Nonparametric High-Dimensional Time Series: Estimation and Prediction

Jiraroj Tosasukul

PhD

UNIVERSITY OF YORK

MATHEMATICS

December 2018

Dedication

To my parents

and

all of my teachers

Abstract

This thesis introduces a new class of functional-coefficient time series models, where the regressors consist of autoregressors and latent factor regressors, and the coefficients are allowed to change with certain index variable. The unobservable factor regressors are estimated through imposing an approximate factor model on very high dimensional exogenous time series variables and subsequently implementing the classical principal component analysis. With the estimated factor regressors, a local linear smoothing method is used to estimate the coefficient functions and obtain a one-step ahead nonlinear forecast of the response variable, and then a wild bootstrap procedure is introduced to construct the prediction interval. The developed methodology is further extended to the case of multivariate response vectors and the model is generalised to the factor-augmented vector time series model with functional coefficients. The latter substantially generalises the linear factor-augmented vector autoregressive model which has been extensively studied in the literature. Under some regularity conditions, the asymptotic properties of the proposed methods are derived. In particular, we show that the local linear estimator and the nonlinear forecast using the estimated factor regressors are asymptotically equivalent to those using the true latent factor regressors. The latter is not feasible in practical applications. This thesis also discusses selection of the numbers of autoregressors and factor regressors and choice of bandwidth in local linear estimation. Some simulation studies and an empirical application to predict the UK inflation are given to investigate the performance of our model and methodology in finite samples.

Contents

Abstract **iii**

List of figures **v**

List of tables **vii**

Acknowledgements **x**

Declaration **xii**

1 Introduction **1**

2 Literature Review **5**

2.1 Univariate Kernel-Based Regression 5

2.1.1 Kernel Regression Estimation 5

2.1.2 Local Polynomial Estimation 9

2.1.3 Bandwidth Selection 11

2.2 Varying-Coefficient Models 12

2.2.1 One-Step Estimation Method 13

2.2.2 Two-Step Estimation Method 14

2.3 Factor Models 16

2.3.1 Estimation of the High-Dimensional Factor Models 18

2.3.2 Determining the Number of Common Factors 20

2.4 Linear Factor-Augmented model 23

3 FA-FCM and Methodology **25**

3.1 FA-FCM with Univariate Response 26
3.2 Estimation of the Rotated Coefficient Functions 29
3.3 One-step Ahead Nonlinear Forecast and Bootstrap Method 31
3.4 Selection of the Bandwidth and the Number of Lags 34
3.4.1 Selection of the Bandwidth 34
3.4.2 Selection of the Number of Lags 35
3.5 FA-FCM with Multivariate Response 36

4 Numerical Studies **39**
4.1 Simulation studies 39
4.1.1 Example 1 with Univariate Response 39
4.1.2 Example 2 with Bivariate Response 48
4.2 An empirical example 59

5 Main Asymptotic Theorems **64**
5.1 Technical Assumptions 64
5.2 Asymptotic Properties 66
5.3 Proofs of Theorems 68
5.3.1 Proofs of the Asymptotic Theorems 68
5.3.2 Some Technical Lemmas and Their Proofs 70

6 Conclusion and Future Research **77**

References **79**

List of Figures

4.1 The solid curve denotes the true values of the response y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 20$ in Example 1. 45

4.2 The solid curve denotes the true values of the response y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 150$ in Example 1. 46

4.3 The solid curve denotes the true values of the response y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 500$ in Example 1. 47

4.4 The solid curve denotes the true values of the response $y_{1,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 20$ in Example 2. 53

4.5 The solid curve denotes the true values of the response $y_{1,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 150$ in Example 2. 54

4.6 The solid curve denotes the true values of the response $y_{1,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 500$ in Example 2. 55

4.7 The solid curve denotes the true values of the response $y_{2,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 20$ in Example 2. 56
4.8 The solid curve denotes the true values of the response $y_{2,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 150$ in Example 2. 57
4.9 The solid curve denotes the true values of the response $y_{2,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 500$ in Example 2. 58
4.10 Plot of the original CPI series from Q1/1997 to Q4/2013. 60
4.11 Plot of the transformed CPI series from Q1/1997 to Q4/2013. 60
4.12 The change of the BIC function. 62
4.13 The solid curve denotes the true values of the transformed CPI y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the transformed CPI from Q1/2013 to Q4/2013. 63

List of Tables

4.1 The percentages of correctly selecting the factor number in 200 replications in Example 1 41
4.2 The average of the bandwidth for the infeasible local linear estimations in Example 1 41
4.3 The average of minimum AMS for the infeasible local linear estimations in Example 1 41
4.4 The average of the bandwidth for the feasible local linear estimations in Example 1 42
4.5 The average of minimum AMS for the feasible local linear estimations in Example 1 42
4.6 The percentages of correctly finding the lag number in 200 replications in Example 1 43
4.7 MSPE of the infeasible one-step ahead forecast in Example 1 43
4.8 MSPE of the feasible one-step ahead forecast in Example 1 44
4.9 The percentages of correctly selecting the factor number in 200 replications in Example 2 49
4.10 The percentages of correctly finding the lag number in 200 replications in Example 2 50
4.11 The average of the bandwidth for the infeasible local linear estimations in Example 2 50
4.12 The average of minimum AMS for the infeasible local linear estimations in Example 2 51
4.13 The average of the bandwidth for the feasible local linear estimations in Example 2 51

4.14 The average of minimum AMS for the feasible local linear estimations in Example 2 51
4.15 MSPE of the infeasible one-step ahead forecast in Example 2 52
4.16 MSPE of the feasible one-step ahead forecast in Example 2 52

Acknowledgements

First and foremost, I would honestly like to express my most profound appreciation and say an enormous thank you to my advisor, Professor Degui Li, for the excellent supervision and his continuous support of my PhD study over four years. He has always been available for any help I have ever wanted. I learned a lot from his deep insights, attention to detail, as well as his dedication and commitment to me. I am truly honoured to have him as my advisor.

Secondly, I would like to express my gratitude to my second advisor, Professor Wenyang Zhang, for his continuous help and support since my application to the PhD programme in the University of York. I also would like to thank my Thesis Advisory Panel member, Dr Marina Knight, for her continuous support and comments.

I am grateful to my financially sponsor: Naresuan University, Phitsanulok, Thailand, for granting me the scholarship and providing me with this great opportunity for completing my studies in the UK. I will apply knowledge and experience that I received during my studies in the UK to my career in the Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, Thailand.

I would like to acknowledge all of the staff at the Department of Mathematics, University of York, UK, for support and help. I would also like to thank the York Advanced Research Computing Cluster (YARCC) for allowing me to use the high-speed computing service in my numerical studies.

My sincere thanks also goes to my lovely friends both in the UK and Thailand for who have assisted me, stood by me and contributed to my educational progress in any way.

Last but not least, I am greatly thankful to my family their supporting, encour-

agement and love.

Declaration

To the best of my knowledge, this thesis does not infringe the copyright of any other person. I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, university. All sources are acknowledged as references. The literature review in Chapter 2 provides main concepts that are related to this thesis:

The literature review in Chapter 2 provides main concepts that are related to this thesis:

- Chapter 2.1 reviews the literature of univariate nonparametric kernel-based regression, including kernel regression estimators and local polynomial regression, which is based on the book by Fan and Gijbels (1996).
- Chapter 2.2 includes the basic idea of the varying-coefficient models and its estimation methods. In order to estimate varying-coefficient functions, one-step estimation method introduced by Cleveland et al. (1992) and two-step estimation method proposed by Fan and Zhang (1999) are mainly considered.
- Chapter 2.3 provides a summary of the high-dimensional factor models and the principal component analysis method for estimating both common factors and factor loadings, including some methods to determine the number of common factors based on the research of Bai and Ng (2002) and Stock and Watson (2002a).
- Chapter 2.4 briefly summaries the linear factor-augmented model and its estimation introduced by Stock and Watson (2002a,b).

A new class of functional-coefficient predictive regression models is proposed in Chapter 3. In particular:

- Chapter 3.1 introduces the *Factor-Augmented Functional-Coefficient Model (FA-FCM)* with univariate response related to the research of Stock and Watson (2002a), Bernanke et al. (2005), Bai and Ng (2006), Pesaran et al. (2011) and Cheng and Hansen (2015).

- Chapter 3.2 proposes the two-stage estimation procedure to estimate the rotated coefficient functions with the first-stage PCA method mainly based on the research of Bai and Ng (2002) and Stock and Watson (2002a).

- Chapter 3.3 introduces the one-step ahead nonlinear forecast method and constructs the prediction interval by using a bootstrap method as in Zhang and Peng (2010) and Chen et al. (2018).

- Chapter 3.4 discusses bandwidth selection and the criteria for selecting the number of lags. A modified multi-fold cross-validation criterion proposed by Cai et al. (2000) and a forward selection method introduced by Wang (2009) and Cheng et al. (2016) are used for choosing the optimal bandwidth and the number of lags respectively.

- Chapter 3.5 considers the extension to FA-FCM with the multivariate response, extending the research of Jiang (2014).

Two simulation studies and one empirical application are provided in Chapter 4 to examine the accuracy of the proposed predictive model and forecasting methodology:

- Chapter 4.1 gives two simulation studies under two different settings for the univariate response and bivariate response. Throughout the simulation studies, the number of factors is chosen by the criterion used by Chen et al. (2018).

- Chapter 4.2 applies the developed predictive model and forecasting method to predict the UK inflation change. The real data set is also analysed by Chen et al. (2018).

The asymptotic results for the developed methodologies for FA-FCM with univariate response are derived in Chapter 5:

- Chapter 5.1-5.3 give the asymptotic results with the technical assumptions, some of which are similar to those in Bai and Ng (2002), Bai and Ng (2006) and Chen et al. (2018).

The Chapter 3-5 are mainly based on my submitted paper: Nonlinear Factor-Augmented Predictive Regression Models with Functional Coefficients, jointly with Professor Degui Li and Professor Wenyang Zhang. This paper is currently under revision for *Journal of Time Series Analysis*.

Chapter 1

Introduction

In recent years, advanced science and technology grow rapidly to facilitate data collecting and computing. Consequently, large data are available in many areas such as finance and economics. An important issue associated with the large data set is that the number of variables becomes very large. It is often the case that the number of variables in big data analysis exceeds the sample size. In this case, traditional statistical tools in data analysis become infeasible, and dimension reduction is needed to effectively extract sample information.

The dimension reduction technique is probably the most commonly-used method for analysing highly-dimensional data. The main idea of the dimension reduction technique is to remove redundant features and identify significant ones in the large data. In the statistical literature, many shrinkage and screening methods have been introduced to select significant variables in the statistical models for independent data (Fan and Lv, 2010; Bühlmann and van de Geer, 2011; Hastie et al., 2015). However, these methods may not perform well in analysing highly-correlated large data set, which is common in finance and economics. The high-dimensional factor model is often preferred to deal with the latter case (Chamberlain, 1983; Fama and French, 1993; Bai and Ng, 2002). When both the cross-sectional size and time series length are large, the principal component analysis is probably the most widely-used method to estimate the factor model (see Section 2.3 for details).

The main focus of this thesis is high-dimensional time series. In traditional

time series models, the coefficients involved are usually assumed to be constants over a long horizon of time (Brockwell and Davis, 1991; Hamilton, 1994; Tsay, 2010). There are many parametric techniques that can be used to estimate these coefficients, e.g., ordinary least squares (OLS) estimation, maximum likelihood estimation (MLE), and Bayesian estimation. However, the assumption of the constant coefficients in time series models is too restrictive and often violated when analysing economic and financial data.

It is sensible to allow the coefficients to change smoothly over time or with an observable time series index variable. In this thesis, we study the varying-coefficient time series model which is a natural extension of linear time series model. The varying-coefficient model is introduced by Hastie and Tibshirani (1993) and has been extensively studied in the past few decades, e.g., Chen and Tsay (1993), Fan and Zhang (1999), Cai et al. (2000), Cai et al. (2000), Xia et al. (2004), Fan and Huang (2005) , Kai et al. (2011), and Jiang et al. (2013).

When analysing large data set, we often apply the high-dimensional varying-coefficient models. In this case, the screening or shrinkage method is usually used to remove the insignificant regressors, and then use the significant ones to build the models. Consequently, the nonlinear model prediction performance can be improved (Wang and Xia, 2009; Lian, 2012; Fan et al., 2014; Liu et al., 2014; Li et al., 2015). However, as pointed out by Fan and Lv (2008) and Chen et al. (2018), existing variable selection approaches may have some problems when irrelevant regressors are highly correlated with some relevant ones. These irrelevant regressors might be selected into the model with higher priority than some other relevant regressors, which would lead to high false positive rates and low true positive rates. This problem could become worse in the time series setting as the regressors usually contain some lags of the response and strong correlation exists among the regressors. To deal with this problem, we need to develop an alternative dimension-reduction technique in order to apply the functional-coefficient model to high-dimensional time series data.

The so-called factor-augmented model provides an effective way to address the above problem. The factor-augmented model is to combine the high-dimensional

factor model with the prediction model in time series analysis. Applying the factor model to high-dimensional exogenous variables, we can extract latent factors which can represent the dynamic feature of the high-dimensional exogenous process. The number of factors is often much smaller than the number of exogenous variables. Thus, the aim of dimension reduction can be achieved. Then the estimated factors as well as the auto-regressors are combined to build the factor-augmented prediction model. Details for recent developments on the factor-augmented model can be found in Section 2.4.

In this thesis, we introduce a new class of functional-coefficient predictive regression models when the regressors consist of auto-regressors and latent factor regressors. Unlike the linear factor-augmented models which have been extensively studied by existing literature, the coefficients are allowed to vary with a random index variable. The proposed model is called the factor-augmented functional-coefficient model (FA-FCM). We propose a two-stage estimation procedure to estimate the coefficient functions (with rotation): first estimate the latent factors (with rotation) and then the rotated coefficient functions are estimated by the local linear smoothing technique (Fan and Gijbels, 1996). In addition, a one-step ahead nonlinear forecasting of the response is constructed by using the estimated coefficient functions. Under regularity conditions, the local linear estimator and the nonlinear forecast using the estimated factor regressors are shown to be asymptotically equivalent to those using the true latent factor regressors. Furthermore, a bootstrap method is introduced to construct the prediction interval. The factor-augmented vector auto-regression is further generalised by allowing the coefficient matrices to change smoothly with an index variable.

The rest of the thesis is organised as follows. Chapter 2 is the literature review, where we briefly review the nonparametric kernel-based regression, varying-coefficient models, high-dimensional factor model, and linear factor-augmented regression. Chapter 3 introduces the univariate FA-FCM model, and proposes the nonlinear estimation and forecasting methodology. The extension to nonlinear vector auto-regression with the multivariate response is also discussed in this chapter. Chapter 4 provides the finite-sample simulation studies and an empirical

illustration to predict the UK inflation using the developed model and method. Chapter 5 derives the asymptotic results. Chapter 6 concludes the thesis.

Chapter 2

Literature Review

In this chapter, we provide a brief overview of concepts, models and methodologies that are closely related to this thesis. This chapter is organised as follows. Section 2.1 contains a review of the literature on univariate nonparametric kernel-based regression, including kernel regression estimation and local polynomial estimation. Section 2.2 provides a review of the varying-coefficient models and its estimation methods. Section 2.3 reviews the factor models and the principal component analysis method for estimating both common factors and factor loadings, and discusses some methods for determining the number of common factors. Section 2.4 briefly reviews the linear factor-augmented model.

2.1 Univariate Kernel-Based Regression

In this section, we review the kernel estimation and local polynomial estimation for a nonparametric regression model, where the regressor is univariate. In addition, we also discuss the choice of optimal bandwidth in kernel-based estimation.

2.1.1 Kernel Regression Estimation

The nonparametric regression modelling approach has become very popular in recent decades. It relaxes some restrictive assumptions imposed on the linear regression models and allows data to "speak for themselves". One of the

most commonly-used nonparametric regression estimation methods is the kernel smoothing technique. In this section, we give a brief review for the kernel-based smoothing method when the regressor is univariate. More details can be found in Wahba (1990), Wand and Jones (1994), Fan and Gijbels (1996), and De Boor (2001).

Consider bivariate data $(X_1, Y_1), \ldots, (X_n, Y_n)$, which form an independent and identically distributed sample from a population (X, Y). A nonparametric regression model is defined as follows:

$$Y = m(X) + \varepsilon, \tag{2.1}$$

where $m(\cdot)$ is known as a nonparametric regression function and the random error ε is independent of X with $\mathsf{E}(\varepsilon) = 0$ and $\mathsf{Var}(\varepsilon) = \sigma_\varepsilon^2$. The main interest in the nonparametric regression is to estimate the regression function $m(\cdot)$, which can be also written as the conditional expectation of Y given $X = x_0$ denoted by $m(x_0) = \mathsf{E}(Y|X = x_0)$.

