tilde{F} = I_r$.

Step 2. Holding the factor loadings $\{\tilde{\lambda}_i : i = 1, \ldots, N\}$ obtained in the previous step fixed, estimate $F(\tau)$ by

$$\hat{F}(\tau) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\lambda}_i \tilde{\lambda}_i' K\left(\frac{\tau_t - \tau}{h_f} \right) \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\lambda}_i X_{it} K\left(\frac{\tau_t - \tau}{h_f} \right) \right), \tag{3.18}$$

where K(.) is the kernel function and h_f is the bandwidth for factor estimation. We refer to these estimated factors as *nonparametric[NP] factors.*

Step 3. Using the estimated factors $\hat{F}(\tau_t)$ $(t = 1, \ldots, T)$ in the previous step, re-estimate the factor loadings by ordinary least squares. Therefore, the estimated factor loadings may be expressed as

$$\hat{\lambda}_i = \left(\sum_{t=1}^{T} \hat{F}(\tau_t)\hat{F}'(\tau_t)\right)^{-1} \left(\sum_{t=1}^{T} \hat{F}(\tau_t)X_{it}\right). \tag{3.19}$$

This nonparametric method of estimating the factors, does not assume that factors are VMA(∞); it is applicable more generally to processes that are locally stationary. Once the factors have been estimated, the FAR model may be estimated by the 3-step method in subsection 3.2.2 with $F(\tau_t)$ therein replaced by the estimates $\hat{F}(\tau_t)$ $(t = 1, \ldots, T)$ proposed in (3.18). This is the method adopted in the simulation and empirical studies of this chapter.

Remark 1: The method described by the aforementioned three steps is feasible because it provides closed form expression for the quantity to be computed in each step. Feasibility in this sense, by itself, does not imply that the method itself is methodologically sound. Although we do not provide a rigorous proofs, we conjecture that the method is sound. Local stationarity of the model suggests that if we estimate it by combining a method for estimating stationary process and a kernel smoothing type nonparametric method, then the estimated model is likely to be close to the true model. Therefore, we conjecture that the qualitative nature of local stationarity is sufficient for the local estimation outlined in the aforementioned three-step procedure to be sound. We emphasize that these arguments do not constitute a rigorous proof, but our conjecture is based on a range of insights. We do not claim that our method has advantages over other methods, since we do not have theory for other methods of estimation for the context studied in this thesis. Our objective was to develop a method that has a sound foundation. The simulations results in the next section corroborate our conjecture.

Remark 2: The assumption that F_t is a smooth function of t imposes more restriction than the assumption that F_t is a time-dependent variable or a process. The former restrictive smoothness assumption is likely to provide stronger support for the suitability of the kernel based estimator, although it is possible that the rates of convergence may possibly be unaffected. Since we do not provide rigorous derivations of the asymptotic properties, we can only conjecture how the estimators might behave. Additionally, it is reasonable to think that the factor loadings also need to be time-varying to capture the structural changes in the panel data set. However, in the literature, it has been discussed that even if the factor loadings are structurally unstable, we can well estimate the factors using full sample and time-invariant factor loadings; see Stock and Watson [2009], Banerjee et al. [2008]

3.3 Simulation study

In this simulation study, we consider one-step ahead forecasting of a univariate time series using special cases of the semi-parametric FAR model in subsection 3.2 with two factors and one observable regressor. Both known and unknown factor cases are considered in this study. The known factor case is useful for evaluating the component of the method proposed in subsection 3.2.2 that forms part of the entire method. Throughout this simulation study we use the Gaussian kernel, $K(u) = (2\pi)^{-1/2} \exp\left(-u^2/2\right)$ for nonparametric estimation.

3.3.1 The DGPs for the regressors and the FAR model

The FAR model contained are two factors and a scalar predictor V_t. Therefore, we may express the FAR model as

$$Y_{t+1} = \alpha_1(\tau_t)F_1(\tau_t) + \alpha_2(\tau_t)F_2(\tau_t) + \beta V_t + \eta_{t+1} \quad (\tau_t = t/T; t = 1, \ldots, T-1)\,, \tag{3.20}$$

where $\eta_t \sim N(0,1)$. We considered VMA(2) and VAR(2) processes for the factors and the regressor V_t.

VMA(2) process:

$$F(\tau_t) = \mu(\tau_t) + \epsilon_t + B_1(\tau_t)\epsilon_{t-1} + B_2(\tau_t)\epsilon_{t-2} \tag{3.21}$$

$$V_t = \gamma + \zeta_t + A_1\zeta_{t-1} + A_2\zeta_{t-2} \tag{3.22}$$

where $(\epsilon_t, \zeta_t) \sim MVN\left(1, \rho|\rho, 0.8\right)$.

VAR(2) structure:

$$F(\tau_t) = \mu(\tau_t) + B_1(\tau_t)F(\tau_{t-1}) + B_2(\tau_t)F(\tau_{t-2}) + \epsilon_t \tag{3.23}$$

$$V_t = \gamma + A_1V_{t-1} + A_2V_{t-2} + \zeta_t, \tag{3.24}$$

where $B_1(\tau_t)$, $B_2(\tau_t)$, and (ϵ_t, ζ_t) are as for the VMA(2) in (3.21) and (3.22). In this design, the factors $F(\tau_t)$ and the regressor V_t are observable.

We considered the four different DGPs in Table 3.1, each with VMA(2) and again each with VAR(2), providing eight DGPs. To provide an over view of the range of settings considered in the study, let us note the following: (a) Both DGP1 and DGP2 have polynomial functions of τ_t for time-varying parameters $\alpha(.)$ while DGP3 and DGP4 have sine/cosine functions. (b)

The coefficient β of the observable regressor is fixed at 0.7 in all DGPs. (c) The VMA(2) and VAR(2) coefficients of factor 1 have sine/cosine functions in DGP1 and DGP3, and exponential functions in DGP2 and DGP4. (d) The VMA(2) and VAR(2) coefficients of factor 2 and V_t in all four DGPs are time-invariant; hence they are stationary.(e) The functions $B_1(u)$ and $B_2(u)$ ($u \in [0,1]$) in the four DGPs of Table 3.1, are continuously differentiable and have bounded first derivative. If we define $Z_t(u) = \mu(u) + \epsilon_t + B_1(u)\epsilon_{t-1} + B_2(u)\epsilon_{t-2}$ then F_t satisfies the definition of local stationarity. Factor 1 is locally stationary while factor 2 and the observable regressor, V_t, are stationary.

Table 3.1: The coefficients of the data generating processes[DGP] for the simulation study

DGP1	**DGP2**
$\alpha_1(\tau_t) = (1/2)(1+\tau_t)$ $\alpha_2(\tau_t) = \tau_t + \tau_t^2$ $\beta = 0.7$	$\alpha_1(\tau_t) = (1/2)(1+\tau_t)$ $\alpha_2(\tau_t) = \tau_t + \tau_t^2$ $\beta = 0.7$
$\mu(\tau_t) = [0.5\sin(2\pi\tau_t), 0.5]$ $B_1'(\tau_t) = [0.5 + 0.3\sin(2\pi\tau_t), 0.8]$ $B_2'(\tau_t) = [-0.5 + 0.3\cos(2\pi\tau_t), -0.2]$	$\mu(\tau_t) = [0.5\sin(2\pi\tau_t), 0.5]$ $B_1'(\tau_t) = [0.7\exp(-0.8 + \tau_t), 0.5]$ $B_2'(\tau_t) = [-0.2\exp(-0.5 + \tau_t), -0.8]$
$\gamma = 0.5$ $A_1 = 0.3$ $A_2 = -0.3$	$\gamma = -0.5$ $A_1 = 0.3$ $A_2 = -0.3$
DGP3	**DGP4**
$\alpha_1(\tau_t) = (1/3)\sin(2\pi\tau_t)$ $\alpha_2(\tau_t) = \cos(2\pi\tau_t)$ $\beta = 0.7$	$\alpha_1(\tau_t) = (1/3)\sin(2\pi\tau_t)$ $\alpha_2(\tau_t) = \cos(2\pi\tau_t)$ $\beta = 0.7$
$\mu(\tau_t) = [0.5\sin(2\pi\tau_t), 0.5]$ $B_1'(\tau_t) = [0.5 + 0.3\sin(2\pi\tau_t), 0.8]$ $B_2'(\tau_t) = [-0.5 + 0.3\cos(2\pi\tau_t), -0.2]$	$\mu(\tau_t) = [0.5\sin(2\pi\tau_t), 0.5]$ $B_1'(\tau_t) = [0.7\exp(-0.8 + \tau_t), 0.5]$ $B_2'(\tau_t) = [-0.2\exp(-0.5 + \tau_t), -0.8]$
$\gamma = 0.5$ $A_1 = 0.3$ $A_2 = -0.3$	$\gamma = -0.5$ $A_1 = 0.3$ $A_2 = -0.3$

Measures of performance

All the simulation estimates are based on 1000 replications. Let $\hat{\alpha}^{(s)}(\tau)$ is the value of $\hat{\alpha}(\tau)$ in the s^{th} replication, $\hat{\beta}^{(s)}$ is the corresponding value of $\hat{\beta}$ in the s^{th} replication ($s = 1, \ldots, S$), and $\bar{\hat{\alpha}}(\tau_t)$ and $\bar{\hat{\beta}}$ are the means of estimated parameters over the S replications, defined as

$$\bar{\hat{\alpha}}(\tau_t) = \frac{1}{S}\sum_{s=1}^{S}\hat{\alpha}^{(s)}(\tau_t), \qquad \bar{\hat{\beta}} = \frac{1}{S}\sum_{s=1}^{S}\widehat{\beta}^{(s)}.$$

Based on $S = 1000$ replications, we estimate the bias, root mean square error [RMSE], and the standard deviation[SD] of the estimates of the coefficients in the FAR model as follows:

$$\text{Bias of } \hat{\alpha} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{S}\sum_{s=1}^{S}(\hat{\alpha}^{(s)}(\tau_t) - \alpha(\tau_t)) = \frac{1}{T}\sum_{t=1}^{T}(\bar{\hat{\alpha}}(\tau_t) - \alpha(\tau_t)) \tag{3.25}$$

$$\text{Bias of } \hat{\beta} = \frac{1}{S}\sum_{s=1}^{S}(\hat{\beta}^{(s)} - \beta) \tag{3.26}$$

$$\text{RMSE of } \hat{\alpha} = \sqrt{\frac{1}{ST}\sum_{t=1}^{T}\sum_{s=1}^{S}\left(\hat{\alpha}^{(s)}(\tau_t) - \alpha(\tau_t)\right)^2} \tag{3.27}$$

$$\text{RMSE of } \beta = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}^{(s)} - \beta\right)^2} \tag{3.28}$$

$$\text{SD of } \hat{\alpha} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\alpha}^{(s)}(\tau_t) - \bar{\hat{\alpha}}(\tau_t)\right)^2} \tag{3.29}$$

$$\text{SD of } \hat{\beta} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\beta}^{(s)} - \bar{\hat{\beta}}\right)^2}, \tag{3.30}$$

Note that

$$\begin{aligned}
[\text{RMSE of } \hat{\alpha}]^2 &= \frac{1}{ST}\sum_{t=1}^{T}\sum_{s=1}^{S}\left(\hat{\alpha}^{(s)}(\tau_t) - \alpha(\tau_t)\right)^2 \\
&= [\text{SD of } \hat{\alpha}]^2 + \frac{1}{T}\sum_{t=1}^{T}(\bar{\hat{\alpha}}(\tau_t) - \alpha(\tau_t))^2 \\
&\neq [\text{SD of } \hat{\alpha}]^2 + [\text{Bias of } \hat{\alpha}]^2,
\end{aligned}$$

where $\neq$ holds, in general. Therefore, RMSE of $\hat{\alpha}$ provides information not contained in the standard deviation or Bias of $\hat{\alpha}$. By contrast, $\hat{\beta}$ is not time varying, and we have

$$[\text{RMSE of } \hat{\beta}]^2 = [\text{SD of } \hat{\beta}]^2 + [\text{Bias of } \hat{\beta}]^2.$$

The next two subsections provide the details for VMA(2) and for VAR(2), respectively.

3.3.2 Simulation with known factors

Simulation results when the regressors are known and are VMA(2)

Let us consider the case when the two factors and the scalar regressor in (3.20) are observable; we assume that and they are VMA(2) with the coefficients being defined by one of the four DGPs in Table 3.1. Heuristic arguments suggest that the optimal bandwidth is $dT^{-1/5}$ for some $d > 0$. In this simulation study, we first estimated a suitable value for d when $T = 100$, and then used the chosen value of d in the other simulations with $T = 300$ and $T = 800$. Since cross-validation is computing intensive, the aforementioned procedure is reasonable for choosing a bandwidth for the simulation study with large values for T.

Let us denote the estimate of $\alpha(\tau_t)$ when the t^{th} observation is deleted from the sample as,

$$\hat{\alpha}_{(-t)}(\tau_t) = \left(\sum_{u=1,u\neq t}^{T-1} F_u K\left(\frac{\tau_u - \tau_t}{h}\right) F_u'\right)^{-1} \sum_{u=1,u\neq t}^{T-1} F_u K\left(\frac{\tau_u - \tau_t}{h}\right)\left(Y_{u+1} - \tilde{\beta}' V_u\right)$$

in which

$$\begin{aligned}
\tilde{\beta} &= \left(\sum_{t=1}^{T-1} \tilde{V}_{(-t)}' \tilde{V}_{(-t)}\right)^{-1} \sum_{t=1}^{T-1} \tilde{V}_{(-t)}' \tilde{Y}_{(-(t+1))}, \\
\tilde{Y}_{-(t+1)} &= Y_{t+1} - F_t'\left(\sum_{u=1,u\neq t}^{T-1} F_u K\left(\frac{\tau_u - \tau_t}{h}\right) F_u'\right)^{-1} \sum_{u=1,u\neq t}^{T-1} F_u K\left(\frac{\tau_u - \tau_t}{h}\right) Y_{u+1}, \\
\tilde{V}_{(-t)} &= V_t - F_t'\left(\sum_{u=1,u\neq t}^{T-1} F_u K\left(\frac{\tau_u - \tau_t}{h}\right) F_u'\right)^{-1} \sum_{u=1,u\neq t}^{T-1} F_u K\left(\frac{\tau_u - \tau_t}{h}\right) V_u.
\end{aligned}$$

Let $\hat{\alpha}_{(-t)}^{(s)}(\tau_t)$ and $\tilde{\beta}^{(s)}$ are the values of $\hat{\alpha}_{(-t)}(\tau_t)$ and $\tilde{\beta}$ in the s^{th} replication ($s = 1, \ldots, S$), respectively. Then,

$$\overline{CV}_1(d) = \frac{1}{TS} \sum_{s=1}^{S} \sum_{t=1}^{T-1} \left(Y_{t+1}^{(s)} - \left(\hat{\alpha}_{(-t)}^{(s)}(\tau_t)' F_t^{(s)} + \tilde{\beta}^{(s)\prime} V_t^{(s)}\right)\right)^2 \qquad (3.31)$$

is the mean cross validation measure taken over $S = 1000$ repeated samples. Values of $\overline{CV}_1(d)$ for ten different values of d when $T = 100$ are provided in Table 3.2. The values in this table show that $\bar{CV}(d)$ decreases with increasing d for each of the four DGPs. For $d > 1.133$, the value of $\overline{CV}_1(d)$ changes hardly for each DGP. Therefore, we chose $d = 1.133$.

With $d = 1.133$ and $h = dT^{-1/5}$, we estimated bias, RMSE, and SD of $\{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}\}$ for the models defined by Table 3.1. The computed values appear in Table 3.3. The results in this table indicate that, for each DGP, both SD and RMSE decrease with increasing T; further, the bias decreases with increasing T, for most cases. The overall performance of the method improves as T increases, which provides confidence in the proposed nonparametric kernel based method.

Table 3.2: Calculated $\overline{CV}(d)$ when the regressors are known and are VMA(2) processes

d	0.050	0.267	0.483	0.7	0.917	**1.133**	1.35	1.567	1.783	2.00
h_p	0.020	0.106	0.192	0.279	0.365	**0.451**	0.537	0.624	0.710	0.796
DGP1	32.69	2.93	2.55	2.27	2.11	2.03	1.99	1.96	1.95	1.95
DGP2	3.59	2.96	2.66	2.41	2.24	2.12	2.10	2.08	2.07	2.06
DGP3	34.04	2.85	2.33	2.13	2.07	2.05	2.03	2.03	2.03	2.02
DGP4	3.64	2.79	2.36	2.20	2.14	2.13	2.12	2.11	2.11	2.11

The bandwidth $h_p = dT^{-1/5}$ and $T = 100$.

Simulation results when the regressors are known and are VAR(2)

The design of the simulation study in this subsection is essentially the same as that in the previous subsection for VMA(2) except that VMA(2) is replaced by VAR(2). The model is defined in (3.23) and (3.24), and the parameter values are provided in Table 3.1. Just as in the previous subsection for VMA(2), we computed the mean of the cross-validation values for ten bandwidths; they are presented in Table 3.4. Based on the results in this table, and by arguments similar to those applied to Table 3.2, we chose $d = 1.35$; therefore, we chose the bandwidth as $h_p = 1.35T^{-1/5}$ for all values of T. Table 3.5 provides estimates of Bias, RMSE, and standard deviation[SD]; the interpretation for this Table is same as that for Table 3.3. Table 3.5 shows that, as T increases, SD decreases, Bias increases for most of the cases, and RMSE decreases for most of the cases. Again, the overall performance of the method improves as T increases, which provides confidence in the proposed nonparametric kernel based method.

The kernel method of estimating factors in Subsection 3.2.3 is new. The simulations in this section with known factors, help us to evaluate the reliability of the method without the confounding effects of factor estimation. If the method were to perform poorly in simulations with known factors, then the chances that the method would not perform well when the factors are estimated. In this sense, the simulations in this section with known factors is a designed experiment to learn about the proposed kernel method. The fact that our method performed well in these simulations provides us the evidence needed to proceed to the next stage of investigating the method for the case when factors are estimated. These positive outcomes do not guarantee that the method will work well with estimated factors, but the indications are promising.

3.3.3 Simulation with estimated factors

We extended the simulation study in the previous subsection to the case when the factors are unobservable, and hence estimated factors are used for estimating the FAR model. Therefore, the extension in this subsection involves the additional steps of generating the panel data and estimating the factors. We also adopted a method similar to that in the previous subsection for choosing the bandwidth for estimating the factors non-parametrically.

In what follows, we first list the simulations steps for a given bandwidth, and then describe the method that we adopted for choosing the bandwidth for the simulation study. In general, the bandwidth performs well if it is data dependent and is chosen in a robust and optimal way. We did not use a data dependent bandwidth in these simulations. The main reason are that we did not study the bandwidth choice rigorously, and we needed to manage the high computational demands that data dependent bandwidth introduce. We chose a bandwidth that is likely to behave robustly by conducting some exploratory simulation studies. Since we do not have a data dependent method that has a sound theoretical basis for optimal choice, we believe that it is sufficient to implement the aforementioned method that may not be optimal but likely to be adequate for our intended purposes. The method of choosing the bandwidth adopted in the empirical example, presented in a later section, is data dependent and hence is different.

Simulation Steps:

Step 1. Generate the two factors $\{F_1(\tau_t), F_2(\tau_t) : t = 1, \ldots, T\}$ as in subsection 3.3.1 using a DGP in Table 3.1.

Step 2. Generate the panel data set $\{X_{it} : i = 1, \ldots, N; t = 1, \ldots, T\}$ using the factor model

$$X_{it} = \lambda_{i1}F_1(\tau_t) + \lambda_{i2}F_2(\tau_t) + e_{it} \qquad (i = 1, \ldots, N; t = 1, \ldots, T),$$

where $\lambda_{i1}, \lambda_{i2}$, and e_{it} are independent and identically distributed with common distribution function $N(0, 1)$ $(i = 1, \ldots, N;\ t = 1, \ldots, T)$.

Step 3. Estimate the factors $F_1(\tau_t)$ and $F_2(\tau_t)$ by the method proposed in Subsection 3.2.3.

Step 4. Estimate the FAR model

$$Y_{t+1} = \alpha_1(\tau_t)\tilde{F}_1(\tau_t) + \alpha_2(\tau_t)\tilde{F}_2(\tau_t) + \beta V_t + \eta_{t+1} \quad (t = 1, \ldots, T-1), \tag{3.32}$$

which is the same as that in (3.20) except that the factors $F_1(\tau_t)$ and $F_2(\tau_t)$ are replaced by the estimates obtained in the previous step.

The foregoing simulation steps assume that the bandwidths are given. In this simulation study, we adopted a method similar to that suggested by Table 3.2. Let $N = 100$ and $T = 100$. Since the nonparametric structure is only on the set of factors, which is a function of time t, we estimate the factors F_t nonparametrically based on the time series dimension. Heuristic arguments suggest that the optimal bandwidth for factor estimation is also $cT^{-1/5}$ for some $c > 0$. Therefore, executed the first three steps of the simulation method introduced in this section with $h_f = cT^{-/5}$ for ten different values of c. The cross-validation measure $CVSE_f$ was computed for these ten values of c as,

$$CVSE_f(h) = \frac{1}{TNS}\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}^{(s)} - \tilde{\lambda}_i^{(s)\prime}\hat{F}_{(-t)}^{(s)}(\tau_t)\right)^2 \tag{3.33}$$

where $\hat{F}_{(-t)}^{(s)}$ is the nonparametrically estimated factor without the t^{th} observation, $\tilde{\lambda}_i^{(s)}$ is the i^{th} factor loading estimated from PCA for s^{th} replication, and S is the number of replications. The results are summarized in Tables 3.6 and 3.7; the definitions of Bias, SD and RMSE are as in the previous subsection. The figures in Table 3.6 show that the cross validation measure is smallest when $c = 0.05$.

The results in Table 3.7 show that nonparametric factor estimators are not that sensitive to the choice of the bandwidth. Results in the columns under the NP factors correspond to the nonparametrically estimated factors, and the results under the PCA column are for the estimated factors and the loadings using PCA. The SD and RMSE of nonparametrically estimated factors are lower compared to those for the PCA estimator. The entries in this table show that as T increases with fixed N, both RMSE and SD decrease for almost all cases. The bias of the estimators of time varying parameters in DGP3 and DGP4 increase as T increases. Overall, the results in Table 3.7 suggest that values of c larger than 0.05 are desirable. Based on the overall performance in terms of the different criteria presented in the two tables, we chose $c = 0.05$. Therefore, we proceeded assuming that $h_f = 0.05T^{-1/5}$ is a suitable way of setting bandwidths for values of T larger than 100.

Next, we performed a similar simulation study to choose a suitable bandwidth for h_p. Let $N = 100$ and $T = 100$. We conducted the simulation experiment for the four DGPs in Table 3.1 with $h_p = dT^{-1/5}$ for ten different values of d. Table 3.8 shows that the $CVSE_p$ decreases as h_p increases for all four DGPs. For $d \geqslant 1.133$, the value of $CVSE_p$ changes hardly. This table shows that $d = 1.133$ is a suitable choice. One of our main objectives of the simulation study is to investigate whether our proposed nonparametric method is stable and performs in a reliable manner. For this purpose, the method that we adopted for choosing the bandwidth serves our purpose well.

For the rest of this simulation study, we fixed the bandwidths as $h_p = 1.133T^{-1/5}$ and $h_f = 0.05T^{-1/5}$ with T and N taking the values 100, 300 and 500. We considered the same DGPs as in Table 3.1 with the factors being VMA(2) and the coefficients as in Table 3.1. As stated in Chapter 2, the coefficients are estimated up to a rotation, H, which is a function of latent factors and factor loadings. Therefore, when factors are estimated, we cannot directly estimate the bias, SD, and RMSE of the estimated coefficients as the rotation matrix is infeasible. Hence, we first estimate the rotation matrix by replacing the estimated factors and factor loadings. The bias, SD, and RMSE of the coefficients are estimated compared to $\alpha\tilde{H}^{-1}$ where $\tilde{H}$ is the estimated rotation matrix. The results are in Table 3.9.

Table 3.3: Estimated bias, RMSE, and standard deviation[SD] for the estimates when the regressors are known and are VMA(2)

		DGP1			DGP2		
	T	100	300	800	100	300	800
$\hat{\alpha}_1$	Bias	-190	-172	-146	6.7	18	15.1
	RMSE	278	206	167	197	126	86
	SD	192	106	71	174	101	64
$\hat{\alpha}_2$	Bias	111	113	97	-27	-25	-21
	RMSE	490	404	332	434	342	272
	SD	191	106	70	153	86	53
$\hat{\beta}$	Bias	-26	-14	-7.5	4.6	1.8	-0.5
	RMSE	114	61	34	111	59	33
	SD	111	60	33	111	59	33
		DGP3			DGP4		
$\hat{\alpha}_1$	Bias	-100	-75	-59	-65	-54	-50
	RMSE	291	210	163	286	224	191
	SD	209	128	88	188	109	72
$\hat{\alpha}_2$	Bias	20.7	-3.3	-21	-41	-48	-51
	RMSE	723	672	616	700	651	593
	SD	228	137	96	173	103	66
$\hat{\beta}$	Bias	5.8	6.8	5.9	-22	-18	-16
	RMSE	137	77	41	136	79	44
	SD	137	77	41	134	77	41

Each entry is multiplied by 1000. The set $\{\hat{\alpha}_1, \hat{\alpha}_2\}$ and $\hat{\beta}$ correspond to the estimated time-varying and constant parameters, respectively. The bandwidth for parameter estimation is $h_p = 1.133T^{-1/5}$.

Table 3.4: Calculated $\overline{CV}_1(d)$ for all four DGPs, when regressors are known and are VAR(2)

d	0.050	0.267	0.483	0.70	0.917	1.133	**1.350**	1.567	1.783	2.00
h_p	0.020	0.106	0.192	0.279	0.365	0.451	**0.537**	0.624	0.710	0.796
DGP1	4.28	3.49	3.13	2.82	2.61	2.49	2.43	2.41	2.39	2.38
DGP2	4.22	3.51	3.12	2.78	2.57	2.45	2.39	2.36	2.35	2.34
DGP3	5.14	3.83	3.14	2.90	2.82	2.79	2.78	2.77	2.77	2.76
DGP4	4.81	3.67	3.02	2.78	2.70	2.67	2.66	2.65	2.64	2.64

The bandwidth $h_p = dT^{-1/5}$ and number of observations, $T = 100$.

Table 3.5: Estimated bias, RMSE, and standard deviation[SD] for the estimates when the regressors are known and are VAR(2)

	T	100	300	800	100	300	800
			DGP1			DGP2	
$\hat{\alpha}_1$	Bias	9.8	-160.0	-223.2	-0.4	-42.0	-43.1
	RMSE	138.5	181.1	229.5	150.3	114.3	92.7
	SD	93.3	75.6	51.2	105.5	67.4	44.0
$\hat{\alpha}_2$	Bias	-18.8	26.4	71.0	-14.9	0.7	-3.0
	RMSE	467.3	414.0	383.6	476.0	401.1	329.7
	SD	133.2	76.0	46.4	153.9	96.8	61.2
$\hat{\beta}$	Bias	7.1	-14.6	-19.3	4.8	14.6	5.5
	RMSE	123.7	66.4	42.5	125.1	67.5	37.0
	SD	123.5	64.7	37.8	125.0	65.9	36.6
			DGP3			DGP4	
$\hat{\alpha}_1$	Bias	69.5	162.9	156.0	16.6	-20.1	-33.5
	RMSE	247.4	241.1	233.5	235.7	188.6	161.4
	SD	121.6	102.3	65.8	125.6	80.3	53.7
$\hat{\alpha}_2$	Bias	-83.1	-117.9	-136.0	-69.0	-64.9	-69.8
	RMSE	716.1	705.5	698.2	722.9	688.8	648.4
	SD	170.3	107.8	71.6	199.2	134.8	93.1
$\hat{\beta}$	Bias	15.7	40.2	30.3	-9.7	-12.1	-17.4
	RMSE	166.0	97.8	59.4	169.3	93.3	55.9
	SD	165.3	89.2	51.1	169.0	92.5	53.2

The entries in the table have been multiplied by 1000. $\{\hat{\alpha}_1, \hat{\alpha}_2\}$ and $\hat{\beta}$ correspond to the estimated time-varying and constant parameters respectively. Bandwidth for parameter estimation, $h_p = 1.133T^{-1/5}$.

Table 3.6: Estimated $CVSE_f$ for 10 different bandwidths, $[T, N] = [100, 100]$

c	**0.050**	0.267	0.483	0.700	0.917	1.133	1.350	1.567	1.783	2.000
h_f	**0.020**	0.106	0.192	0.279	0.365	0.451	0.537	0.624	0.710	0.796
DGP1, DGP3	313.9	407.9	416.4	424.6	425.3	428.0	430.0	432.5	427.2	436.5
DGP2, DGP4	345.4	415.8	422.9	427.0	429.9	432.0	433.6	434.7	435.4	436.0

The bandwidth for factor estimation is $h_f = cT^{-1/5}$. The DGP1 and DGP3 have the same factors, while DGP2 and DGP4 has the same factors. Highlighted c, h_f values provied the least estimated $CVSE_f$ values.

Table 3.7: Estimated bias, RMSE, and standard deviation of the estimated NP and PCA factor and their loadings for $[T, N] = [100, 100]$

	DGP1 and DGP3											
	NP factors											PCA
	c	0.050	0.267	0.483	0.700	0.917	1.133	1.350	1.567	1.783	2.000	
	h_f	0.020	0.106	0.192	0.279	0.365	0.451	0.537	0.624	0.710	0.796	
$\hat{F}_1$	Bias	200.8	201.1	201.2	201.2	201.2	201.1	201.1	201.1	201.0	201.0	200.1
	SD	406.7	185.2	146.0	127.7	117.9	112.7	110.0	108.7	107.9	107.5	494.3
	RMSE	1149.7	1276.4	1291.2	1300.2	1306.3	1310.3	1313.0	1314.8	1316.0	1316.9	970.6
$\hat{F}_2$	Bias	95.1	95.6	96.7	97.4	97.4	97.1	96.8	96.5	96.3	96.1	95.3
	SD	307.3	115.9	79.2	63.0	54.2	49.3	46.7	45.2	44.3	43.8	668.0
	RMSE	1302.4	1342.1	1332.5	1318.2	1308.3	1303.2	1300.7	1299.6	1299.0	1298.7	1477.7
$\hat{\lambda}_1$	Bias	-4.6	-4.2	-7.0	-16.9	-13.7	-25.8	-35.6	-46.4	-50.0	-37.9	-3.8
	SD	2059.7	2019.7	2423.9	3946.8	6231.0	10149.5	13939.7	19443.7	26940.9	37299.5	1368.5
$\hat{\lambda}_2$	Bias	2.0	1.8	2.1	4.5	2.7	4.1	5.5	4.7	1.4	-5.5	2.3
	SD	634.5	634.7	722.4	1031.1	1472.2	2240.3	3059.4	4292.1	6038.2	8505.5	575.8
	DGP2 and DGP4											
	NP factors											PCA
$\hat{F}_1$	Bias	217.0	217.3	217.3	217.3	217.3	217.2	217.2	217.2	217.2	217.1	217.1
	SD	327.2	140.9	110.9	96.8	89.1	85.0	82.9	81.8	81.2	80.8	488.5
	RMSE	1042.3	1186.0	1203.2	1213.2	1219.8	1224.1	1226.8	1228.7	1230.0	1231.0	968.4
$\hat{F}_2$	Bias	26.6	26.6	27.3	27.7	27.8	27.6	27.4	27.2	27.1	27.0	26.6
	SD	371.4	161.3	118.3	100.3	91.5	86.9	84.5	83.2	82.4	82.0	714.1
	RMSE	1562.3	1462.1	1432.2	1410.4	1396.9	1389.8	1386.3	1384.4	1383.4	1382.8	1557.8
$\hat{\lambda}_1$	Bias	-4.8	-4.0	-3.9	-0.3	5.8	-3.1	6.6	-9.9	-52.7	-67.2	-4.0
	SD	2083.9	1803.2	1985.8	2592.1	4531.6	7048.4	11904.7	16914.7	22729.3	28693.3	1435.2
$\hat{\lambda}_2$	Bias	1.1	1.4	1.3	-0.3	-4.0	-1.7	-5.9	-1.1	9.8	10.1	1.3
	SD	649.9	690.4	791.5	1070.6	2008.3	2883.0	4298.1	5592.3	7023.0	8786.3	597.6

All the results are multiplied by 1000. Results under the NP panel are corresponding to the nonparametrically estimated factors and the PCA column is corresponding to the PCA estimated factors. Bandwidth in factor estimation for NP factors is $h_f = cT^{-1/5}$.

Table 3.8: Estimated $CVSE_p$ for all four DGPs when factors are estimated, $[T, N] = [100, 100]$

d	0.050	0.267	0.483	0.700	0.917	**1.133**	1.35	1.567	1.783	2.00
h_p	0.020	0.106	0.192	0.279	0.365	**0.451**	0.537	0.624	0.710	0.796
DGP1	19.14	2.44	2.13	1.99	1.95	1.93	1.93	1.92	1.92	1.92
DGP2	5.87	2.26	2.09	1.99	1.96	1.95	1.95	1.94	1.94	1.94
DGP3	19.32	2.57	2.21	2.12	2.10	2.08	2.08	2.08	2.08	2.08
DGP4	6.05	2.37	2.17	2.10	2.07	2.06	2.05	2.05	2.05	2.05

The factor estimation bandwidth is fixed to be $h_f = 0.05T^{-1/5}$ and the bandwidth for model estimation, $h_p = dT^{-1/5}$.

After choosing the two optimal bandwidths, $\hat{h}_f = 0.05T^{-1/5}$ and $\hat{h}_p = 1.133T^{-1/5}$, the semiparametric model was estimated for the nine combinations of $[T, N]$ using kernel estimation. The results, presented in Table 3.9, show that as T increases (with fixed N) both RMSE and SD decrease for most of the cases. Also, bias of the time-varying parameters in DGP3 and DGP4 increase as T increases while bias of the constant parameter decreases in all four DGPs. According to the simulation design in Table 3.1, ranges of the coefficients are: DGP 1 and 2: $\alpha_1 = [0, 1]$, $\alpha_2 = [0, 2]$, $\beta = 0.7$, DGP 3 and 4: $\alpha_1 = [-0.3, 0.3]$, $\alpha_2 = [-1, 1]$, $\beta = 0.7$. The scale of maximum bias and RMSE of the estimated coefficients in Table 3.9 for all four DGPs are: DGP 1 and 3: $\alpha_1 : 0.3, \alpha_2 : 0.6$ and $\beta : 0.04$. The results in Table 3.9 are of the same scale as the result in Table 3.3, for known factor case. Additionally, as a ratio, the estimated bias and RMSE of the coefficients are not that high.

Table 3.9: Estimated bias, RMSE, and standard deviation of the estimated coefficients when factors are unknown, $T = [100, 300, 500]$

N			DGP1			DGP2			DGP3			DGP4		
		T	100	300	500	100	300	500	100	300	500	100	300	500
100	$\hat{\alpha}_1$	Bias	-196.48	-236.99	-253.60	60.73	14.80	-26.37	58.57	93.86	98.67	-12.38	41.90	48.30
		std	300.58	288.81	268.21	320.36	287.76	252.10	350.92	382.96	376.78	363.09	366.24	342.81
		RMSE	369.18	379.17	375.62	345.08	301.72	266.84	397.62	420.93	411.81	414.38	409.13	381.76
	$\hat{\alpha}_2$	Bias	431.90	402.26	381.05	341.58	339.86	333.80	-111.47	-147.08	-166.66	-67.98	-116.23	-135.31
		std	163.20	104.92	84.14	165.91	111.24	94.04	195.69	121.13	98.31	210.33	148.25	124.23
		RMSE	694.02	635.07	601.57	616.17	574.67	553.64	737.52	722.93	716.99	730.23	711.42	701.98
	$\hat{\beta}$	Bias	-18.25	-8.54	-4.39	6.63	6.41	5.63	7.81	7.00	7.65	-7.61	-5.61	-3.33
		std	103.21	58.39	44.07	107.95	59.60	44.38	109.39	62.74	46.27	113.19	62.25	46.23
		RMSE	104.81	59.01	44.29	108.15	59.94	44.73	109.67	63.13	46.90	113.44	62.51	46.35
300	$\hat{\alpha}_1$	Bias	-164.23	-239.28	-238.54	95.79	13.83	-25.47	86.52	110.61	121.06	27.92	54.83	66.29
		std	304.61	287.53	275.44	330.30	280.49	254.11	353.24	376.99	392.60	378.40	357.25	352.30
		RMSE	356.48	379.78	369.65	362.39	295.36	265.49	404.87	419.95	434.61	428.37	403.34	395.17
	$\hat{\alpha}_2$	Bias	431.49	403.63	385.42	337.30	339.74	332.34	-113.42	-146.12	-160.66	-77.22	-123.51	-140.60
		std	162.67	101.63	83.11	165.95	107.33	90.68	192.77	123.40	96.70	214.49	142.81	126.03
		RMSE	694.77	635.74	604.69	614.94	576.65	553.36	737.09	724.12	716.67	731.90	711.68	704.24
	$\hat{\beta}$	Bias	-18.52	-11.61	-8.20	9.21	4.01	3.08	10.41	5.10	4.38	-4.46	-7.03	-5.62
		std	101.82	58.37	45.05	106.72	59.77	46.30	110.57	61.36	47.98	115.34	62.20	47.57
		RMSE	103.49	59.51	45.80	107.12	59.90	46.41	111.06	61.57	48.18	115.42	62.60	47.90
500	$\hat{\alpha}_1$	Bias	-157.54	-234.41	-253.82	81.58	16.71	-25.07	84.60	111.45	111.40	16.75	57.64	74.53
		std	308.29	285.06	268.61	334.30	284.92	253.42	357.49	380.48	381.73	381.29	367.77	351.21
		RMSE	356.51	373.90	374.93	362.31	297.97	264.71	408.89	424.41	420.83	429.81	413.71	396.15
	$\hat{\alpha}_2$	Bias	423.99	407.63	382.46	329.97	341.95	333.62	-117.22	-143.87	-162.18	-84.76	-124.39	-142.38
		std	165.43	101.75	80.00	167.36	105.86	89.15	196.14	117.81	94.95	214.06	144.49	120.69
		RMSE	687.82	638.57	602.00	610.82	576.85	554.07	739.07	722.88	716.58	732.81	713.38	703.36
	$\hat{\beta}$	Bias	-17.71	-9.46	-4.34	8.05	7.03	6.10	11.08	6.59	8.48	-7.80	-4.34	-1.48
		std	107.37	57.22	44.35	108.28	58.95	45.19	114.18	61.42	47.30	114.90	62.28	47.17
		RMSE	108.82	58.00	44.56	108.58	59.36	45.60	114.71	61.77	48.06	115.17	62.43	47.19

All the values are multiplied by 1000. The semiparametric FAR model is considered with nonparametrically estimated factors. Two optimal bandwidths considered in NP factor estimation and model estimation are $\{h_f, h_p\} = \{0.05T^{-1/5}, 1.133T^{-1/5}\}$.

3.4 Empirical application

In this section we apply the semi-parametric FAR model for forecasting the three key economic variables, *log(GDP), GDP growth rate,* and *inflation.* Since the data set contains quarterly data, we start with a basic AR(4) (and AR(1)) model and augmented it with the nonparametric FAR model. The coefficients of the lag terms in the autoregressive part of the model are assumed to be time invariant in every model, and the coefficients of the factors are allowed to be varying with time unless the contrary is made clear. We compare and contrast the in-sample and out-of-sample forecast performance of the semi-parametric models with AR models and the mean model. For the semi-parametric FAR models, we consider both PCA and nonparametric estimates of the factors. To quantify forecast performance of different models, we use in-sample mean squared error (MSE in-sample) and out-of-sample R-square R^2_{os} defined as,

$$R^2_{os} = 1 - \left(\sum_{t=T_1+1}^{T} (Y_t - \hat{Y}_t)^2 \right) \left(\sum_{t=T_1+1}^{T} (Y_t - \tilde{Y}_t)^2 \right)^{-1}, \tag{3.34}$$

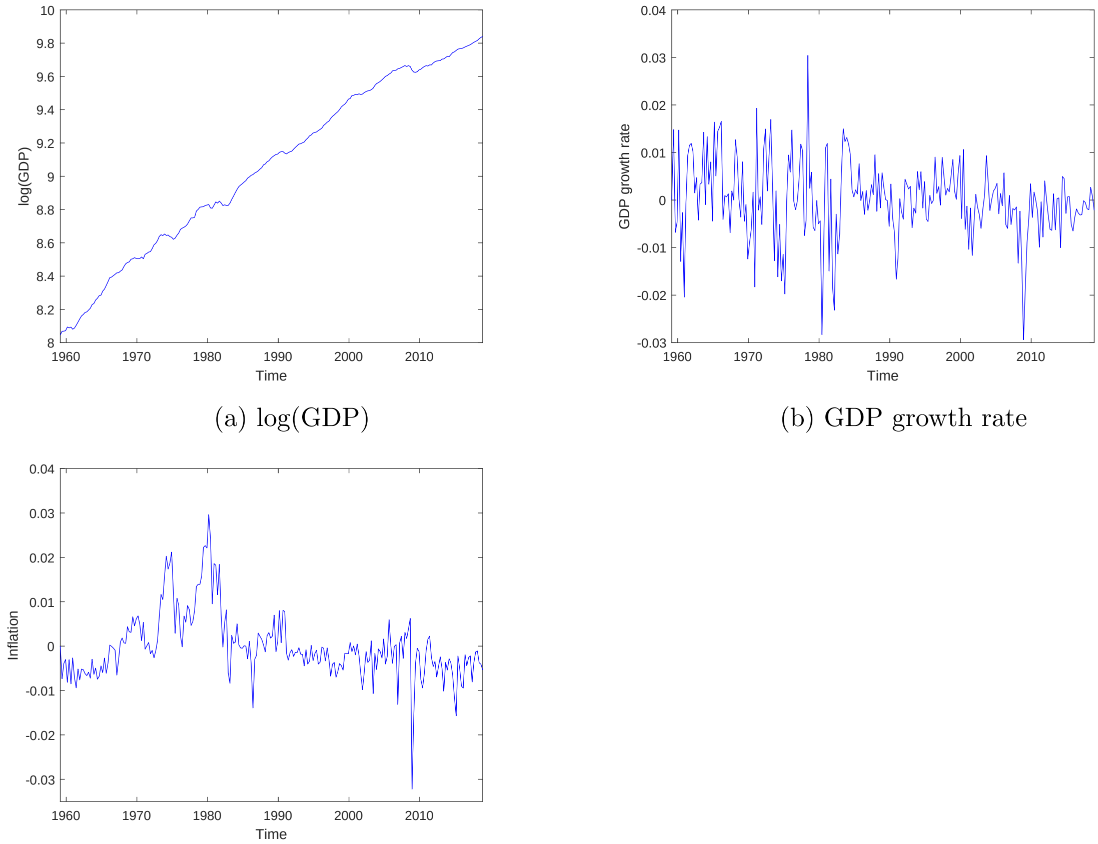
where T_1 is the end of the estimation period, and $\hat{Y}_t$ and $\tilde{Y}_t$ are the predictions based on the semi-parametric FAR model and the benchmark model, respectively. In our comparisons, we consider AR(1) to be the benchmark model. The Gaussian kernel, denoted K, is used throughout.

3.4.1 Data description

We used the same FRED-QD panel data set as that in Chapter 2; therefore, $T = 240$ and $N = 100$. The variables in the set of panel data are categorized into two aggregation levels called " high-aggregates" containing 110 variables and "sub-aggregates" containing 100 variables. The variables in sub-aggregates are used for estimating the factors and the resulting estimated factors are used as predictors to forecast the response variables in the high-aggregates. The three variables that we wish to forecast, namely *log(GDP), GDP growth rate* and *inflation* are in the high-aggregate group. Time series plots of these variables are presented in Figure 3.1.

3.4.2 Nonparametric factor estimation

The purpose of this subsection is to compare and contrast the nonparametric method of estimating factors introduced in Subsection 3.2.3 with the PCA estimates. We choose that $r = 8$, where r is the number of factors. The previous Chapter provide the rational for choosing 8 factors by applying Bai and Ng [2002] information criteria (see Chapter 2, section 4). We estimated



(a) log(GDP)

(b) GDP growth rate

(c) Inflation

Figure 3.1: Time series plots of the three response variables

the eight factors and their corresponding loadings by PCA, and treated them as the initial estimates of the factors and factor loadings required in Step 1 of the nonparametric estimation procedure in Subsection 3.2.3. To implement Step 2 of the procedure therein, we need to choose a bandwidth h_f. To this end, we assumed that the optimal bandwidth satisfies $h_f = cT^{-1/5}$, for some unknown c. We chose the value of c and hence the bandwidth, $\hat{h}_f$, by minimizing the leave-one-out cross-validation criterion,

$$CVSE_f(h) = \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \tilde{\lambda}_i' \hat{F}_{(-t)}(\tau_t) \right)^2, \tag{3.35}$$

where $\tilde{\lambda}_i$ is the i^{th} factor loading estimated using PCA and $\hat{F}_{(-t)}(\tau_t)$ is the set of nonparametrically estimated factors (NP factors) by leaving the t^{th} observation out. We computed the cross-validation criterion for a set of grid values of c, and chose the value c for which $CVSE_f(h)$ was the minimum; this involved, estimating the eight factors nonparametrically for each bandwidth using the panel data on 100 variables. The chosen value of the bandwidth was $h_f = 0.017$. To provide some insight into the suitability of the bandwidth, recall that $T = 240$. Therefore,

$h_f = 0.017$ corresponds to $T \times h_f = 4.1$; thus the kernel, the standard normal density, is applied with a bandwidth of approximately 4 quarters. Therefore,in the context of these data, $h_f = 0.017$ does not appear to be too small or too large. The estimated PCA and NP factors are shown in Figure 3.2. This figure shows that for each factor, the estimate obtained by the proposed nonparametric method has the same trend as the one estimated by PCA, but the former is smoother compared to the latter.

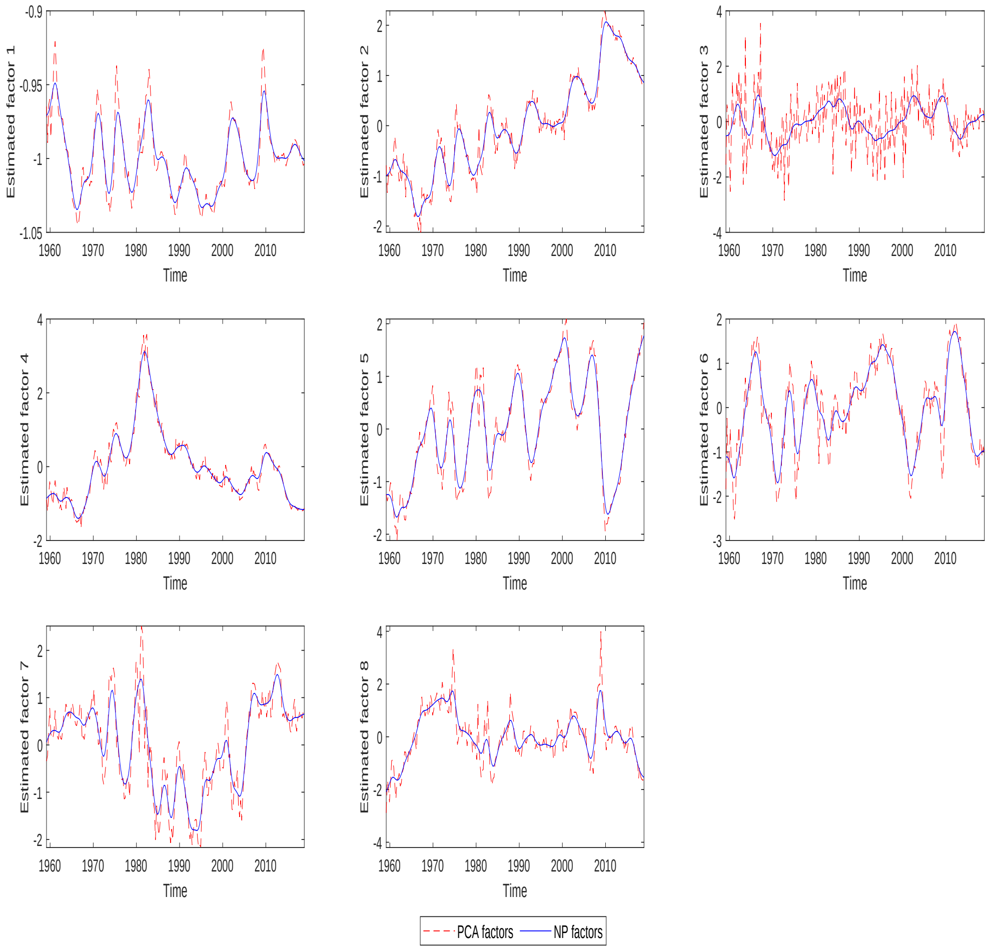


Figure 3.2: The plot of 8 estimated NP and PCA factors using the panel dataset of 100 variables

The blue solid lines are corresponding to the nonparametrically estimated factors and red dash lines are corresponding to the PCA factors. The bandwidth for the factor estimation is 0.017.

3.4.3 One-step ahead forecasting evaluation

The FAR models that we considered are formed by including the factors as additional predictors to AR(1) and AR(4) models. Therefore, a suitable way to assess whether the inclusion of factors improve the performance of predictors is to use the models without the factors as benchmarks. Therefore, we considered AR(1) and AR(4) as benchmarks. Furthermore, AR models and mean models have been used in the literature as benchmark models for evaluating the forecasting models, see Stock and Watson [1998a, 2005, and 2012]. The model specifications are as follows:

FAR models,

$$\textit{Model 1: } Y_{t+1} = \alpha_t' \hat{F}_t + \beta Y_t + \eta_{t+1}, \tag{3.36}$$

$$\textit{Model 2: } Y_{t+1} = \alpha_t' \tilde{F}_t + \beta Y_t + \eta_{t+1}, \tag{3.37}$$

$$\textit{Model 3: } Y_{t+1} = \alpha_t' \hat{F}_t + \sum_{i=0}^{3} \beta_i Y_{t-i} + \eta_{t+1}, \tag{3.38}$$

$$\textit{Model 4: } Y_{t+1} = \alpha_t' \tilde{F}_t + \sum_{i=0}^{3} \beta_i Y_{t-i} + \eta_{t+1}, \tag{3.39}$$

AR models and mean model

$$\textit{Model 5: } Y_{t+1} = \beta Y_t + \eta_{t+1}, \tag{3.40}$$

$$\textit{Model 6: } Y_{t+1} = \sum_{i=0}^{3} \beta_i Y_{t-i} + \eta_{t+1}, \tag{3.41}$$

$$\textit{Model 7: } Y_{t+1} = \bar{Y} + \eta_{t+1}, \tag{3.42}$$

where $\hat{F}$ is the set of 8 estimated factors using the nonparametric estimation and $\tilde{F}$ is the set of estimated factors using PCA, α_t is the time-varying coefficients of factors, β is the set of coefficients of lagged terms, and $\bar{Y}$ is the in-sample mean of the response variable.

Assessing the in-sample performance

By examining the sensitivity of in-sample performance to bandwidth using arguments similar to those in Appendix for out-of-sample sensitivity, we chose $\hat{h}_f = 0.011$. For bandwidth h_p we assumed that the optimal value satisfies $h_p = dT^{-1/5}$ for some unknown d. We considered 10 different values for d in the interval $(0.05, 2.00)$ and estimated the model using semi-parametric estimation. Table 3.10 provides the in-sample MSE of prediction for all the model specifications for the three response variables. Table 3.10 shows that the FAR models have the least MSE when $h_p = 0.017$. Moreover, the FAR model with nonparametrically estimated factors and four lags of the response variable, performed better in in-sample forecasting for log(GDP) and GDP growth rate. For inflation, Model 4, which contains PCA estimated factors, has the least MSE compared to the other models. Overall, for all bandwidths h_p, FAR model specifications perform better than the AR models and mean model in in-sample performance for the three response variables.

Table 3.10: In-sample performance of the models in terms of MSE for different bandwidths h_p

h_p	0.017	0.089	0.162	0.234	0.306	0.379	0.451	0.524	0.596	0.668
					log(GDP)					
Model 1	2.59	8.16	10.74	11.72	12.11	12.35	12.50	12.60	12.68	12.73
Model 2	3.54	9.06	10.30	11.00	11.36	11.61	11.77	11.88	11.95	12.01
Model 3	**1.72**	7.76	10.08	11.00	11.34	11.52	11.63	11.69	11.74	11.78
Model 4	2.98	8.59	9.90	10.50	10.79	10.98	11.10	11.18	11.23	11.26
AR(1)	16.46									
AR(4)	13.91									
Mean model	-									
					GDP growth rate					
Model 1	5.38	10.74	11.75	12.14	12.36	12.51	12.62	12.69	12.74	12.78
Model 2	3.29	9.41	10.53	11.13	11.46	11.68	11.81	11.89	11.95	11.99
Model 3	**2.67**	9.98	11.13	11.58	11.81	11.95	12.04	12.10	12.14	12.17
Model 4	3.09	8.99	10.10	10.66	10.95	11.15	11.26	11.33	11.38	11.41
AR(1)	14.70									
AR(4)	13.76									
Mean model	16.07									
					Inflation					
Model 1	2.07	3.55	4.44	4.73	4.86	4.92	4.96	4.99	5.00	5.01
Model 2	1.31	3.79	4.49	4.77	4.91	4.99	5.04	5.07	5.09	5.10
Model 3	1.07	3.30	4.18	4.44	4.54	4.60	4.63	4.65	4.66	4.67
Model 4	**1.00**	3.58	4.14	4.37	4.50	4.58	4.62	4.66	4.68	4.69
AR(1)	5.72									
AR(4)	4.98									
Mean model	13.53									

All the entries in this table have been multiplied by 1000. The bandwidth for factor estimation is $h_f = 0.011$. For log(GDP), the mean model is not considered as the series is nonstationary. Highlighted values correspond to the least MSE value for the response variable.

Assessing the out-of-sample performance

In this section, we consider the expanding window one-step ahead out-of-sample forecasting. The out-of-sample forecasting performance is evaluated in terms of R^2_{os} defined in (3.34). For both stationary and nonstationary response variables, we consider the basic AR(1) model as the benchmark to calculate R^2_{os}. Here, we consider the same ten bandwidth values for h_p. Unlike in the in-sample evaluation, we considered different bandwidths for factor estimation for the three response variables. In expanding window forecast, 1959:Q1 - 1998:Q4 is considered as the first in-sample period and forecast one-step ahead for 1999:Q1 - 2018:Q4 time period. For h_f, we chose the values 0.011, 0.024, and 0.04 for log(GDP), GDP growth rate, and inflation, respectively. A justification for these bandwidth choices is given in the Appendix under sensitivity analysis.

Table 3.11 provides the calculated R^2_{os} compared to basic AR(1) model for all model specifications. According to the results in the table, except for first two h_p, all four FAR models have better performance compared to the AR models for forecasting log(GDP). For forecasting GDP growth rate, the two FAR models with nonparametrically estimated factors have negative R^2_{os} for all the bandwidths, and the two FAR models with PCA factors have positive R^2_{os} for $h_p \geqslant 0.0234$. This implies, for forecasting GDP growth rate, that the semi-parametric FAR models with PCA factors have improved forecast performance compared to the AR(1) model. Model 4, the FAR model augmented with PCA factors, has better performance in out-of-sample forecasting for log(GDP) and GDP growth rate when $h_p = 0.0379$. For inflation forecasting, the two FAR models with AR(4) terms have negative values for R^2_{os} for all the bandwidth options. For forecasting inflation, Model 1, FAR model augmented with nonparametrically estimated factors, has better forecast performance compared to the PCA analogue with every bandwidth, except for $h_p = 0.017$. Model 1 with $h_p = 0.668$ has the highest calculated R^2_{os} for forecasting inflation, hence, this model has the best out-of-sample forecast performance for in forecasting inflation. Overall, with some specific bandwidths, the FAR models outperform the basic AR models and the mean model in out-of-sample forecasting for both stationary and nonstationary response variables.

The plots in Figures 3.3 - 3.5 show the one-step ahead out-of-sample (expanding window) forecasting for log(GDP), GDP growth rate and inflation, for the period 1999:Q1 - 2018:Q4 with the initial estimation period being 1959:Q1 - 1998:Q4. Figure 3.3 indicates that the prediction series of log(GDP) using the four FAR models are very close to each other and they appear to be close to the original time series. The predicted series of GDP growth rate and inflation by four FAR models are comparably close to each other and seem to pick the fluctuations of the original series.