We start with a traditional kernel regression estimators called the Nadaraya-Watson regression estimator that is proposed by Nadaraya (1964) and Watson (1964). It is defined as:

$$\hat{m}_{NW}(x_0) = \sum_{i=1}^{n} w_i(x_0) Y_i, \tag{2.2}$$

where the weights $w_i(x_0)$ are given by

$$w_i(x_0) = \frac{K_b\left(X_i - x_0\right)}{\sum_{j=1}^{n} K_b\left(X_j - x_0\right)}, \tag{2.3}$$

in which $K_b(\cdot) = K(\cdot/b)/b$, $K(\cdot)$ is a kernel function and b is a bandwidth controlling the level of smoothing. The bias of the Nadaraya-Watson regression estimator is

$$\mathsf{bias}(\hat{m}_{NW}(x_0)) = \left(m''(x_0) + \frac{2m'(x_0)f'(x_0)}{f(x_0)}\right)\frac{b^2}{2}\int_{-\infty}^{\infty} u^2 K(u)du, \tag{2.4}$$

where $f(x_0)$ is the density function of X at the point x_0, $f'(x_0)$ is the first derivative of $f(x_0)$, $m'(x_0)$ and $m''(x_0)$ are the first and second derivatives of $m(x_0)$,

respectively, and the variance of the Nadaraya-Watson regression estimator is

$$\sigma_n^2(\hat{m}_{NW}(x_0)) = \frac{1}{nbf(x_0)} \int_{-\infty}^{\infty} K^2(u)du, \tag{2.5}$$

by noting that ε is independent of X and $\mathsf{Var}(\varepsilon) = 1$.

We next discuss the properties and choice of the kernel function. It is usually a non-negative smooth function satisfying the following properties:

i) symmetry: $K(-u) = K(u)$,

ii) normalisation: $\int_{-\infty}^{\infty} K(u)du = 1$,

iii) finite second moment: $\int_{-\infty}^{\infty} u^2 K(u)du < \infty$.

The following lists are some commonly-used kernel functions:

i) Uniform or Boxcar kernel:

$$K(u) = \frac{1}{2} I(\mid u \mid \leqslant 1),$$

ii) Epanechnikov or Parabolic kernel:

$$K(u) = \frac{3}{4}(1 - u^2) I(\mid u \mid \leqslant 1),$$

iii) Gaussian kernel:

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}) I(\mid u \mid \leqslant 1),$$

iv) Biweight or Quartic kernel:

$$K(u) = \frac{15}{16}(1 - u^2)^2 I(\mid u \mid \leqslant 1),$$

v) Triweight kernel:

$$K(u) = \frac{35}{32}(1 - u^2)^3 I(\mid u \mid \leqslant 1),$$

where $I(\cdot)$ is an indicator function.

It is well-known that the choice of kernel function does not strongly affect the performance of the nonparametric regression estimation (Fan and Gijbels, 1996; Fan et al., 1998). As pointed out by Fan (1992), the Epanechnikov kernel function achieves the optimal efficiency in the class of all regression functions whose second derivative is bounded by a constant in the neighbourhood of the point x_0. Thus, we employ the Epanechnikov kernel function in the rest of this thesis.

The choice of bandwidth is more crucial than the choice of the kernel function because it determines the smoothness of the estimated regression function. As the bandwidth increases, the estimated regression function goes from undersmoothing to oversmoothing. As the bandwidth diverges to infinity, the regression function is approximated by a constant. In the kernel-based estimation, we usually let the bandwidth tend to zero, which can help capture a complex pattern of the nonparametric regression function. The details on how to determine an optimal value of bandwidth will be given in Section 2.1.3.

Another commonly-used local kernel smoothing estimator is the Gasser-Müller regression estimator proposed by Gasser and Müller (1984), which is defined as

$$\hat{m}_{GM}(x_0) = \sum_{i=1}^{n} \left[\int_{s_{i-1}}^{s_i} K_b\left(u - x_0\right) du \right] Y_i, \tag{2.6}$$

with $s_0 = -\infty$, $s_n = +\infty$, and $s_i = (X_i + X_{i+1})/2$ for $i = 1, \ldots, n-1$, where $K_b(\cdot) = K(\cdot/b)/b$, $K(\cdot)$ is a kernel function and b is a bandwidth.

The bias of the Gasser-Müller regression estimator is

$$\mathsf{bias}(\hat{m}_{GM}(x_0)) = \frac{b^2 m''(x_0)}{2} \int_{-\infty}^{\infty} u^2 K(u) du, \tag{2.7}$$

and the variance of the Gasser-Müller regression estimator is

$$\sigma_n^2(\hat{m}_{GM}(x_0)) = 1.5 \left[\frac{1}{nbf(x_0)} \int_{-\infty}^{\infty} K^2(u) du \right]. \tag{2.8}$$

By comparing with the Nadaraya-Watson regression estimator, the bias of the Gasser-Müller regression estimator is simpler, but its variance is larger. Both

the Nadaraya-Watson and the Gasser-Müller regression estimators are based on local constant approximation, so they suffer from the boundary effect, i.e., the nonparametric regression estimates usually have large bias near the boundary points. The so-called local polynomial regression provides an effective way to address this problem. This will be reviewed in the next section.

2.1.2 Local Polynomial Estimation

The local polynomial regression estimation is more general than the kernel regression estimations, giving estimates for both the regression function and its derivatives. As in Section 2.1.1, suppose that we have bivariate data (X_i, Y_i), $i = 1, \ldots, n$, which form an independent and identically distributed from a population (X, Y) and consider estimating the regression function $m(x)$ at the point of $x = x_0$.

A local polynomial approximation techniqe is applied to the unknown regression function $m(x)$ in a neighbourhood of x_0. Suppose that the $(p+1)$th derivative of $m(x)$ at the point x_0 exists. We consider the following Taylor expansion:

$$\begin{aligned} m(x) \quad \approx \quad & m(x_0) + m^{(1)}(x_0)(x - x_0) + \frac{m^{(2)}(x_0)}{2!}(x - x_0)^2 \\ & + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p, \end{aligned} \tag{2.9}$$

where $m^{(k)}(x)$ is the k-th order derivative of $m(x)$. We can rewrite the approximation in (2.9) as

$$m(x) = \sum_{j=0}^{p} \beta_j^0 (x - x_0)^j, \tag{2.10}$$

where

$$\beta_j^0 = \frac{m^{(j)}(x_0)}{j!} \quad \text{for} \quad j = 0, 1, \ldots, p.$$

We next estimate β_j^0, $j = 0, 1, \ldots, p$, using the local polynomial method. Consider the weighted least squares objective function:

$$Q\left(\beta_0, \beta_1, \ldots, \beta_p\right) = \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right\}^2 K_b(X_i - x_0), \tag{2.11}$$

where $K_b(\cdot) = K(\cdot/b)/b$ with $K(\cdot)$ being a kernel function and b being a bandwidth. Minimise the objective function $Q(\beta_0, \beta_1, \ldots, \beta_p)$ in (2.11) with respect to β_j, $j = 0, 1, \ldots, p$, and denote the estimators by $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$. Then, we can obtain the estimators of the unknown regression function $m(x_0)$ and its derivatives:

$$\hat{m}^{(j)}(x_0) = j!\hat{\beta}_j \quad \text{for} \quad j = 0, 1, \ldots, p.$$

For convenience, the weighted least squares objective function in (2.11) is usually written in the matrix form. As in Fan and Gijbels (1996), we write

$$Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\intercal}\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{2.12}$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{pmatrix},$$

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^{\intercal},$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\intercal},$$

and $\mathbf{W}$ is the $n \times n$ diagonal weight matrix with $K_b(X_i - x_0)$ being the diagonal components, i.e.,

$$\mathbf{W} = \mathsf{diag}\left\{K_b(X_1 - x_0), \ldots, K_b(X_n - x_0)\right\}.$$

If the $(p+1) \times (p+1)$ matrix $\mathbf{X}^{\intercal}\mathbf{W}\mathbf{X}$ is invertible, the minimiser to $Q(\boldsymbol{\beta})$ can be obtained as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\intercal}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\intercal}\mathbf{W}\mathbf{Y}, \tag{2.13}$$

and

$$\hat{m}^{(j)}(x_0) = j!\mathbf{e}_{j+1}^{\intercal}\hat{\boldsymbol{\beta}},$$

where $j = 0, 1, \ldots, p$ and $\mathbf{e}_{j+1}$ is a $(p+1)$-dimensional vector with the $(j+1)$th entry being one and the others zeros.

Note that when the order of the local polynomial regression is set to be zero, it reduces to the Nadaraya-Watson regression estimator. The local linear regression is a special case of the local polynomial regression when the order of the local polynomial regression is equal to one. The local linear estimation is probably the most commonly-used local polynomial estimation. More details on local polynomial regression estimation including its asymptotic theory and empirical application can be found in Fan and Gijbels (1996).

2.1.3 Bandwidth Selection

The bandwidth selection plays an important role in implementing the kernel regression estimators. The bandwidth affects the level of smoothness, through controlling the size of the local neighbourhood in kernel estimation. The trade-off between bias and variance is usually considered when selecting an optimal bandwidth. If we choose a small size of bandwidth, then it leads to small bias but large variance for the kernel estimation. When we increase the size of bandwidth, the kernel estimation variance can be reduced but the bias would be larger.

In practice, the value of optimal bandwidth is not available, and depends on the unknown regression function. Several methods for bandwidth selection have been proposed in the literature. One of the most widely-used methods is the cross-validation (CV) method introduced by Stone (1974). For simplicity of exposition, we consider only the cross-validation criteria for the Nadaraya-Watson kernel regression estimator. The same idea can also be applied to other kernel-based estimation methods.

The main idea of the cross-validation method is described as follows. The CV objective function is defined by

$$\mathsf{CV}(b) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{m}_{b,-i}(X_i)\right]^2 w(X_i), \tag{2.14}$$

where $\hat{m}_{b,-i}(X_i)$ is a leave-one-out estimator of $m(X_i)$ with the i-th pair of ob-

servations (X_i, Y_i) left out in estimation, and $w(\cdot)$ is a weight function. For the Nadaraya-Watson kernel regression estimator, the leave-one-out estimate $\hat{m}_{-i}(X_i)$ is defined as

$$\hat{m}_{b,-i}(X_i) = \frac{\sum_{j \neq i} K_b(X_j - X_i) Y_i}{\sum_{j \neq i} K_b(X_j - X_i)}, \tag{2.15}$$

where $K_b(\cdot) = K(\cdot/b)/b$, $K(\cdot)$ is a kernel function and b is a bandwidth. Subsequently, the optimal bandwidth can be chosen by minimising the CV function $\mathsf{CV}(b)$ defined in (2.14) with respect to b.

2.2 Varying-Coefficient Models

In the traditional linear regression and time series models, it is well-known that the coefficients involved are often assumed to be constants over a long time period. There are several parametric methods which can be applied to estimate the constant coefficients, for instance, ordinary least squares (OLS) estimation, maximum likelihood estimation (MLE) and Bayesian estimation. However, the assumption of the constant coefficients in the models is often too restrictive and violated in many practical applications, e.g., analysis of financial and economic data collected over a long time horizon. Hence, it may be more appropriate to allow the coefficients in the models to change over time or with an observable time series index variable.

The varying-coefficient model proposed by Hastie and Tibshirani (1993) is a natural extension of the classical linear models by allowing the coefficients in the models to vary with a random index variable. It can be used to explore the nonlinear dynamic pattern in univariate or multivariate time series data analysis. Instead of estimating joint multivariate nonparametric regression function directly, in the varying-coefficient model, we nonparametrically estimate smooth coefficient functions, each of which is univariate nonparametric regression function. Thus, the curse of dimensionality issue can be avoided. The varying-coefficient models and its generalised version have been extensively developed and received much attention in the statistics literature, e.g., Cleveland et al. (1992), Chen and Tsay (1993), Fan and Zhang (1999), Cai et al. (2000), Cai et al. (2000), Xia et al. (2004),

Fan and Huang (2005) , Kai et al. (2011), and Jiang et al. (2013).

In this section, we give a brief review of the varying-coefficient models and two estimation methods: one-step estimation and two-step estimation. Both of these two methods are based on the local polynomial smoothing technique introduced in Section 2.1.

2.2.1 One-Step Estimation Method

Let Y be a response variable and $X_1, \ldots, X_p$ be covariates. The varying-coefficient model can be formulated in the following form:

$$Y = \sum_{j=1}^{p} \alpha_j(U) X_j + \varepsilon, \tag{2.16}$$

where U is an observable random index variable, $\alpha_j(\cdot)$ is an unknown coefficient function and ε is the error term of the model with

$$\mathsf{E}(\varepsilon | U, X_1, \ldots, X_p) = 0 \qquad \text{a.s.},$$

and

$$\mathsf{Var}(\varepsilon | U, X_1, \ldots, X_p) = \sigma^2(U) \qquad \text{a.s.}$$

Here, we use a.s. to denote almost surely.

Note that model (2.16) allows the unknown coefficient functions $\alpha_1(U), \ldots, \alpha_p(U)$ to vary with the random index variable U. If U is replaced by scaled time points, model (2.16) becomes the time- varying coefficient model.

We next introduce the one-step local linear estimation method proposed by Cleveland et al. (1992) to estimate the smooth coefficient function in (2.16). Assume that $\alpha_j(\cdot)$ has continuous second-order derivatives. For each given u_0, we apply the Taylor expansion to approximate the varying-coefficient functions locally:

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha_j{}'(u_0)(u - u_0), \tag{2.17}$$

for u in a neighbourhood of u_0, $\alpha_j{}'(\cdot)$ is the first-order derivative function of $\alpha_j(\cdot)$,

$j = 1, \ldots, p$. Consider the following local least-squares objective function:

$$Q(A, B) = \sum_{i=1}^{n} \left(\left[Y_i - \sum_{j=1}^{p} \alpha_j + \beta_j (U_i - u_0) X_{ij} \right]^2 K_b(U_i - u_0) \right), \tag{2.18}$$

where $A = (\alpha_1, \ldots, \alpha_p)^\intercal$, $B = (\beta_1, \ldots, \beta_p)^\intercal$, $K_b(\cdot) = K(\cdot/b)/b$, $K(\cdot)$ is a kernel function and b is a bandwidth.

By minimising the objective function in (2.18), the local linear estimators of the coefficient functions can be estimated as

$$\hat{\alpha}(u_0) = (\mathbf{I}_p, \mathbf{O}_p) \left[\mathbb{X}^\intercal \mathbf{W} \mathbb{X} \right]^{-1} \left[\mathbb{X}^\intercal \mathbf{W} \mathbf{Y} \right], \tag{2.19}$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix, and $\mathbf{O}_p$ is a $p \times p$ null matrix with each entry being 0,

$$\mathbb{X} = \begin{pmatrix} X_{11} & X_{11}(U_1 - u_0) & X_{12} & X_{12}(U_1 - u_0) & \cdots & X_{1p} & X_{1p}(U_1 - u_0) \\ X_{21} & X_{21}(U_2 - u_0) & X_{22} & X_{22}(U_2 - u_0) & \cdots & X_{2p} & X_{2p}(U_2 - u_0) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n1} & X_{n1}(U_n - u_0) & X_{n2} & X_{n2}(U_n - u_0) & \cdots & X_{np} & X_{np}(U_n - u_0) \end{pmatrix},$$

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^\intercal,$$

$$\mathbf{W} = \mathsf{diag}\left\{ K_b(U_1 - u_0), \ldots, K_b(U_n - u_0) \right\}.$$

2.2.2 Two-Step Estimation Method

The one-step local linear estimation introduced in Section 2.2.1 is not efficient when different coefficient functions have different degrees of smoothness and cannot achieve the optimal estimation convergence rate. To address this problem, Fan and Zhang (1999) introduce the so-called two-step estimation procedure, which we review in this section.

To avoid confusion, we let the subscripts "0", "1", and "2" denote those related to the initial, one-step, and two-step estimators, respectively. We next review the two-step estimation procedure.

We first obtain the initial estimators of the coefficient function $\alpha_j(u)$, $j = 1, \ldots, p$, by minimising the local least-squares objective function:

$$Q_0(A, B) = \sum_{i=1}^{n} \left(\left[Y_i - \sum_{j=1}^{p} \{\alpha_j + \beta_j(U_i - u_0)\} X_{ij} \right]^2 K_{b_0}(U_i - u_0) \right), \tag{2.20}$$

where $A = (\alpha_1, \ldots, \alpha_p)^\intercal$, $B = (\beta_1, \ldots, \beta_p)^\intercal$, $K_{b_0}(\cdot) = K(\cdot/b_0)/b_0$, $K(\cdot)$ is the kernel function, and b_0 is an initial bandwidth. The obtained initial estimators of the coefficient function are denoted by $\hat{\alpha}_{1,0}(u_0), \hat{\alpha}_{2,0}(u_0), \ldots, \hat{\alpha}_{p,0}(u_0)$.

Without loss of generality, assume that the coefficient function $\alpha_p(u)$ is smoother than the rest of the coefficient functions $\alpha_j(u)$, $j = 1, \ldots, p-1$, which have the same degree of smoothness. Specifically, the coefficient functions $\alpha_j(u)$ are assumed to have second-order derivatives, and the coefficient function $\alpha_p(u)$ has bounded fourth-order derivative. Therefore, for each given point u_0, the coefficient function $\alpha_p(u)$ can be approximated via

$$\alpha_p(u) \approx \alpha_p(u_0) + \alpha_p'(u_0)(u - u_0) + \frac{\alpha_p^{(2)}(u_0)}{2!}(u - u_0)^2 + \frac{\alpha_p^{(3)}(u_0)}{3!}(u - u_0)^3, \tag{2.21}$$

for u in a neighbourhood of u_0, where $\alpha_p^{(k)}(\cdot)$ is the k-th order derivative of $\alpha_p(\cdot)$ and $\alpha_p'(u_0) = \alpha_p^{(1)}(u_0)$.