Table 3.11: Out-of-sample performance of the models relative to AR(1), in terms of R^2_{os} for expanding window forecast

h_p	0.017	0.089	0.162	0.234	0.306	0.379	0.451	0.524	0.596	0.668
					log(GDP)					
Model 1	-1.22	-0.05	0.01	0.03	0.05	0.03	0.01	-0.01	-0.01	-0.02
Model 2	-4.69	0.01	0.20	0.24	0.26	0.25	0.25	0.24	0.23	0.23
Model 3	-0.37	-0.06	0.02	0.06	0.08	0.07	0.06	0.05	0.05	0.046
Model 4	-4.11	-0.03	0.21	0.27	0.29	**0.29**	**0.29**	0.29	0.29	0.29
AR(4)	-1.05									
Mean	–									
					GDP growth rate					
Model 1	-15.23	-0.31	-0.24	-0.20	-0.14	-0.11	-0.09	-0.08	-0.08	-0.07
Model 2	-7.19	-0.26	-0.01	0.04	0.06	0.06	0.06	0.06	0.06	0.05
Model 3	-8.71	-0.27	-0.23	-0.20	-0.13	-0.10	-0.08	-0.07	-0.06	-0.05
Model 4	-7.04	-0.29	-0.02	0.05	0.07	**0.07**	0.07	0.06	0.05	0.05
AR(4)	-0.02									
Mean	-0.26									
					Inflation					
Model 1	-80.83	-0.23	0.03	0.05	0.06	0.07	0.07	0.07	0.07	**0.07**
Model 2	-5.05	-0.47	-0.03	0.02	0.02	0.03	0.03	0.03	0.03	0.04
Model 3	-98.45	-0.31	-0.10	-0.06	-0.03	-0.02	-0.01	-0.01	-0.01	-0.01
Model 4	-3.38	-0.80	-0.20	-0.12	-0.10	-0.10	-0.10	-0.11	-0.11	-0.11
AR(4)	-0.22									
Mean	-0.29									

The h_f for the three response variables log(GDP), GDP growth rate, and inflation are 0.011,0.024, and 0.040 respectively. The first in-sample size period is 1959:Q1-1998:Q4 and the out-of-sample period is 1999:Q1-2018:Q4. Mean model results for log(GDP) are not included as the series is non-stationary. The highlighted values correspond to the highest R^2_{os}.

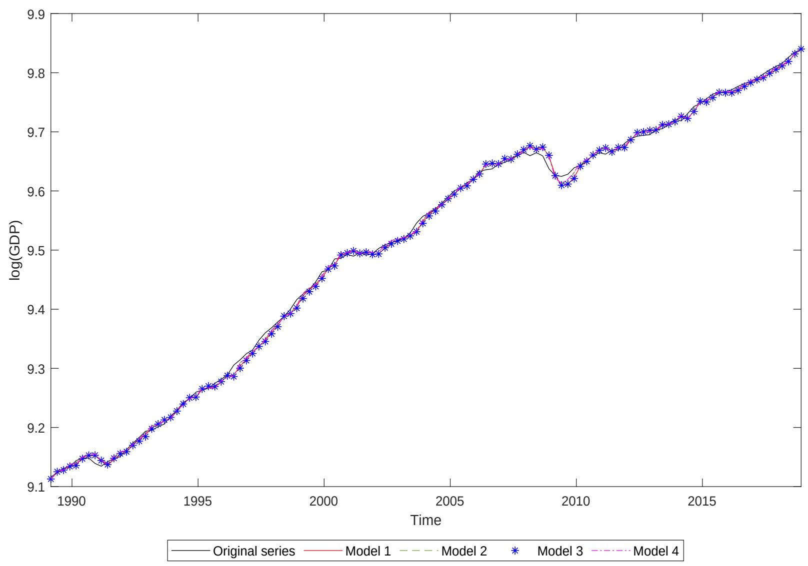


Figure 3.3: The observed log(GDP) and one-step ahead out-of-sample expanding window forecast for log(GDP) using semi-parametric FAR models

The black solid line: observed time series, red solid line: prediction series from Model 1, green dash line: prediction series from Model 2, blue *:prediction series from Model 3, magenta -.: prediction series from Model 4. The bandwidth for the factor estimation is 0.011 and bandwidth for model estimation is 0.379.

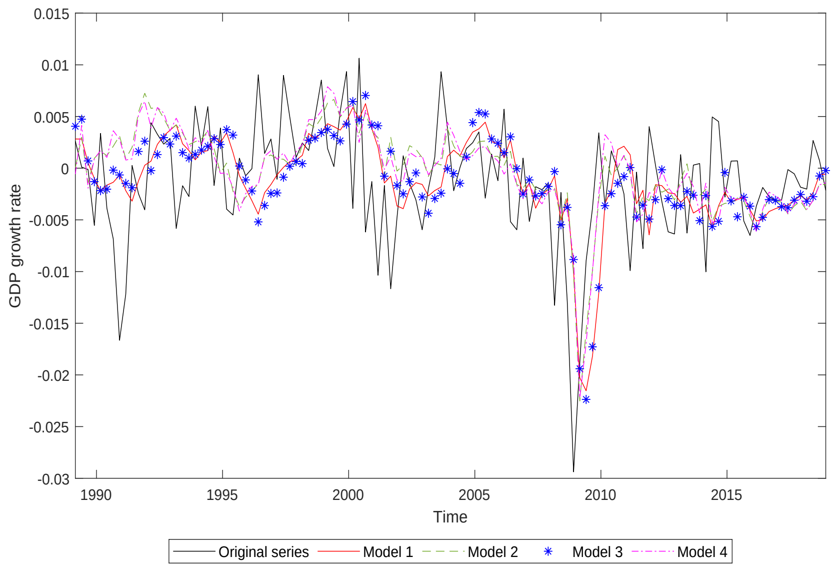


Figure 3.4: The observed GDP growth rate and one-step ahead out-of-sample expanding window forecast for GDP growth rate using semi-parametric FAR models

The black solid line: observed time series, red solid line: prediction series from Model 1, green dash line: prediction series from Model 2, blue *:prediction series from Model 3, magenta -.: prediction series from Model 4. The bandwidth for the factor estimation is 0.024 and bandwidth for model estimation is 0.379

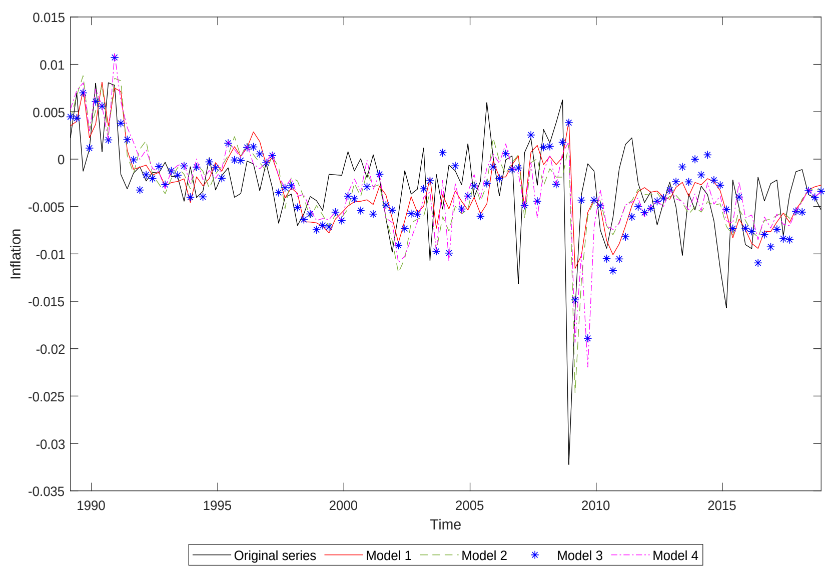


Figure 3.5: The observed inflation series and one-step ahead out-of-sample expanding window forecast for inflation

The black solid line: observed time series, red solid line: prediction series from Model 1, green dash line: prediction series from Model 2, blue *:prediction series from Model 3, magenta -.: prediction series from Model 4. The bandwidth for the factor estimation is 0.040 and bandwidth for model estimation is 0.071.

Remark 3: In this empirical example, we observe that the best model in terms of in-sample performance is not the same as that in terms of out-of-sample performance. This is consistent with what we often observe in practice, mainly because of over-fitting, but it may also be because the chosen model is not appropriate. To address some of these issues, it would help to use various specification tests, a topic outside the scope of this paper, and hardly any methods are available at this stage. Nevertheless, it is important to know how to specify the best model in practice to avoid over-fitting. One possible way is to adopt the technique in this empirical study, by evaluating the out-of-sample forecast performance over a period of time. If it performs well in out-of-sample forecasting for some adequate number of times in the recent past, then that would be an indication of the reliability of the method. Whether or not the model over-fits is not that important for forecasting if it forecasts reliably. If the objective is an inference about individual coefficients, then the issues are different and the role of out-of-sample forecast performance would be different. However, as our model is not fully nonparametric, only some of the coefficients are time-varying, the model is still linear in variables. Hence, the over-fitting issue may not be serious as the fully nonparametric model.

3.5 Conclusion

This chapter proposed a new approach to forecast a univariate time series using a semi-parametric model combined with a factor model for a large set of panel data. We assumed that the large number of variables in the panel data are also potential predictors. A factor model is used for reducing the dimension of the panel data and thus to reduce variability in the estimates and forecasts. Since there are a large number of potential predictors in the panel data, it is very likely that these variables, and hence the factors extracted from them, are a mixture of stationary and nonstationary variables. Further, we also assumed that the panel data spans a long time period and hence econometric models in this setting are likely to contain parameters that change over time to represent structural changes. The methodology developed in this chapter is specifically designed to accommodate the aforementioned type of mixture of stationary and nonstationary predictors and time-varying parameters. The simulations studies show that the proposed method makes significant improvements over its competitors. The empirical example, on forecasting log(GDP), GDP growth rate, and Inflation illustrates the relevance and contribution of the proposed method for forecasting a macroeconomic variable. In summary this makes a significant contribution by proposing a a new method for forecasting a macroeconomic variable. The numerical studies, albeit limited by necessity, demonstrates that the proposed method makes a significant contribution to advance econometric methodology.

3.6 Appendices

3.6.1 Appendix A: Comparison of methods for estimating $\alpha(\tau_t)$

In this appendix, we present the results of our evaluations to compare Hermite polynomial, trigonometric polynomial, and kernel methods for estimating the time-varying parameter $\alpha(\tau_t)$ in the FAR model. We also compare the parametric method with the aforementioned three methods.

Estimation method

First, let us recall the semi-parametric FAR model,

$$Y_{t+1} = \alpha(\tau_t)'\tilde{F}_t + \beta' V_t + \eta_{t+1}, \qquad (\tau_t = t/T; t = 1, \ldots, T), \tag{A2.1}$$

where $\tilde{F}_t$ is the r-dimensional estimated factor and V_t is the m-dimensional observable regressor. We estimate the time-varying parameter, $\alpha(\tau_t)$, and constant parameter β by the following 3-step algorithm:

Step 1: First, approximate $\alpha_j(\tau_t) \quad (j = 1, \ldots, r)$ by the sum of the first few terms in a representation of $\alpha_j(\tau_t)$ using an orthonormal basis of the function space. We may express such an approximation by

$$\alpha_j(\tau_t) \approx \sum_{i=1}^{k} Z_i(\tau_t) c_{ij},$$

where $\{Z_i(.); 1 \leqslant i \leqslant k\}$ are the first k elements of the basis, $c_{ij} \quad (i = 1, \ldots, k; j = 1, \ldots, r$ are unknown coefficients. Let

$$C = \begin{pmatrix} c_{1,1} & \cdots & c_{1,r} \\ \vdots & \ddots & \vdots \\ c_{k,1} & \cdots & c_{k,r} \end{pmatrix};$$

k is usually called the *truncation parameter*. Then, we obtain,

$$Y_{t+1} = \left(\tilde{F}_t' \otimes Z_k(\tau_t)'\right) vec(C) + \beta' V_t + \eta_{t+1}; \tag{A2.2}$$

where we have used η_{t+1} as a generic error term; strictly speaking, the η_{t+1} in (A2.2) and that in (A2.1) are not equal.

Step 2: Let $\delta = \begin{pmatrix} (vec(C))' & \beta' \end{pmatrix}'$ and $\tilde{W}_t = \begin{pmatrix} \tilde{F}_t' \otimes Z_k(\tau_t)' & V_t' \end{pmatrix}'$ for $(t = 1, \ldots, T)$. Then, we have

$$Y_{t+1} = \tilde{W}_t'\delta + \eta_{t+1}. \tag{A2.3}$$

Let $\hat{\delta} = \begin{pmatrix} [vec(\hat{C})]' & \hat{\beta}' \end{pmatrix}'$ denote the OLS estimator of δ; therefore,

$$\hat{\delta} = \left(\sum_{t=1}^{T-1} \tilde{W}_t\tilde{W}_t'\right)^{-1} \left(\sum_{t=1}^{T-1} \tilde{W}_t Y_{t+1}\right).$$

Step 3: Estimate $\alpha_j(\tau_t)$ by $\hat{\alpha}_j(\tau_t) = \sum_{i=1}^{k} Z_i(\tau_t)\hat{c}_{ij}$, $(j = 1, \ldots, r)$, where $\hat{C} = (\hat{c}_{ij})$.

Model estimation using Hermite polynomials

Let $L^2[\mathbb{R}, exp(-\omega^2/2)]$ denote the Hilbert space of continuous functions on $\mathbb{R}$ with the inner product defined by $< g_1, g_2 >= \int g_1(\omega)g_2(\omega)\exp(-\omega^2/2)\, d\omega$. Let $H_i(\cdot)$ $(i = 0, 1, \ldots)$ denote the sequence of Hermite polynomials defined by

$$H_i(\omega) = (-1)^i exp(\omega^2/2)\frac{d^i}{d\omega^i} exp(-\omega^2/2) \quad (i = 0, 1, \ldots,).$$

Then, the first few Hermite polynomials are $H_0(\omega) = 1$, $H_1(\omega) = \omega$, $H_2(\omega) = \omega^2 - 1$, and $H_3(\omega) = \omega^3 - 3\omega$. These polynomials are orthogonal in the sense $\int H_i(\omega)H_j(\omega)exp(-\omega^2/2)d\omega = \sqrt{\pi}2^i i!\delta_{ij}$, where δ_{ij} is the Kronecker delta. Let $\mathscr{H}_i(\omega) = \left(\sqrt{2\pi} i!\right)^{-1/2} H_i(\omega)$. Then, $\{\mathscr{H}_i(\omega), i = 0, 1, \ldots\}$ is an orthonormal basis of the Hilbert space $L^2(\mathbb{R}, exp(-\omega^2/2))$.

Let $g(\omega) \in L^2[\mathbb{R}, exp(-\omega^2/2)]$. Then the orthogonal series expansion for the unknown function is

$$g(\omega) = \sum_{i=0}^{\infty} c_i \mathscr{H}_i(\omega) = g_k(\omega) + \gamma_k(\omega).$$

where

$$c_i = \langle g(\omega), \mathscr{H}_i(\omega)\rangle, \qquad g_k(\omega) = \sum_{i=0}^{k-1} c_i\mathscr{H}_i(\omega), \qquad \gamma_k(\omega) = \sum_{i=k}^{\infty} c_i\mathscr{H}_i(\omega).$$

Therefore,

$$g(\omega) = Z_k(\omega)'c + \gamma_k(\omega), \tag{A2.4}$$

where $c_{[k\times 1]} = (c_0, c_1, \ldots, c_{k-1})'$ and $Z_k(\omega) = (\mathscr{H}_0(\omega), \mathscr{H}_1(\omega), \ldots, \mathscr{H}_{k-1}(\omega))'$.

Now, using (A2.4), the semi-parametric FAR model (A2.1) can be expressed as,

$$Y_{t+1} = \left(\tilde{F}_t' \otimes Z_k(\tau_t)'\right) vec(C) + \beta' V_t + \eta_{t+1},$$

where $C_{[k\times r]} = \{c_{ij}; i = 1, \ldots, k, j = 1, \ldots, r\}$. Let $(\hat{C}, \hat{\beta})$ denote the OLS estimate of (C, β); also let $\hat{C} = (\hat{c}_{ij})$. Then, we have $\hat{\alpha}(\tau_t) = \sum_{i=1}^{k} Z_i(\tau_t)\hat{c}_k$, where $\hat{c}_k = (\hat{c}_{k1}, \ldots, \hat{c}_{kr})'$.

Model estimation using trigonometric polynomials

In trigonometric polynomial estimation method we approximate the $\alpha(\tau_t)$ by the a trigonometric series; for example $Z_k = (1, cos(v), sin(v); cos(2v), sin(2v); \ldots, cos(kv), sin(kv))'$ with $v = 2\pi\tau_t$. Then, again for $\tilde{W}_t = \begin{pmatrix} \tilde{F}_t' \otimes Z_k(\tau_t)' & V_t' \end{pmatrix}$ and $\delta = \begin{pmatrix} (vec(C))' & \beta' \end{pmatrix}'$, we have

$$Y_{t+1} = \tilde{W}_t'\delta + \eta_{t+1}.$$

by which we can estimate $(\hat{\alpha}(\tau_t), \hat{\beta})$.

Choice of truncation parameter

In practice, the truncation parameter is an unknown integer, and estimated using the dataset. In the literature on non-parametric and semi-parametric model estimation using orthogonal polynomials, it has been discussed that the truncation parameter $k \sim T^{1/c}$ where $c \in [1/a, 1/4)$ and $a \geqslant 7$ (see Gao et al. 2002, Gao 2007, Hall et al. 2007, Dong et al. 2015, Dong et al. 2019, Zhou et al. 2020, and Dong et al. 2021) The following two criteria have been proposed for obtaining an optimal truncation parameter, $\hat{k}$, for stationary time series models, which can also be used in the panel data models.

1)GCV method:

Gao et al. [2002] proposed the generalized cross-validation method to obtain the estimator $\hat{k}$:

$$\hat{k}_{GCV} = \arg\min_k \left(1 - \frac{k}{T}\right)^{-2} \frac{1}{T} \sum_{t=1}^{T} \left(Y_{t+1} - \hat{Y}_{t+1,k}\right)^2. \tag{A2.5}$$

2) AIC method:

Cai [2007] proposed the following estimator based on the Akaike information criterion (AIC):

$$\hat{k}_{AIC} = \arg\min_k \log\left(\frac{1}{T}\left(Y_{t+1} - \hat{Y}_{t+1,k}\right)^2\right) + 2\frac{T_{\lambda,k} + 1}{T - T_{\lambda,k} - 2}, \tag{A2.6}$$

where $\hat{Y}_{t+1,k}$ is the one-step ahead forecast using the truncation parameter k and $T_{\lambda,k}$ is the trace of $\tilde{W}_{t,k}\left(\tilde{W}'_{t,k}\tilde{W}_{t,k}\right)^{-1}\tilde{W}_{t,k}$ and $\tilde{W}_{t,k} = \tilde{W}_t$ is as in (A2.3).

Rigorous methodology has not yet been developed for the optimal choice of the truncation parameter in semi-parametric models with both stationary and nonstationary processes. Since our main focus is to obtain better forecasting performances, we shall choose the truncation parameter for the model through in-sample and out-of-sample forecasting performance in terms of mean squared error (Dong et al. [2018]).

Simulation study with different methods for estimating $\alpha(\tau_t)$

This section reports the results of preliminary simulation studies to different methods for estimating $\alpha(\tau_t)$.

Design of the Simulation

We considered the following DGP:

$$Y_{t+1} = \alpha_{1t}F_{1t} + \alpha_{2t}F_{2t} + \beta Y_t + \eta_{t+1}, \quad \eta_t \sim N(0,1), \qquad (t = 1, \ldots, T-1), \tag{A2.7}$$

$$F_{1t} = F_{1,t-1} + v_t, \qquad (v_t, F_{2t}) \sim MVN(0,\rho), \qquad \rho = (1,0|0,1), \tag{A2.8}$$

$$X_{it} = \lambda_i^{(1)}F_{1t} + \lambda_i^{(2)}F_{2t} + e_{it}, \quad \lambda_i \sim N(0,1), \quad e_{it} \sim N(0,1), \quad (t = 1, \ldots, T; i = 1, \ldots, N). \tag{A2.9}$$

The following two DGP s were considered for the time varying parameters (α_{1t} and α_{2t}):

DGP1: $\alpha_{1t} = (1/2)(1+\tau_t), \alpha_{2t} = \tau_t + \tau_t^2$, and $\beta = 0.7$.

DGP2: $\alpha_{1t} = (1/3)sin(2\pi\tau_t), \alpha_{2t} = cos(2\pi\tau_t)$, and $\beta = 0.7$, $(\tau_t = t/T; t = 1, \ldots, T)$.

We considered the following seven different values for the truncation constant k in the polynomial approximations for the time varying parameters: $k = \lceil cT^{1/7} \rceil$, $c = [1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.4]$; see Dong et al. [2015] for a discussion on the rationale for these choices; these values of k lie in the interval $[4, 10]$. For a given value of the truncation parameter k, we chose the corresponding bandwidth h as $1/k$ for the kernel estimation method. Since the FAR model contains a lag term of the response variable as a regressor, we included a burn-in period of 100 time units. Hence, the first 100 observations of each data sample were discarded. We chose the length of time period $T = 300$, and the number of variables in the panel data $N = 100$. The estimated in-sample sum of squared error (SSE) and out-of-sample mean square forecast errors (MSFE) are reported in Tables 3.12 and 3.13. Table 3.12 shows that the Hermite polynomial estimation has better in-sample performance (except when $k = 9$). Overall, the in-sample performances of the three estimation methods for the semi-parametric model are better than that of the parametric method. The out-of-sample performance reported in Table 3.13 shows that the semi-parametric model estimated by the kernel method is not very sensitive to the choice of the bandwidth compared to the choice of the truncation parameter in the polynomial estimation methods. Further, the forecasting performance of the former is also is better compared to other two estimation methods. Overall, kernel estimation method performed better in both in-sample and out-of sample forecasting.

Table 3.12: In-sample sum of squares of error[SSE] for the estimation methods when the factors are estimated

k	DGP1				DGP2			
	Hermite	Kernel	Trig.	Para.	Hermite	Kernel	Trig.	Para.
4	0.99	1.28	1.13	1.93	1.03	1.43	1.04	1.82
5	0.98	1.18	1.09		0.97	1.35	1.01	
6	0.97	1.12	1.05		0.96	1.29	0.98	
7	0.97	1.08	1.02		0.95	1.24	0.96	
8	0.96	1.05	0.99		0.94	1.21	0.94	
9	1.42	1.02	0.97		1.01	1.14	0.93	
10	1.01	1.01	0.95		1.01	1.12	0.91	

Notes: 1)The values under Trig. and Para. columns correspond to the results from trigonometric estimation and parametric estimation respectively. 2) Bandwidth for kernel estimation is $h = 1/k$.

Table 3.13: Out-of-sample mean square of forecast error[MSFE] for the estimation methods when factors are estimated

k	DGP1				DGP2			
	Hermite	Kernel	Trig.	Para.	Hermite	Kernel	Trig.	Para.
4	1.10	1.27	3.23	2.33	1.09	1.40	2.01	2.07
5	1.11	1.15	3.17		1.12	1.30	2.18	
6	1.19	1.09	3.13		1.15	1.22	2.51	
7	1.44	1.06	3.29		1.40	1.16	2.71	
8	1.67	1.05	3.53		1.64	1.12	2.95	
9	2.19	1.03	3.74		3.22	1.09	3.02	
10	5.58	1.02	3.88		4.21	1.06	3.00	

Notes: 1)Full sample size is $T = 300$, and $T1 = 250$, where $T1$ is the end of first in-sample period. 2) One-step ahead out-of-sample forecasting was conducted with expanding window. 3) For the kernel estimation method, bandwidth $h = 1/k$. 4) The values under Trig. and Para. columns correspond to the results from trigonometric estimation and parametric estimation respectively.

Empirical Application

We use the setting of the empirical example in the empirical study section, Section 3.4, of this Chapter to evaluate the three estimation methods. In this section we restrict to the FAR model with the unknown factors replaced by the PCA factors, more specifically

$$Y_{t+1} = \alpha_t' \tilde{F}_t + \beta Y_t + \eta_{t+1}, \tag{A2.10}$$

where $\tilde{F}_t$ is the factor estimated by PCA. For this particular model, we estimated the time varying parameter α_t by three different methods: Hermite polynomial, trigonometric polynomial, and the kernel method. We chose the truncation parameter k in the range $[1, 10]$ for Hermite and trigonometric polynomial estimations; for each value of the truncation parameter k, the corresponding bandwidth for the kernel method was chosen as $h = 1/k$ Table 3.14 provides the estimated sum of squares of error[SSE] for the three estimation methods. The results in this table show that the methods based on Hermite polynomial, trigonometric polynomial, and the kernel method have approximately the same in-sample performance in terms of SSE.

Table 3.14: Estimated SSE (in-sample) using three different estimation methods

log(GDP)					GDP growth rate					Inflation				
k	Hermite	Kernel	Trig.	Para	k	Hermite	Kernel	Trig.	Para	k	Hermite	Kernel	Trig.	Para
1	1.86	1.84	1.89	1.86	1	1.77	1.75	1.72	1.77	1	2.48	2.47	1.90	2.48
2	1.76	1.80	1.59		2	1.75	1.71	1.50		2	2.21	2.42	1.87	
3	1.79	1.76	1.45		3	1.66	1.67	1.32		3	1.99	2.34	1.70	
4	1.80	1.74	1.41		4	1.63	1.65	1.21		4	1.84	2.26	1.34	
5	1.86	1.74	1.27		5	1.76	1.65	1.16		5	1.83	2.19	1.09	
6	1.55	1.75	0.93		6	1.44	1.64	0.90		6	1.75	2.14	0.91	
7	1.41	1.74	0.76		7	1.32	1.62	0.73		7	1.51	2.09	0.69	
8	1.30	1.73	0.58		8	1.26	1.60	0.60		8	1.49	2.04	0.55	
9	1.26	1.72	0.41		9	1.22	1.58	0.46		9	1.39	2.00	0.41	
10	6.67	1.72	0.30		10	4.80	1.57	0.30		10	1.27	1.96	0.31	

Notes: 1) The SSFE values have been multiplied by 1000. 2) The figures under Trig. column correspond to the results for trigonometric estimation method. 3) The figures under para. column correspond to the parametric FAR estimation.

Evaluation in terms of out-of-sample forecast performance

We evaluate the one-step ahead out of sample forecast performance of the methods for four different initial in-sample periods. The four initial in-sample periods are defined in Table 3.15.

The symbol T_2 in Table 3.15 indicates that it refers to the initial estimation and the corresponding forecast periods being 1959:Q1-2003:Q4 and 2004:Q1-2018:Q4, respectively. The

Table 3.15: Initial in-sample and out-of-sample periods

First estimation period (T_i)		Out-of-sample period
T_1	1959:Q1-1998:Q4	1999:Q1-2018:Q4
T_2	1959:Q1-2003:Q4	2004:Q1-2018:Q4
T_3	1959:Q1-2008:Q4	2009:Q1-2018:Q4
T_4	1959:Q1-2013:Q4	2014:Q1-2018:Q4

estimated out-of-sample SSFE for the estimation methods are reported in the Table 3.16. In contrast to the results based on in-sample SSE, the out-of-sample SSFE for both Hermite and trigonometric polynomials are comparably higher than that for kernel estimation method. Furthermore, the out-of-sample SSFE for the two polynomial estimations are sensitive to the choice of the truncation parameters. Although the foregoing evaluations considered only a small number of scenarios, the results are presented are sufficient for us to prefer the kernel method to the two polynomial methods.

3.6.2 Appendix B: Additional results for bandwidth selection

In this appendix, we report the results on bandwidth selection for factor and parameter estimations. These results are based on a simulation study and an empirical application.

Additional results from the simulation study

Simulation results when factors are known (observed)

The estimated Bias, RMSE, and standard deviation[SD] for $\hat{\beta}$ and for the local estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are reported in Tables 3.17-3.20. Table 3.17 provides the results for ten bandwidths when the factors are known and are VMA(2). Both tables are for $T = 100$. Table 3.17 for VMA(2) factors show that for most cases Bias of every estimate strictly increases with increasing bandwidth. The RMSE of each estimate has a single minimum as a function of bandwidth. For each row of RMSE in Table 3.17, the local minimum occurs for bandwidth between 0.483 and 1.133 (between 3rd and 6th columns).