By using the local cubic approximation in (2.21), we consider the following local least-squares objective function,

$$\begin{aligned} Q_1\left(A_1, B_1, \alpha_p, \beta_p, \gamma_p, \delta_p\right) \;=\; & \sum_{i=1}^{n} \Bigg[Y_i - \sum_{j=1}^{p-1} \{\alpha_j + \beta_j(U_i - u_0)\} X_{ij} - \{\alpha_p + \beta_p(U_i - u_0) \\ & + \gamma_p(U_i - u_0)^2 + \delta_p(U_i - u_0)^3\} X_{ip} \Big]^2 K_{b_1}(U_i - u_0), \end{aligned} \tag{2.22}$$

where $A_1 = (\alpha_1, \ldots, \alpha_{p-1})^\intercal$, $B_1 = (\beta_1, \ldots, \beta_{p-1})^\intercal$, $K_{b_1}(\cdot) = K(\cdot/b_1)/b_1$, $K(\cdot)$ is the kernel function K and b_1 is the bandwidth in the first step. Note that b_1

may be different from b_0 used in the initial estimation. By minimising the local least-squares objective function $Q_1\left(A_1, B_1, \alpha_p, \beta_p, \gamma_p, \delta_p\right)$, we obtained the estimates $\hat{\alpha}_j, \hat{\beta}_j$, $j = 1, \ldots, p-1$, and $\hat{\alpha}_p, \hat{\beta}_p, \hat{\gamma}_p, \hat{\delta}_p$. Then, let $\hat{\alpha}_{p,1}(u_0) = \hat{\alpha}_p$, which is called the one-step estimator.

Finally, we substitute the initial estimates $\hat{\alpha}_{1,0}(\cdot), \hat{\alpha}_{2,0}(\cdot), \ldots, \hat{\alpha}_{p-1,0}(\cdot)$, and minimise the following objective function:

$$\begin{aligned} Q_2\left(\alpha_p, \beta_p, \gamma_p, \delta_p\right) &= \sum_{i=1}^{n}\Bigg[Y_i - \sum_{j=1}^{p-1}\hat{\alpha}_{j,0}(U_i)X_{tj} - \{\alpha_p + \beta_p(U_i - u_0) + \gamma_p(U_i - u_0)^2 \\ &\quad + \delta_p(U_i - u_0)^3\}X_{ip}\Bigg]^2 K_{b_2}(U_i - u_0), \end{aligned} \tag{2.23}$$

where $K_{b_2}(\cdot) = K(\cdot/b_2)/b_2$, $K(\cdot)$ is the kernel function and b_2 is the bandwidth. Let $\tilde{\alpha}_p, \tilde{\beta}_p, \tilde{\gamma}_p$, and $\tilde{\delta}_p$ be defined by minimising $Q_2\left(\alpha_p, \beta_p, \gamma_p, \delta_p\right)$ with respect to $\alpha_p, \beta_p, \gamma_p$ and δ_p. Denote $\hat{\alpha}_{p,2}(u_0) = \tilde{\alpha}_p$ as the two-step estimator of $\alpha_p(u_0)$.

Note that the bandwidth b_0 for the initial estimates of the coefficient function is usually chosen to be small for reducing bias in the first step of this procedure. On the other hand, we consider higher-order smoothing in the second step with a larger bandwidth b_2 to reduce the variance of the final estimator of the coefficient functions $\alpha_p(u_0)$. Consequently, the two-step estimator often outperforms the one-step estimator in numerical studies.

2.3 Factor Models

With rapid developments in data collection techniques and information technology over the past decade, a large amount of data can now be observed and usually characterised by a large number of variables over a long time span. Many traditional statistical analysis techniques fail in cases where the number of variables in data analysis exceeds the sample size. Hence, it is important to develop effective statistical tools for high-dimensional data analysis. Dimension reduction is one of the most commonly-used techniques to deal with this issue.

The key idea of the dimension reduction is to reduce a large number of initial variables in the high-dimensional data. It can be done either by removing redun-

dant variables and keeping only the most relevant variables, or by finding a small set of features that contain the same (or similar) fundamental information as the initial variables.

Among the dimension reduction techniques, the high-dimensional factor model introduced by Chamberlain and Rothschild (1983) is very commonly used. The main idea of the high-dimensional factor models is to reduce a large number of initial variables to a small number of common factors that summarise most of the information contained by the whole data set. It often assumes that both the cross-sectional dimension p and the time-series length n jointly tend to infinity in the factor model estimation.

In recent years, the factor models have been extensively studied in financial and economic data analyses. Some well-known examples of factor models are the arbitrage pricing theory (ART) (Ross, 1976) and the Fama-French three-factor model (Fama and French, 1993). As shown in Ross (1976), ART assumes that a few number of factors can capture the variation in a large number of asset returns. In the Fama-French three-factor model, it is assumed that the three observed factors including the market risk, the company size and the value of a company (the book-to-market ratio) to explain the behaviour of vast stock returns. In reality, it is often the case that the common factors and the factor number are unknown. It is crucial to estimate the common factors and determine the factor number in high-dimensional factor models.

In general, there are two different modelling approaches for the high-dimensional factor models: the dynamic factor model and the static factor model. In the dynamic factor model, the common factors are assumed to follow the autoregressive process. Moreover, the observed variables in the dynamic factor model are influenced by the common components as well as their lagged values. Although the dynamic factor model is more general than the static factor model, it is more difficult to estimate the dynamic factors and it is not the main focus of this PhD thesis. Hence, in this section, we only focus on a review of the static factor model and its estimation, including some selection criteria for the number of factors.

2.3.1 Estimation of the High-Dimensional Factor Models

Let x_{it} denote the response variables for unit i observed at a time point t, $i = 1, \ldots, p$ and $t = 1, \ldots, n$. The static factor model is defined by,

$$x_{it} = \lambda_i^\intercal \mathbf{F}_t + u_{it}, \tag{2.24}$$

where $\mathbf{F}_t = (F_{t1}, \ldots, F_{tk_0})^\intercal$ is a k_0-dimensional vector of the common factors, λ_i is a vector of factor loadings, and u_{it} is the idiosyncratic errors. Here, we assume that the number of the common factors k_0 is known and will discuss how to estimate it in Section 2.3.2.

The static factor model (2.24) can also be written in the vector form:

$$\mathbf{X}_t = \mathbf{B}\mathbf{F}_t + \mathbf{U}_t, \tag{2.25}$$

where $\mathbf{X}_t = (x_{1t}, \ldots, x_{pt})^\intercal$ is a p-dimensional vector of observations at time t, $\mathbf{B} = (\lambda_1, \ldots, \lambda_p)^\intercal$ is a $p \times k_0$ matrix of factor loadings, $\mathbf{F}_t$ is a k_0-dimensional vector of common factors, and $\mathbf{U}_t = (u_{1t}, \ldots, u_{pt})^\intercal$ is a p-dimensional vector of idiosyncratic errors. For further notational simplicity, the static factor model can be written in the matrix form:

$$\mathbf{X} = \mathbf{F}\mathbf{B}^\intercal + \mathbf{U}, \tag{2.26}$$

where $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^\intercal$ is an $n \times p$ matrix of the observations, $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_n)^\intercal$ is an $n \times k_0$ matrix of common factors, and $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_n)^\intercal$ is an $n \times p$ matrix of the idiosyncratic errors of the model.

There are several methods that can be used to estimate both the common factors and factor loadings. One of the most commonly-used methods is the principal components analysis. The classical principal components analysis in multivariate setting is done by transforming the initial multiple variables to a new set of variables, which are ordered so that the first few retain most of the variation that present in all the initial variables.

Connor and Korajczyk (1986) are among the first to propose that the common

factors can be estimated by applying the principal component analysis method under the assumption that the number of the time dimension n is fixed whereas the number of cross-sectional dimension p tends to infinity. The idiosyncratic errors in Connor and Korajczyk (1986) are assumed to be independent and identically distributed over cross-sectional and time series dimensions. Stock and Watson (2002a,b) study the approximate factor models by relaxing the restrictive assumption on idiosyncratic errors. Some further developments on deriving asymptotic properties of the estimators in factor models using the principal component analysis can be found in Bai and Ng (2002) and Bai (2003, 2004).

We next introduce the principal components method to estimate common factors and factor loadings in the approximate factor model. The estimated factor loadings $\hat{\lambda}_i$ and the estimated common factors $\hat{\mathbf{F}}_t$ are obtained by minimising the least squares objective function:

$$V(\mathbf{F}, \mathbf{B}) = \sum_{i=1}^{p} \sum_{t=1}^{n} (x_{it} - \lambda_i^{\intercal} \mathbf{F}_t)^2 = tr\left[\left(\mathbf{X} - \mathbf{F}\mathbf{B}^{\intercal} \right) \left(\mathbf{X} - \mathbf{F}\mathbf{B}^{\intercal} \right)^{\intercal} \right], \tag{2.27}$$

where $tr(\cdot)$ is a matrix trace. Clearly, the solution for the objective function in (2.27) is not unique. If $(\hat{\mathbf{B}}, \hat{\mathbf{F}})$ is a solution to minimise $V(\mathbf{F}, \mathbf{B})$, then $(\hat{\mathbf{B}}\mathbf{Q}, \hat{\mathbf{F}}\mathbf{Q})$ is also a solution for any $k_0 \times k_0$ orthogonal matrix $\mathbf{Q}$, satisfying $\mathbf{Q}^{\intercal}\mathbf{Q} = \mathbf{I}_{k_0}$. The solution $(\hat{\mathbf{B}}, \hat{\mathbf{F}})$ would be unique by assuming the following restrictions:

$$\hat{\mathbf{F}}^{\intercal}\hat{\mathbf{F}}/n = \mathbf{I}_{k_0} \quad \text{and} \quad \hat{\mathbf{B}}^{\intercal}\hat{\mathbf{B}} \quad \text{is diagonal.}$$

The following procedure can be used to minimise the least squares objective function in (2.27). By concentrating out $\hat{\mathbf{B}}$, the minimisation in (2.27) is equivalent to maximising $tr[\mathbf{F}^{\intercal}(\mathbf{X}\mathbf{X}^{\intercal})\mathbf{F}]$ with respect to $\mathbf{F}$. By using the normalisation restriction that $\hat{\mathbf{F}}^{\intercal}\hat{\mathbf{F}}/n = \mathbf{I}_{k_0}$, the estimated matrix of common factors $\hat{\mathbf{F}}$ is obtained using the k_0 eigenvectors (multiplied by $\sqrt{n}$) of the $n \times n$ matrix $\mathbf{X}\mathbf{X}^{\intercal}$ associated with the k_0 largest eigenvalues (ranked in the descending order), and the factor loadings can be estimated as $\hat{\mathbf{B}} = \mathbf{X}^{\intercal}\hat{\mathbf{F}}/n$.

As the non-zero eigenvalues of $\mathbf{X}\mathbf{X}^{\intercal}$ and $\mathbf{X}^{\intercal}\mathbf{X}$ are the same, an alternative procedure can be used to minimise the objective function in (2.27). By concentrating

out $\hat{\mathbf{F}}$, the minimisation in (2.27) is equivalent to maximising $tr[\hat{\mathbf{B}}^\intercal(\mathbf{X}^\intercal\mathbf{X})\hat{\mathbf{B}}]$. Here, we assume that the normalisation:

$$\hat{\mathbf{B}}^\intercal\hat{\mathbf{B}}/p = \mathbf{I}_{k_0} \quad \text{and} \quad \hat{\mathbf{F}}^\intercal\hat{\mathbf{F}} \quad \text{is diagonal.}$$

Subsequently, the estimated factor loadings $\hat{\mathbf{B}}$ is given by the k_0 eigenvectors (multiplied by $\sqrt{p}$) of the $p \times p$ matrix $\mathbf{X}^\intercal\mathbf{X}$ associated with the k_0 largest eigenvalues (ranked in the descending order). Finally, we can construct the estimation of common factors by $\hat{\mathbf{F}} = \mathbf{X}\hat{\mathbf{B}}/p$.

2.3.2 Determining the Number of Common Factors

It is well-known that the number of common factors k_0 is usually unspecified. Accurate selection of factor number is an important issue in factor model estimation. Several methods have been proposed in the literature for selecting a true number of common factor. As the principal component analysis is used to estimate the factor models, we first review some classical methods to determine the number of components. These methods have been commonly used in multivariate data analysis.

i) Cumulative percentage of total variation (Jolliffe, 2002)
Let $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_p)^\intercal$ be a vector of eigenvalues of a sample covariance matrix $\frac{1}{n}\mathbf{X}^\intercal\mathbf{X}$, which are arranged in descending order and δ be a pre-specified value of percentage. In this method, the optimal number of common factors k_0 is determined by selecting the smallest number of components to achieve a certain percentage of the total variance, i.e.,

$$\frac{\hat{\lambda}_1 + \cdots + \hat{\lambda}_k}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_p} \geqslant \delta.$$

In practical applications, δ is usually chosen as a value between 70% and 90%.

ii) Kaiser's rule

Kaiser (1960) introduces a different idea by comparing the eigenvalue of each principal component with a pre-specified value. In practice, we may retain the components with eigenvalues greater than the average of the total variance, which is defined by $\frac{1}{p}\sum_{i=1}^{p}\hat{\lambda}_i$. When the principal components are obtained from the sample correlation matrix, the total variance equals to p. As a result, the number of common factors can be selected as the number of components whose corresponding eigenvalues are greater than one.

iii) Scree plot

The scree plot introduced by Cattell (1966) is a plot of the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ (in descending order). For determining the number of components, we seek an elbow in the scree plot. If the elbow position is between the k-th and $(k+1)$-th eigenvalues, the number of components is chosen as k.

In the high-dimensional approximate factor models with both p and n divergent to infinity, some information criteria can be used to select the number of factors. We next briefly review these criteria which are introduced and symmetrically studied by Bai and Ng (2002).

When the number of common factors is k, we let $V(k)$ be the sum of squared residuals defined by

$$V(k) = (np)^{-1}\sum_{i=1}^{p}\sum_{t=1}^{n}(x_{it} - (\hat{\lambda}_i^k)^{\intercal}\hat{\mathbf{F}}_t^k)^2, \tag{2.28}$$

where $\hat{\lambda}_i^k$ and $\hat{\mathbf{F}}_t^k$ are the estimated factor loading and common factors by principal components analysis when the factor number is k.

The objective function for the information criterion is defined by

$$PC(k) = V(k) + k \cdot g(p,n),$$

where $g(p,n)$ is a penalty function satisfying the following conditions:

i) $g(p,n) \to 0$,

ii) $C_{n,p}^2 \cdot g(p,n) \to \infty$ with $C_{n,p} = \min\{\sqrt{p}, \sqrt{n}\}$.

The above two conditions ensure that $\hat{k} = \operatorname{argmin} PC(k)$ is a consistent estimator of k_0, see, for example, Theorem 2 in Bai and Ng (2002).

Bai and Ng (2002) suggest the following choices of $g(p,n)$ in the penalty term:

$$\begin{aligned} g_1(p,n) &= k\hat{\sigma}^2 \left(\frac{n+p}{np}\right) \ln\left(\frac{np}{n+p}\right), \\ g_2(p,n) &= k\hat{\sigma}^2 \left(\frac{n+p}{np}\right) \ln C_{n,p}^2, \\ g_3(p,n) &= k\hat{\sigma}^2 \frac{\ln C_{n,p}^2}{C_{n,p}^2}, \end{aligned}$$

where $\hat{\sigma}^2$ is a consistent estimate of $(np)^{-1}\sum_{i=1}^{p}\sum_{t=1}^{n} \mathsf{E}(u_{it})^2$. Bai and Ng (2002) also suggest replacing $\hat{\sigma}^2$ with $V(k_{\max})$, where $k_{\max}$ is a bounded positive integer such that $k_0 \leqslant k_{\max}$, and

$$V(k_{\max}) = (np)^{-1}\sum_{i=1}^{p}\sum_{t=1}^{n} \hat{e}_{it}^2, \qquad \hat{e}_{it} = x_{it} - (\hat{\lambda}_i^{k_{\max}})^{\intercal}\hat{\mathbf{F}}_t^{k_{\max}},$$

with $\hat{\lambda}_i^{k_{\max}}$ and $\hat{\mathbf{F}}_t^{k_{\max}}$ being the estimated factor loading and common factors using $k_{\max}$ factors.

In practice, Bai and Ng (2002) introduce the following three information criteria objective functions with the logarithmic transformation on $V(k)$:

$$\begin{aligned} IC_1(k) &= \ln(V(k)) + k\left(\frac{n+p}{np}\right) \ln\left(\frac{np}{n+p}\right), \\ IC_2(k) &= \ln(V(k)) + k\left(\frac{n+p}{np}\right) \ln C_{n,p}^2, \\ IC_3(k) &= \ln(V(k)) + k\frac{\ln C_{n,p}^2}{C_{n,p}^2}. \end{aligned}$$

The estimated number of factors is obtained such that $IC_j(k)$ $(j = 1,2,3)$ is minimised over $1 \leqslant k \leqslant k_{\max}$. The number $k_{\max}$ is often chosen as $k_{\max} = 8\left\lfloor (\min\{n,p\}/100)^{(1/4)} \right\rfloor$ where $\lfloor\cdot\rfloor$ denotes the floor function. The numerical studies in Bai and Ng (2002) show that their information criteria have reliable performance in finite samples.

2.4 Linear Factor-Augmented model

In this section, we briefly review the estimation and forecasting methodology for the linear factor-augmented prediction model. In macroeconomic time series, one of the most commonly-used techniques is the factor-augmented, or diffusion index, model (c.f., Stock and Watson, 2002a; Bernanke et al., 2005; Bai and Ng, 2006; Cheng and Hansen, 2015). The linear factor-augmented model is a natural combination of the high-dimensional factor model and the linear forecasting model. Unlike the traditional time series models, the factor-augmented model includes both lags of response and some latent common factors. Through the factor model, we can estimate factor regressors which describe the dynamic feature of the high-dimensional exogenous process. The number of factors is usually much smaller than the number of exogenous variables. Hence, the use of the factor-augmented model achieves dimension reduction.

We first introduce estimation of the linear factor-augmented prediction model, and then review h-step ahead prediction of y_{t+h} where the positive integer h denotes the forecast horizon. The linear factor-augmented model is defined by

$$y_{t+h} = \boldsymbol{\beta}_1^\intercal \mathbf{F}_t + \boldsymbol{\beta}_2^\intercal \mathbf{Y}_t + \epsilon_{t+h}, \tag{2.29}$$

where $\mathbf{Y}_t = (y_t, \cdots, y_{t-d_0+1})^\intercal$ is a $d_0 \times 1$ vector of lags of response variables, $\boldsymbol{\beta}_1 = (\beta_{11}, \cdots, \beta_{1k_0})^\intercal$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \cdots, \beta_{2d_0})^\intercal$ are two column vectors of coefficients, and ϵ_{t+h} is a predictive model error. The factor regressor $\mathbf{F}_t$ is assumed to satisfy the factor model (2.24) in Section 2.3.1.

The main procedure to estimate the factor-augmented prediction model (2.29) can be divided into two stages. In the first stage, we estimate the common factors $\hat{\mathbf{F}}_t$ by using the principal components method with some restrictions of factors and factor loading as mentioned in Section 2.3.1. For simplicity, we assume that the number of factors k_0 is pre-specified. It can be estimated by using the information criteria in Section 2.3.2. In the second stage, the estimated factors $\hat{\mathbf{F}}_t$ from the first stage are used as regressors. In this stage, we estimate the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in the factor-augmented model (2.29) via the ordinary least squares.

For notational simplicity, the factor-augmented model in (2.29) can be rewritten as follows:

$$y_{t+h} = \boldsymbol{\beta}_H^\intercal \mathbf{X}_t + \epsilon_{t+h}, \tag{2.30}$$

where $\mathbf{X}_t = \left[(\mathbf{H}\mathbf{F}_t)^\intercal, \mathbf{Y}_t^\intercal\right]^\intercal$, $\boldsymbol{\beta}_H = \left[\boldsymbol{\beta}_1^\intercal \mathbf{H}^{-1}, \boldsymbol{\beta}_2^\intercal\right]^\intercal$ and $\mathbf{H}$ is a $k_0 \times k_0$ rotation matrix such that $\hat{\mathbf{F}}_t$ estimates $\mathbf{H}^\intercal \mathbf{F}_t$ consistently. The specific definition of $\mathbf{H}$ will be given in Section 3.2.

The estimated coefficients of $\boldsymbol{\beta}_H$, denoted by $\hat{\boldsymbol{\beta}}_H$, can be constructed as

$$\hat{\boldsymbol{\beta}}_H = \left(\hat{\mathbf{X}}^\intercal \hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}^\intercal \mathbf{Y},$$

where $\mathbf{Y} = (y_{h+1}, \cdots, y_n)^\intercal$ and $\hat{\mathbf{X}} = \left(\hat{\mathbf{X}}_1, \cdots, \hat{\mathbf{X}}_{n-h}\right)^\intercal$. By using $\hat{\mathbf{X}}_t$ and $\hat{\boldsymbol{\beta}}_H$, we can construct the h-step ahead prediction of the response variable:

$$\hat{y}_{t+h} = \hat{\boldsymbol{\beta}}_H^\intercal \hat{\mathbf{X}}_t. \tag{2.31}$$

According to the factor-augmented model in (2.29), it is worth to point out that when $h = 1$ and $y_{t+1} = (\mathbf{F}_{t+1}^\intercal, \mathbf{Y}_{t+1}^\intercal)^\intercal$, the model becomes the factor-augmented vector autoregressive (FAVAR) model introduced by Bernanke et al. (2005). Bernanke et al. (2005) also suggest two techniques to estimate the FAVAR model: the two-step method based on the principal component estimation of factors, and the Bayesian method based on Gibbs sampling.

Chapter 3

FA-FCM and Methodology

In this chapter, we propose a new class of functional-coefficient predictive regression models, i.e., the factor-augmented functional-coefficient models (FA-FCM), where the regressors consist of auto-regressors and latent factor regressors, and the coefficients are allowed to vary with a random index variable. A two-stage estimation procedure is introduced to estimate the rotated coefficient functions. In the first stage, the latent factor regressors can be estimated by implementing the principal component analysis as introduced in Section 2.3. In the second stage, we use a local linear smoothing method to estimate the rotated coefficient functions using the estimated factor regressors and auto-regressors. For simplicity, the same bandwidth is employed for all rotated coefficient functions. A one-step ahead nonlinear forecast of the response variable and a bootstrap procedure are applied to construct prediction of response and the prediction interval. In addition, the developed methodology is further extended to the case of multivariate response, and the model is generalised to the factor-augmented vector predictive regression with functional coefficients. The main focus of this chapter is to introduce the FA-FCM structure and the relevant methodology. The numerical studies to examine finite-sample performance will be given in Chapter 4. The asymptotic properties as well as their proofs will be given in Chapter 5.

The structure of this chapter is organised as follows. Section 3.1 introduces FA-FCM with univariate response. Section 3.2 proposes the two-stage estimation procedure to estimate the rotated coefficient functions. Section 3.3 introduces the

one-step ahead nonlinear forecast method and constructs the prediction interval by using a bootstrap method. Section 3.4 discusses bandwidth selection and the criteria for selecting the number of lags. Finally, Section 3.5 considers the extension to FA-FCM with multivariate response.

3.1 FA-FCM with Univariate Response

As introduced in Section 2.2, the varying-coefficient regression/auto-regression models are a natural extension of the traditional parametric linear regression/auto-regression models, and they can be used to explore nonlinear dynamic pattern in univariate or multivariate time series data analysis. In the last two decades, the varying-coefficient modelling approach and its generalised version have experienced rapid development (c.f., Chen and Tsay, 1993; Hastie and Tibshirani, 1993; Fan and Zhang, 1999; Cai et al., 2000; Xia et al., 2004; Kai et al., 2011; Jiang et al., 2013). Fan and Zhang (2008) and Park et al. (2015) give an extensive review of some recent developments in the field.

In this thesis, we consider the dynamic time series setting. Let us start with an introduction of functional-coefficient predictive regression model defined by

$$\begin{aligned} y_{t+1} &= \sum_{i=1}^{q_n} \alpha_{1i}(u_t) z_{ti} + \sum_{j=1}^{d_0} \alpha_{2j}(u_t) y_{t+1-j} + \varepsilon_{t+1} \\ &= \mathbf{Z}_t^{\intercal} \boldsymbol{\alpha}_1(u_t) + \mathbf{Y}_t^{\intercal} \boldsymbol{\alpha}_2(u_t) + \varepsilon_{t+1}, \quad t = 1, \cdots, n, \end{aligned} \tag{3.1}$$

where y_{t+1} is a response variable, $\mathbf{Z}_t = (z_{t1}, \cdots, z_{tq_n})^{\intercal}$ is a q_n-dimensional column vector of random covariates, $\mathbf{Y}_t = (y_t, \cdots, y_{t-d_0+1})^{\intercal}$ is a column vector of d_0 lags of the response, u_t is a univariate index variable, $\boldsymbol{\alpha}_1(\cdot) = [\alpha_{11}(\cdot), \cdots, \alpha_{1q_n}(\cdot)]^{\intercal}$ and $\boldsymbol{\alpha}_2(\cdot) = [\alpha_{21}(\cdot), \cdots, \alpha_{2d_0}(\cdot)]^{\intercal}$ are two column vectors of coefficient functions, and ε_{t+1} is the model error. If the index component $u_t = t/n$ with n as the time series length, the model (3.1) is known as the time-varying coefficient time series model (c.f., Robinson, 1989; Cai, 2007). In the econometric terminology, the components of $\mathbf{Z}_t$ are exogenous variables which are usually determined by the factors outside of our models, whereas the components of $\mathbf{Y}_t$ are determined within our model.

In this thesis, we assume that q_n, the number of exogenous covariates $\mathbf{Z}_t$, diverges to infinity as the sample size n grows, but d_0, the number of lags, is fixed. The divergence rate of q_n will be specified in Chapter 5 via some technical assumptions. When the dimension of regressors in the models is ultra large or moderately large, a commonly-used approach is to apply certain shrinkage estimation or screening method to remove the insignificant regressors and then use the significant regressors to build the varying-coefficient models, enhancing the nonlinear model prediction accuracy (c.f., Wang and Xia, 2009; Lian, 2012; Fan et al., 2014; Liu et al., 2014; Li et al., 2015). However, as briefly mentioned in Chapter 1 and pointed out in some variable selection literature such as Fan and Lv (2008), when irrelevant regressors are highly correlated with some relevant ones, through the shrinkage or screening approach, these irrelevant regressors might be selected into the model with higher priority than some other relevant regressors, leading to high false positive rates and low true positive rates in the variable selection of the varying-coefficient models. This problem could become worse in the time series setting as the regressors usually contain some lags of the response and it is not uncommon to find some strong correlations among the regressors, see Chen et al. (2018) for some numerical evidences. Therefore, to address this problem, we need to develop an alternative dimension-reduction technique for the high-dimensional functional-coefficient predictive regression model (3.1).

As the number of lags is assumed to be fixed, we only need to consider the dimension reduction on the high-dimensional exogenous regressors $\mathbf{Z}_t$. This will be done by imposing an approximate factor modelling structure which have been commonly used in the analysis of economic and financial time series data:

$$\mathbf{Z}_t = \mathbb{B}_n \mathbf{F}_t + \mathbf{V}_t, \tag{3.2}$$

where $\mathbb{B}_n$ is a $q_n \times k_0$ matrix of factor loadings, $\mathbf{F}_t = (F_{t1}, \cdots, F_{tk_0})^\intercal$ is a k_0-dimensional latent common factor which is stationary and weakly dependent over time, and $\mathbf{V}_t$ is a q_n-dimensional column vector of idiosyncratic errors. Note that the notations used here for the approximate factor model (3.2) are different

from those used in Section 2.3. The number of factors k_0 is unknown, and it can be determined via some criteria.

The approximate factor model (3.2) has been extensively studied in the literature, see, for example, Chamberlain and Rothschild (1983), Fama and French (1993), Bai and Ng (2002), Fan et al. (2013) and the references therein. Through the factor model (3.2), the latent factor time series process may carry a large proportion of the "dynamic information" contained in the high-dimensional observable time series $\mathbf{Z}_t$. But k_0, the number of factors is typically much smaller than q_n, the dimension of $\mathbf{Z}_t$. So, we can achieve dimension reduction. Hence, in this thesis, instead of directly estimating the dynamic relationship between y_{t+1} and $\mathbf{Z}_t$, we consider the following functional-coefficient predictive model using the latent factor regressors:

$$y_{t+1} = \mathbf{F}_t^\intercal \boldsymbol{\beta}_1(u_t) + \mathbf{Y}_t^\intercal \boldsymbol{\beta}_2(u_t) + \epsilon_{t+1}, \tag{3.3}$$

where $\boldsymbol{\beta}_1(\cdot) = [\beta_{11}(\cdot), \cdots, \beta_{1k_0}(\cdot)]^\intercal$ and $\boldsymbol{\beta}_2(\cdot) = [\beta_{21}(\cdot), \cdots, \beta_{2d_0}(\cdot)]^\intercal$ are two column vectors of coefficient functions, and ϵ_{t+1} is the error term. A significant difference between models (3.1) and (3.3) is that the factor regressors $\mathbf{F}_t$ in the latter model (3.3) are unobservable, while all of the regressors in the former model (3.1) are observable. Furthermore, the number of regressors in the model (3.3) is $k_0 + d_0$, which is fixed and much smaller than that in the model (3.1).

Through a combination of (3.2) and (3.3), we obtain the functional-coefficient predictive regression model with latent factor regressors and call it as the *Factor-Augmented Functional-Coefficient Model (FA-FCM)* throughout the thesis. The FA-FCM can be viewed as a generalisation of the linear factor-augmented regression or auto-regression models (c.f., Stock and Watson, 2002a; Bernanke et al., 2005; Bai and Ng, 2006; Pesaran et al., 2011; Cheng and Hansen, 2015) in which the factor regressors $\mathbf{F}_t$ can be regarded as the "proxy" when we aim to describe the dynamic relationship between y_{t+1} and $\mathbf{Z}_t$. Through the dimension reduction, it is expected that the nonlinear forecast using the FA-FCM in (3.3) could be more accurate than that using the conventional functional-coefficient time series model (3.1).

3.2 Estimation of the Rotated Coefficient Functions

As the factor regressors in the model (3.3) are unobservable, they must be determined from the observed time series data. In this section, we propose a two-stage estimation procedure to develop a feasible nonparametric estimation approach: in the first stage, we determine the rotated factor regressors by applying the principal component analysis (PCA) method as introduced in Section 2.3; and in the second stage, we estimate the rotated coefficient functions by using the local linear smoothing technique. We next give a detailed description of the estimation procedure.

STAGE 1: As mentioned in Section 2.3.1, the PCA is one of the most commonly-used techniques for dimension reduction. For simplicity, we now assume that k_0 is pre-specified. Chapter 4 will discuss how to determine the number k_0 in practice. In this stage, we obtain the estimated factor regressors by using the principal components analysis approach. Specifically, letting $\mathbb{Z}_n = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^\intercal$, an $n \times q_n$ matrix of observations, we conduct an eigenanalysis on the $n \times n$ (normalised) matrix $\mathbb{Z}_n\mathbb{Z}_n^\intercal/(nq_n)$, and obtain

$$\hat{\mathbb{F}}_n = \left(\hat{\mathbf{F}}_1, \ldots, \hat{\mathbf{F}}_n\right)^\intercal,$$

where $\hat{\mathbb{F}}_n$ is an $n \times k_0$ matrix which consists of the k_0 eigenvectors (multiplied by $\sqrt{n}$) associated with the k_0 largest eigenvalues of the matrix $\mathbb{Z}_n\mathbb{Z}_n^\intercal/(nq_n)$ (ranked in the descending order). In addition, we may further construct the estimation of factor loading matrix (with rotation) by

$$\hat{\mathbb{B}}_n = \mathbb{Z}_n^\intercal\hat{\mathbb{F}}_n/n,$$

where we have used the fact of $\hat{\mathbb{F}}_n^\intercal\hat{\mathbb{F}}_n/n = \mathbf{I}_{k_0}$ with $\mathbf{I}_{k_0}$ being a $k_0 \times k_0$ identity matrix. Some existing literature (c.f., Bai and Ng, 2002; Stock and Watson, 2002a) shows that, under some mild conditions, $\hat{\mathbf{F}}_t$ *is a consistent estimate of the rotated*

common factor $\mathbf{HF}_t$ (see also the proof in Chapter 5), where

$$\mathbf{H} = \mathbf{Q}_n^{-1}\left(\hat{\mathbb{F}}_n^{\intercal}\mathbb{F}_n/n\right)\left(\mathbb{B}_n^{\intercal}\mathbb{B}_n/q_n\right), \ \ \mathbb{F}_n = (\mathbf{F}_1, \ldots, \mathbf{F}_n)^{\intercal},$$

and $\mathbf{Q}_n$ is a $k_0 \times k_0$ diagonal matrix of the first k_0 largest eigenvalues of $\mathbb{Z}_n\mathbb{Z}_n^{\intercal}/(nq_n)$ arranged in the descending order. Furthermore, one may prove that the rotation matrix $\mathbf{H}$ is asymptotically invertible, indicating the existence of the inverse matrix $\mathbf{H}^{-1}$ with probability approaching one.

STAGE 2: We estimate the rotated coefficient functions by employing the local linear smoothing technique. Specifically, letting $\mathbf{X}_t = \left[(\mathbf{HF}_t)^{\intercal}, \mathbf{Y}_t^{\intercal}\right]^{\intercal}$, $\boldsymbol{\beta}_H(\cdot) = \left[\boldsymbol{\beta}_1^{\intercal}(\cdot)\mathbf{H}^{-1}, \boldsymbol{\beta}_2^{\intercal}(\cdot)\right]^{\intercal}$ and noting that $\mathbf{H}^{-1}\mathbf{H} = \mathbf{I}_{k_0}$, we may rewrite model (3.3) as

$$y_{t+1} = \boldsymbol{\beta}_H^{\intercal}(u_t)\mathbf{X}_t + \epsilon_{t+1}. \tag{3.4}$$

As in the literature, we assume that the coefficient functions $\boldsymbol{\beta}_1(\cdot)$ and $\boldsymbol{\beta}_2(\cdot)$ have continuous second-order derivatives (see Assumption 7(ii) in Chapter 5), implying that $\boldsymbol{\beta}_H(\cdot)$ have continuous second-order derivatives as well for given $\mathbf{H}$. Instead of estimating $\boldsymbol{\beta}_1(\cdot)$ and $\boldsymbol{\beta}_2(\cdot)$, we next use the local linear smoothing method to estimate the rotated coefficient functions $\boldsymbol{\beta}_H(\cdot)$. This is the same as the one-step local linear estimation introduced in Section 2.2.1.