Table 3.18 reports the corresponding results for the case when factors follow VAR(2) processes. The main observations for this case are very similar to those for VMA(2) in the previous paragraph. Except for $\hat{\alpha}_2$ in DGP3, each row of RMSE, as a function of bandwidth, has one minimum. Therefore, the optimal bandwidth, with respect to RMSE, lies in the range of bandwidths

Table 3.16: One-step ahead out-of-sample forecasting performance of three model estimation methods, in terms of SSFE

k	Hermite				Kernel				Trigonometric			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
	log(GDP)											
1	2.80	2.31	0.51	0.06	2.74	2.23	0.51	0.06	3.27	2.09	0.93	0.19
2	3.38	2.52	0.96	0.44	2.67	2.07	0.53	0.07	3.50	2.56	1.22	0.15
3	4.37	3.29	0.99	0.20	2.64	1.97	0.55	0.09	5.69	4.59	1.97	0.34
4	7.32	4.33	1.60	0.37	2.69	1.99	0.56	0.09	6.64	5.53	2.50	0.83
5	10.58	7.60	4.00	1.46	2.76	2.02	0.58	0.09	7.65	6.34	2.98	0.81
6	15.28	12.26	3.50	0.55	2.83	2.05	0.61	0.08	10.58	7.76	3.87	0.81
7	23.85	18.52	8.16	1.26	2.90	2.09	0.64	0.07	28.74	12.06	6.58	1.68
8	53.87	32.69	20.25	2.71	2.99	2.15	0.67	0.07	33.70	19.81	6.09	2.07
	GDP growth rate											
1	2.53	1.83	0.53	0.06	2.49	1.75	0.52	0.05	3.41	2.25	1.34	0.23
2	3.35	2.42	0.97	0.45	2.45	1.62	0.53	0.05	4.01	3.12	1.30	0.17
3	3.93	2.80	0.85	0.11	2.47	1.57	0.54	0.07	5.33	4.39	1.75	0.43
4	8.45	4.68	1.88	0.26	2.51	1.61	0.54	0.07	5.98	4.96	2.24	0.71
5	10.97	8.37	3.73	1.54	2.54	1.66	0.56	0.08	7.99	5.05	2.14	0.41
6	12.99	9.81	3.62	0.78	2.58	1.71	0.59	0.08	14.68	11.10	4.80	0.35
7	21.68	16.94	9.45	1.37	2.64	1.77	0.61	0.08	28.74	14.50	8.49	2.77
8	1835	540	19.40	1.86	2.74	1.84	0.64	0.08	24.03	17.98	5.69	2.54
	Inflation											
1	2.86	2.51	0.47	0.28	2.87	2.51	0.47	0.27	2.83	2.56	0.71	0.33
2	3.40	2.88	0.82	0.36	2.90	2.52	0.46	0.27	3.75	3.37	0.88	0.49
3	4.11	3.70	1.22	0.52	2.90	2.53	0.45	0.26	4.76	4.40	1.16	0.52
4	5.96	5.42	1.40	0.42	2.92	2.57	0.46	0.26	5.99	5.38	1.61	0.55
5	9.54	8.58	2.31	0.61	2.95	2.63	0.48	0.26	6.61	5.90	1.98	0.44
6	13.63	12.38	5.20	1.90	3.04	2.73	0.51	0.27	9.43	7.83	3.00	0.53
7	20.58	19.08	9.04	1.69	3.18	2.88	0.55	0.28	11.30	9.56	3.87	1.10
8	33.28	29.08	19.80	4.55	3.37	3.06	0.59	0.29	14.42	11.26	4.43	1.61

Notes: 1) The SSFE entries in the table have been multiplied by 1000. 2) One-step ahead out-of-sample forecasting is conducted with expanding window. 3) For the kernel estimation method, bandwidth $h = 1/k$.

chosen for this simulation study; it appears that a value between 0.483 and 1.133 (between 3rd and 6th columns) is suitable.

Simulation with estimated factors

Table 3.19 provides the Bias, RMSE, and standard deviation[SD]of the estimates when the FAR model uses estimated factors, instead of known factors. The results in these simulations are for $T = 100$ and $N = 100$. Table 3.19 shows that, for DGP1 and DGP2, the Bias and SD of the estimates strictly decrease as functions of bandwidth. For every DGP, RMSE has a minimum between the two extremes of the bandwidth in the table; with respect to RMSE, an optimal choice of the bandwidth could be a value between 0.05 and 2.0. In Section 3.4, the Empirical

Application section of the chapter, we used cross validation as the criterion for choosing the bandwidth.

Table 3.20 provide the bias, RMSE and standard deviation of the non-parametrically and PCA estimated factors for nine combinations of $[T, N]$. Results in the columns under the NP factors are corresponding to the non-parametrically estimated factors, and the results under the PCA column are for the estimated factors and the loadings using PCA.

Table 3.17: Estimated Bias, RMSE, and standard deviation[SD] of estimates for the case when factors are known and are VMA(2).

	d	0.050	0.267	0.483	0.700	0.917	1.133	1.350	1.567	1.783	2.000
						DGP1					
$\hat{\alpha}_1$	Bias	4.48	-32.13	-87.58	-143	-174	-190	-199	-204	-208	-210
	RMSE	6596	471	315	257	263	277	289	298	304	309
	SD	6592	458	293	210	194	192	192	194	195	195
$\hat{\alpha}_2$	Bias	-14.80	13.63	47.83	82.95	100	111	117	122	125	127
	RMSE	6554	456	344	373	439	490	525	550	567	580
	SD	6550	443	284	207	192	190	191	192	194	195
$\hat{\beta}$	Bias	-5.40	-1.27	-2.81	-9.50	-18.35	-26.14	-32.11	-36.50	-39.73	-42.13
	RMSE	120	103	102	103	107	113	119	125	129	132
	SD	120	103	102	102	105	110	115	119	123	125
						DGP2					
$\hat{\alpha}_1$	Bias	-2.79	0.20	3.23	2.42	4.00	6.72	9.18	11.08	12.51	13.58
	RMSE	754	283	221	196	192	197	203	208	212	215
	SD	753	282	217	189	178	174	173	173	173	174
$\hat{\alpha}_2$	Bias	-5.66	-11.27	-19.87	-23.66	-25.80	-27.08	-27.77	-28.14	-28.35	-28.46
	RMSE	586	242	242	303	375	434	475	505	526	541
	SD	585	231	182	161	154	152	152	153	154	155
$\hat{\beta}$	Bias	-2.19	-2.73	-3.44	-1.36	1.84	4.61	6.65	8.10	9.13	9.88
	RMSE	120	103	101	101	105	111	116	121	125	128
	SD	120	103	101	101	105	110	116	121	125	128
						DGP3					
$\hat{\alpha}_1$	Bias	18.96	-53.71	-54.55	-66.89	-89.15	-100	-105	-107	-108	-108
	RMSE	6665	566	365	286	281	290	299	305	310	314
	SD	6637	519	356	251	218	208	205	203	203	202
$\hat{\alpha}_2$	Bias	-28.06	23.27	-19.87	-30.10	-4.19	20.75	39.46	52.96	62.77	70.01
	RMSE	6622	550	557	657	703	722	732	737	740	741
	SD	6595	507	362	268	237	227	224	222	221	221
$\hat{\beta}$	Bias	-5.31	-1.81	-0.99	2.21	4.43	5.84	6.84	7.58	8.12	8.52
	RMSE	120	103	108	121	131	136	139	140	141	142
	SD	120	103	108	121	131	136	139	140	141	141
						DGP4					
$\hat{\alpha}_1$	Bias	-6.44	-28.59	-45.11	-54.62	-61.36	-64.80	-66.34	-67.04	-67.37	-67.53
	RMSE	755	305	271	273	280	286	290	293	296	297
	SD	755	293	237	211	195	188	184	182	181	181
$\hat{\alpha}_2$	Bias	-2.25	-6.35	-41.72	-62.07	-55.81	-40.81	-26.39	-14.73	-5.73	1.18
	RMSE	587	299	463	604	671	699	712	718	721	723
	SD	586	244	206	188	177	172	170	169	168	168
$\hat{\beta}$	Bias	-2.67	-4.03	-11.04	-16.97	-20.23	-22.05	-23.17	-23.88	-24.37	-24.71
	RMSE	120	104	106	119	130	135	138	139	140	140
	SD	120	103	106	118	128	133	136	137	138	138

Notes: (1) The estimates in the table have been multiplied by 1000. (2) $\hat{\beta}$ is the profile least squares estimate; $\{\hat{\alpha}_1, \hat{\alpha}_2\}$ are kernel based estimates of the time-varying coefficients of the factors. 3) Bandwidth for parameter estimation is $h_p = dT^{-1/5}; T = 100$.

Table 3.18: Estimated bias, RMSE, and standard deviation for the estimated coefficients with 10 bandwidths when factors are known and have VAR(2) structure, $T = 100$

	d	0.050	0.267	0.483	0.700	0.917	1.133	1.350	1.567	1.783	2.000
						DGP1					
$\hat{\alpha}_1$	Bias	-4.60	2.69	5.84	7.63	8.71	9.39	9.83	10.13	10.34	10.49
	RMSE	363	140	112	109	117	128	138	145	151	155
	std	363	140	110	99.29	94.89	93.51	93.35	93.59	93.93	94.26
$\hat{\alpha}_2$	Bias	-1.72	-8.80	-14.72	-18.39	-19.46	-19.32	-18.85	-18.36	-17.95	-17.61
	RMSE	293	137	188	274	358	422	467	498	520	536
	std	293	120	111	116	122	128	133	136	138	140
$\hat{\beta}$	Bias	-2.05	1.46	2.30	3.28	4.66	6.03	7.15	8.02	8.67	9.16
	RMSE	126	102	101	103	109	116	123	130	135	139
	std	126	102	101	103	108	116	123	129	135	139
						DGP2					
$\hat{\alpha}_1$	Bias	-3.69	-1.99	0.01	0.20	-0.01	-0.27	-0.48	-0.66	-0.79	-0.90
	RMSE	413	156	125	122	130	141	150	157	162	166
	std	413	156	121	109	105	104	105	106	107	108
$\hat{\alpha}_2$	Bias	-0.77	-6.29	-12.72	-16.03	-16.49	-15.81	-14.92	-14.13	-13.48	-12.97
	RMSE	328	153	200	285	368	431	476	506	528	543
	std	328	137	129	134	142	149	153	157	159	161
$\hat{\beta}$	Bias	3.11	4.10	3.38	3.36	3.81	4.36	4.84	5.22	5.51	5.73
	RMSE	123	101	100	102	108	116	125	132	137	142
	std	123	101	100	102	108	116	125	132	137	142
						DGP3					
$\hat{\alpha}_1$	Bias	-0.88	28.79	54.60	63.42	66.36	68.19	69.57	70.61	71.38	71.96
	RMSE	364	168	185	207	225	238	247	253	258	261
	std	364	151	132	126	123	122	121	121	121	121
$\hat{\alpha}_2$	Bias	-1.73	-30.57	-83.33	-111	-109	-96.56	-83.09	-71.91	-63.17	-56.43
	RMSE	295	239	456	606	675	704	716	721	724	725
	std	295	160	176	179	175	172	170	169	169	168
$\hat{\beta}$	Bias	-2.53	2.87	8.70	13.47	15.48	15.91	15.76	15.47	15.19	14.94
	RMSE	126	104	114	137	153	161	166	168	169	170
	std	126	104	114	136	152	161	165	167	168	169
						DGP4					
$\hat{\alpha}_1$	Bias	-0.28	15.08	24.09	22.29	19.13	17.37	16.58	16.25	16.12	16.08
	RMSE	414	179	181	196	212	225	235	242	247	251
	std	414	167	142	133	128	126	125	125	125	124
$\hat{\alpha}_2$	Bias	-1.33	-29.26	-79.49	-103	-98.13	-83.39	-69.03	-57.40	-48.43	-41.57
	RMSE	330	254	470	618	684	711	722	728	730	732
	std	329	180	204	209	205	201	199	198	197	197
$\hat{\beta}$	Bias	1.91	1.02	-4.70	-8.61	-10.04	-10.11	-9.74	-9.30	-8.92	-8.60
	RMSE	123	103	113	137	155	164	169	171	173	174
	std	123	103	113	136	154	164	169	171	173	174

Notes: (1) The estimates in the table have been multiplied by 1000. (2) $\hat{\beta}$ is the profile least squares estimate of β; $\{\hat{\alpha}_1, \hat{\alpha}_2\}$ are kernel based estimates of the time-varying coefficients of the factors. 3) Bandwidth for parameter estimation is $h_p = dT^{-1/5}; T = 100$

Table 3.19: Estimated bias, RMSE, and standard deviation for 10 bandwidths when factors are estimated and have VMA(2) structure, $[T, N] = [100, 100]$

	d	0.050	0.267	0.483	0.700	0.917	1.133	1.350	1.567	1.783	2.000
						DGP1					
$\hat{\alpha}_1$	Bias	-54.44	-47.55	-91.12	-140	-167.	-180.	-187	-191	-193	-195
	RMSE	2469	557	426	376	361	360	362	365	368	370
	SD	2468	548	409	343	311	298	292	290	289	289
$\hat{\alpha}_2$	Bias	22.71	79.12	222	345	403	431	446	455	461	464
	RMSE	4772	574	476	576	651	691	714	729	738	745
	SD	4770	539	289	189	165	158	156	155	154	154
$\hat{\beta}$	Bias	-3.81	-5.01	-7.49	-12.25	-16.35	-19.30	-21.39	-22.87	-23.95	-24.75
	RMSE	123	104	103	102	102	103	103	104	104	104
	SD	123	104	102	101	101	101	101	101	101	101
						DGP2					
$\hat{\alpha}_1$	Bias	-2.76	-0.80	27.20	55.34	71.55	79.29	83.05	85.01	86.12	86.78
	RMSE	2291	552	423	374	356	349	347	347	347	348
	SD	2291	551	418	360	333	321	315	312	311	310
$\hat{\alpha}_2$	Bias	-10.05	48.50	152	253	310	337	352	360	365	368
	RMSE	2073	402	383	482	564	612	641	659	672	680
	SD	2072	377	243	187	168	162	160	159	159	159
$\hat{\beta}$	Bias	-3.71	-2.04	-0.62	1.17	3.20	4.75	5.82	6.56	7.09	7.48
	RMSE	123	109	108	108	108	108	109	109	110	110
	SD	123	109	108	107	108	108	109	109	110	110
						DGP3					
$\hat{\alpha}_1$	Bias	-55.15	-2.47	15.77	35.47	49.77	59.46	65.85	70.14	73.10	75.22
	RMSE	2490	639	515	441	405	394	393	394	396	398
	SD	2488	631	509	422	370	346	335	330	327	326
$\hat{\alpha}_2$	Bias	-7.80	-47.98	-152	-170	-136	-108	-90.14	-77.66	-69.03	-62.86
	RMSE	4776	608	616	718	735	735	734	734	733	733
	SD	4773	549	320	219	196	190	188	187	187	187
$\hat{\beta}$	Bias	-3.81	-2.13	-0.21	4.49	7.20	8.57	9.38	9.89	10.24	10.48
	RMSE	123	104	104	106	107	108	109	109	109	110
	SD	123	104	104	105	107	108	108	109	109	109
						DGP4					
$\hat{\alpha}_1$	Bias	-13.44	-8.24	-11.31	-3.96	7.50	15.50	20.46	23.61	25.71	27.17
	RMSE	2313	626	516	458	431	420	416	415	415	415
	SD	2312	617	496	427	388	368	358	353	351	349
$\hat{\alpha}_2$	Bias	-27.02	-38.32	-102	-119	-95.68	-71.63	-54.00	-41.59	-32.77	-26.34
	RMSE	2079	466	573	685	719	729	732	734	734	735
	SD	2077	410	304	240	216	209	207	206	206	206
$\hat{\beta}$	Bias	-3.91	-3.34	-2.93	-4.41	-6.12	-7.08	-7.60	-7.90	-8.08	-8.20
	RMSE	123	110	110	112	114	115	116	117	117	117
	SD	123	110	110	112	114	115	116	116	117	117

Note: 1) The entries in the table have been multiplied by 1000. 2) Model is considered with nonparametrically estimated factors. 3) The bandwidth for factor estimation is $h_f = 0.02$.

Table 3.20: Bias and standard deviations[SD] of the estimated factors and the factor loadings; $h_f = 0.05T^{-1/5}$

			DGP1 and DGP3						DGP2 and DGP4					
			NP factors			PCA factors			NP factors			PCA factors		
N		T	100	300	500	100	300	500	100	300	500	100	300	500
100	$\hat{F}_1$	Bias	200	200	200	217	200	200	217	215	216	217	215	216
		SD	406	256	207	488	486	486	327	195	156	488	480	481
		RMSE	1149	1244	1264	968	970	970	1042	1147	1170	968	968	968
	$\hat{F}_2$	bias	95.2	94.5	95.3	26.6	94.5	95.3	26.7	26.8	28.0	26.6	26.8	28.0
		SD	307	198	163	714	665	667	371	248	202	714	711	708
		RMSE	1302	1331	1345	1557	1473	1473	1562	1506	1492	1557	1554	1556
	$\hat{\lambda}_1$	bias	-4.7	9.6	3.5	-4.1	6.5	2.7	-4.8	8.6	3.7	-4.1	6.2	3.4
		SD	2059	1815	1707	1435	1358	1358	2083	1693	1606	1435	1424	1426
	$\hat{\lambda}_2$	bias	2.0	-5.0	-0.8	1.3	-4.5	-0.7	1.1	-5.5	-0.7	1.3	-4.8	-0.5
		SD	634	589	578	597	572	569	649	629	618	597	591	589
300	$\hat{F}_1$	bias	200	200	200	200	200	200	214	213	215	214	213	215
		SD	402	254	205	491	486	482	324	192	153	481	478	477.
		RMSE	1151	1247	1264	970	969	969	1041	1150	1170	968	968	968
	$\hat{F}_2$	bias	96.7	95.9	100.4	96.7	95.9	100.4	28.5	27.7	30.5	28.4	27.7	30.6
		SD	306	198	160	659	656	656	372	249	202	707	705	703
		RMSE	1301	1335	1346	1474	1475	1470	1564	1512	1494	1557	1556	1555
	$\hat{\lambda}_1$	bias	2.3	3.4	2.7	1.7	2.0	2.1	2.3	2.7	2.4	1.8	1.9	2.1
		SD	2059	1817	1708	1366	1363	1358	2076	1695	1606	1431	1427	1426
	$\hat{\lambda}_2$	bias	-0.5	-1.4	-1.5	-0.5	-1.2	-1.3	-0.2	-1.5	-1.6	-0.4	-1.4	-1.5
		SD	631	583	574	573	568	566	640	623	614	589	585	585
500	$\hat{F}_1$	bias	200	202	201	200	202	201	211	214	215	211	214	215
		SD	404	255	205	490	485	481	325	192	154	479	476.5	476
		RMSE	1149	1248	1264	969	971	970	1039	1149	1171	967	968	968
	$\hat{F}_2$	bias	91.8	100	103	91.9	100	103	21.9	31.7	33.8	21.9	31.7	33.9
		SD	304	196	161	653	653	654	373	249	203	707	705	702
		RMSE	1302	1334	1346	1475	1473	1470	1565	1510	1492	1557	1556	1554
	$\hat{\lambda}_1$	bias	-0.4	0.1	4.3	-0.3	0.5	3.1	-0.1	-0.1	3.6	-0.4	0.6	2.8
		SD	2052	1829	1712	1367	1363	1357	2072	1700	1607	1431	1428	1424
	$\hat{\lambda}_2$	bias	0.2	0.2	-2.6	-0.1	0.2	-2.5	0.0	0.1	-2.9	-0.1	0.1	-2.8
		SD	629	585	574	573	570	566	639	623	614	588	585.1	584

All the values are multiplied by 1000. The NP factors are estimated using $h_f = 0.05T^{-1/5}$

Additional results from the empirical application

We evaluated 100 different values for the pair (h_f, h_p) in terms of out-of-sample forecast error when expanding windows are used. We carried out separate evaluations for $\log(GDP)$, GDP Growth Rate, and Inflation. The initial estimation period is 1959:Q1-1998:Q4, and hence the corresponding forecast period is 1999:Q1-2018:Q4. Tables 3.21 to 3.23 provide the out-of-sample sum of squared forecasting error [SSFE]. For forecasting log(GDP), we chose $h_f = 0.011$ because Table 3.21 shows that $h_f = 0.011$ provides the best overall performance in terms of SSFEs. Similarly, for forecasting GDP Growth Rate we chose $h_f = 0.024$ because Table 3.22 shows that $h_f = 0.024$ provides the best overall performance in terms of SSFEs. Finally, or forecasting inflationf we chose $h_f = 0.04$ because Table 3.23 shows that the chosen figure provides the best overall performance in terms of SSFEs.

Table 3.21: Sum of squared forecast error for one-step ahead forecast of *log(GDP)*

	Model 1										Model 2
$h_p \backslash h_f$	0.011	0.014	0.017	0.020	0.024	0.027	0.030	0.033	0.037	0.040	
0.017	8.12	13.04	15.66	15.93	65.15	554	190	336	277	8067	20.82
0.089	3.86	3.83	3.68	3.54	3.47	3.36	3.20	3.07	3.00	2.96	3.65
0.162	3.65	3.66	3.66	3.69	3.71	3.73	3.73	3.72	3.70	3.68	2.93
0.234	3.55	3.62	3.66	3.71	3.76	3.79	3.80	3.80	3.78	3.74	2.78
0.306	3.49	3.61	3.70	3.78	3.86	3.92	3.96	3.98	3.98	3.96	2.73
0.379	3.56	3.71	3.83	3.94	4.04	4.12	4.19	4.23	4.26	4.27	2.73
0.451	3.63	3.79	3.93	4.05	4.16	4.25	4.33	4.40	4.45	4.47	2.75
0.524	3.68	3.85	3.99	4.11	4.23	4.33	4.42	4.50	4.56	4.59	2.78
0.596	3.71	3.88	4.02	4.15	4.27	4.37	4.47	4.56	4.63	4.67	2.81
0.668	3.73	3.90	4.04	4.17	4.29	4.40	4.51	4.60	4.67	4.72	2.83
h_p	Model 3										Model 4
0.017	5.00	7.30	9.13	20.24	181	586	167	225	654	77.83	18.70
0.089	3.89	3.71	3.53	3.42	3.38	3.31	3.12	2.90	2.75	2.65	3.76
0.162	3.58	3.51	3.50	3.61	3.80	3.96	4.03	4.00	3.88	3.73	2.89
0.234	3.43	3.46	3.55	3.71	3.93	4.15	4.34	4.44	4.44	4.36	2.67
0.306	3.37	3.47	3.58	3.73	3.91	4.10	4.29	4.46	4.57	4.61	2.61
0.379	3.41	3.54	3.67	3.81	3.99	4.19	4.41	4.63	4.81	4.92	2.59
0.451	3.45	3.59	3.71	3.86	4.04	4.25	4.49	4.73	4.95	5.11	2.59
0.524	3.47	3.61	3.73	3.87	4.05	4.27	4.52	4.78	5.02	5.20	2.59
0.596	3.49	3.61	3.74	3.88	4.06	4.27	4.53	4.80	5.05	5.25	2.60
0.668	3.49	3.62	3.74	3.88	4.06	4.27	4.53	4.81	5.07	5.27	2.61

Notes: 1) All the values have been multiplied by 1000. 2) First in-sample period is 1959:Q1-1998:Q4 and the corresponding forecast period is 1999:Q1-2018:Q4. 3) Models 1 and 3 contain NP estimated factors and Models 2 and 4 contain PCA estimated factors. 4) We use these results to select the optimal $\hat{h}_f$. 5) These forecasts used expanding window.

Table 3.22: Sum of squared (expanding window) forecasting error for the one-step ahead forecasting, *GDP growth rate*

	Model 1										Model 2
$h_p \backslash h_f$	0.011	0.014	0.017	0.020	0.024	0.027	0.030	0.033	0.037	0.040	
0.017	23.62	36.33	38.06	40.19	43.50	43.84	56.84	64.02	138	5725	21.96
0.089	4.23	4.17	3.95	3.67	3.52	3.50	3.54	3.60	3.62	3.61	3.38
0.162	3.41	3.34	3.28	3.28	3.31	3.36	3.40	3.42	3.44	3.44	2.70
0.234	3.16	3.17	3.19	3.21	3.22	3.22	3.22	3.22	3.21	3.20	2.58
0.306	2.96	2.99	3.02	3.04	3.05	3.05	3.06	3.06	3.06	3.05	2.53
0.379	2.89	2.92	2.94	2.95	2.96	2.97	2.97	2.98	2.98	2.99	2.51
0.451	2.87	2.90	2.91	2.92	2.92	2.93	2.93	2.94	2.95	2.96	2.51
0.524	2.87	2.89	2.90	2.90	2.90	2.90	2.91	2.92	2.93	2.94	2.52
0.596	2.88	2.89	2.89	2.89	2.88	2.89	2.89	2.90	2.92	2.94	2.53
0.668	2.88	2.89	2.89	2.88	2.87	2.87	2.88	2.89	2.91	2.93	2.54
h_p	Model 3										Model 4
0.017	9.52	14.47	18.84	21.81	26.01	27.78	53.01	44.62	58.30	136	21.55
0.089	3.99	3.96	3.77	3.55	3.41	3.38	3.41	3.47	3.52	3.53	3.47
0.162	3.37	3.31	3.27	3.27	3.31	3.35	3.40	3.42	3.44	3.43	2.72
0.234	3.14	3.17	3.19	3.20	3.21	3.22	3.21	3.20	3.19	3.17	2.54
0.306	2.93	2.98	3.01	3.03	3.03	3.04	3.03	3.03	3.02	3.01	2.49
0.379	2.84	2.89	2.92	2.93	2.94	2.94	2.94	2.94	2.94	2.94	2.49
0.451	2.82	2.86	2.88	2.88	2.89	2.89	2.89	2.89	2.89	2.90	2.50
0.524	2.81	2.84	2.86	2.86	2.85	2.85	2.86	2.86	2.87	2.88	2.52
0.596	2.81	2.83	2.84	2.84	2.83	2.83	2.84	2.84	2.85	2.87	2.54
0.668	2.81	2.83	2.83	2.83	2.82	2.82	2.82	2.83	2.84	2.86	2.56

Notes: 1) All the values are multiplied by 1000. 2) First in-sample period is 1959:Q1- 1998:Q4 and forecasting period is 1999:Q1- 2018:Q4. 3) Model 1 and 3 contain NP estimated factors and Model 2 and 4 contain PCA estimated factors. 4) We use these reults to select the optimal h_f, $\hat{h}_f$, to the main part of the chapter.

Table 3.23: Sum of squared (expanding window) forecasting error for the one-step ahead forecasting, *inflation*

	Model 1										Model 2
$h_p \backslash h_f$	0.011	0.014	0.017	0.020	0.024	0.027	0.030	0.033	0.037	0.040	
0.017	20.93	43.29	49.81	42.97	44.01	46.22	59.42	65.87	284	246	18.22
0.089	4.48	4.44	4.23	4.02	3.85	3.71	3.62	3.62	3.66	3.71	4.43
0.162	3.26	3.14	3.06	3.01	2.97	2.95	2.93	2.92	2.93	2.93	3.09
0.234	3.04	3.01	2.98	2.96	2.94	2.92	2.90	2.89	2.88	2.87	2.97
0.306	2.95	2.94	2.93	2.91	2.89	2.88	2.86	2.85	2.84	2.83	2.94
0.379	2.91	2.90	2.89	2.88	2.86	2.85	2.84	2.83	2.82	2.81	2.92
0.451	2.89	2.88	2.87	2.86	2.85	2.84	2.83	2.82	2.81	2.80	2.92
0.524	2.87	2.87	2.86	2.85	2.84	2.83	2.82	2.81	2.81	2.80	2.91
0.596	2.86	2.86	2.85	2.85	2.84	2.83	2.82	2.81	2.80	2.80	2.91
0.668	2.86	2.85	2.85	2.84	2.83	2.83	2.82	2.81	2.80	2.80	2.90
h_p	Model 3										Model 4
0.017	7.51	17.38	24.16	21.85	21.03	26.41	30.90	39.43	80.71	299	13.19
0.089	4.55	4.26	4.03	3.95	3.91	3.84	3.78	3.79	3.86	3.94	5.41
0.162	3.90	3.67	3.50	3.40	3.34	3.31	3.30	3.30	3.30	3.30	3.62
0.234	3.54	3.45	3.39	3.35	3.31	3.29	3.27	3.24	3.22	3.20	3.35
0.306	3.40	3.35	3.31	3.27	3.24	3.21	3.18	3.16	3.13	3.11	3.31
0.379	3.34	3.30	3.26	3.22	3.18	3.15	3.13	3.10	3.08	3.06	3.31
0.451	3.31	3.27	3.23	3.19	3.15	3.12	3.10	3.08	3.06	3.04	3.32
0.524	3.30	3.25	3.21	3.17	3.14	3.11	3.08	3.06	3.04	3.03	3.33
0.596	3.29	3.24	3.20	3.16	3.13	3.10	3.07	3.05	3.04	3.02	3.33
0.668	3.28	3.23	3.19	3.15	3.12	3.09	3.07	3.05	3.03	3.02	3.34

Notes: 1) All the values are multiplied by 1000. 2) First in-sample period is 1959:Q1- 1998:Q4 and forecasting period is 1999:Q1- 2018:Q4. 3) Model 1 and 3 contain NP estimated factors and Model 2 and 4 contain PCA estimated factors. 4) We use these reults to select the optimal h_f, $\hat{h}_f$, to the main part of the chapter.

We use these three different bandwidths to estimate the factors in out-of-sample (expanding window) forecasting in the main chapter.

3.6.3 Appendix C: Rolling window forecast

This section provides results to support our choice to use expanding window, instead of the rolling window, in the empirical application presented in Section 3.4. We considered rolling window forecast for the following cases. Window sizes: 80 and 120; bandwidth for factor estimation: $h_f = 0.011$; and ten values for h_p. Table 3.24 provides the values of R^2_{os} with AR(1) as the benchmark model. The results in Table 3.24 show that FAR models with non-parametrically estimated factors have negative values for R^2_{os} for forecasting each of the three variables. The FAR models with PCA factors have better performance with $h_p = 0.668$ for forecasting log(GDP). For forecasting GDP Growth Rate, except for Model 2, every model specification with any of the bandwidths considered have lower predictive performance compared to the AR models. For Inflation forecasting, all the model specifications have negative R^2_{os} with the window size 80. Model 2 with $h_p = 0.451$ has better forecast performance than with window size 120. Overall, for a few bandwidth options, semi-parametric FAR models with PCA factors have better out-of-sample forecasting performance compared to the AR models and the mean model. Hence, for forecasting any of the three response variables by a semi-parametric FAR model, the performance with expanding window is better than that with the rolling window.

Table 3.24: Out-of-sample performance of the models relative to AR(1), in terms of R^2_{os} for rolling window forecast

Window size	h_p	0.017	0.089	0.162	0.234	0.306	0.379	0.451	0.524	0.596	0.668
log(GDP)											
80	Model 1	-121	-0.50	-0.19	-0.10	-0.10	-0.09	-0.08	-0.07	-0.06	-0.05
	Model 2	-632	-0.78	-0.27	0.01	0.08	0.11	0.12	0.14	0.15	**0.15**
	Model 3	-114	-0.42	-0.18	-0.11	-0.14	-0.16	-0.16	-0.15	-0.15	-0.14
	Model 4	-848	-0.91	-0.32	-0.03	0.03	0.04	0.05	0.06	0.06	0.07
	AR(4)	-1.50									
120	Model 1	-4.33	-0.26	-0.19	-0.15	-0.19	-0.21	-0.20	-0.19	-0.18	-0.17
	Model 2	-12.73	-0.27	-0.04	0.07	0.09	0.09	0.09	0.09	0.09	0.09
	Model 3	-2.66	-0.21	-0.29	-0.27	-0.32	-0.33	-0.33	-0.32	-0.31	-0.30
	Model 4	-21.17	-0.41	-0.09	0.04	0.09	0.10	0.10	0.10	0.10	**0.10**
	AR(4)	-1.44									
GDP growth rate											
80	Model 1	-126	-1.29	-0.54	-0.32	-0.26	-0.22	-0.18	-0.16	-0.14	-0.13
	Model 2	-1072	-1.28	-0.54	-0.19	-0.08	-0.04	-0.02	0.00	0.01	**0.02**
	Model 3	-34.54	-1.09	-0.51	-0.35	-0.30	-0.27	-0.24	-0.22	-0.20	-0.19
	Model 4	-1132	-1.57	-0.63	-0.23	-0.12	-0.09	-0.07	-0.06	-0.05	-0.05
	AR(4)	-0.01									
	Mean model	-0.15									
120	Model 1	-28.25	-0.88	-0.44	-0.29	-0.23	-0.18	-0.15	-0.12	-0.10	-0.09
	Model 2	-22.58	-0.64	-0.22	-0.03	**0.02**	0.01	0.00	-0.01	-0.01	-0.02
	Model 3	-5.14	-0.63	-0.44	-0.33	-0.29	-0.24	-0.20	-0.18	-0.17	-0.16
	Model 4	-27.87	-0.77	-0.29	-0.08	-0.01	-0.01	-0.03	-0.04	-0.05	-0.05
	AR(4)	0.01									
	Mean model	-0.18									
Inflation											
80	Model 1	-1619	-1.67	-0.82	-0.58	-0.41	-0.31	-0.26	-0.23	-0.22	-0.21
	Model 2	-188	-3.01	-1.29	-0.53	-0.24	-0.15	-0.12	-0.10	-0.10	-0.09
	Model 3	-382	-1.22	-0.76	-0.62	-0.52	-0.43	-0.37	-0.33	-0.31	-0.29
	Model 4	-81.86	-2.52	-1.33	-0.71	-0.39	-0.30	-0.26	-0.24	-0.23	-0.22
	AR(4)	-0.13									
	Mean model	**0.10**									
120	Model 1	-9.23	-0.61	-0.41	-0.19	-0.08	-0.05	-0.03	-0.03	-0.03	-0.03
	Model 2	-13.12	-0.85	-0.33	-0.02	0.04	0.05	**0.06**	0.06	0.05	0.05
	Model 3	-2.16	-0.35	-0.42	-0.30	-0.22	-0.18	-0.16	-0.15	-0.14	-0.13
	Model 4	-7.94	-0.82	-0.53	-0.20	-0.09	-0.06	-0.05	-0.04	-0.04	-0.04
	AR(4)	-0.12									
	Mean model	-0.17									

Notes: 1) The h_f for the three response variable is 0.011. 2) Mean model results for log(GDP) is not included as the series is non-stationary. 3) The highlighted values are corresponding to the highest R^2_{os}.

Chapter 4

Forecasting Univariate Time Series using a Semiparametric Multi-level FAR model

4.1 Introduction

Accurate forecasting of key economic variables, such as GDP growth and inflation, is central to making economic policy decisions. Therefore, forecasting GDP growth and inflation is a popular topic among applied researchers. In a method that has attracted considerable attention in the recent literature, an approximate factor model and a regression model are used for predicting macroeconomic variables; this method has two steps for using the two models jointly in a sequential manner. In the first step, the approximate factor model is used for estimating a small number of global (pervasive) factors to extract a large proportion of the information contained in a set of panel data on a large number of economic variables that are potential predictors. In the second step, the generated factors are included in a separate model, called a factor augmented regression (FAR) model, for forecasting the desired macroeconomic variable. This method was reviewed and extended in the previous two chapters. The objective of this Chapter is to build on the previous chapters and develop a new improved method for forecasting macroeconomic variables.

Albeit recent, there is a growing literature on what are called *multi-level factor models* (for example, Wang 2008, Boivin and Ng 2006, Beck et al. 2009, Breitung and Eickmeier 2016, and Rodríguez-Caballero 2021). In the factor models studied in the previous Chapters and those studied in the literature, the factors are permitted to impact all the variables in the high-dimensional panel data. Consequently, a large number of factor loadings need to be estimated. The idea that underlies the multi-level panel model is that researchers are able to use their prior knowledge and partition the panel data into several distinct groups; as an example, the groups may be economic sectors, or geographical regions. The factors generated using the variables in one particular group are assumed to impact only the variables in that particular group, not the variables in the other groups. These restrictions were exploited in the proposition of the two-level factor model, in which the global (pervasive) factors affect all the variables in the panel and the group (non-pervasive) factors impact only the specific group.

The aim of this Chapter is to combine the literature on the aforementioned two-level factor models and the new developments in the previous two chapters of this thesis, and propose a two-level FAR model that allows for a mixture of I(1) and I(0) level-1 and level-2 factors and time varying coefficients for the factors. We call this model the *semi-parametric two-level FAR model*. We evaluate the accuracy of proposed model for out-of-sample predictions of log(GDP), GDP growth, and inflation. We compare our proposed method with its competitors that include the mixture-FAR model and the semi-parametric one-level mixture FAR model studied in Chapters 2 and 3 respectively, and some time series models.

The two-level factor model is estimated by the method introduced by Breitung and Eickmeier [2016], where the panel of only I(0) variables and thus I(0) factors are studied. We adapt this method to our panel data model setup with I(1) and I(0) variables and hence a mixture of I(0) and I(1) global and group factors. A kernel method that is similar to the one proposed in Chapter 3 is used for the estimation of semi-parametric two-level FAR model.

Although there are advances made on the development of the estimation of multi-level factor models (Wang 2008, and Breitung and Eickmeier 2016), empirical applications of this method were largely focused on explaining the underlying behavior of business cycles, international business cycles and international trades. Beck et al. [2009] used the two-level factor model (with national factors and regional factors) and showed that they play a major role in explaining inflation variability in the regional inflation from six Euro area countries. To our knowledge, the two-level factors are not yet used to improve the FAR model which is widely used for forecasting macroeconomic variables.

For forecasting I(0) variables, inflation and GDP growth rate, using FAR models and time series models, the aforementioned studies recommend the rolling window-sampling scheme over the recursive (expanding) window sampling scheme. Further, Rossi and Inoue [2012] derived an optimum rolling widow size for forecasting GDP growth and inflation using time series models with observable variables. Recently, Inoue et al. [2017] studied such time series models with time-varying parameters and derived the optimal rolling window size for forecasting stationary macroeconomic variables. They found that the proposed method works well for forecasting inflation but not for GDP growth. The model that we study in this Chapter has several additional features: two-level factor structure in the FAR, factors are mixture of I(0) and I(1), and factor parameters are time-varying. In this Chapter, we will consider both the expanding (recursive) window and the rolling window sampling schemes and assess the sensitivity of forecasts to the two sampling schemes as well to the window size.

The main contribution of this Chapter is to develop a method for forecasting univariate macroeconomic variables by extending the well-known factor model to include the following three new features: the factor model and thus the FAR model include two-level factors, the factor parameters are time-varying, and the FAR includes mixture of I(1) and I(0) two-level factors. As will be seen later in this Chapter, based on prior knowledge or selection criteria such as goodness-of-fit and correlation measures, we can select the appropriate number of global and group factors as predictors in the FAR model for the desired macroeconomic variable. Thus, this Chapter builds three separate semi-parametric two-level FAR models for the three variables log(GDP), GDP and inflation.

The rest of this Chapter is organized as follows. Section 4.2 introduces the model setup and notations, and proposes a kernel method for the estimation of the semi-parametric mixture-FAR model with two-level factor structure. Section 4.3 reports the results of the estimation and the selection of global and group factors as predictors. Section 4.4 evaluates the accuracy of the models' forecasts; the results are analysed in these sections. Section 4.5 concludes the Chapter. The results of the empirical analysis are reported in tables and plots in the Appendix to this Chapter.

Related literature

Inflation and GDP growth forecasts are frequently produced and used to improve decision-making at the micro and macro levels. Therefore, generating accurate forecasts of GDP growth and inflation of major countries and regions has been the main focus of a vast number of studies in the economic and econometric literature. For example, see Banerjee and Marcellino [2006] for forecasting GDP growth and inflation for the US, Demers and Cheung [2007] for Canada, Barnett et al. [2014] for the UK, and Banerjee et al. [2005] for Euro area, and Lahiri and Sheng [2010] for G7 countries. See also the recent survey papers Eickmeier and Ziegler [2008] for forecasting GDP growth and inflation and Kavtaradze and Mokhtari [2018] for inflation. Forecasting inflation was shown to be difficult (Stock and Watson 2007) which led to the emergence of huge literature in theory and applications on modeling and assessing the accuracy of forecasting inflation by applied researchers and central bankers.

Abbate and Marcellino [2014] showed that the main reason for predictive failure is the use of single models that by necessity can only incorporate a small subset of the variables. Thus, predictive failure or inaccurate forecast is the result of (1) not taking account of all information in the data, and (2) not taking account of model uncertainty.

Since the advent of the factor models by Stock and Watson [2002a] and Stock and Watson [2007], Bai and Ng [2002] and Bai and Ng [2006] and thus the FAR model, many of the aforementioned studies assessed the accuracy of forecasts by the univariate FAR model relative to standard time series models and those based on economic theory such as Phillips curve. Evidence in the huge literature on this topic indicate that the FAR model captures a high proportion of the information content in the large panel data through only a few factors. Such FAR models out-perform the time series models and economic theory based model in the out-of-sample predictions.

The standard FAR model frequently ignores structural changes over time. When the relationship between a macroeconomic variable and its fundamentals changes over time, the underlying parameters and structure might change with time as well. Models with time varying parameters (TVP) explicitly allow non-linear reactions to the structural changes. These models were found to produce more precise estimates over other econometric models. The most important advantage of the model with time-varying coefficients is that it corrects specification errors as a result of incorrect functional forms, omitted variables and measurement errors in models. Moreover, the TVP models are able to show improved forecasting ability even when dealing with non-stationary variables and seasonally unadjusted data (Abbate and Marcellino 2014 and Kavtaradze and Mokhtari [2018]). In this Chapter, we allow the parameters of FAR model to be time-varying and evidence in the literature indicates that such model tends to improve out-of-sample predictions.

Testing and modeling structural instabilities in the factor loadings and the FAR model parameters were given attention in the recent literature. For example, see Banerjee et al. [2008]

and Stock and Watson [2009], Corradi and Swanson [2014], and Su and Wang [2017]. Albeit the development of methods allows structural instabilities in both panel data model and FAR model, only the in-sample properties of these models were studied and the performance of such an FAR model in the out-of-sample prediction is largely unknown. Wei and Zhang [2020] used a time-varying diffusion index model for forecasting stationary variables. Furthermore, the aforementioned methods are developed only for the I(0) factors and the FAR model for I(0) variables. In this thesis, we assume constant factor loadings and allow TVP in the FAR model.

Recently, a methodology for the multi-level factor model of I(0) variables was developed and studied the movements of international business cycle (nationally and regionally) and international trades. Several approaches have been proposed for estimating multi-level factors. Wang [2008] developed an iterative principal component (PC) method for a large dimensional multi-level factor model with stationary factor structure, and showed consistency and asymptotic normality of the factor estimators. In a comparative study of alternative multi-level factor estimation methods, such as the two-step PCA estimator, a sequential PCA approach, and a quasi-ML approach, Breitung and Eickmeier [2016] recommended the use of a sequential least squares (LS) algorithm for the estimation of multi-level factors and implemented them to estimate two-level and three-level factors. This LS method is a two-step approach based on a canonical correlation analysis (CCA). Through a simulation study, Breitung and Eickmeier [2016] showed that, under certain conditions, sequential LS estimators tend to outperform the quasi ML estimator and two-step PCA estimator. Choi et al. [2018] proposed a sequential PCA estimation approach to consistently estimate global and regional stationary factors and derived the asymptotic normality of the PCA estimators. Choi et al. [2018] used the CCA, which is similar to Breitung and Eickmeier [2016], to estimate the initial global factors and multi-level factors. Under the assumption that the number of global factors are known, Choi et al. [2018] also proposed several information criteria to estimate the optimal number of regional factors.

In this Chapter, we adapt the LS method introduced by Breitung and Eickmeier [2016], which was proposed for the panel with I(0) variables, to the panel with both I(0) and I(1) variables.

4.2 Methodology

4.2.1 Model setup and notations

Let $X_{[T\times N]} = \{X_{it}; t = 1, \ldots, T, i = 1, \ldots, N\}$ be the panel of observable time series that has a common factor structure, and let $Y_{[T\times 1]} = \{Y_t; t = 1, \ldots, T\}$ be the observable series that we wish to forecast. The factor model with one-level (global) factor structure given as,

$$\text{one-level factor model:} \qquad X_{it} = \lambda_i' F_t + e_{it}, \qquad (i = 1, \ldots, N, t = 1, \ldots, T), \tag{4.1}$$

where $F = \{F_t; t = 1, \ldots T\}$ is the $T \times r$ matrix of unobservable global (pervasive) factors, $\Lambda = \{\lambda_i; i = 1, \ldots, N\}$ is the $N \times r$ matrix of factor loadings, and $e = \{e_{it}; i = 1, \ldots, N, t = 1, \ldots, T\}$ is the $T \times N$ matrix of errors of the factor model.

Let us consider a two-level factor model which consists of a set global (level-1) factors and another set of group (level-2) factors. The global factors affect all the variables in the panel model, whereas each of the group factors affect only a group of variables in the panel.

Remark 1: In the empirical application that we study in this paper, we develop a semi-parametric two-level FAR model for forecasting three macroeconomic variables, GDP in level (I(1)) and GDP growth and inflation (both are (I(0)). We use the widely studied FRED-QD data set in the factor model and thus the construction of FAR model for forecasting the three variables. Stock and Watson (2002) categorised 100 *sub-aggregate* variables into 12 groups; see the Appendix for details. In our applications, we estimate level-1 factors and level-2 factors from 12 groups. Therefore, we refer to these factors as group factors. We will use this terminology throughout the paper. Further we use global factor and level-1 as well as group factor and level-2 factor synonymously.

Remark 2: To our knowledge, formal methods are not yet developed for estimating the number of levels in multi-level factors. The choice of the number of (level 1 and level 2) factors depends on the problem in hand. Studies by Breitung and Eickmeier [2016], Beck et al. [2009], and Hirata et al. [2013] provide heuristic methods to determine the number of levels, the size of S, and the number of level 1 and level 2 factors in their empirical analyses. We proposed heuristic methods that suit our empirical setting, which involves generating two-level factors for forecasting GDP,

GDP growth and inflation. Stock and Watson categorized the panel of variables into 12 groups. Therefore, we set $S = 12$ in our empirical study.

Let s indicate group s and $x_{s,it}$ be the i^{th} variable in group s observed at time t. Then, the two-level factor model can be specified as,

$$\text{Two-level factor model:} \quad x_{s,it} = \gamma'_{s,i} H_t + \lambda'_{s,i} R_{s,t} + e_{s,it} \qquad (i = 1, \ldots, N_s, s = 1, \ldots, S), \tag{4.2}$$

where H_t is an $r_G \times 1$ vector of global factors, $R_{s,t}$ is an $r_s \times 1$ vector of group factors, $\gamma_{s,i}$ and $\lambda_{s,i}$ respectively are their factor loadings, $e_{s,it}$ is the set of idiosyncratic errors, S is the number of groups and N_s is number of variables in group s such that $N = N_1 + \ldots + N_S$.

Let us express model (4.2) in a matrix form as,

$$X_{s,t} = \begin{pmatrix} \Gamma_s & \Lambda_s \end{pmatrix} \begin{pmatrix} H_t \\ R_{s,t} \end{pmatrix} + e_{s,t}, \tag{4.3}$$

where Γ_s and Λ_s are factor loadings corresponding to global and group factors for the group s, and $e_{s,t}, X_{s,t}$ are the sets of N_s idiosyncratic errors and variables in group s respectively. The two-level factor model (4.3) with S groups with necessary block restrictions can be specified as,

$$\begin{pmatrix} X_{1,t} \\ \vdots \\ X_{S,t} \end{pmatrix} = \begin{pmatrix} \Gamma_1 & \Lambda_1 & 0 & \ldots & 0 \\ \Gamma_2 & 0 & \Lambda_2 & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ \Gamma_S & 0 & 0 & \ldots & \Lambda_S \end{pmatrix} \begin{pmatrix} H_t \\ R_{1,t} \\ \vdots \\ R_{S,t} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{S,t} \end{pmatrix} \tag{4.4}$$

$$X_t = \Lambda^m F_t^m + e_t, \tag{4.5}$$

where $X_t = (X'_{1,t}, \ldots, X'_{S,t})'$, $F_t^m = (H'_t, R'_{1,t}, \ldots, R'_{S,t})'$ and $\Lambda^m = (\Gamma', \Lambda)$ for $\Lambda = diag(\Lambda_1, \ldots, \Lambda_S)$, $\Gamma = (\Gamma'_1, \Gamma'_2, \ldots, \Gamma'_S)'$, Λ is a block diagonal matrix and each of its elements is the factor loadings for variables in each group. The number of global factors, r_G, and group factors $r_R = \sum_{s=1}^{S} r_s$ are assumed to be known. Thus, the two-level factor model can be succinctly defined as,

$$X = F^m \Lambda^{m'} + e. \tag{4.6}$$

where $X = (X_1, \ldots, X_T)'$. If $\Gamma_s H_t$ is known for $(s = 1, \ldots, S)$, then the group specific factors $R_{s,t}$ and their factor loadings Λ_s can be estimated from the factor model, $X_{s,t} - \Gamma_s H_t = \Lambda_s R_{s,t} + e_{s,t}$ using data from group s. On the other hand, if ΛR_t is known, then H_t and Γ can be obtained from factor model $X_t - \Lambda R_t = \Gamma H_t + e_t$ using the data from all the S groups. However, in practice, both global and group specific factors and their factor loadings are unknown and they need to be estimated jointly.

In this model, we allow both global and group factors to be a mixture of I(0) and I(1) latent time series. So, let us define $F_t^m = \left(E_t^{m'}, G_t^{m'}\right)'$, where E_t^m is the set of non-stationary global and group specific factors and $G_t^{m'}$ is the set of such stationary factors. Therefore,

$$E_t^m = E_{t-1}^m + u_t, \tag{4.7}$$

where u_t is stationary. Then, the two-level factor model takes the form,

$$X_t = \Lambda^{(m1)} E_t^m + \Lambda^{(m2)} G_t^m + e_t, \quad (t = 1, \dots, T), \tag{4.8}$$

where $\Lambda^{(m1)}$ and $\Lambda^{(m2)}$ are the sets of factor loadings corresponding to non-stationary and stationary factors respectively.

Let us assume that the factor parameters are time-varying. Then, the semi-parametric two-level FAR model for forecasting Y_t at t+h, can be specified as,

$$Y_{t+h} = \alpha_t' E_t^m + \beta_t' G_t^m + \omega' W_t + \epsilon_{t+h} = \theta_t' F_t^m + \omega' W_t + \epsilon_{t+h}, \qquad (t = 1, \dots, T), \tag{4.9}$$

where W_t is a set of observable regressors, ω is $n \times 1$ vector of unknown constant parameters. The set of $\theta_t = (\alpha_t', \beta_t')'$ is an $r \times 1$ set of unknown time-varying parameters which are functions of time and take the following form,

$$\alpha_t = \alpha(\tau_t) \text{ and } \beta_t = \beta(\tau_t) \text{ for } \tau_t = t/T, t = 1, \dots, T,$$

where $\alpha(.)$ and $\beta(.)$ are unknown smoothing functions.

The novelty of this model is that it is a flexible and richer semi-parametric FAR model with two-level factors which are in turn a mixture of I(0) and I(1) factors. We refer to this model as *semi-parametric two-level FAR model.*

4.2.2 Two-level factor estimation and identification

Recently, Breitung and Eickmeier [2016] proposed a sequential least squares (LS) approach for estimating two-level and three-level factors from the factor model with stationary variables.

In this paper, we extend this method to the two-level factor model with stationary and non-stationary factors. Therefore, we need to justify if some conditions are met in applying this method to empirical problems.

As the coefficients in model (4.9) that correspond to nonstationary and stationary time series are expected to have different convergence rates, we derive two new scaling matrices. Let p_1 and q_1 be the number of I(1) and I(0) global factors respectively, and let $p_{s,2}$ and $q_{s,2}$ be the number of I(1) and I(0) group factors generated from group s. Let us define two different scaling matrices for global and group factors as,

$$D_{1T} = diag(TI_{p_1}, T^{1/2}I_{q_1})_{r_G\times r_G}, \qquad D_{2T} = diag(D_{1,2T}, \ldots, D_{S,2T})_{r_R\times r_R}, \tag{4.10}$$

where $D_{s,2T} = diag(TI_{p_{s,2}}, T^{1/2}I_{q_{s,2}})_{r_s\times r_s}$ for $s = 1, \ldots, S$.

We impose the following restrictions in order to estimate and identify the global and group factors using the PCA estimation methods. Two normalization and diagonalization restrictions are imposed to estimate global factors, and another two that ensure the identification of the regional factors. These restrictions are:

i) $\Sigma_H = D_{1T}^{-1}\sum_{t=1}^{T} H_t H_t'$ and $\Sigma_s = D_{s,2T}^{-1}\sum_{t=1}^{T} R_{s,t}R_{s,t}'$ for $s = 1, \ldots, S$ are orthonormal matrices.

ii) The two matrices $N^{-1}\Gamma_s'\Gamma_s$ and $N^{-1}\lambda_s'\lambda_s$ are diagonal for $s = 1, \ldots, S$.

iii) S blocks of group factors are uncorrelated with the block of global factors.

Sequential least-square estimation of two-level factor model

Let us define the objective function, which is the sum of square residuals (SSR) of model (4.2), as,

$$SSR(F^m, \Lambda^m) = \sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{t=1}^{T}\left(x_{s,it} - \gamma_{s,i}' H_t - \lambda_{s,i}' R_{s,t}\right)^2. \tag{4.11}$$

The approach to the estimation of two-level factor involves the following steps:

Step 1: Find a set of initial estimators for both global and group factors, which we denote by, $\widehat{H}^{(0)} = \left(\widehat{H}_1^{(0)}, \widehat{H}_2^{(0)}, \ldots, \widehat{H}_T^{(0)}\right)$ and $\widehat{R}_s^{(0)} = \left(\widehat{R}_{s,1}^{(0)}, \widehat{R}_{s,2}^{(0)}, \ldots, \widehat{R}_{s,T}^{(0)}\right)$ for all $s = 1, \ldots, S$ respectively.

Breitung and Eickmeier [2016] and Wang [2008] have shown that initial values for two level factors can be consistently estimated using Canonical Correlation Analysis (CCA), Maximum

Likelihood Estimating (MLE), or Principal Components (PC) analysis. Employing a consistent estimator for initial values in the algorithm, we can ensure that the estimation starts in a neighbourhood of the global minimum and it will avoid unnecessary iterations.

Step 2: Estimate the factor loadings of global and group factors using the factor model with known initial factors, defined as,

$$x_{s,it} = \gamma_{s,i}' \widehat{H}_t^{(0)} + \lambda_{s,i}' \widehat{R}_t^{(0)} + \tilde{e}_{s,it}. \tag{4.12}$$

Let the estimators of factor loading be $\hat{\gamma}_{s,i}^{(0)}$ and $\hat{\lambda}_{s,i}^{(0)}$ for S groups. In matrix form, $\hat{\gamma}_s^{(0)} = (\hat{\gamma}_{s,1}^{(0)}, \hat{\gamma}_{s,2}^{(0)}, \ldots, \hat{\gamma}_{s,N_s}^{(0)})'$ and $\hat{\Lambda}_s^{(0)} = \left(\hat{\lambda}_{s,1}^{(0)}, \ldots, \hat{\lambda}_{s,N_s}^{(0)}\right)'$.

Step 3: Construct the matrix of (estimated) factor loadings with block-diagonal matrix for the group factor loadings as follows:

$$\hat{\Lambda}^{m(0)} = \begin{pmatrix} \hat{\Gamma}_1^{(0)} & \hat{\Lambda}_1^{(0)} & 0 & \cdots 0 \\ \hat{\Gamma}_2^{(0)} & 0 & \hat{\Lambda}_2^{(0)} & \cdots 0 \\ \vdots & & \ddots & \vdots \\ \hat{\Gamma}_S^{(0)} & 0 & 0 & \hat{\Lambda}_S^{(0)} \end{pmatrix} \tag{4.13}$$

Step 4: Re-estimate the factors using updated factor loading matrix $\hat{\Lambda}^{m(0)}$.

$$\hat{F}_t^{m(1)} = \begin{pmatrix} \widehat{H}_t^{(1)} \\ \widehat{R}_{1,t}^{(1)} \\ \vdots \\ \widehat{R}_{S,t}^{(1)} \end{pmatrix} = \left(\hat{\Lambda}^{m(0)\prime} \hat{\Lambda}^{m(0)}\right)^{-1} \hat{\Lambda}^{m(0)\prime} X_t. \tag{4.14}$$

Step 5: Replace the factors in (4.12) with the updated factors, and re-estimate the factor loadings.

Step 6: Repeat step 2 to step 5 until the factor and loading matrices converge with a norm 1×10^{-4} (or 1000 iterations).

It was shown that this LS approach consistently estimates the two-level factors and the factor loadings only up to a rotation, Q. See Wang [2008] and Breitung and Eickmeier [2016] for details. The semi-parametric FAR model (4.9) two-level factor structure can be specified as,

$$\begin{aligned}
Y_{t+h} = \theta_t' F_t^m + \omega' W_t + \epsilon_{t+h} &= \theta_t' Q^{-1} \left(Q F_t^m - \hat{F}_t^m + \hat{F}_t^m \right) + \omega' W_t + \epsilon_{t+h} \\
&= \theta_t' Q^{-1} \hat{F}_t^m + \omega' W_t + \theta_t' Q^{-1} \left(Q F_t^m - \hat{F}_t^m \right) + \epsilon_{t+h} \\
&= \theta_t' Q^{-1} \hat{F}_t^m + \omega' W_t + error.
\end{aligned} \tag{4.15}$$

Remark 3: In the empirical application presented in Section 4.3, we forecast GDP, GDP growth rate and inflation. The level 2 factor structure in the two-level factor model depends on the practical problem in hand. In the context of improving the widely studied FAR model with global factors forecasting, we aim to chose a set of global and group factors collectively the information contents in the factors may enrich the FAR model for forecasting a specific macroeconomic variable. To this end, the main question we seek to answer is: what is the optimum number of global factors and the optimum number of group factors should be included in the FAR model for forecasting GDP growth? In Chapter 2, we have eight global factors estimated from the panel of 100 sub-aggregate variables by the PCA method. Of these eight factors, we choose the ones that are highly correlated with GDP growth and also based on the increase in R^2 goodness-of-fit measure by each factor. Additionally, we use the plots of these global factors with GDP growth superimposed and visualization would indicate how well these factors track the behavior of GDP growth. Using this approach, we find the appropriate number of factors to be three. We use these three global factors as the initial values in the rest of the estimation process. Moreover, we choose the groups[1] that contain variables that are drivers of GDP growth, Group 1 with employment and unemployment, Group 2 with various interest rates, and so on. Thus, we chose seven groups; details are given in Section 4.3.