Let

$$\begin{aligned}
\mathbb{X}_n(u) =& \begin{pmatrix} \mathbf{X}_1^{\intercal} & \mathbf{X}_1^{\intercal}(u_1 - u) \\ \vdots & \vdots \\ \mathbf{X}_n^{\intercal} & \mathbf{X}_n^{\intercal}(u_n - u) \end{pmatrix}, \\
\mathbb{Y}_n =& (y_2, \cdots, y_{n+1})^{\intercal}, \\
\mathbb{W}_n(u) =& \mathsf{diag}\left\{K_b(u_1, u), \cdots, K_b(u_n, u)\right\},
\end{aligned}$$

with $K_b(u_t, u) = K\left((u_t - u)/b\right)$, where $K(\cdot)$ is a kernel function and b is a bandwidth. Then a local linear estimate of $\boldsymbol{\beta}_H(u)$ can be constructed as

$$\tilde{\boldsymbol{\beta}}_H(u) = (\mathbf{I}_{k_0+d_0}, \mathbf{O}_{k_0+d_0})\left[\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{X}_n(u)\right]^{-1}\left[\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{Y}_n\right], \tag{3.5}$$

where u is on the support of the index variable u_t, $\mathbf{I}_{k_0+d_0}$ and $\mathbf{O}_{k_0+d_0}$ are identity and zero matrices of dimension $(k_0+d_0)\times(k_0+d_0)$, respectively. However, the local linear estimation in (3.5) is infeasible and cannot be implemented directly as the factor regressors involved in $\mathbf{X}_t$ are not observable. In practice, we have to replace $\mathbf{X}_t$ by $\hat{\mathbf{X}}_t = \left(\hat{\mathbf{F}}_t^\intercal, \mathbf{Y}_t^\intercal\right)^\intercal$ with $\hat{\mathbf{F}}_t$ obtained in Stage 1, and let $\hat{\mathbb{X}}_n(u)$ be defined as $\mathbb{X}_n(u)$ with $\mathbf{X}_t$ replaced by $\hat{\mathbf{X}}_t$. Then, we obtain the following feasible local linear estimate of $\boldsymbol{\beta}_H(u)$:

$$\hat{\boldsymbol{\beta}}_H(u) = (\mathbf{I}_{k_0+d_0}, \mathbf{O}_{k_0+d_0}) \left[\hat{\mathbb{X}}_n^\intercal(u)\mathbb{W}_n(u)\hat{\mathbb{X}}_n(u)\right]^{-1} \left[\hat{\mathbb{X}}_n^\intercal(u)\mathbb{W}_n(u)\mathbb{Y}_n\right]. \tag{3.6}$$

In Chapter 4, we will compare the finite-sample performance between the feasible and infeasible local linear estimators. In Chapter 5, we will show that the feasible local linear estimator $\hat{\boldsymbol{\beta}}_H(u)$ has the same asymptotic distribution as the infeasible one $\tilde{\boldsymbol{\beta}}_H(u)$.

3.3 One-step Ahead Nonlinear Forecast and Bootstrap Method

Construction of one-step or multi-step ahead forecasting and the relevant prediction interval is an important issue in time series analysis. In this section, we will first introduce a one-step ahead nonlinear forecasting technique by using the estimates of the rotated coefficient functions, and then we construct the one-step-ahead prediction interval by using the wild bootstrap procedure.

Given the observations $(y_{t+1}, u_t, \mathbf{Z}_t)$ with $t = 1, \cdots, n-1$ and $(u_n, \mathbf{Z}_n)$, with the feasible local linear estimation introduced in Section 3.2, we may obtain the one-step ahead predicted value of y_{n+1}:

$$\hat{y}_{n+1|n} = \hat{\boldsymbol{\beta}}_{H,n-1}^\intercal(u_n)\hat{\mathbf{X}}_n, \tag{3.7}$$

where $\hat{\boldsymbol{\beta}}_{H,n-1}(\cdot)$ is the local linear estimate as in (3.6) using only the sample $(y_{t+1}, u_t, \mathbf{Z}_t)$, $t = 1, \cdots, n-1$. In fact, $\hat{y}_{n+1|n}$ defined in (3.7) can be regarded

as an estimate of

$$y_{n+1|n} = \mathbf{F}_n^\intercal \boldsymbol{\beta}_1(u_n) + \mathbf{Y}_n^\intercal \boldsymbol{\beta}_2(u_n) = \boldsymbol{\beta}_H^\intercal(u_n)\mathbf{X}_n.$$

The asymptotic property of $\hat{y}_{n+1|n}$ will be given in Theorem 2 in Chapter 5. In practice, it is often of interest to further construct the prediction interval of $y_{n+1|n}$, which is helpful to examine prediction stability. For given $0 < \alpha < 1$, the $(1-\alpha)$ prediction interval of $y_{n+1|n}$ can be defined by

$$\left[\hat{y}_{n+1|n} - c_{\alpha/2} \times \sqrt{\hat{\mathsf{var}}(\hat{y}_{n+1|n})},\ \hat{y}_{n+1|n} + c_{\alpha/2} \times \sqrt{\hat{\mathsf{var}}(\hat{y}_{n+1|n})}\right], \tag{3.8}$$

where $c_{\alpha/2}$ is the upper $\alpha/2$-percentile of $\left(\hat{y}_{n+1|n} - y_{n+1|n}\right)/\sqrt{\hat{\mathsf{var}}(\hat{y}_{n+1|n})}$ and $\hat{\mathsf{var}}(\hat{y}_{n+1|n})$ is the estimate of the variance of $\hat{y}_{n+1|n}$. However, (3.8) cannot be directly used as neither $c_{\alpha/2}$ nor $\hat{\mathsf{var}}(\hat{y}_{n+1|n})$ is known. We next use a wild bootstrap procedure to estimate $c_{\alpha/2}$ and $\hat{\mathsf{var}}(\hat{y}_{n+1|n})$, and then proceed to construct a feasible prediction interval.

We next describe five steps to determine $c_{\alpha/2}$ and $\hat{\mathsf{var}}(\hat{y}_{n+1|n})$, which are crucial to construct the one-step ahead prediction interval of $y_{n+1|n}$.

STEP 1: Using the observations $(y_{t+1}, u_t, \mathbf{Z}_t)$, $t = 1, \cdots, n-1$, we estimate the rotated coefficient functions $\boldsymbol{\beta}_H(\cdot)$ by the local linear smoothing method (3.6), and denote the resulting estimates by $\hat{\boldsymbol{\beta}}_{H,n-1}(u_t)$ for $t = 1, \cdots, n$. Let $\hat{\mathbf{X}}_t$, $t = 1, \cdots, n$, be defined as in Section 3.2, where $\hat{\mathbf{F}}_t$ is obtained by implementing PCA on the observations of $\mathbf{Z}_t$, $t = 1, \cdots, n$. Construct the one-step ahead forecast $\hat{y}_{n+1|n}$ as in (3.7).

STEP 2: Generate the bootstrap sample:

$$y_{t+1}^* = \hat{\mathbf{X}}_t^\intercal \hat{\boldsymbol{\beta}}_{H,n-1}(u_t) + \epsilon_{t+1}^*, \quad t = 1, \cdots, n-1,$$

where $\epsilon_{t+1}^* = \tilde{\epsilon}_{t+1} \cdot \eta_{t+1}$, $\{\eta_t\}$ is a sequence of independent and identically distributed (*i.i.d.*) random variables drawn from a pre-specified distribution with mean zero and unit variance, such as $\mathsf{N}(0,1)$, and $\tilde{\epsilon}_{t+1} = \hat{\epsilon}_{t+1} - \sum_{t=1}^{n-1} \hat{\epsilon}_{t+1}/(n-1)$ with $\hat{\epsilon}_{t+1} = y_{t+1} - \hat{\mathbf{X}}_t^\intercal \hat{\boldsymbol{\beta}}_{H,n-1}(u_t)$.

STEP 3: As in Step 1, use the generated data set $\{(y^*_{t+1}, u_t, \hat{\mathbf{X}}_t) : t = 1, \cdots, n-1\}$ to re-estimate the rotated coefficient functions at u_t, $t = 1, \cdots, n$, and denote the resulting estimates as $\hat{\boldsymbol{\beta}}^*_{H,n-1}(u_t)$. Construct the one-step ahead forecast:

$$\hat{y}_{n+1|n}(1) = \left[\hat{\boldsymbol{\beta}}^*_{H,n-1}(u_n)\right]^{\intercal} \hat{\mathbf{X}}_n.$$

STEP 4: Repeat Steps 2 and 3 for M times and obtain M bootstrap one-step ahead predicted values, $\hat{y}_{n+1|n}(i)$, $i = 1, \cdots, M$. The estimate of the variance of $\hat{y}_{n+1|n}$ is obtained via the sample variance of $\left\{\hat{y}_{n+1|n}(i) : i = 1, \cdots, M\right\}$ and is denoted by $\hat{\mathsf{var}}^*(\hat{y}_{n+1|n})$.

STEP 5: For each $i = 1, \cdots, M$, use the sequence $\hat{y}_{n+1|n}(i)$ and $\hat{\mathsf{var}}^*(\hat{y}_{n+1|n})$ to compute $q^*_n(i) = \left[\hat{y}_{n+1|n}(i) - \hat{y}_{n+1|n}\right] / \sqrt{\hat{\mathsf{var}}^*(\hat{y}_{n+1|n})}$, and then obtain the estimate of $c_{\alpha/2}$ by calculating the upper $\alpha/2$-percentile of $\{q^*_n(i) : i = 1, \cdots, M\}$. We denote the estimate of $c_{\alpha/2}$ by $\hat{c}^*_{\alpha/2}$.

Using $c^*_{\alpha/2}$ and $\mathsf{var}^*(\hat{y}_{n+1|n})$ obtained in the above wild bootstrap procedure, we can construct the feasible $(1-\alpha)$ prediction interval of $y_{n+1|n}$ by

$$\left[\hat{y}_{n+1|n} - c^*_{\alpha/2} \times \sqrt{\hat{\mathsf{var}}^*(\hat{y}_{n+1|n})},\ \hat{y}_{n+1|n} + c^*_{\alpha/2} \times \sqrt{\hat{\mathsf{var}}^*(\hat{y}_{n+1|n})}\right]. \tag{3.9}$$

The above bootstrap method to construct the prediction interval can be seen as a nonparametric generalisation of the bootstrap prediction interval introduced by Gonçalves et al. (2017) for the parametric linear factor-augmented regression model. As in Teräsvirta et al. (2010), here we let the auto-regressors (lags of the response) in the time series model be invariant when generating the bootstrap samples, although other bootstrap methods may also be applicable in our setting. We refer to Zhang and Peng (2010) and Chen et al. (2018) for some recent development on constructing the point-wise or simultaneous confidence bands in the varying-coefficient models.

3.4 Selection of the Bandwidth and the Number of Lags

When the coefficient functions in model (3.3) are estimated by using the local linear smoothing method, the accuracy of the resulting estimates depends on the choice of bandwidth and the number of lags, both of which are unknown in practice. In this section, we first introduce a bandwidth selection method proposed by Cai et al. (2000) for selecting the optimal bandwidth b, and then a forward selection criterion to determine the number of lags d_0.

3.4.1 Selection of the Bandwidth

An important issue in the developed local linear estimation and one-step ahead nonlinear forecasting approaches is the choice of bandwidth b. Hence, we need to carefully choose an appropriate bandwidth to implement our proposed method. As the time series process is assumed to be stationary and weakly dependent (see Assumption 1 in Chapter 5), the classical cross-validation method introduced in Section 2.1.3 is not applicable in our setting. We next use a modified multi-fold cross-validation criterion proposed by Cai et al. (2000) to determine an appropriate bandwidth for time series data. This criterion will be used in the numerical studies in Chapter 4.

Let m and Q be two positive integers such that $n > mQ$. To determine an optimal bandwidth, we consider using Q sub-samples of time series, each of length $n - qm$, $q = 1, 2, \ldots, Q$, to estimate the coefficient functions. Then, we construct the one-step ahead nonlinear forecast for each "out-sample" with length m by using the estimated FA-FCM based on the time series sub-sample observed before the out-sample, and calculate the mean squared prediction errors. Specifically, we defined the following average mean squared error:

$$\mathsf{AMS}(b) = \sum_{q=1}^{Q} \mathsf{AMS}_q(b), \tag{3.10}$$

where

$$\mathsf{AMS}_q(b) = \frac{1}{m}\sum_{t=n-qm+1}^{n-qm+m}\left[y_{t+1} - \hat{\boldsymbol{\beta}}_{H,q}^{\intercal}(u_t)\hat{\mathbf{X}}_t\right]^2,$$

and $\hat{\boldsymbol{\beta}}_{H,q}(\cdot)$ and $\hat{\mathbf{X}}_t$ are computed using the sub-sample of time series observations $(y_{t+1}, u_t, \mathbf{Z}_t)$, $1 \leqslant t \leqslant n - qm$ in (3.6) with the bandwidth re-scaled to be $b \cdot [n/(n-qm)]^{1/5}$. In our numerical studies, as suggested by Cai et al. (2000), we use $m = \lfloor 0.1n \rfloor$ and $Q = 4$, where $\lfloor \cdot \rfloor$ denotes the floor function.

3.4.2 Selection of the Number of Lags

In order to implement the proposed estimation and forecasting method, we need to specify d_0, the number of lags in the predictive regression model (3.3). The model selection issue in parametric linear factor-augmented models is recently studied by Djogbenou (2017). However, his selection criterion is not applicable to our nonparametric model setting. In this section, We use the forward selection criterion as a screening tool and employ the *Bayesian Information Criteria (BIC)* as the stopping rule to estimate d_0, motivated by the forward selection method introduced by Wang (2009) and Cheng et al. (2016) for the high-dimensional regression models without any latent factor regressor.

If the number of lags is assumed to be d, a positive integer, we let $\hat{\boldsymbol{\beta}}_H(u|d)$ be the feasible local linear estimate of the rotated coefficient function $\boldsymbol{\beta}_H(u)$, similar to the definition of $\hat{\boldsymbol{\beta}}_H(u)$ given in (3.6). Let $\hat{\mathbf{X}}_t(d)$ be defined as $\hat{\mathbf{X}}_t$ when the number of lags is d. In the forward selection procedure, we examine the change of residual sum of squares defined by

$$\hat{\sigma}_n^2(d) = \frac{1}{n}\sum_{t=1}^{n}\left[y_{t+1} - \hat{\boldsymbol{\beta}}_H^{\intercal}(u_t|d)\hat{\mathbf{X}}_t(d)\right]^2,$$

and compute the BIC value as in Wang and Xia (2009):

$$\mathsf{BIC}(d) = \log\hat{\sigma}_n^2(d) + d \cdot \frac{\log(nb)}{nb}, \tag{3.11}$$

when the lags of response are sequentially added to the FA-FCM. Specifically,

we start with the predictive regression model without any lag ($d = 0$) in which case $\mathsf{BIC}(0) = \log \hat{\sigma}_n^2(0)$. In the second step, we add the first lag to the model and compute $\mathsf{BIC}(1)$. Continue this forward procedure by adding one lag in each step, and determine the number of lags in the proposed FA-FCM as $\hat{d}$ so that $\mathsf{BIC}(\hat{d}+1) > \mathsf{BIC}(\hat{d})$. Our numerical studies show that the above forward procedure selects the number of lags accurately in finite samples.

3.5 FA-FCM with Multivariate Response

The parametric linear vector auto-regressive models have been commonly applied in analysing multiple macroeconomic time series data (Sims, 1980; Lütkepohl, 2005). In recent years, to deal with high-dimensional time series and achieve dimension reduction, there has been increasing interest on combining the approximate factor model and linear vector auto-regression and studying the so-called factor-augmented vector auto-regressive models. The latter modelling approach is first introduced by Bernanke et al. (2005), and has been extensively studied in recent years (c.f., Bai and Ng, 2006; Bai et al., 2016).

In this section, we aim to make a further extension of the factor-augmented vector auto-regression by allowing the coefficient matrices to vary with an index variable, and generalise the FA-FCM and the relevant methodologies developed in the Sections 3.2 and 3.3 to the case of multiple response variables. Specifically, suppose (3.2) and

$$\bar{\mathbf{y}}_{t+1} = \mathbf{B}_0^{\intercal}(u_t)\mathbf{F}_t + \sum_{k=1}^{d_1} \mathbf{B}_k^{\intercal}(u_t)\bar{\mathbf{y}}_{t+1-k} + \bar{\boldsymbol{\epsilon}}_{t+1}, \tag{3.12}$$

where $\bar{\mathbf{y}}_{t+1}$ is p_0-dimensional column vector of response variables, $\mathbf{B}_0(\cdot)$ is a $k_0 \times p_0$ matrix of coefficient functions and $\mathbf{B}_k(\cdot)$ is a $p_0 \times p_0$ matrix of coefficient functions, $k = 1, \cdots, d_1$, and $\bar{\boldsymbol{\epsilon}}_t$ is a p_0-dimensional column vector of errors.

Model (3.12) generalises the multivariate functional-coefficient time series model proposed by Jiang (2014) which excludes the latent factor regressors in the predictive model. To simplify the discussion, we assume that the dimension p_0 is

fixed.

Consider estimating the matrices of coefficient functions by using the local linear method as in Section 3.2. Let

$$\begin{aligned}\bar{\mathbf{X}}_t &= \left[(\mathbf{H}\mathbf{F}_t)^\intercal, \bar{\mathbf{y}}_t^\intercal, \cdots, \bar{\mathbf{y}}_{t-d_1}^\intercal\right]^\intercal, \\ \bar{\mathbf{B}}_H(\cdot) &= \left[\mathbf{B}_0^\intercal(\cdot)\mathbf{H}^{-1}, \mathbf{B}_1^\intercal(\cdot), \cdots, \mathbf{B}_{d_1}^\intercal(\cdot)\right],\end{aligned}$$

and then rewrite (3.12) as

$$\bar{\mathbf{y}}_{t+1} = \bar{\mathbf{B}}_H(u_t)\bar{\mathbf{X}}_t + \bar{\boldsymbol{\epsilon}}_{t+1}. \tag{3.13}$$

As the rotated factor regressors $\mathbf{H}\mathbf{F}_t$ are latent, to develop a feasible nonparametric estimation method, we have to replace $\mathbf{H}\mathbf{F}_t$ by the PCA estimate $\hat{\mathbf{F}}_t$ defined in Stage 1 of the estimation procedure introduced in Section 3.2. Let $\tilde{\mathbb{Y}}_n = (\bar{\mathbf{y}}_2, \cdots, \bar{\mathbf{y}}_{n+1})$ and

$$\tilde{\mathbb{X}}_n(u) = \begin{pmatrix} \tilde{\mathbf{X}}_1^\intercal & \tilde{\mathbf{X}}_1^\intercal(u_1 - u) \\ \vdots & \vdots \\ \tilde{\mathbf{X}}_n^\intercal & \tilde{\mathbf{X}}_n^\intercal(u_n - u) \end{pmatrix},$$

where $\tilde{\mathbf{X}}_t$ is defined as $\bar{\mathbf{X}}_t$ but with $\mathbf{H}\mathbf{F}_t$ replaced by $\hat{\mathbf{F}}_t$. Then, similarly to (3.6), we can obtain the following feasible local linear estimates of $\bar{\mathbf{B}}_H(u)$ and its derivative $\bar{\mathbf{B}}'_H(u)$:

$$\left[\tilde{\mathbf{B}}_H(u), \tilde{\mathbf{B}}'_H(u)\right] = \left[\tilde{\mathbb{Y}}_n\mathbb{W}_n(u)\tilde{\mathbb{X}}_n(u)\right]\left[\tilde{\mathbb{X}}_n^\intercal(u)\mathbb{W}_n(u)\tilde{\mathbb{X}}_n(u)\right]^{-1}, \tag{3.14}$$

where $\mathbb{W}_n(u)$ is defined in Section 3.2. As in Theorem 1 to be given in Chapter 5, under some regularity conditions, we may prove that the above local linear estimators are asymptotically equivalent to those directly using the unobservable rotated factor regressors $\mathbf{H}\mathbf{F}_t$.

Finally, the one-step ahead nonlinear forecast of $\bar{\mathbf{y}}_{n+1}$ can be constructed by

following that in Section 3.3, i.e.,

$$\tilde{\mathbf{y}}_{n+1|n} = \tilde{\mathbf{B}}_{H,n-1}(u_n)\tilde{\mathbf{X}}_n, \tag{3.15}$$

where $\tilde{\mathbf{B}}_{H,n-1}(\cdot)$ is the local linear estimate as constructed in (3.14) using the sample $(\bar{\mathbf{y}}_{t+1}, u_t, \mathbf{Z}_t)$, $t = 1, \cdots, n-1$. The wild bootstrap method introduced in Section 3.3 can be used to construct the prediction intervals for the multiple response.

Chapter 4

Numerical Studies

In this chapter, we provide two simulated examples and one empirical application to evaluate the finite-sample performance of the methods proposed in Chapter 3.

4.1 Simulation studies

In this section, we give the simulation studies under two different settings. The first simulated example is used to examine the accuracy of the proposed model with univariate response. The second simulated example is conducted to examine the performance of the model with bivariate response. Throughout the simulation studies, the kernel function is chosen as the Epanechnikov kernel introduced in Section 2.1.1.

4.1.1 Example 1 with Univariate Response

In this section, we consider the following univariate FA-FCM:

$$y_{t+1} = \mathbf{F}_t^\intercal \boldsymbol{\beta}_1(u_t) + \mathbf{Y}_t^\intercal \boldsymbol{\beta}_2(u_t) + \sigma \cdot \epsilon_{t+1}, \;\; t = 1, \cdots, n, \tag{4.1}$$

where $\boldsymbol{\beta}_1(u) = [\beta_{11}(u), \cdots, \beta_{1k_0}(u)]^\intercal$ and $\boldsymbol{\beta}_2(u) = [\beta_{21}(u), \cdots, \beta_{2d_0}(u)]^\intercal$ are two column vectors of coefficient functions, $k_0 = 4$, $d_0 = 3$, $\beta_{11}(u) = \sin(u)$, $\beta_{12}(u) = \cos(u)$, $\beta_{13}(u) = \sqrt{u}$, $\beta_{14}(u) = \log(1+u)$, $\beta_{21}(u) = \frac{1}{4}\sin(u)$, $\beta_{22}(u) = \frac{1}{4}\cos(u)$ and

$\beta_{23}(u) = \frac{1}{4}$. The observations of the index variable u_t are independently generated from $\mathsf{U}(0,1)$, the model errors ϵ_t are independently generated from $\mathsf{N}(0,1)$, σ is chosen as 0.2, $\mathbf{F}_t = (F_{t1}, \cdots, F_{t4})^\intercal$ is a four-dimensional latent common factor vector with each factor component generated from the following $\mathsf{AR}(1)$ process:

$$F_{tk} = 0.5 \cdot F_{t-1,k} + z_{tk}, \quad k = 1, 2, 3, 4, \tag{4.2}$$

where z_t are *i.i.d.* and follow the standard normal distribution. In addition, we use the factor model structure (3.2) to generate the exogenous observations $\mathbf{Z}_t$, where each row of the $q_n \times k_0$ factor loading matrix $\mathbb{B}_n$ is independently generated from $\mathsf{N}(\mathbf{0}_4, \mathbf{I}_4)$ and the idiosyncratic error vector $\mathbf{V}_t \sim \mathsf{N}(\mathbf{0}_{q_n}, \mathbf{I}_{q_n})$. The sample size n is set as 200, 500 and 1000, whereas the dimension q_n is set as 20, 150 and 500. The replication number is 200.

In the simulation, we use the observations $(y_{t+1}, u_t, \mathbf{Z}_t)$ to construct the local linear estimates of the rotated coefficient functions $\boldsymbol{\beta}_1(u)$ and $\boldsymbol{\beta}_2(u)$ in the predictive model (4.1), and then obtain the one-step ahead nonlinear forecast of the response as introduced in Section 3.3.

The common factors $\mathbf{F}_t$ and its number are usually unknown in practice. In this simulation, the latent factors are estimated by the PCA technique introduced in Section 2.3.1, and the number of factors is determined by choosing the first few eigenvectors of $\mathbb{Z}_n\mathbb{Z}_n^\intercal/(nq_n)$ (corresponding to the first few largest eigenvalues) such that 80% of the total variation is accounted for. The latter criterion is also used by Chen et al. (2018) and performs well in our simulation. Table 4.1 shows the estimation result for the factor number. We find that when the dimension q_n is as large as 150, the number of common factors can be very accurately estimated over 200 replications.

The bandwidth in the developed local linear smoothing method is determined by the modified multi-fold cross-validation method introduced in Section 3.4.1. Furthermore, to save the computational time, for each combination of n and q_n, we only compute the average of the bandwidths (minimising the AMS function) over

Table 4.1: The percentages of correctly selecting the factor number in 200 replications in Example 1

q_n \ n	200	500	1000
20	80.0%	83.5%	87.0%
150	100.0%	100.0%	100.0%
500	100.0%	100.0%	100.0%

20 replications for both the infeasible and feasible local linear estimations. Then we use this average value as the optimal bandwidth in our simulation. The simulated results for the optimal bandwidths and minimum of AMS are summarised in the following tables.

Table 4.2: The average of the bandwidth for the infeasible local linear estimations in Example 1

q_n \ n	200	500	1000
20	0.5650	0.3425	0.3150
150	0.6025	0.4325	0.2825
500	0.5750	0.4100	0.2850

Table 4.3: The average of minimum AMS for the infeasible local linear estimations in Example 1

q_n \ n	200	500	1000
20	0.189653	0.169835	0.163798
150	0.191634	0.165929	0.167936
500	0.199697	0.167018	0.165778

Table 4.4: The average of the bandwidth for the feasible local linear estimations in Example 1

q_n \ n	200	500	1000
20	0.7800	0.5925	0.4325
150	0.6400	0.4375	0.2975
500	0.6000	0.4225	0.2925

Table 4.5: The average of minimum AMS for the feasible local linear estimations in Example 1

q_n \ n	200	500	1000
20	0.670779	0.575377	0.514620
150	0.232345	0.211111	0.212695
500	0.215317	0.181714	0.177993

From the results in Tables 4.2-4.5, the optimal bandwidths and the corresponding minimum AMS values are similar between the feasible local linear estimation and the infeasible when both the sample size n and the dimension q_n are large. For both the infeasible and feasible local linear estimations, we find that as the sample size n increases, the values of the optimal bandwidths decrease.

The forward selection procedure with the BIC stopping rule introduced in Section 3.4.2 is used to determine the number of lags in the predictive model. From Table 4.6, we find that the overall performance of the forward selection method is satisfactory, and it improves as either the dimension q_n or the sample size n increases.

Table 4.6: The percentages of correctly finding the lag number in 200 replications in Example 1

q_n \ n	200	500	1000
20	92.5%	95.5%	98.5%
150	92.5%	98.5%	100.0%
500	97.0%	99.5%	100.0%

We next examine the finite-sample performance of the one-step ahead nonlinear forecasting method constructed in Section 3.3. The simulated sample is split into two parts: the "in-sample" (containing the first 90% of the time series observations) used for estimation, and the "out-sample" (containing the last 10% of the time series observations) used for prediction. The forecasting performance is measured via the following Mean Squared Prediction Error (MSPE):

$$\mathsf{MSPE} = \frac{1}{\lfloor 0.1n \rfloor} \sum_{t=n-\lfloor 0.1n \rfloor}^{n} \left(\hat{y}_{t+1|t} - y_{t+1} \right)^2, \tag{4.3}$$

where $\hat{y}_{t+1|t}$ is defined as in (3.7). For the aim of comparison, we also consider the infeasible local linear estimation defined in (3.5) and use it to construct the infeasible one-step ahead nonlinear forecast.

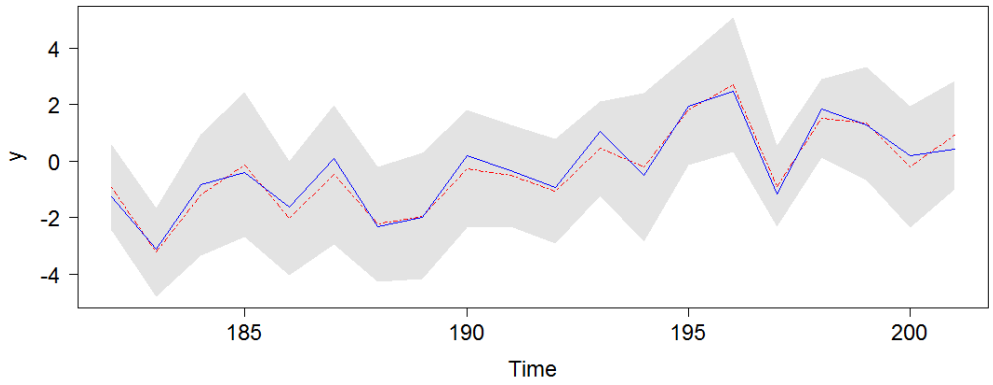
Table 4.7: MSPE of the infeasible one-step ahead forecast in Example 1

q_n \ n	200	500	1000
20	0.046389 (0.015)	0.042255 (0.009)	0.041290 (0.007)
150	0.047819 (0.014)	0.043122 (0.009)	0.041390 (0.006)
500	0.046349 (0.013)	0.042422 (0.009)	0.041388 (0.006)

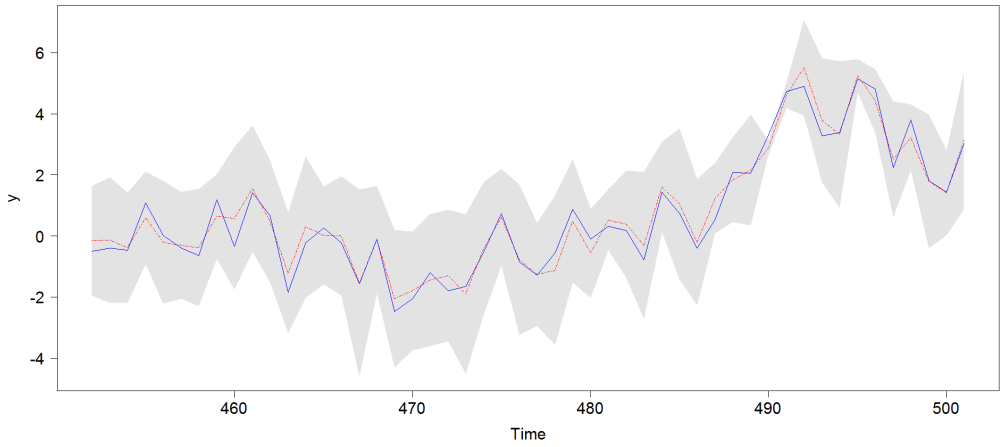
Table 4.8: MSPE of the feasible one-step ahead forecast in Example 1

q_n \ n	200	500	1000
20	0.160698 (0.060)	0.142234 (0.044)	0.141089 (0.041)
150	0.061223 (0.019)	0.054664 (0.011)	0.052289 (0.007)
500	0.049660 (0.015)	0.045416 (0.010)	0.044540 (0.006)

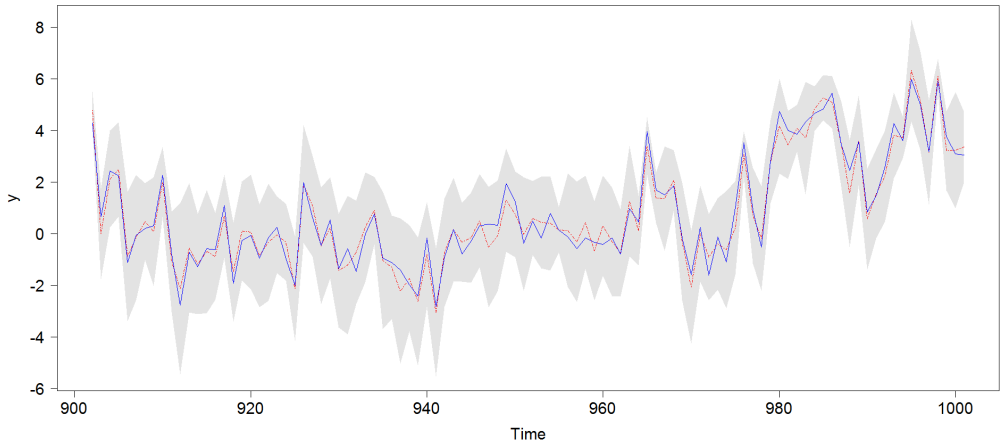
Tables 4.7 and 4.8 give the mean and standard error (in parentheses) of the MSPE values over 200 replications for the infeasible and feasible nonlinear forecasts, respectively. By comparing the MSPE values in the two tables, we may find that although the infeasible nonlinear forecast outperforms the feasible one (which is unsurprising and mainly due to the estimation error in the PCA estimation of the latent factors), the difference becomes very small when the dimension q_n increases to 500. In addition, Figures 4.1-4.3 give the 95% prediction interval in the out-sample forecasting by using the wild bootstrap procedure introduced in Section 3.3 with $q_n = 20, 150$, and 500, respectively.