4.2.3 Semi-parametric estimation of FAR model with two-level factor structure

Let us consider the semi-parametric FAR model with two-level factor structure:

$$Y_{t+h} = \theta_t' \hat{F}_t^m + \omega' W_t + \epsilon_{t+h}, \tag{4.16}$$

[1]See the Appendix for a brief description of 12 groups.

where $\hat{F}_t^m$ is the set of estimated global and group factors, and W_t is the set of observable regressors. We estimate the model by non-parametric kernel smoothing method that was developed for models with stationary predictors; see Gao and Hawthorne [2006], Li et al. [2011a], Chen et al. [2012], and Wei and Zhang [2020]. We adapt these methodologies for estimating the coefficients in the semi-parametric FAR model with non-stationary and stationary predictors.

Let us define a weight function $V_T(\tau_t, t)$ of the form,

$$V_T(\tau_t, \tau) = \frac{K\left(\frac{\tau_t - \tau}{h_w}\right)}{\sum_{u=1}^{T} K\left(\frac{\tau_u - \tau}{h_w}\right)}, \tag{4.17}$$

where $K(.)$ is the kernel smoothing function, and h_w is the bandwidth. Generally, the kernel function is a continuous non-negative smoothing function that satisfies the following properties: i) $\int K(u)du = 1$, ii) $\int uK(u)du = 0$, and iii) $\int u^2 k(u)du = \kappa < \infty$. In this paper, we use the Gaussian kernel function defined as,

$$K(u) = \frac{1}{\sqrt{2\pi}} exp^{-u^2/2}, \tag{4.18}$$

for estimating the time-varying factor coefficients in model (4.16), bandwidth h_w satisfies the conditions that $h_w \to 0$ and $Th_w \to \infty$ as $T \to \infty$. Bandwidth selection is a crucial part of non-parametric estimation, and is typically data-driven. Therefore, in our empirical study, we chose the bandwidth accordingly. See, Section 4.3 for details.

The following three-step method is used in the semi-parametric estimation.

Step 1: *Nonparametric estimation of time-varying factor parameters*

For a given ω, we estimate the time-varying coefficients $\theta(\tau_t)$ by minimizing the loss function,

$$\sum_{t=1}^{T-h} \left(Y_{t+h} - \omega' W_t - \theta(\tau)' \hat{F}_t^m\right)^2 K\left(\frac{\tau_t - \tau}{h_w}\right). \tag{4.19}$$

Therefore, the estimated $\theta(\tau)$ is given as,

$$\begin{aligned}\tilde{\theta}(\tau) = &\left(\sum_{t=1}^{T-h} \hat{F}_t^m K\left(\frac{\tau_t - \tau}{h_w}\right) \hat{F}_t^{m'}\right)^{-1} \sum_{t=1}^{T-h} \hat{F}_t^m K\left(\frac{\tau_t - \tau}{h_w}\right) Y_{t+h} \\ &- \left(\sum_{t=1}^{T-h} \hat{F}_t^m K\left(\frac{\tau_t - \tau}{h_w}\right) \hat{F}_t^{m'}\right)^{-1} \sum_{t=1}^{T-h} \hat{F}_t^m K\left(\frac{\tau_t - \tau}{h_w}\right) \omega' W_t.\end{aligned}$$

Step 2: *Estimation constant parameters*

Replace $\theta(\tau_t)$ by its estimator $\tilde{\theta}(\tau_t)$ in (4.16), and obtain,

$$\begin{aligned}
Y_{t+h} &= \tilde{\theta}(\tau_t)'\hat{F}_t^m + \omega' W_t + \epsilon_{t+h} \\
&= \hat{F}_t^{m\prime} \left(\sum_{s=1}^{T-h} \hat{F}_s^m K\left(\frac{\tau_s - \tau_t}{h_w}\right) \hat{F}_s^{m\prime} \right)^{-1} \sum_{s=1}^{T-h} \hat{F}_s^m K\left(\frac{\tau_s - \tau_t}{h_w}\right) \left(Y_{s+h} - \omega' V_s\right) + \omega' W_t + \epsilon_{t+h}.
\end{aligned}$$

Let

$$\tilde{Y}_{t+h} = Y_{t+h} - \hat{F}_t^{m\prime} \left(\sum_{s=1}^{T-h} \hat{F}_s^m K\left(\frac{\tau_s - \tau_t}{h_w}\right) \hat{F}_s^{m\prime} \right)^{-1} \sum_{s=1}^{T-h} \hat{F}_s^m K\left(\frac{\tau_s - \tau_t}{h_w}\right) Y_{s+h}, \tag{4.20}$$

$$\tilde{W}_t = W_t - \hat{F}_t^{m\prime} \left(\sum_{s=1}^{T-h} \hat{F}_s^m K\left(\frac{\tau_s - \tau_t}{h_w}\right) \hat{F}_s^{m\prime} \right)^{-1} \sum_{s=1}^{T-h} \hat{F}_s^m K\left(\frac{\tau_s - \tau_t}{h_w}\right) W_s. \tag{4.21}$$

Then, consider the linear model $\tilde{Y}_{t+h} = \omega' \tilde{W}_t + \epsilon_{t+h}$. The least-square estimate of ω is given by

$$\hat{\omega} = \left(\sum_{t=1}^{T-h} \tilde{W}_t' \tilde{W}_t \right)^{-1} \left(\sum_{t=1}^{T-h} \tilde{W}_t' \tilde{Y}_{t+h} \right)^{-1}. \tag{4.22}$$

Step 3: *Semi-parametric estimation of FAR model*

Replace ω with $\hat{\omega}$ in the equation of $\tilde{\theta}(\tau)$ (infeasible estimate) and, obtain the feasible estimate $\hat{\theta}(\tau)$ as,

$$\hat{\theta}(\tau) = \left(\sum_{t=1}^{T-h} \hat{F}_t^m K\left(\frac{\tau_t - \tau}{h_w}\right) \hat{F}_t^{m\prime} \right)^{-1} \sum_{t=1}^{T-h} \hat{F}_t^m K\left(\frac{\tau_t - \tau}{h_w}\right) \left(Y_{t+h} - \hat{\omega} W_t\right). \tag{4.23}$$

4.2.4 Bandwidth selection

Silverman's (1986) rule-of-thumb bandwidth is defined as $h_w \approx 1.06 \times \hat{\sigma} T^{-1/5}$, where $\hat{\sigma}$ is the standard deviation of the sample of size T. Chen et al. [2012] introduced a correction factor to this bandwidth, which is defined as $h_w \propto T^{-(1+c)/5}$ for $c < 4 - 10/\delta$, where $\delta > 10/3$. Therefore, we consider a range of bandwidths for the optimal bandwidth selection in the non-parametric kernel smoothing estimation of models. Since the FAR models under study consist of a mixture of nonstationary and stationary regressors, we consider three initial values for the bandwidth $h_w \sim T^{-c_1/5}$ with $c_1 = \{0.5, 1.0, 1.5\}$, where T is the number of observations. In a preliminary analysis, we find that, except for $c_1 = 1$, the time-varying coefficient estimates corresponding to the other two bandwidths seem to have high volatility and some boundary issues. Therefore, we consider the Gaussian kernel function together with the bandwidth $h_{opt} \sim T^{-1/5}$ as the initial value, and some points in the neighbourhood of $c_2 \times T^{-1/5}$ with $c_2 \in (0.5, 1.15)$.

By minimizing the asymptotic mean integrated squared error (AMISE) of the time-varying coefficients, we find an optimal bandwidth $h_{opt} = 0.9 \times T^{-1/5} \approx 0.3$. We also considered two bandwidths for the two sets of coefficients, α_t and β_t, of the FAR model as they correspond to nonstationary and stationary factors respectively. However, the lowest AMISE was achieved for the same bandwidth of $h_{opt} = 0.9 \times T^{-1/5}$.

4.3 Empirical application

4.3.1 Data description

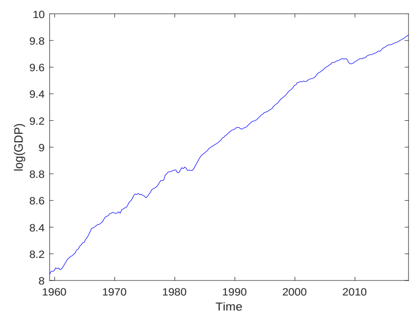
We use the widely studied database FRED-QD, which consists of 240 quarterly macroeconomic time series in the US for the period from 1959:Q1 to 2018:Q4. A balanced panel data of 210 time series over 60 years ($T = 240$) is used in this empirical study, which excludes 36 variables with missing observations. According to the level of aggregation, the data set is categorized into 110 "high-aggregate" variables and 100 "sub-aggregate" variables which are further categorized into 12 groups.[2] We use the panel of 100 sub-aggregates for estimating two-level factors which are used in the FAR model for forecasting three key macroeconomic variables. The method proposed in Chapter 2 showed a mixture of I(0) and I(1) factors can be used for forecasting I(1) response variables such as GDP. The results established in (Phillips [2015]) show that a mixture of I(0) and I(1) series can also be applied for forecasting I(0) response variables such as GDP growth and inflation. Figure 4.1 exhibits the plots of these three series.

4.3.2 Estimation of two-level factors

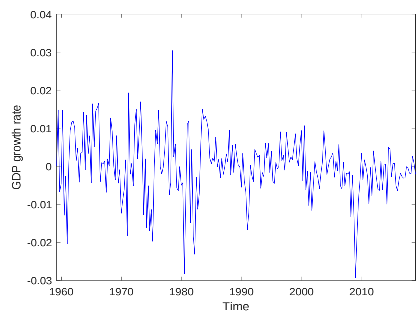
Estimation of level-1 factors

We use both Panel Information criterion (IC) and Integrated Panel Criterion (IPC) introduced by Bai and Ng [2002] and Bai [2004], respectively, for estimating the optimum number of global factors and the factors themselves. These factors were estimated from the panel of 100 variables which are a mixture of I(0) and I(1) variables. The panel criterion selected eight global factors which is the optimal number of factors (See Chapter 2, Section 2.4). They are sufficient to

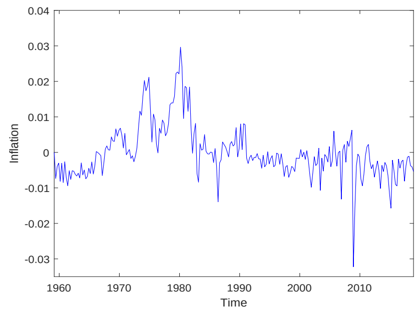
[2] There are 12 groups listed in the updated Appendix of FRED-QD. In this thesis we use only 12 groups and omitted the group listed as "other". *https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QDappendix.pdf*.



(a) log(GDP) - $I(1)$ series



(b) GDP growth rate - $I(0)$ series



(c) Inflation - $I(0)$ series

Figure 4.1: Plots of log(GDP), GDP growth and inflation

explain the large variation (95%) in the panel of 100 sub-aggregate variables. The number of non-stationary factors are identified using IPC and found that it contains four I(1) and four I(0) factors. Figure 4.2 shows the plots of eight factors. The Augmented-Dickey fuller (ADF) test results provide additional evidence that the factors $\{1, 2, 4, 5\}$ are $I(1)$ and the factors $\{3, 6, 7, 8\}$ are $I(0)$ series. The estimates of idiosyncratic errors, $\{\hat{e}_{it}; t = 1, \ldots, T\}$ of all 100 variables are also found to be $I(0)$. The one-level (global) factor model with estimated factors is specified as,

$$x_{it} = \tilde{\lambda}_i^{(1)'} \tilde{E}_t + \tilde{\lambda}_i^{(2)'} \tilde{G}_t + \hat{e}_{it} \qquad (t = 1, \ldots, T, i = 1, \ldots, N),$$

where $\tilde{E}_t$ is the set of four generated non-stationary factors and $\tilde{G}_t$ is the set of four generated stationary factors.

Selection of global factors

As our interest lies in forecasting GDP, GDP growth, and inflation, we use some basic criteria to select the number of global factors that would improve the FAR model for forecasting each of the three response variables. We consider the correlations between the response variable and the estimated eight global factors that appear in Table 4.1. Furthermore, we superimpose the plot of the response variable on those of the global factors and visualize how many factors can closely

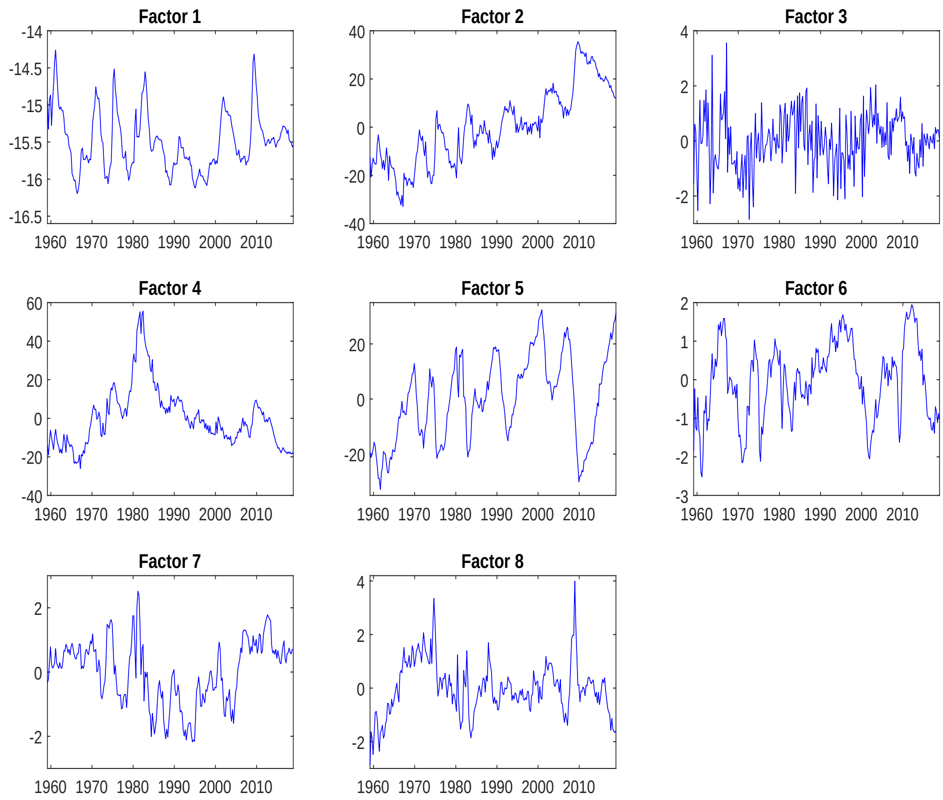


Figure 4.2: Generated eight factors from the panel of 100 variables

Notes: Generated factors 1,2,4 and 5 are I(1) & 3,6,7 and 8 are I(0).

track the response variable. Figures 4.6, 4.7, and 4.8 (in Appendix 4.6.1) present such plots for log(GDP), GDP growth and inflation respectively. For log(GDP), we select global factors 2, 5 and 6; for GDP growth, global factors 1, 7 and 8; and for inflation, global factors 1, 2 and 4. These factors constitute initial values required for the two-level factor estimation. Let us denote the initial global factors by $\widehat{H}_t^{(0)}$.

Estimation of level-2 factors

We consider the two-level factor model, which is defined in model (4.3), of 100 variables in FRED-QD data which consists of 12 groups. For example, group 1 consists of 23 economic variables such as consumption expenditure, exports, imports, and real disposable personal income; and group 2 includes 50 variables related to employment and unemployment. See Table 4.11 for a brief description of the groups.

Table 4.1: Correlation between estimated 8 global factors and the response variables

Response variable	Estimated factors							
	1	2	3	4	5	6	7	8
log(GDP)	-0.13	**0.83**	0.10	-0.06	**0.49**	**0.21**	-0.02	0.05
GDP growth rate	**-0.24**	-0.23	0.10	-0.19	-0.14	0.17	**-0.26**	**-0.29**
Inflation	**-0.21**	**-0.40**	-0.04	**0.51**	0.19	0.09	0.15	0.15

Notes: 1) The values in bold correspond to high correlations. 2) Hence, we select the estimated global factors 2,5, and 6 for log(GDP); 1,7, and, 8 factors for GDP growth rate; and 1, 2, and 4 for inflation in the semi-parametric FAR model.

Following the estimation of the initial global factors $\widehat{H}_t^{(0)}$, we remove effects of these global factors from the panel variables in model (4.2) and then categorize the purged panel data into the 12 groups according to the group code in FRED-QD data set.

Let us denote the estimated group factors by $\widehat{R}_{s,t}^{(0)}$ $(s = 1, \ldots, 12)$. Thus, we have the required initial values for both level-1 and level-2 factors. Let us denote the set of factors by $\hat{F}_t^{(0)m} = \left(\widehat{H}_t^{(0)\prime}, \widehat{R}_{1,t}^{(0)\prime}, \ldots, \widehat{R}_{13,t}^{(0)\prime}\right)^\prime$.

Estimation of two-level factor model

We use the initial estimates of global and group factors obtained in the previous sections and the sequential LS based algorithm outlined in Section 4.2 for estimating the optimum global and group factors from the panel of 100 variables. Let us denote the set of final estimates of the factors as $\hat{F}_t^m = \left(\widehat{H}_t^\prime, \widehat{R}_{1,t}^\prime, \ldots, \widehat{R}_{12t}^\prime\right)^\prime$. ADF test results indicate that the set of estimated group factors is a mixture of $I(0)$ and $I(1)$ series.

Selection of level-2 factors

From the estimated 12 group factors, we select the level-2 factors that would improve the out-of-sample predictability of semi-parametric two-level factor FAR models for forecasting each of the three variables log(GDP), GDP growth and inflation. To select suitable level-2 factors, we use prior knowledge of potential drivers of these three variables. Additionally, we assess the strength of the relationship based on the correlations between these 12 group factors and the variable, say GDP growth, and the increase in R^2 in the goodness of fit of the FAR model when each group factor is added. The results are presented in Tables 4.2 and 4.3, respectively.

The selected group factors for the three variables are:

Table 4.2: Correlation of the estimated 12 group specific factors and the response variables

Response variable	Group specific factors											
	1	2	3	4	5	6	7	8	9	10	11	12
log(GDP)	**-0.17**	-0.10	-0.11	-0.09	**-0.14**	-0.09	**-0.15**	0.05	**-0.17**	**-0.16**	-0.07	**-0.11**
GDP growth rate	0.14	**-0.27**	-0.16	-0.21	**-0.28**	**-0.27**	**-0.27**	-0.04	**-0.24**	**-0.26**	-0.20	-0.18
Inflation	-0.05	0.07	**0.18**	0.06	**0.16**	0.13	0.12	**0.22**	**0.26**	**0.14**	**0.18**	0.11

Notes: 1) The values in bold correspond high correlations. 2) These highly correlated groups specific factors are taken into account in choosing the level-2 factors.

Table 4.3: In-sample performance of the original semi-parametric two-level FAR model with three global factors and one group factor, in terms of R^2 for the three response variables

Response variable	Group specific factors											
	1	2	3	4	5	6	7	8	9	10	11	12
log(GDP)	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.998
GDP growth rate	0.22	0.21	0.21	**0.26**	0.22	**0.23**	**0.23**	**0.25**	0.22	0.21	**0.24**	0.21
Inflation	**0.65**	0.65	**0.66**	0.65	0.65	0.65	0.65	**0.66**	**0.66**	0.65	**0.66**	0.65

Notes: 1) The semi-parametric FAR model (4.16) augmented with AR(4) is considered with generated three global factors selected from the Table 4.1 and one group specific factor. 2) The figures for log(GDP) indicate that irrespective to the selection of level 2 factors, the model has good in-sample performance with level 1 factors. 3) Highlighted figures are corresponding to the highest R^2 values. 4) We use this information, among others, in the selection of group factors for each response variable.

- Group factors for log(GDP): 2, 3, 4, 5, 7, 9 and 12

- Group factors for GDP growth: the same group factors as for log(GDP)

- Group factors for Inflation: 3, 4, 6, 8, 9, 10, and 11

The two-level factor model can be specified as:

$$x_{s,it} = \hat{\gamma}'_{s,i}\widehat{H}_t + \hat{\lambda}'_{s,i}\widehat{R}_{s,t} + \hat{e}_{s,it} \qquad (i = 1, \ldots, N_s, s = 1, \ldots, S, t = 1, \ldots, T).$$

Clearly, two-level factors are a mixture of I(1) and I(0) series. Therefore, we classify the factors into two sets: $\hat{E}_t^m$ is the set of non-stationary global and group factors, whereas $\hat{G}_t^m$ is the set of stationary global and group factors.

Hence, the fitted two-level factor model and the semi-parametric two-level factor FAR model can be given as,

$$X_t = \hat{\Lambda}^{(1)m}\hat{E}_t^m + \hat{\Lambda}^{(2)m}\hat{G}_t^m + \hat{e}_t, \tag{4.24}$$

$$Y_{t+h} = \alpha_t'\hat{E}_t^m + \beta_t'\hat{G}_t^m + \omega_1 Y_t + \omega_2 Y_{t-1} + \omega_3 Y_{t-2} + \omega_4 Y_{t-3} + \epsilon_{t+h} \quad (h > 0). \tag{4.25}$$

4.3.3 Forecasting macroeconomic variables

Model selection criteria

We compare the in-sample predictive performance of the semi-parametric two-level FAR model using the sum of square errors (in-sample SSE). In order to evaluate the out-of-sample predictability of the models relative to the benchmark model, we use the out-of-sample R-square (R^2_{os}) defined as:

$$R^2_{os} = 1 - \left(\sum_{t=T_1+1}^{T}(Y_t - \hat{Y}_t)^2\right)\left(\sum_{t=T_1+1}^{T}(Y_t - \tilde{Y}_t)^2\right)^{-1}, \tag{4.26}$$

where $\hat{Y}_t$ is the prediction of Y_t by the semi-parametric two-level FAR, $\tilde{Y}_t$ is the prediction of Y_t by the benchmark model, and T_1 is the initial in-sample size for the expanding widow method. Since we consider one-step ahead forecasting, the $T_1 + j$ observations are used for estimation of the models and one-step ahead predictions are made at $T_1 + j + 1$ ($j = 0, \ldots, T - (T_1 + 1)$).

FAR and time series model specifications

We evaluate the in-sample and out-of-sample predictability of parametric and semi-parametric two-level factor FAR models. For the purpose of forecast comparison, we consider both one-level FAR and two-level FAR models. Note that, in Chapter 3, we proposed semi-parametric one-level FAR model and assessed the accuracies of the models in forecasting the same three variables, GDP, GDP growth and inflation. In this Chapter, we will assess accuracy of the improved semi-parametric two-level factor FAR in comparison to the one-level factor FAR counterparts and some time series models.

Consider the model specifications given as:

$$
\begin{aligned}
\text{Model 1 } &: Y_{t+1} = \omega Y_t + \epsilon_{t+1},\\
\text{Model 2 } &: Y_{t+1} = \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+1},\\
\text{Model 3 } &: Y_{t+1} = \alpha' \hat{E}_t^m + \beta' \hat{G}_t^m + \omega Y_t + \epsilon_{t+1},\\
\text{Model 4 } &: Y_{t+1} = \alpha' \hat{E}_t^m + \beta' \hat{G}_t^m + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+1},\\
\text{Model 5 } &: Y_{t+1} = \alpha_t' \hat{E}_t^m + \beta_t' \hat{G}_t^m + \omega Y_t + \epsilon_{t+1},\\
\text{Model 6 } &: Y_{t+1} = \alpha_t' \hat{E}_t^m + \beta_t' \hat{G}_t^m + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+1},\\
\text{Model 7 } &: Y_{t+1} = \alpha' \hat{E}_t^m + \beta_t' \hat{G}_t^m + \omega Y_t + \epsilon_{t+1},\\
\text{Model 8 } &: Y_{t+1} = \alpha' \hat{E}_t^m + \beta_t' \hat{G}_t^m + \sum_{i=0}^{3} \omega_{1+i} Y_{t-i} + \epsilon_{t+1},\\
\text{Model 9 } &: Y_{t+1} = c + \epsilon_{t+1},
\end{aligned}
$$

where $\hat{E}_t^m, \hat{G}_t^m$ are the sets of nonstationary and stationary two-level factors, α and α_t are the sets of coefficients of nonstationary factors, β and β_t are the coefficients of stationary factors, and ω is the set of coefficients of lagged dynamics.

Models 1 and 2 are the basic autoregressive models, Models 3 and 4 are the two-level FAR models with constant parameters, Models 5–8 are two-level FAR models with time-varying parameters, and Model 9 is the simple mean model. Moreover, we include economic policy uncertainty index (EUI) in the models (3)–(8) as a predictor and consider the models with and without EUI as a predictor.

Remark 4: Recall that the FAR models 3-8 include both global (pervasive) factors and group (non-pervasive) factors. For comparison purpose, we also consider the same FAR model specifications 3–8 but only with global factors, which we studied in the previous Chapter. The FAR models with global factors with constant and/or time varying parameters are not explicitly specified here to save space. In these specifications, coefficients $(\hat{E}_t^m, \hat{G}_t^m)$ in Models 3–8 are replaced by $(\tilde{E}_t, \tilde{G}_t)$ respectively.

In-sample predictability of the models

The semi-parametric two-level FAR models given in Models 3–8 were estimated by the Gaussian kernel method with a bandwidth of $h_{opt} = 0.3$. We assessed the in-sample fit of the models using the sum of squares of errors (SSE) and the results are reported in Table 4.4. The values in columns 2 and 3 are SSEs of models for log(GDP) without and with EUI in the model, respectively. Clearly, Model 6 with EUI is the winner in terms of in-sample predictability of log(GDP) relative to other models.

Table 4.4: In-sample performance of the two-level FAR models, in terms of sum of squared errors, SSE

	log(GDP)		GDP growth rate		Inflation	
EUI	0	1	0	1	0	1
Model 3	10.96	10.89	10.75	10.74	4.20	4.13
Model 4	10.50	10.45	10.32	10.32	3.94	3.87
Model 5	10.14	10.06	9.78	9.76	4.08	4.02
Model 6	9.71	**9.64**	9.45	**9.44**	3.84	3.76
Model 7	10.95	10.89	10.40	10.39	4.11	4.04
Model 8	10.49	10.44	10.00	10.00	3.84	**3.76**

Notes: 1) All the values are multiplied by 1000. 2) Models 1 to 8 are defined in Section 4.3. 3) Models 1-2 are basic AR models, models 3-4 are parametric FAR, and models 5-8 are semiparametric FAR models. 3) "1" indicates that presence of EUI in the model and "0" indicate absence of EUI. 4) The figures in bold indicates the least SSE. 5) EUI variable has consistently improped the in-sample predictability for all the model specifications.

Furthermore, for comparison, we estimated the semi-parametric one-level FAR Models 3–8 and the results are reported in Table 4.7 (in Appendix 4.6.2). Again the modified Model 6, which is the semi-parametric one-level FAR, outperforms the other one-level FAR models for log(GDP). By contrast, for log(GDP) and the two I(0) variables GDP growth and inflation, the semi-parametric two-level FAR models perform better than the one-level FAR counterparts in in-sample predictability. For GDP growth, Model 6 (semi-parametric two-level FAR) outperforms the other models, while Model 8 is the winner for inflation relative to others.

4.4 Evaluation of out-of-sample predictability of models

4.4.1 Forecasts of log(GDP), GDP growth and inflation

To assess the out-of-sample predictability of the models, we generate one-step ahead point forecasts from Models 3–8 and compute R^2_{os} (defined in (4.26)) against the benchmark models;

when forecasting I(1) variables the benchmark is Model 2 and when forecasting I(0) variables the benchmark is Model 2 or Model 9.

In the expanding (recursive) window sampling design, we consider four sets of initial estimation window sizes to evaluate the one-step ahead out-of-sample forecasts of models. The one-step ahead forecasts were generated by expanding the end of the sample (window) period by one quarter. The four initial window sizes used in the assessment of out-of-sample forecasts are defined in Table 4.5.

Table 4.5: Initial window sizes and out-of-sample forecasting time periods

T_i	First estimation period	Out-of-sample period
T_1	1959:Q1 - 1998:Q4	1999:Q1 - 2018:Q4
T_2	1959:Q1 - 2003:Q4	2004:Q1 - 2018:Q4
T_3	1959:Q1 - 2008:Q4	2009:Q1 - 2018:Q4
T_4	1959:Q1 - 2013:Q4	2014:Q1 - 2018:Q4

GDP forecast evaluation

The out-of-sample predictive performance of semi-parametric two-level FAR Models 3–8 for log(GDP), GDP growth rate and inflation were assessed relative to Model 2 in terms of R^2_{os}. The results are reported in Table 4.6. The values in the first panel show that the semi-parametric two-level Models 3–7 do not perform as well as the AR(4) model in forecasting log(GDP). The best performer is Model 8 (without EUI) which has the only positive R^2_{os}, while they are negative for other models.

Furthermore, we estimated the semi-parametric one-level FAR Models 3–8. The calculated R^2_{os} are reported in Table 4.8 (in Appendix 4.6.2). In contrast to the findings of two-level FAR models, semi-parametric one-level FAR models have better performance against AR(4) model for the initial expanding window sizes T_3 and T_4. The results in the first panel indicate that the Model 3 performs the best in the out-of-sample prediction of log(GDP) for the window size T_4. The forecast performance of the models appear to be sensitive to initial window size.

We use Models 3, 5 and 7 (with one-level factors only) to generate one-step-ahead forecasts of log(GDP) for the window size T_1. Plots of the forecasts along with the observed log(GDP) series in Figure 4.3 indicate that these forecasts are capture the trend of log(GDP) and fluctuations around that trend reasonably well.

Table 4.6: Out-of-sample expanding window forecasting performance of the two-level FAR models, in terms of R^2_{os}

Models	Without EUI				With EUI			
	T1	T2	T3	T4	T1	T2	T3	T4
	log(GDP)							
Model 3	-0.164	-0.234	-0.278	-1.068	-0.222	-0.284	-0.282	-0.709
Model 4	-0.239	-0.323	-0.269	-1.257	-0.293	-0.369	-0.271	-0.849
Model 5	-0.107	-0.143	-0.168	-0.049	-0.143	-0.167	-0.215	-0.030
Model 6	-0.149	-0.215	-0.169	-0.088	-0.185	-0.244	-0.212	-0.055
Model 7	0.010	-0.031	-0.279	-1.443	-0.046	-0.046	-0.300	-1.097
Model 8	**0.028**	-0.017	-0.193	-0.923	-0.038	-0.050	-0.237	-0.751
	GDP growth rate							
Model 3	0.036	0.130	-0.058	-0.740	0.042	0.148	-0.031	-0.726
Model 4	0.026	0.096	-0.077	-1.076	0.029	0.098	-0.058	-1.057
Model 5	0.094	0.121	0.053	-0.102	0.098	0.128	0.082	-0.104
Model 6	0.088	0.104	0.071	-0.143	0.097	0.110	0.095	-0.137
Model 7	0.166	0.211	0.075	**0.366**	0.172	0.219	0.097	0.331
Model 8	0.170	0.185	0.080	0.266	0.172	0.191	0.110	0.245
	Inflation							
Model 3	0.210	0.219	0.212	0.230	0.220	0.228	0.219	0.196
Model 4	0.071	0.078	0.115	0.237	0.094	0.103	0.071	0.188
Model 5	0.140	0.140	0.160	**0.251**	0.150	0.152	0.173	0.210
Model 6	0.001	-0.005	0.017	0.212	0.023	0.021	-0.013	0.146
Model 7	-0.021	-0.017	0.091	0.124	-0.015	-0.010	0.078	0.100
Model 8	-0.141	-0.154	-0.027	0.156	-0.119	-0.127	-0.094	0.106

Notes: 1) All the R^2_{os} values are calculated by considering AR(4) as the bench-mark model. 2) T_1, T_2, T_3, and T_4 are the first in-sample periods as stated in Table 4.5. 3) Model specifications 3 to 8 are defined in the previous section. 4) Highlighted values provide the models that have highest R^2_{os} values for each response variable for all four different in-sample periods.

GDP growth forecast evaluation

The semi-parametric two-level FAR model with various specifications given in Models 3–8 for GDP growth were estimated by the Gaussian kernel with the bandwidth, $h_{opt} = 0.3$. The out-of-sample predictive performance of Models 3–8 for GDP growth were assessed against the benchmark AR(4) Model 2 in terms of out-of-sample R^2_{os}. The results in the second panel of Table 4.6 show that overall, the semi-parametric two-level FAR Models 5–8 perform better than the AR(4) model across all four window sizes and that the best performer is Model 7 (without EUI) with the highest R^2_{os} relative to other models.

Further analysis of these results for GDP growth indicate that, for smaller window sizes T_1 and T_2, the semi-parametric two-level FAR models with EUI outperform the models without

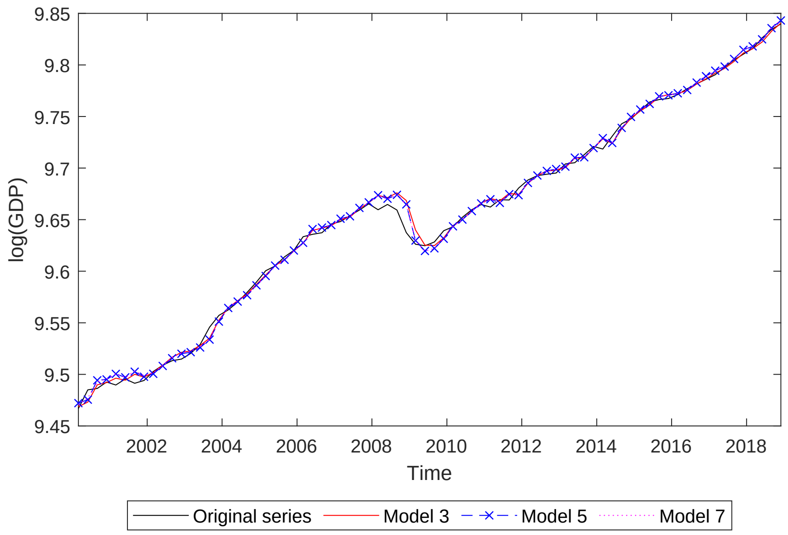


Figure 4.3: The observed log(GDP), and plots of one-step ahead out-of-sample forecasts of log(GDP) for 1999:Q1- 2018:Q4. Red solid line: predicted series with Model 3. Blue crosses: predicted series with Model 5. Magenta dotted line: predicted series with Model 7. Black solid line: observed data.

EUI. This observation is consistent across all the Models 3–8. Furthermore, the semi-parametric one-level Models 3–8 are estimated, and the calculated R^2_{os} are reported in Table 4.8 (in Appendix 4.6.2). We generated one-step-ahead GDP growth forecasts with Models 3, 5 and 7 for the window size T_1 and the plots along with GDP growth are shown in Figure 4.4. They indicate some important features of the semi-parametric two-level FAR models with EUI in their out-of-sample predictive performance. For T_1 (1959:Q1-1998:Q4) and the out-of-sample period 1999:Q1-2018:Q4, Model 5 with all factor parameters time-varying and EUI performs the best in comparison with the other models. This model's forecasts clearly capture the downturn in 2003 due to war and the deep downturn in GDP growth the US experienced during the GFC.

Inflation forecast evaluation

The semi-parametric two-level FAR model with various specifications given in Models 3–8 were estimated by the Gaussian kernel method with bandwidth, $h_{opt} = 0.3$ for the four initial expanding window sizes listed in Table 4.5. The out-of-sample predictive performance of Models 3–8 for inflation were assessed relative to Model 2 in terms of the R^2_{os} measure. The results reported in the third panel of Table 4.6 show that overall, the semi-parametric two-level Models 3-8 perform better than the AR(4) model across all four window sizes, with Model 5 being the

Figure 4.4: The observed GDP growth rate, and plots of one-step ahead out-of-sample forecasts of GDP growth rate for 1999:Q1- 2018:Q4. Red solid line: predicted series with Model 3. Blue crosses: predicted series with Model 5. Magenta dotted line: predicted series with Model 7. Black solid line: observed data.

best performer. Moreover, the results in Table 4.6 indicate that the best performer is Model 5 (without EUI) with the highest R^2_{os} relative to other models.

Further analysis of the results for inflation forecasts indicate that, for smaller window sizes T_1 and T_2, the semi-parametric two-level FAR models with EUI outperforms the model without EUI. This observation is consistent across Models 3–8. These observations are very similar to what was observed for GDP growth forecasts. We generated one-step-ahead inflation forecasts with Models 3, 5 and 7 for the window size T_1. The plots along with observed inflation in Figure 4.5 indicate some important features of the semi-parametric two-level FAR Model 7 with EUI and its out-of-sample forecast performance. Specifically, Model 7 with time-varying coefficients of stationary factors, and with EUI performs the best in comparison with the other models. This Model's forecasts clearly capture the low inflation level experienced in the US caused by the GFC.

4.4.2 Sensitivity analysis

Sensitivity analysis of expanding window size

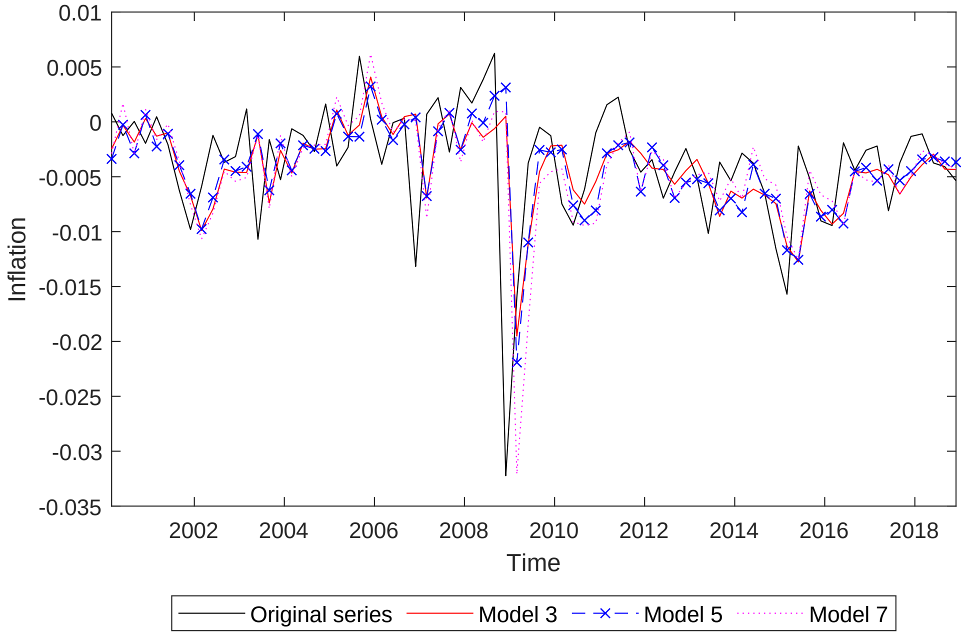


Figure 4.5: The observed inflation, and plots of one-step ahead out-of-sample forecasts of inflation for 1999:Q1- 2018:Q4. Red solid line: predicted series with Model 3. Blue crosses: predicted series with Model 5. Magenta dotted line: predicted series with Model 7. Black solid line: observed data.

The R^2_{os} measures in Table 4.6 indicate that there is no single FAR model that outperforms the rest of the models across all four initial expanding window sizes. That is, the out-of-sample forecast performance of the models is sensitive to window size. To further understand the nature of the sensitivity to the initial expanding window size, we consider 61 different initial window sizes, T_i ($i = 1, \ldots, 61$). We start with a first window size of T_1 with 40 years, from 1959:Q1 to 1999:Q4 and then extend the end of sample period by one quarter until we reach the final expanding window size of 55 years from 1959:Q1 to 2013:Q4. Thus, we consider 61 different window sizes that provide initial estimation periods.

In this sensitivity analysis, we examine the effect of the initial expanding window size on the out-of-sample predictability of the models under investigation. The plots of out-of-sample R^2_{os} values for the Model 3–8 appear in Figures 4.9, 4.10, 4.11 (in Appendix 4.6.3). They show that, for forecasting GDP with smaller window sizes, the semi-parametric FAR models perform better than the other models, and with the large window sizes, the parametric models outperform the semi-parametric counterparts. For forecasting GDP growth rate with smaller window sizes, on the other hand, the two-level semi-parametric FAR models perform better than other models and with large window sizes one-level semi-parametric models outperform other models. The two-level semi-parametric FAR models show better forecastability throughout the different in-sample periods for forecasting Inflation.

Comparison of expanding window and rolling window schemes

Albeit limited, we generated out of sample forecasts with one-level FAR models using the rolling window sampling scheme with the initial window size of 40 years from 1959:Q1-1998:Q4. The rolling window size of 160 quarters moved forward by one quarter until the end of the sampling period is 2013:Q4. The out-of-sample R^2_{os} values of forecasts of log (GDP), GDP growth and inflation are reported in Table 4.9 (in Appendix 4.6.3). We have used two more rolling window sizes of 45 years and 50 years. Generated forecasts and R^2_{os} values show that the forecast performance of these model declined for the two larger window sizes (see Table 4.9). The out-of-sample R^2_{os} values show that the expanding window sampling scheme generates better forecasts relative to the rolling window sampling scheme. Our finding is in contrast to recent findings of several studies (mentioned in the introduction) which find that the rolling widow scheme is better than the expanding (recursive) window scheme for generating out-of-sample forecasts.

However, our findings are not surprising because previous studies mostly used stationary FAR models and time series models and generated forecasts of stationary time series. The models used in this thesis are nonlinear FAR models, which include three distinct features: the generated factors in the FAR model are a mixture of I(1) and I(0) variables; the factors can be two-level factors; and the factor parameters are time varying. To have a deep understanding on how the two sampling scheme will work for models, a large scale simulation study and methodological developments along the line of Inoue et al. [2017] are required. We leave them for the future research.

Remark 5: As discussed in Chapter 3, there maybe an over-fitting problem when the semi-parametric two-level FAR model is used for forecasting. However, the semi-parametric model that we use is linear in regressors and the coefficients of the factors vary over time. Thus, although present, the over-fitting problem in this case may not be as serious as in fully nonparametric models. To circumvent the over-fitting problem, various model specifications, different in-sample sizes to estimate the models in both recursive window and rolling window setting in order to generate out-of-sample predictions. Based on these out-of sample predictions, we have selected the best model that predict the response variable well relative to other models.

4.5 Conclusion

This Chapter proposes a new semi-parametric two-level factor structure to the widely studied FAR model for forecasting univariate time series GDP, GDP growth and inflation. We relax three main assumptions that underlie the standard one-level (global) FAR model. The features of the improved method includes two-level factor structure, factors are a mixture of I(0) and I(1) series and the parameters of factors are time-varying.

There are several stages involved in estimation of the new models for forecasting macroeconomic variables. First, we setup a two-level factor data model, which include global (pervasive) factors and group (non-pervasive) factors. We consider 12 groups of economic variables defined in Stock and Watson [2002a] and include them in the two-level factor model through block restrictions. The two-level factor model is estimated by the sequential algorithm-based LS method proposed by Breitung and Eickmeier [2016]. An advantage of this method is that a set of global (level-1) factors and group (level-2) factors can be selected as predictors in the FAR model for desired macroeconomic variables. Such a FAR model is likely to generate more accurate forecasts relative to its competitors.

To further improve out-of-sample predictability, we include the economic policy uncertainty index (EUI) in the semi-parametric two-level FAR model, which is known to reduce the instability in forecast errors. The proposed semi-parametric two-level factor FAR model generates better forecasts for the two stationary variables, GDP growth and inflation, relative to its competitors. By contrast, the semi-parametric one-level factor FAR model generates better forecasts for the non-stationary variable, GDP, in comparison to other models.

We find that models' forecast performance largely depend on the sampling scheme used in the forecast evaluation. The expanding window works better than the rolling window scheme for the semi-parametric one-level FAR model forecasts of log(GDP), while rolling window works better than the expanding window for the proposed semi-parametric two-level FAR model forecasts of GDP growth and inflation. Further research is required to assess the validity of the results of the comparison between the two sampling schemes, which will be undertaken in future research.

4.6 Appendices

4.6.1 Appendix A: Initial Global Factors

Figures 4.6– 4.8 report the data visualizations used in the selection of global factors for the FAR specifications.

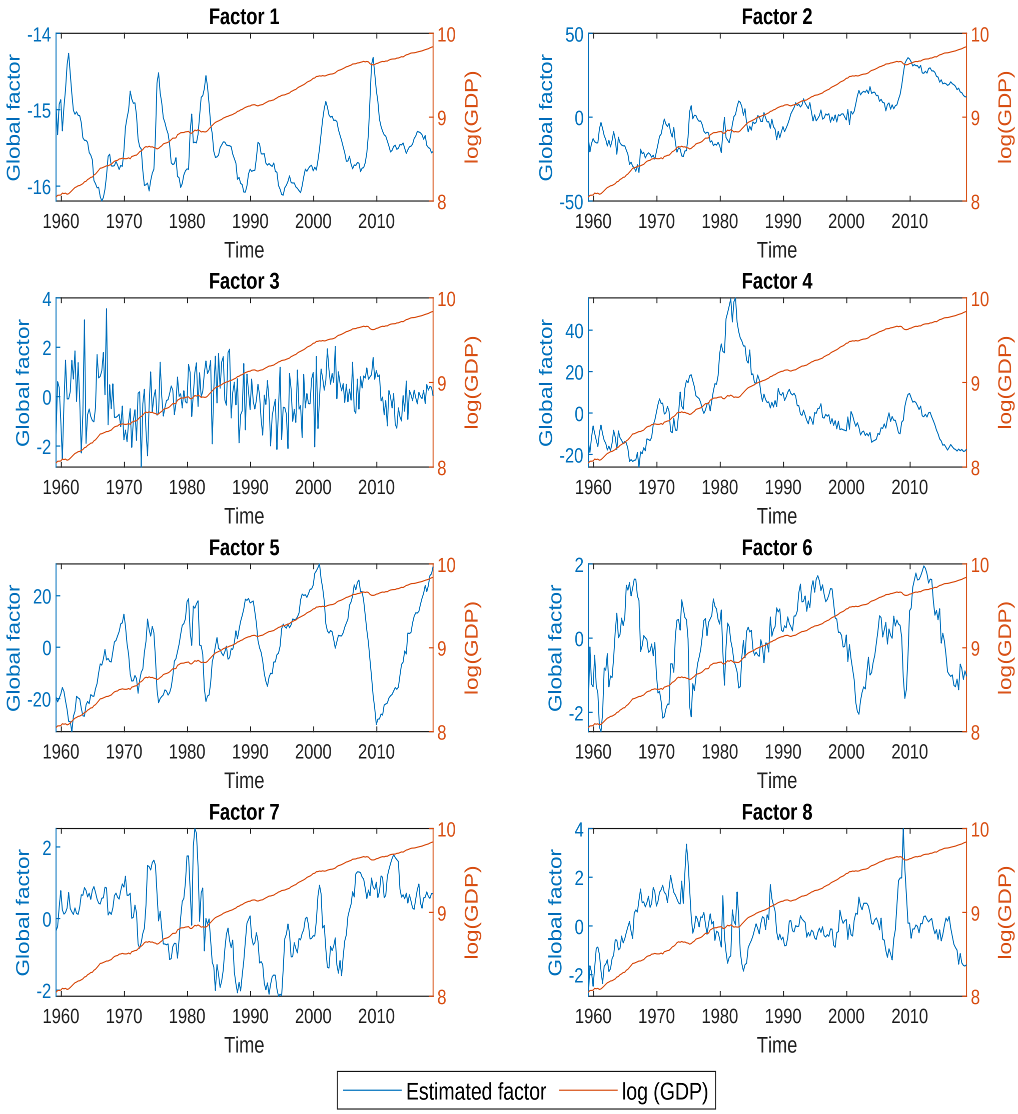


Figure 4.6: Plot of the generated 8 global factors and observed log(GDP) series

Notes: 1)The blue solid lines are corresponding to the estimated global factors and orange line is corresponding to the observed response series. 2) Using these plots, we visualize well the response variables can tract these factors. 3)These plots were exploited along with the correlation measures to select the number global factors for each response variable, which in turn used as the initial values required for the estimation of the two-level factor model.

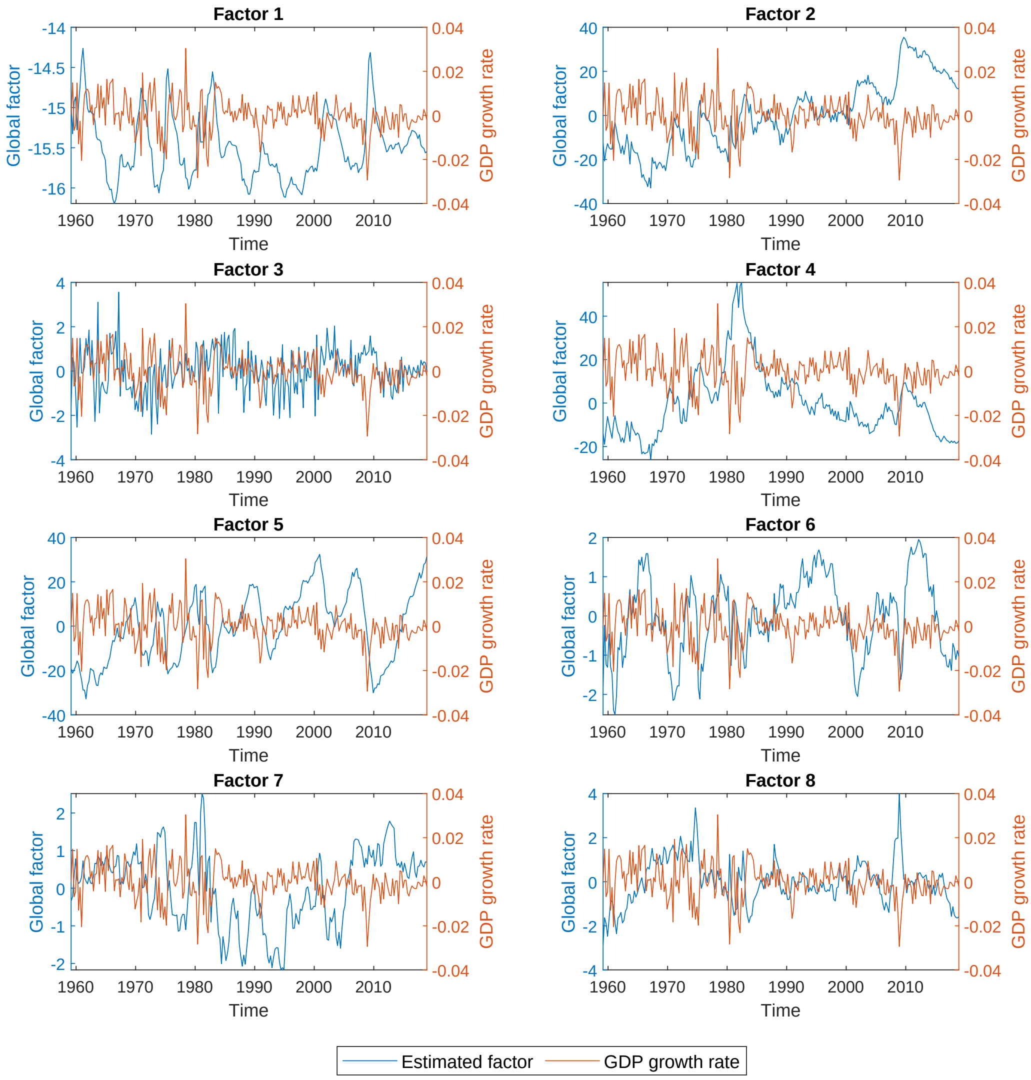


Figure 4.7: Plot of the generated 8 global factors and observed GDP growth rate series

Notes: 1)The blue solid lines are corresponding to the estimated global factors and orange line is corresponding to the observed response series. 2) Using these plots, we visualize well the response variables can tract these factors. 3)These plots were exploited along with the correlation measures to select the number global factors for each response variable, which in turn used as the initial values required for the estimation of the two-level factor model.

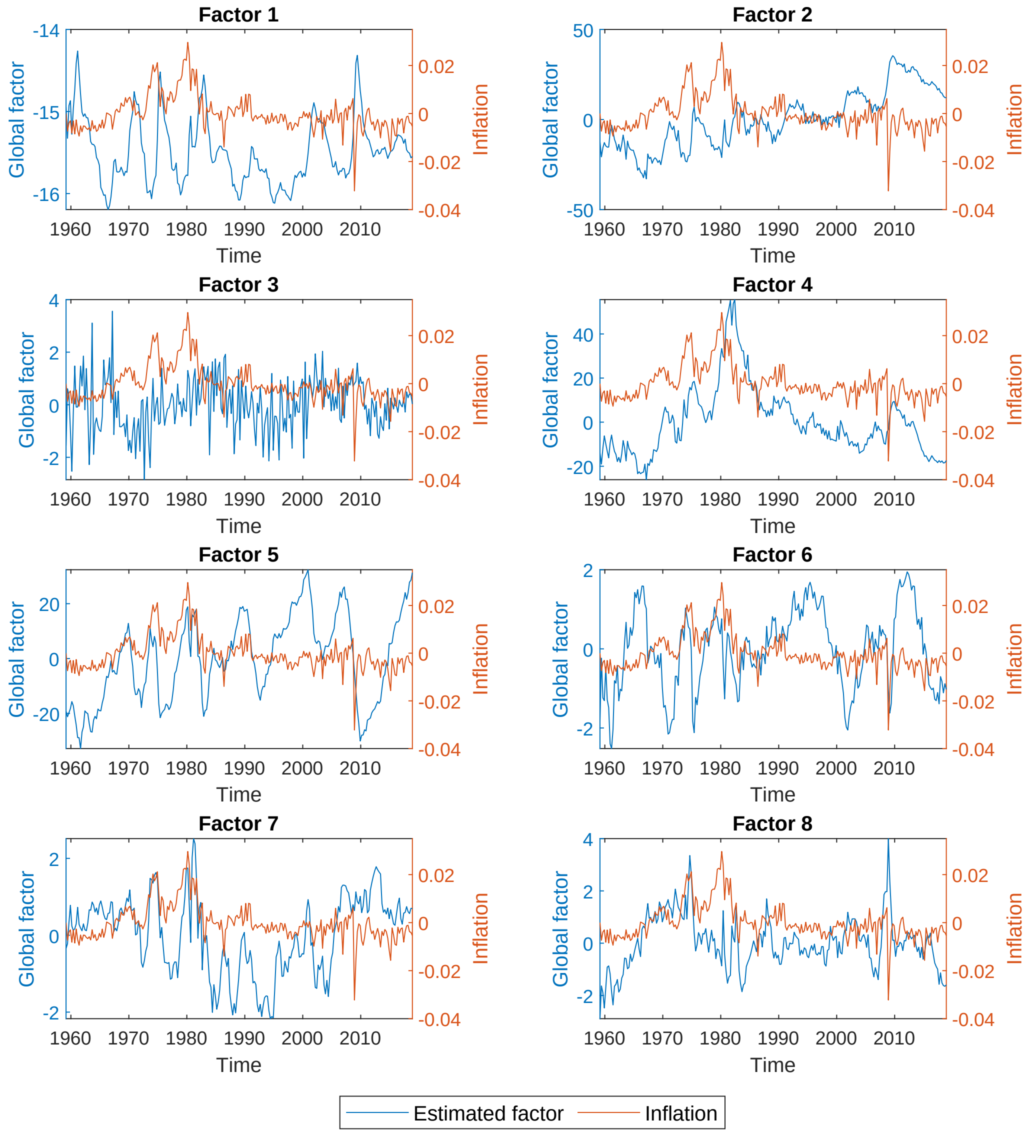


Figure 4.8: Plot of the generated 8 global factors and observed inflation series

Notes: 1)The blue solid lines are corresponding to the estimated global factors and orange line is corresponding to the observed response series. 2) Using these plots, we visualize well the response variables can tract these factors. 3)These plots were exploited along with the correlation measures to select the number global factors for each response variable, which in turn used as the initial values required for the estimation of the two-level factor model.

4.6.2 Appendix B: Results for Semiparametric One-level FAR model

In-sample forecasting performance

Table 4.7: In-sample performance of the one-level FAR models, in terms of sum of squared errors, SSE

	log(GDP)		GDP growth rate		Inflation	
EUI	0	1	0	1	0	1
Model 1	16.46	15.39	14.30	14.67	5.70	5.67
Model 2	13.91	13.65	13.80	13.73	5.00	4.91
Model 3	12.14	12.12	12.07	12.04	5.16	5.08
Model 4	11.33	11.32	11.45	11.43	4.76	4.67
Model 5	11.22	11.21	11.33	11.32	4.90	4.85
Model 6	10.67	**10.67**	10.83	**10.82**	4.49	**4.42**
Model 7	11.35	11.34	11.53	11.52	5.03	4.96
Model 8	10.79	10.78	11.03	11.02	4.60	4.53

Notes: 1) All the values are multiplied by 1000. 2) Models 1 to 8 are defined in Section 3.3. 3) Models 1-2 are basic AR models, models 3-4 are parametric FAR, and models 5-8 are semiparametric FAR models. 3) "1" indicates that presence of EUI in the model and "0" indicate absence of EUI. 4) The figures in bold indicates the least SSE. 5) EUI variable has consistently improped the in-sample predictability for all the model specifications.