(a) $n = 200$

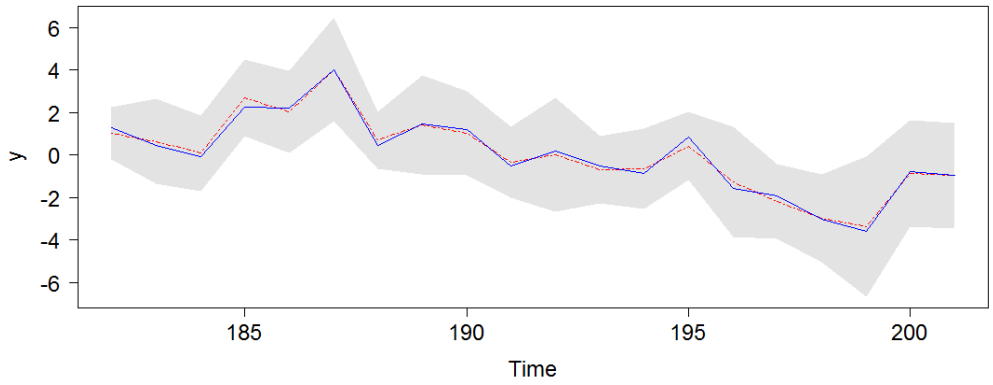


(b) $n = 500$

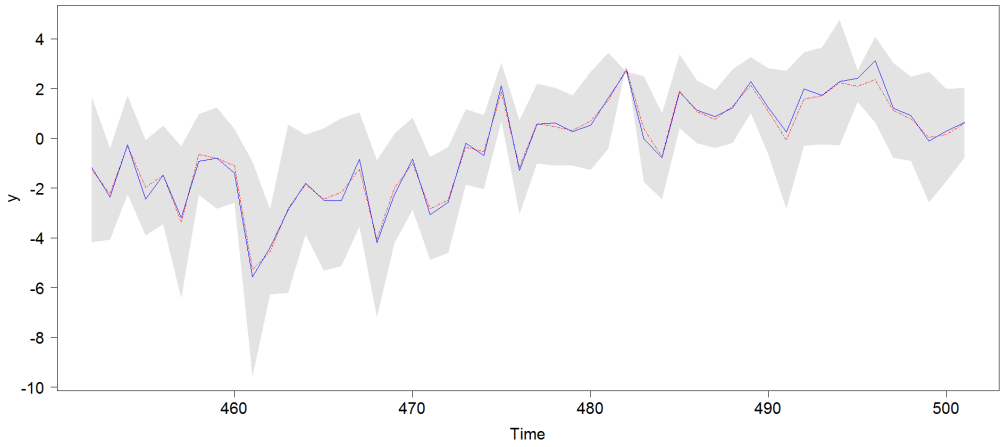


(c) $n = 1000$

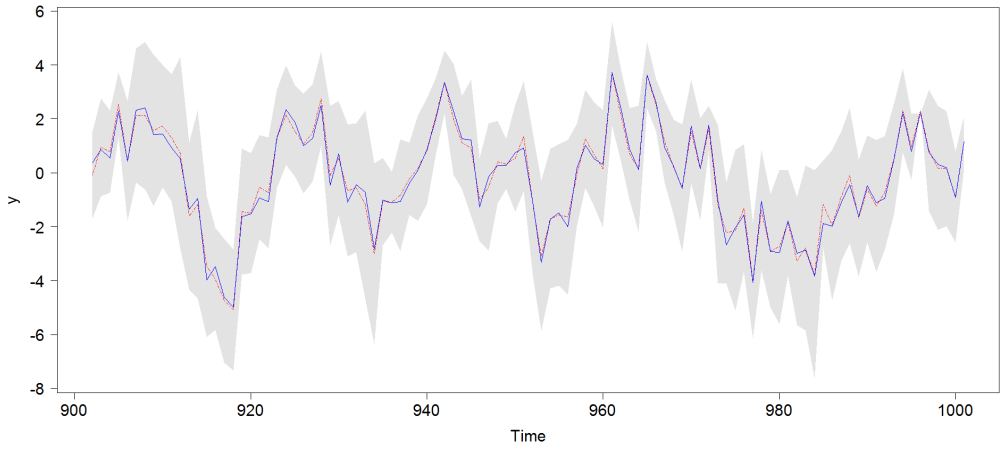
Figure 4.1: The solid curve denotes the true values of the response y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 20$ in Example 1.



(a) $n = 200$

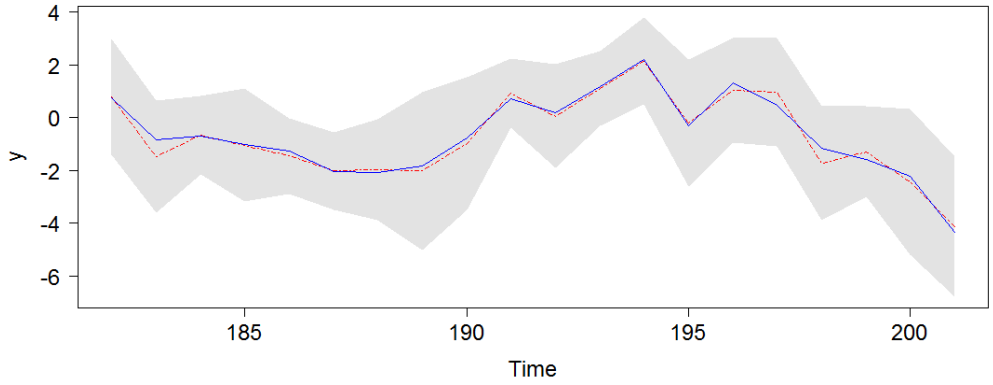


(b) $n = 500$

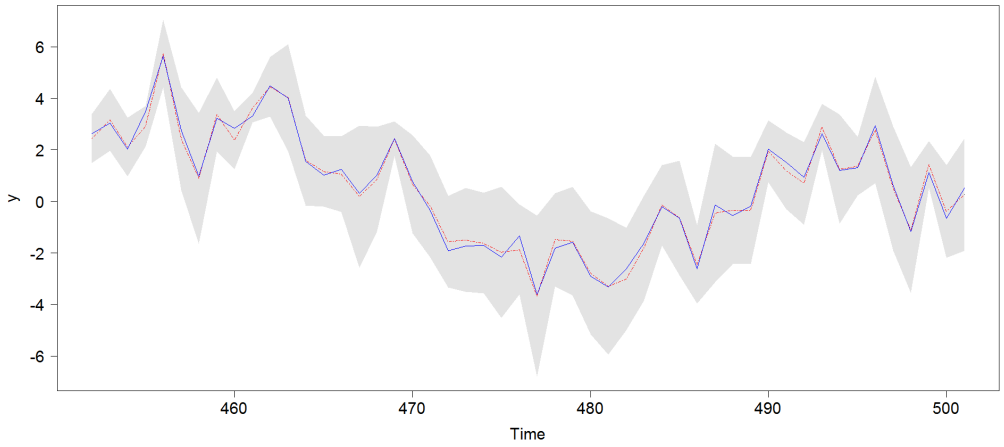


(c) $n = 1000$

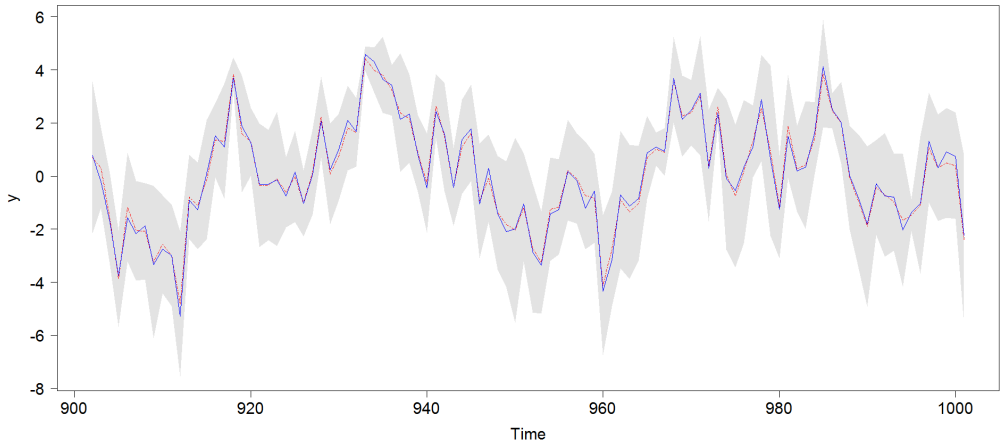
Figure 4.2: The solid curve denotes the true values of the response y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 150$ in Example 1.



(a) $n = 200$



(b) $n = 500$



(c) $n = 1000$

Figure 4.3: The solid curve denotes the true values of the response y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 500$ in Example 1.

4.1.2 Example 2 with Bivariate Response

In this section, we consider the bivariate factor-augmented functional-coefficient model with the following form:

$$\bar{\mathbf{y}}_{t+1} = \mathbf{B}_0^{\intercal}(u_t)\mathbf{F}_t + \sum_{k=1}^{2} \mathbf{B}_k^{\intercal}(u_t)\bar{\mathbf{y}}_{t+1-k} + \bar{\boldsymbol{\epsilon}}_{t+1}, \tag{4.4}$$

where $\bar{\mathbf{y}}_t = (y_{1t}, y_{2t})^{\intercal}$ is a bivariate response vector, $\mathbf{F}_t$ is a four-dimensional latent factor vector generated in the same way as in Example 1, $\bar{\boldsymbol{\epsilon}}_t = (\epsilon_{1t}, \epsilon_{2t})^{\intercal}$ with ϵ_{1t} and ϵ_{2t} being independently generated from $\mathsf{N}(0,1)$, $\mathbf{B}_0(\cdot)$ is a 4×2 matrix of coefficient functions and $\mathbf{B}_k(\cdot)$ is a 2×2 diagonal matrix of coefficient functions, $k = 1, 2$. Specifically,

$$\begin{aligned}
\mathbf{B}_0^{\intercal}(u) &= \begin{bmatrix} B_{0,11}(u) & B_{0,12}(u) & B_{0,13}(u) & B_{0,14}(u) \\ B_{0,21}(u) & B_{0,22}(u) & B_{0,23}(u) & B_{0,24}(u) \end{bmatrix} \\
\mathbf{B}_1^{\intercal}(u) &= \begin{bmatrix} B_{1,11}(u) & B_{1,12}(u) \\ B_{1,21}(u) & B_{1,21}(u) \end{bmatrix}, \quad \mathbf{B}_2^{\intercal}(u) = \begin{bmatrix} B_{2,11}(u) & B_{2,12}(u) \\ B_{2,21}(u) & B_{2,21}(u) \end{bmatrix},
\end{aligned}$$

where $B_{0,11}(u) = \sin(u)$, $B_{0,12}(u) = \cos(u)$, $B_{0,13}(u) = \sqrt{u}$, $B_{0,14}(u) = \log(1+u)$, $B_{0,21}(u) = 2u$, $B_{0,22}(u) = (1+u)^2$, $B_{0,23}(u) = 1/\exp(u)$ and $B_{0,24}(u) = \cos^2(u)$; for $k = 1, 2$, $B_{k,11}(u) = B_{k,22}(u) = 0.2 \cdot I(u \leqslant 0.5) - 0.4 \cdot I(u > 0.5)$ and $B_{k,12}(u) = B_{k,21}(u) = 0.3 \cdot I(u \leqslant 0.5) + 0.2 \cdot I(u > 0.5)$, $I(\cdot)$ denotes the indicator function. The definitions of $\mathbf{B}_1(u)$ and $\mathbf{B}_2(u)$ ensure that the generated bivariate observations $\bar{y}_t$ have a stationary pattern over time. In addition, the generating scheme for the index variable u_t and the exogenous variables $\mathbf{Z}_n$ is the same as that in Example 1. The sample size n is set to be 200, 500 and 1000, and the dimension q_n of $\mathbf{Z}_n$ is set to be 20, 150 and 500.

Tables 4.9 and 4.10 give the percentages of correctly choosing 4 latent factors (which account for at least 80% of the total variation) and the percentages of accurately identifying 2 lags in the model, respectively. Like in the case of univariate response (Example 1), the proposed methods have reliable numerical performance in specifying the predictive model structure.

Table 4.9: The percentages of correctly selecting the factor number in 200 replications in Example 2

q_n \ n	200	500	1000
20	83.5%	85.0%	91.0%
150	100.0%	100.0%	100.0%
500	100.0%	100.0%	100.0%

Table 4.10: The percentages of correctly finding the lag number in 200 replications in Example 2

q_n \ n	200	500	1000
20	74.5%	86.5%	97.5%
150	71.0%	78.5%	89.0%
500	73.0%	81.5%	90.5%

As in Example 1, we apply the multi-fold cross-validation criterion proposed in Section 3.4.1 to select the optimal bandwidth. The optimal bandwidth can be chosen by minimising the AMS function in (3.10). The estimated bandwidths and the minimum values of AMS are summarised in Tables 4.11-4.14.

Table 4.11: The average of the bandwidth for the infeasible local linear estimations in Example 2

q_n \ n	200	500	1000
20	0.1665	0.0795	0.0400
150	0.1645	0.0730	0.0385
500	0.1830	0.0745	0.0430

Table 4.12: The average of minimum AMS for the infeasible local linear estimations in Example 2

q_n \ n	200	500	1000
20	0.795209	0.470228	0.317070
150	0.795775	0.411576	0.338338
500	0.766907	0.494591	0.353485

Table 4.13: The average of the bandwidth for the feasible local linear estimations in Example 2

q_n \ n	200	500	1000
20	0.2105	0.0910	0.0545
150	0.1735	0.0745	0.0405
500	0.1855	0.0745	0.0440

Table 4.14: The average of minimum AMS for the feasible local linear estimations in Example 2

q_n \ n	200	500	1000
20	1.485806	1.018779	0.876313
150	0.877193	0.476728	0.397177
500	0.803621	0.514640	0.370274

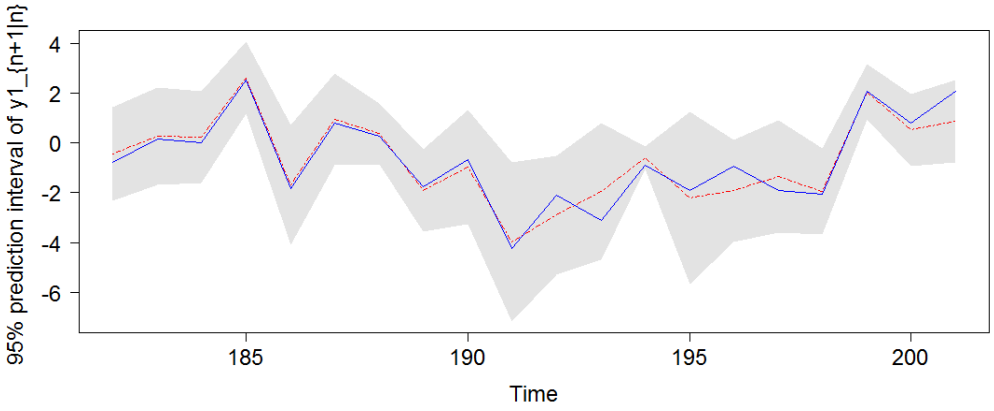
Table 4.15: MSPE of the infeasible one-step ahead forecast in Example 2

q_n \ n	200	500	1000
20	0.703400 (0.281)	0.649507 (0.190)	0.624566 (0.136)
150	0.744730 (0.368)	0.648674 (0.177)	0.628194 (0.133)
500	0.703471 (0.324)	0.646440 (0.191)	0.614113 (0.131)

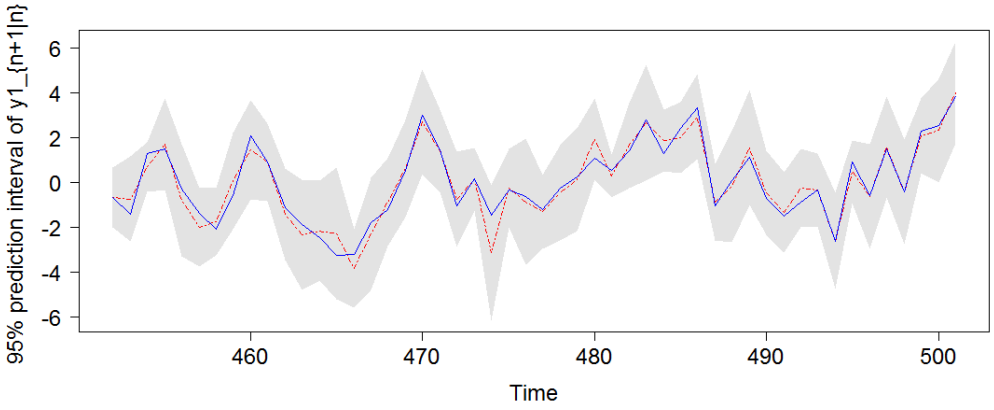
Table 4.16: MSPE of the feasible one-step ahead forecast in Example 2

q_n \ n	200	500	1000
20	0.832918 (0.325)	0.751301 (0.212)	0.717344 (0.144)
150	0.761248 (0.369)	0.659639 (0.178)	0.641022 (0.135)
500	0.709743 (0.325)	0.650851 (0.191)	0.617325 (0.131)

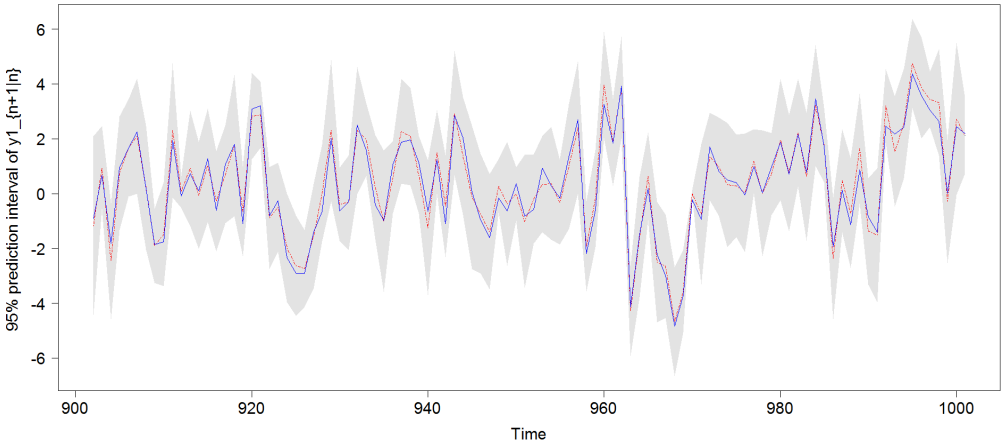
As in Example 1, we also split the simulated sample into the in-sample (with the first 90% of the time series observations) used for estimation and out-sample (with the last 10% of the time series observations) used for prediction. To measure the forecasting accuracy, we compute the accumulated MSPE for both $y_{1,t+1}$ and $y_{2,t+1}$ within the out-sample, where both the infeasible and feasible local linear estimation of the rotated coefficient functions are considered. The relevant MSPE values are given in Tables 4.15 and 4.16, from which we may find that the nonlinear forecast using the feasible local linear estimation has similar forecasting accuracy to the infeasible one assuming the latent factors are known a priori (in particular when the dimension p_n is as large as 150). In addition, Figures 4.4-4.9 give the 95% prediction interval for $y_{1,t+1}$ and $y_{2,t+1}$, respectively, in the out-sample forecasting, using the bootstrap procedure.