Out-of-sample forecasting performance

Table 4.8: Out-of-sample expanding window forecasting performance of the one-level FAR models, in terms of R^2_{os}

Models	Without EUI				With EUI			
	T1	T2	T3	T4	T1	T2	T3	T4
	log(GDP)							
Model 3	-0.063	-0.145	0.297	0.455	-0.065	-0.155	0.292	**0.498**
Model 4	0.030	-0.029	0.215	0.352	0.005	-0.058	0.208	0.436
Model 5	-0.007	0.026	0.253	0.189	0.028	0.044	0.233	0.305
Model 6	0.042	0.068	0.191	0.076	0.043	0.059	0.177	0.197
Model 7	0.019	0.042	0.242	0.171	0.004	0.029	0.243	0.322
Model 8	0.062	0.080	0.187	0.040	0.026	0.049	0.181	0.226
	GDP growth rate							
Model 3	-0.038	0.028	0.160	**0.416**	-0.042	0.026	0.171	0.403
Model 4	-0.065	0.014	0.115	0.327	-0.074	0.008	0.131	0.321
Model 5	-0.020	0.160	0.188	0.297	-0.061	0.158	0.170	0.317
Model 6	0.005	0.146	0.146	0.202	-0.056	0.131	0.134	0.220
Model 7	-0.050	0.160	0.159	0.334	-0.029	0.158	0.194	0.284
Model 8	-0.039	0.139	0.119	0.230	-0.011	0.139	0.158	0.195
	Inflation							
Model 3	0.023	0.030	0.161	0.001	0.019	0.023	0.179	0.010
Model 4	-0.135	-0.141	0.056	0.026	-0.124	-0.129	0.025	0.008
Model 5	0.007	0.017	0.169	0.107	-0.031	-0.029	0.170	0.067
Model 6	-0.119	-0.124	-0.045	0.058	-0.164	-0.172	-0.002	0.027
Model 7	-0.032	-0.028	0.139	0.064	-0.001	0.007	**0.208**	0.108
Model 8	-0.184	-0.193	0.000	0.048	-0.115	-0.119	-0.018	0.037

Notes: 1) All the R^2_{os} values are calculated by considering AR(4) as the bench-mark model. 2) T_1, T_2, T_3, and T_4 are the first in-sample periods as stated in Table 4.5. 3) Model specifications 3 to 8 are defined in the previous section. 4) Highlighted values provide the models that have highest R^2_{os} values for each response variable for all four different in-sample periods.

4.6.3 Appendix C: Results for Sensitivity Analysis

Rolling window forecasting

Table 4.9: Out-of-sample rolling window forecasting performance of the one-level FAR models, in terms of R^2_{os}

	log(GDP			GDP growth rate			Inflation		
Window size(years)	40	45	50	40	45	50	40	45	50
Model 3	0.027	-0.081	-0.275	**0.043**	-0.090	-0.220	-0.050	-0.008	-0.002
Model 4	**0.233**	0.226	0.194	0.097	-0.052	-0.151	-0.047	0.018	-0.248
Model 5	-0.030	0.002	0.007	-0.049	-0.005	0.010	**0.018**	-0.005	-0.078
Model 6	-0.053	-0.032	-0.018	-0.044	-0.065	-0.009	-0.136	-0.147	-0.839
Model 7	0.054	-0.021	-0.045	-0.100	-0.035	-0.031	-0.002	-0.069	-0.163
Model 8	-0.011	-0.043	-0.061	-0.078	-0.088	-0.049	-0.153	-0.226	-1.029

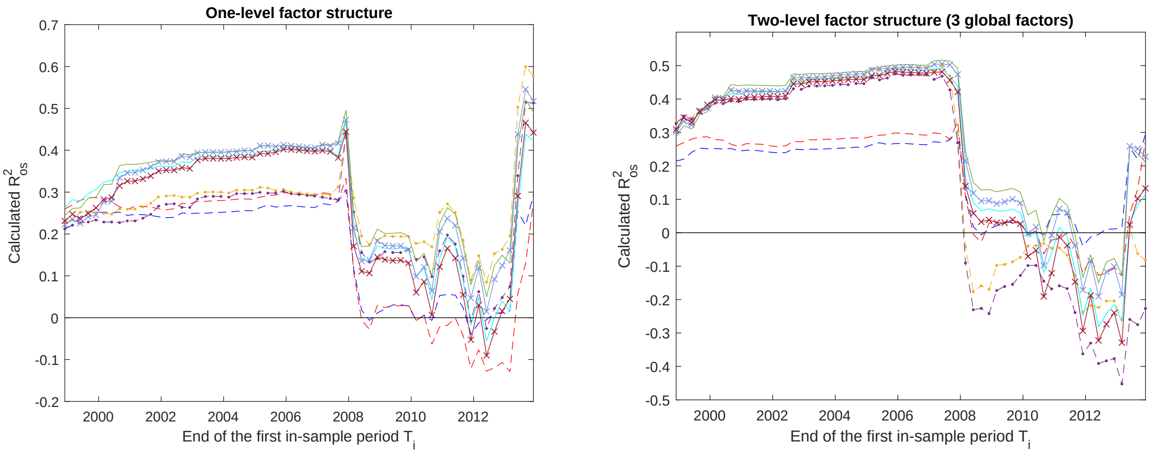
Notes: 1) One-step ahead forecasts are generated with rolling window sampling scheme with the 40,45 and 50 years of window sizes. 2) All the R^2_{os} values are calculated by considering AR(4) as the bench-mark model. 3) Model specifications 3 to 8 are defined in Section 4.3.2. 4) Highlighted values provide the models that have highest R^2_{os} values for each response variable for all four different in-sample periods.

Table 4.10: Out-of-sample rolling window forecasting performance of the two-level FAR models, in terms of R^2_{os}

	log(GDP			GDP growth rate			Inflation		
Window size(years)	40	45	50	40	45	50	40	45	50
Model 3	-0.015	-0.056	0.199	0.220	0.303	0.348	-0.049	-0.065	0.272
Model 4	0.200	0.231	**0.294**	0.264	0.227	0.285	-0.045	-0.053	0.093
Model 5	0.091	0.060	-0.028	0.147	0.203	0.014	0.101	0.090	**0.354**
Model 6	0.080	0.020	-0.107	0.164	0.207	0.030	-0.030	-0.062	-0.024
Model 7	0.165	0.084	0.103	0.246	0.295	**0.355**	-0.048	-0.112	-0.034
Model 8	0.092	0.106	0.163	0.250	0.256	0.323	-0.217	-0.271	-0.550

Notes: 1) One-step ahead forecasts are generated with rolling window sampling scheme with the 40,45 and 50 years of window sizes. 2) All the R^2_{os} values are calculated by considering AR(4) as the bench-mark model. 3) Model specifications 3 to 8 are defined in Section 4.3.2. 4) Highlighted values provide the models that have highest R^2_{os} values for each response variable for all four different in-sample periods.

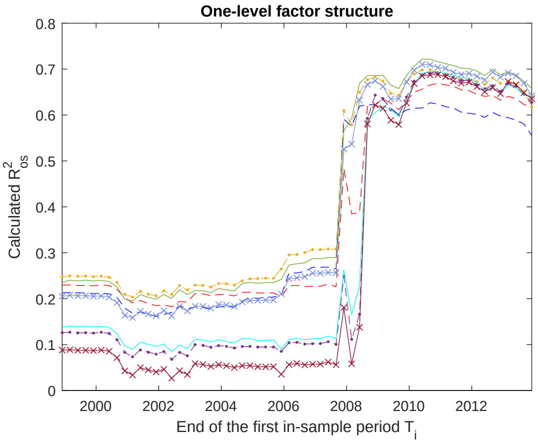
Results for sensitivity analysis of expanding window size



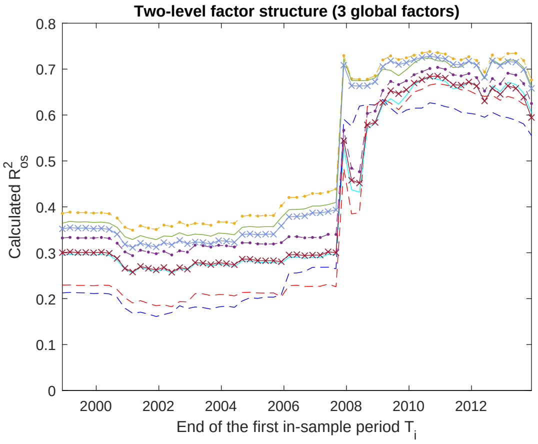
(a) (b)

Figure 4.9: Out-of-sample R^2_{os} of all the model specifications for forecasting GDP growth with 61 expanding window sizes

Notes: 1). Plot (a) represents R^2_{os} of the FAR models with one-level factor structure. Plots (b) show R^2_{os} for the two-level FAR models. 2) R^2_{os} caluclated compared to mean model. 3) All the 61 expanding window sample size periods start from 1959:Q1, with T_1 = 1959:Q1-1998:Q4 and T_{61} = 1959:Q1-2013:Q4. The x-axis in the plot is denoting the time T_i, end of the first estimation period. 4) Blue and red dash lines represent the AR(1) and AR(4) models, the purple and yellow lines with dots represent the parametric FAR augmented with AR(1) and AR(4). The green and cyan solid lines represent the two semi-parametric FAR models with (α_t, β_t) augmented with AR(1) and AR(4) respectively. Blue and brown lines with 'x' represent the semi-parametric FAR (α, β_t) augmented with AR(1) and AR(4) respectively. 5) The plots conclude that the two-level FAR models show higher predictability for small first in-sample periods (such as T_1– T_{37}) and the one-level FAR models show better predicabilty for larger first in-sample periods. 6) There is an acute structural difference between the first estimation periods 1959:Q1-2007:Q2 and 1959:Q1-2008:Q2.



(a)



(b)

Figure 4.10: Out-of-sample $R^2_{os}s$ of models for forecasting inflation with expanding window sizes

Notes: 1) Plot (a) represent $R^2_{os}s$ of the FAR models with one-level factor structure. Plots (b) represent R^2_{os} for the two-level FAR models with 3 global factors. 2) $R^2_{os}s$ measures are calculated against the mean model as benchmark. 3) The 61 expanding window sizes start from 1959:Q1, with T_1 = 1959:Q1-1998:Q4, T_{61} = 1959:Q1-2013:Q4. The T_i appears on the x-axis denotes the end of window size i. Legend are same as for Figure 4.9. 4) The plots conclude that irrespective to the choice of the first in-sample period, all the one-level and two-level FAR models show better forecastability compared to the mean Model 9. 5) There is an acute structural difference between the first estimation periods 1959:Q1-2007:Q2 and 1959:Q1-2008:Q2.

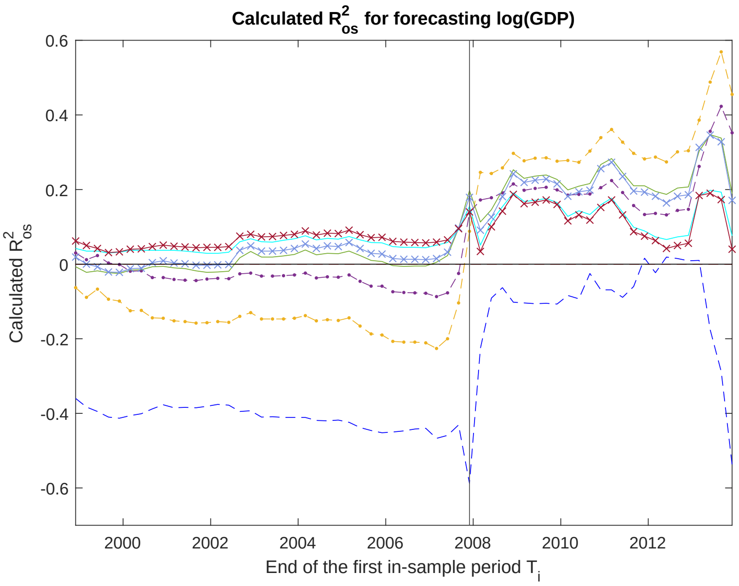


Figure 4.11: Calculated R^2_{os} for 61 different first in-sample periods, log(GDP)

Notes: 1) Plots represent R^2_{os} for the one-level FAR models. 2) The 61 expanding window sizes start from 1959:Q1, with T_1 = 1959:Q1-1998:Q4, T_{61} = 1959:Q1-2013:Q4. The T_i appears on the x-axis denotes the end of window size i. The FAR models with one-level factor structure 3) $R^2_{os}s$ measures are calculated against the AR(4) model as benchmark. 4) Legend are same as for Figure 4.9. 5) The plots conclude that the semi-parametric FAR models show higher predictability for small first in-sample periods (such as T_1– T_{37}) and the parametric FAR models show better predicabilty for larger first in-sample periods

4.6.4 Appendix D: Panel variables

Table 4.11 provides a brief information about the 12 groups that we used as potential level-2 factors in this study.

Table 4.11: The 12 groups of the panel data of 100 variables and some economic variables

Number	Group	Variables
1	NIPA	Consumption:Durable Real private fixed investment Real Exports of Goods & Services ...
2	Industrial Production	IP: Durable Materials IP:Nondurable Materials Capacity Utilization: Total Industry ...
3	Employment and Unemployment	All Employees: Durable goods All Employees: Education & Health Services All Employees: Government ...
4	Housing	New Private Housing Units Authorized All-Transactions House Price Index for the United States ...
5	Inventories, Orders, and Sales	Real Retail and Food Services Sales Real Value of Manufacturers' New Orders ...
6	Prices	Business Sector: Implicit Price Deflator Personal consumption expenditures: Durable goods ...
7	Earnings and Productivity	Manufacturing Sector: Real Compensation Per Hour Business Sector: Real Compensation Per Hour ...
8	Interest Rates	Effective Federal Funds Rate 3-Month Treasury Bill: Secondary Market Rate ...
9	Money and Credit	Real Commercial and Industrial Loans Total Real Nonrevolving Credit Owned and Securitized ...
10	Household Balance Sheets	Real Total Liabilities of Households and Nonprofit Organizations Real Net Worth of Households and Nonprofit Organizations ...
11	Exchange Rates	Trade Weighted U.S. Dollar Index: Major Currencies U.S. / Euro Foreign Exchange Rate ...
12	Stock Markets	CBOE S&P 100 Volatility Index: VXO S&P's Common Stock Price Index: Composite ...

For more details, see the updated appendix of FRED-QD $https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QDappendix.pdf$.

Chapter 5

Conclusion and Future Direction

5.1 Conclusion

Accurate forecasting of key macroeconomic variables such as GDP growth and inflation is central to making economic policy decisions. Therefore, a huge literature emerged on forecasting key macroeconomic variables using improved time series models and economic theory based models for generating accurate forecasts of macroeconomic variables. A method that has attracted considerable attention in the literature (in theory and empirical application) is the FAR model. Despite its popularity, in light of recent developments, we can identify the three main assumptions that underpin the FAR model. The objective of this thesis is to relax the these assumptions and propose improved methods for forecasting macroeconomic variables such as GDP, GDP growth rate and inflation. The accuracy of the proposed models is assessed against its competitors.

This thesis contains three main Chapters. The three assumptions that underlie the FAR model are progressively relaxed in each Chapter and new FAR methods are developed and used for forecasting macroeconomic variables and assessed the accuracy of these models' forecasts relative to its competitors. In this final Chapter, the objective of the three Chapters are briefly stated and discuss the main findings. We also provide directions for potential future research topics.

In Chapter 2, we relax the assumption that the FAR model consists of only I(1) factors when forecasting I(1) variables and allow the model to consist a mixture of I(0) and I(1) factors as predictors. The mixture of I(0) and I(1) factors are estimated from the large panel data typically contains a mixture of stationary and non-stationary economic variables and include them in the FAR model. This Chapter derives asymptotic results for a method for estimating the FAR model when the set of predictors includes a mixture of I(0) and I(1) factors and constructing an asymptotically valid prediction interval.

In a simulation study, we find that parameter estimates of the improved FAR model have favourable properties in moderate samples. The out-of-sample prediction intervals have the coverage probabilities that are close to nominal probabilities in moderate samples. The mixture FAR model was used to generate out-of-sample predictions of non-stationary variables, GDP and Industrial Production [IP], using the quarterly panel data, FRED-QD, on US macroeconomic variables. In an evaluation of accuracy of forecasts using the out-of-sample R^2_{os}, we find that the proposed model outperforms the standard FAR model and time series models. We have highlighted the complexities involved in deriving the asymptotic results for the mixture-FAR, and shown the potential gains in statistical efficiency. In both empirical application and simulations study, we have shown that the accuracy of the mixture FAR is better than its competitors.

In Chapter 3, we relax the assumption that parameters of the proposed FAR model in the previous Chapter are constant over time and allow the key parameters of the mixture-FAR model to be time varying, albeit in a controlled manner. This Chapter develops a method of estimating proposed semi-parametric FAR model and then uses the estimated model for forecasting three key macroeconomic variables GDP, GDP growth and inflation.

In a simulation study, we explored two sperate methods for estimation of the latent factors: the conventional Principal Components Analysis (PCA) and a nonparametric local estimation method. We observed that the factors estimated by a non-parametric method were not very sensitive to different bandwidths, but the estimated coefficients in the FAR model were sensitive to the bandwidth choice. Therefore, we studied the cross-validation for choosing suitable bandwidths for both factor estimation and parameter estimation. Using the FRED-QD data set, we evaluated the performance of the new method for forecasting the I(1) log(GDP), and two I(0) GDP growth rate and inflation. The results show that our proposed semi-parametric FAR method forecasts all three aforementioned variables better than the competing models. Moreover, we observe that the model with nonparametric estimates of factors forecasts inflation better than the semi-parametric FAR model with PCA factors, whereas the model with PCA factors generated log(GDP) and GDP growth rate forecasts better than the model with nonparametric counterparts. Since the proposed FAR include two new features, I(1) and I(0) factors and the key parameters are time varying, the derivation of the asymptotic properties for the proposed model is very demanding. We would approach this in two stages: (i) we assume that the factors in the FAR are known; and (ii) we assume that the factors unknown, which is the case in practice, and estimated. Work is in progress to derive the asymptotic properties of the method under assumption (i) and will complete the derivations under assumption (ii) in the future.

In Chapter 4, we allow the semi-parametric FAR model to include a mixture of I(0) and I(1) two-level factors, called *global factors* and *group factors*. This proposed FAR model is an extension of the one studied in Chapter 3. We refer to the proposed model as *semi-parametric two-level FAR model*. To improve the out-of-sample predictive performance of the method further, we included the economic policy uncertainty index (EUI) as an additional predictor in the two-level FAR model.

We estimated a mixture of stationary and non-stationary eight global and 12 group factors (Stock and Watson 2002a) from the panel of 100 economic variables by the sequential least squares algorithm based method proposed by Breitung and Eickmeier [2016]. This method is developed for estimating I(0) two-level factors from the panel of I(0) variables, which we adapt to our model setup. Since the results from kernel estimation method outperformed the Hermite and trigonometric polynomial estimations for the same FRED-QD panel dataset in Chapter 3, the semi-parametric two-level FAR model was estimated by a kernel method. The novelty of the proposed model is that, based on prior knowledge about the relationship between the two-level factors the desired macroeconomic variable and selection criteria such as goodness-of-fit and correlation measures, we can select the number of global and group factors to be included in each of the three semi-parametric two-level FAR models for log(GDP), GDP growth rate and inflation. For example, we select three global and seven group factors as predictors in the FAR model for inflation. In contrast, the FAR models that studied in the literature, the same set of global factors was used for forecasting macroeconomic variables. Thus, the proposed method in this Chapter is likely to improve the accuracy of the forecasts relative to its competitors. Furthermore, the models with EUI perform better than those without EUI.

We observed that the semi-parametric two-level FAR model forecast the two stationary variables GDP growth rate and inflation better than the semi-parametric one-level FAR model proposed in Chapter 3 and other time series models. For forecasting the non-stationary variable log(GDP), on the other hand, the semi-parametric one-level FAR performs better than the semi-parametric two-level FAR model. Furthermore, the semi-parametric two-level FAR model's forecasts largely depend on the initial expanding window size and that, under this sampling scheme, the proposed model generates better out-of-sample forecasts in small window sizes. Moreover, we find that the semi-parametric two-level FAR models more accurate out-of sample predictions for the rolling window sampling scheme than for the expanding window scheme. The reverse is true for the semi-parametric one-level FAR model. This finding is in contrast to the recommendations by previous studies to use rolling window scheme for forecasting stationary variables such as GDP growth and inflation.

5.2 Future research direction

From our experience on working on the methodological developments and empirical applications studied in this thesis, we present some research topics that could provide directions for future research.

In order to obtain the asymptotic distributions of the parametric and nonparametric coefficients in the semi-parametric FAR model proposed in Chapter 3, we assumed that the factors are known and follows a time-varying $VMA(\infty)$ process. Work is in progress to derive the asymptotic properties of the proposed method. However, these factors are unknown in practice and they need to be estimated from the panel data set and the estimated factors appear in the semi-parametric FAR model. For this case, the derivation of asymptotic properties of the parameter estimates and the prediction intervals is very demanding and will be undertaken in the future.

In this thesis, we have adapted the estimation methods proposed in the literature for the estimation of the both two-level factor panel model, which consists of I(0) and I(1) variables and the semi-parametric two-level FAR model. A methodological development involving derivations of asymptotic results for the flexible FAR models' parameter estimates and prediction intervals would advance the research in the area. Once we complete the methodological developments proposed in Chapter 3, the derivations of asymptotic results for the method proposed in Chapter 4 will become manageable.

The two information criteria proposed by Bai and Ng [2002] and Bai [2004] do not provide a consistent estimator of the optimal number of factors in the multi-level factor model. Thus, a new information criterion to estimate the optimal number of stationary and non-stationary factors in the multi-level factor model needs to be introduced.

Bibliography

Abbate, A. and Marcellino, M. (2014). Modelling and forecasting exchange rates with time-varying parameter models. *Journal of Economic Literature*.

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

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):137–183.

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

Bai, J. and Ng, S. (2005). *A new look at panel testing of stationarity and the PPP hypothesis.* Cambridge University Press.

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

Banerjee, A. and Marcellino, M. (2006). Are there any reliable leading indicators for us inflation and gdp growth? *International Journal of Forecasting*, 22(1):137–151.

Banerjee, A., Marcellino, M., and Masten, I. (2005). Leading indicators for euro-area inflation and gdp growth. *Oxford Bulletin of Economics and Statistics*, 67:785–813.

Banerjee, A., Marcellino, M., and Masten, I. (2008). Chapter 4 forecasting macroeconomic variables using diffusion indexes in short samples with structural change. *Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization)*, 3:149–194.

Barnett, A., Mumtaz, H., and Theodoridis, K. (2014). Forecasting uk gdp growth and inflation under structural change. a comparison of models with time-varying parameters. *International Journal of Forecasting*, 30(1):129–143.

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

Beck, G. W., Hubrich, K., and Marcellino, M. (2009). Regional inflation dynamics within and across euro-area countries and a comparison with the united states. *Economic Policy*, 24(57):142–184.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

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

Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84.

Breitung, J. and Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches. In *Dynamic factor models*, volume 35 of *Advances in Econometrics*, pages 177–214. Emerald Group Publishing Limited, Bingley.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.

Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chen, X. B., Gao, J., Li, D., and Silvapulle, P. (2018). Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business & Economic Statistics*, 36(1):88–100.

Cheng, X. and Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2):280–293.

Choi, I. (2017). Efficient estimation of nonstationary factor models. *Journal of Statistical Planning and Inference*, 183:18–43.

Choi, I., Kim, D., Kim, Y. J., and Kwark, N.-S. (2018). A multilevel factor model: Identification, asymptotic theory and applications. *Journal of Applied Econometrics*, 33(3):355–377.

Corradi, V. and Swanson, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, 182(1):100–118.

Demers, F. and Cheung, C. (2007). Evaluating forecasts from factor models for canadian gdp growth and core inflation. Technical report, Bank of Canada.

Dette, H. and Wu, W. (2022). Prediction in locally stationary time series. *Journal of Business & Economic Statistics*, 40(1):370–381.

Dong, C., Gao, J., and Linton, O. (2021). High dimensional semiparametric moment restriction models. *Journal of Econometrics*.

Dong, C., Gao, J., and Peng, B. (2015). Partially linear panel data models with cross-sectional dependence and nonstationarity. *Available at SSRN 2585334*.

Dong, C., Gao, J., and Peng, B. (2018). Varying-coefficient panel data models with partially observed factor structure. *Available at SSRN 3102631*.

Dong, C., Gao, J., and Peng, B. (2019). Series estimation for single-index models under constraints. *Australian & New Zealand Journal of Statistics*, 61(3):299–335.

Eickmeier, S. (2005). Common stationary and non-stationary factors in the euro area analyzed in a large-scale factor model. Discussion paper 2, Studies of the Economic Research Centre.

Eickmeier, S. and Ziegler, C. (2008). How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach. *Journal of Forecasting*, 27(3):237–265.

Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6):1031–1057.

Gao, J. (2007). *Nonlinear time series: semiparametric and nonparametric methods*. CRC Press.

Gao, J. and Hawthorne, K. (2006). Semiparametric estimation and testing of the trend of temperature series. *The Econometrics Journal*, 9(2):332–355.

Gao, J., Tong, H., and Wolff, R. (2002). Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis*, 83(2):324–359.

Gonçalves, S. and Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, 182(1):156–173.

Hall, P., Li, Q., and Racine, J. S. (2007). Nonparametric estimation of regression functions in the presence of irrelevant regressors. *The Review of Economics and Statistics*, 89(4):784–789.

Hannadige, S., Gao, J., Silvapulle, M. J., and Silvapulle, P. (2021). Forecasting a nonstationary time series using a mixture of stationary and nonstationary predictors. Working paper.

Hansen, B. E. (2001). The new econometrics of structural change: dating breaks in us labour productivity. *Journal of Economic perspectives*, 15(4):117–128.

Hirata, M. H., Kose, M. M. A., and Otrok, M. C. (2013). Regionalization vs. globalization. Working paper WP/13/19, International Monetary Fund.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

Kavtaradze, L. and Mokhtari, M. (2018). Factor models and time-varying parameter framework for forecasting exchange rates and inflation: A survey. *Journal of Economic Surveys*, 32(2):302–334.

Lahiri, K. and Sheng, X. (2010). Learning and heterogeneity in gdp and inflation forecasts. *International Journal of Forecasting*, 26(2):265–292.

Li, D., Chen, J., and Gao, J. (2011a). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, 14(3):387–408.

Li, D., Chen, J., and Lin, Z. (2011b). Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference*, 141(2):995–1013.

Li, D., Phillips, P. C., and Gao, J. (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics*, 215(2):607–632.

Li, D., Qian, J., and Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516):1804–1819.

Ludvigson, S. C. and Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222.

Maciejowska, K. (2010). Common factors in nonstationary panel data with a deterministic trend–estimation and distribution theory. Working paper, European University Institute.

Moon, H. R. and Perron, B. (2007). An empirical analysis of nonstationarity in a panel of interest rates with factors. *Journal of Applied Econometrics*, 22(2):383–400.

Motta, G., Hafner, C. M., and von Sachs, R. (2011). Locally stationary factor models: Identification and nonparametric estimation. *Econometric Theory*, 27(6):1279–1319.

Phillips, P. C. (2015). Halbert white jr. memorial jfec lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13(3):521–555.

Rodríguez-Caballero, C. V. (2021). Energy consumption and gdp: a panel data analysis with multi-level cross-sectional dependence. *Econometrics and Statistics*.

Rossi, B. and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453.

Silvapulle, P. and Jayasuriya, S. (2018). Nonparametric time varying coefficient panel model for climate change and rice yield in two states in india. *Available at SSRN 3233473*.

Silvapulle, P., Smyth, R., Zhang, X., and Fenech, J.-P. (2017). Nonparametric panel data model for crude oil and stock market prices in net oil importing countries. *Energy Economics*, 67:255–267.

Smeekes, S. and Wijler, E. (2019). High-dimensional forecasting in the presence of unit roots and cointegration. Working paper, Department of Quantitative Economics, Maastricht University.

Stock, J. H. and Watson, M. (2009). 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.

Stock, J. H. and Watson, M. W. (1998a). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Technical report, National Bureau of Economic Research.

Stock, J. H. and Watson, M. W. (1998b). Diffusion indexes. Technical report, National Bureau of Economic Research.

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

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

Stock, J. H. and Watson, M. W. (2005). An empirical comparison of methods for forecasting using many predictors. *Manuscript, Princeton University*, 46.

Stock, J. H. and Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33.

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

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, volume 2, pages 415–525. Elsevier.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633.

Wang, P. (2008). Large dimensional factor models with a multi-level factor structure: identification, estimation and inference. *Unpublished manuscript, New York University.*

Wei, J. and Zhang, Y. (2020). A time-varying diffusion index forecasting model. *Economics Letters*, 193:109337.

Yan, Y., Gao, J., and Peng, B. (2021). On time-varying var models: Estimation, testing and impulse response analysis. *Testing and Impulse Response Analysis (October 31, 2021).*

Zhang, Y., Su, L., and Phillips, P. C. (2012). Testing for common trends in semi-parametric panel data models with fixed effects. *The Econometrics Journal*, 15(1):56–100.

Zhou, W., Gao, J., Harris, D., and Kew, H. (2020). Semi-parametric single-index predictive regression models with cointegrated regressors. *Available at SSRN 3694950*.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):513–531.