(a) $n = 200$

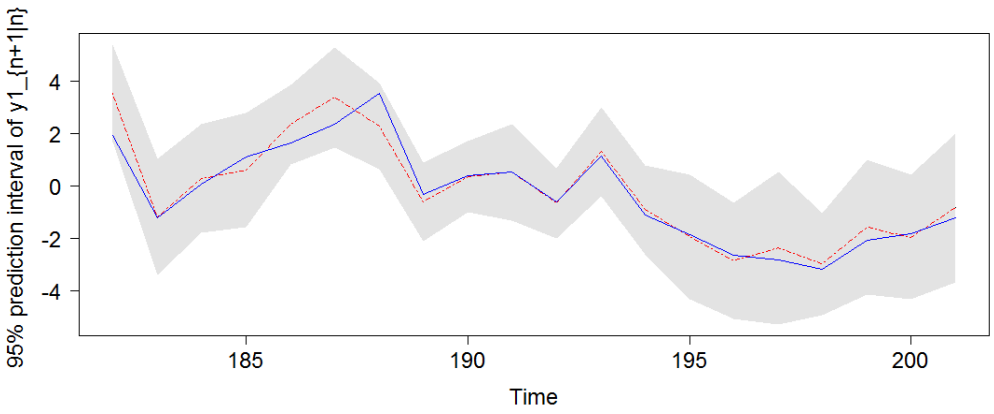


(b) $n = 500$

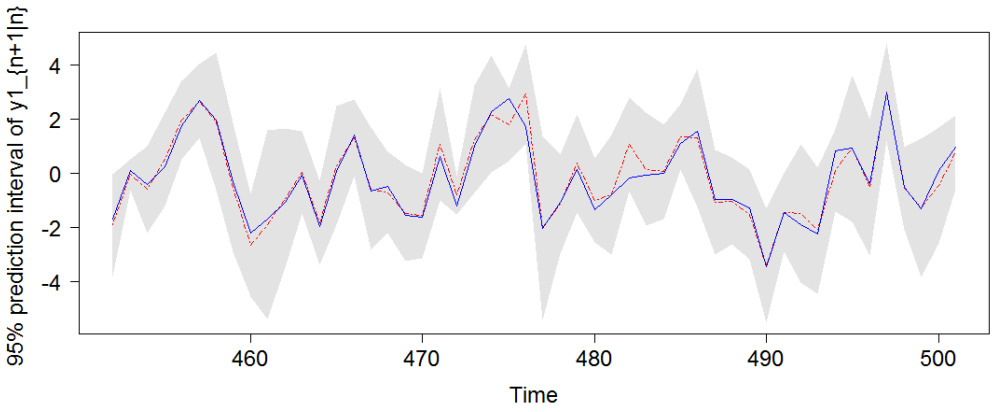


(c) $n = 1000$

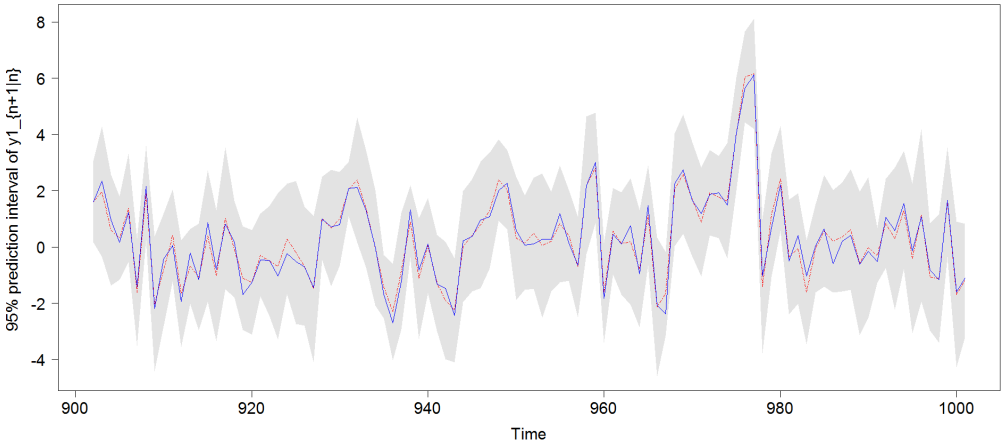
Figure 4.4: The solid curve denotes the true values of the response $y_{1,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 20$ in Example 2.



(a) $n = 200$

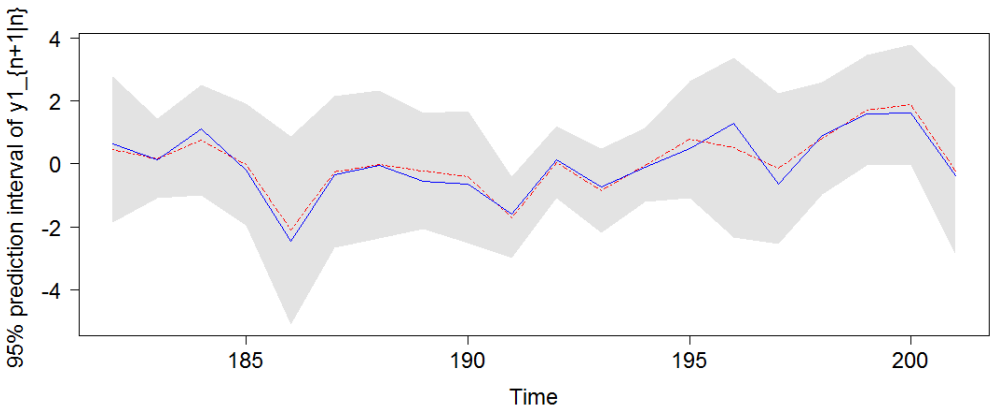


(b) $n = 500$

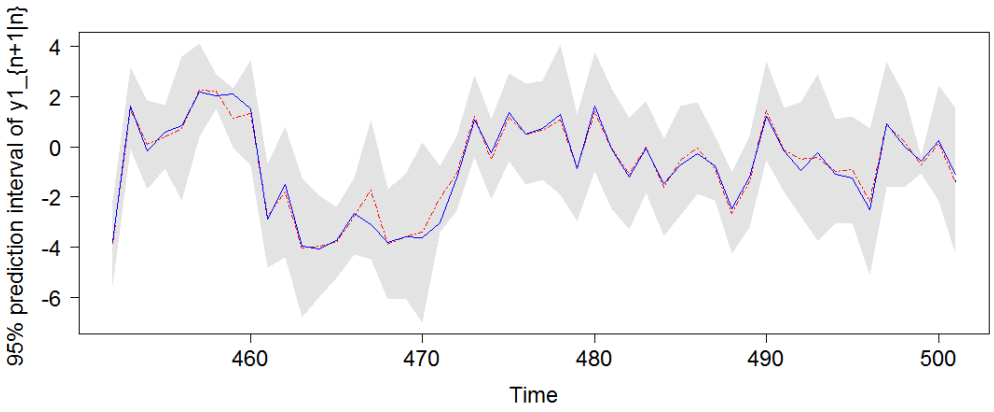


(c) $n = 1000$

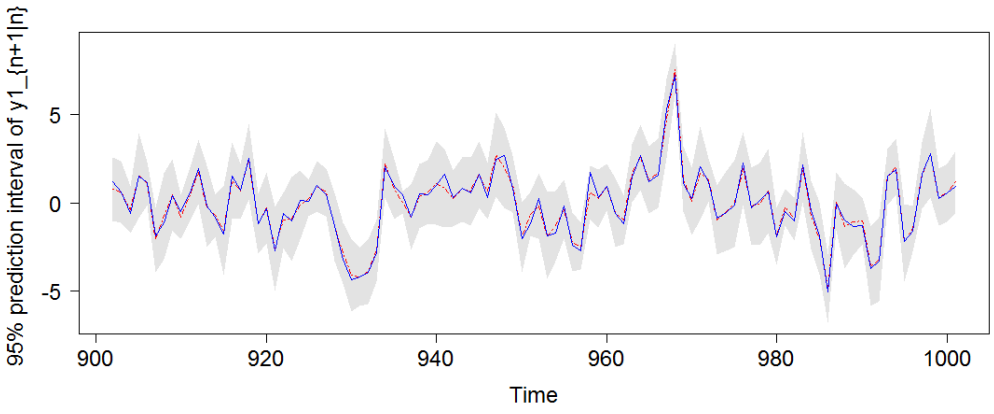
Figure 4.5: The solid curve denotes the true values of the response $y_{1,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 150$ in Example 2.



(a) $n = 200$

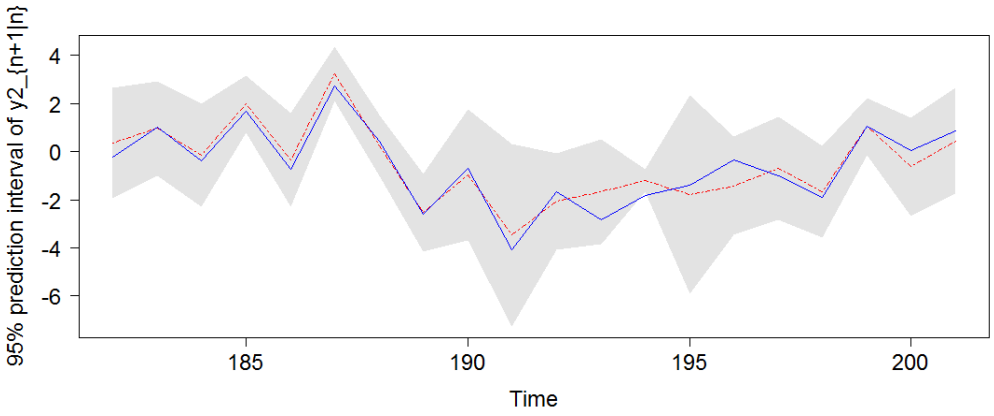


(b) $n = 500$

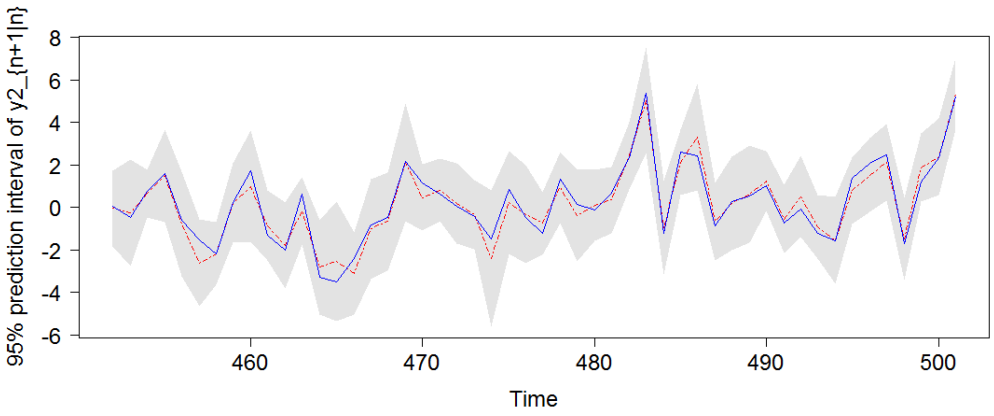


(c) $n = 1000$

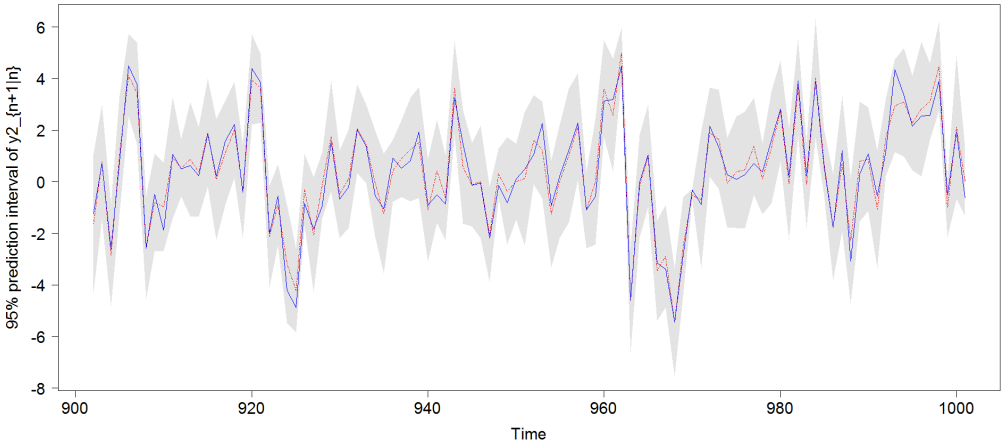
Figure 4.6: The solid curve denotes the true values of the response $y_{1,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 500$ in Example 2.



(a) $n = 200$

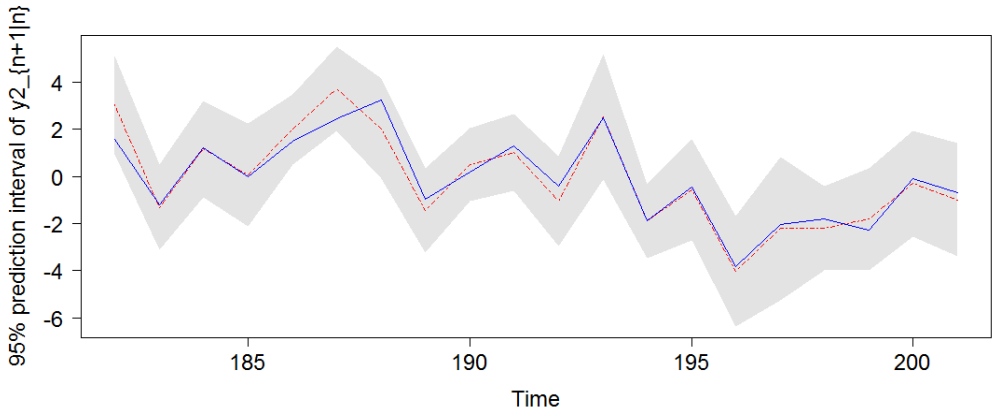


(b) $n = 500$

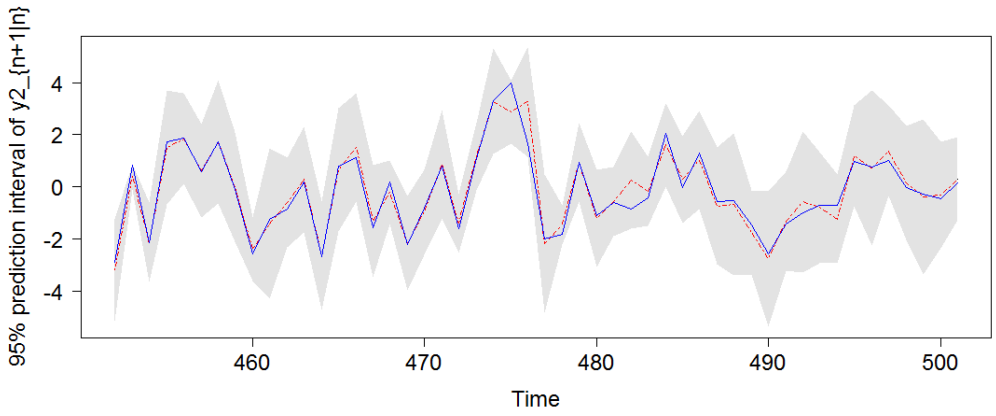


(c) $n = 1000$

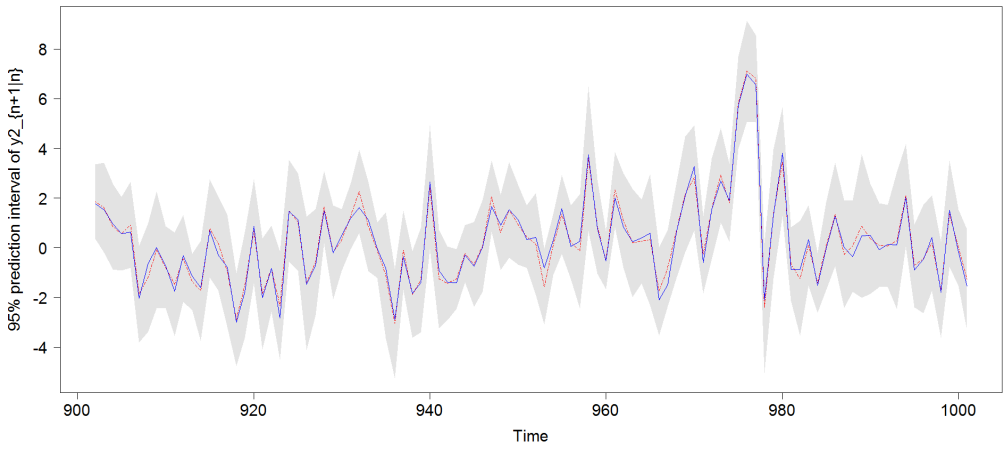
Figure 4.7: The solid curve denotes the true values of the response $y_{2,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 20$ in Example 2.



(a) $n = 200$

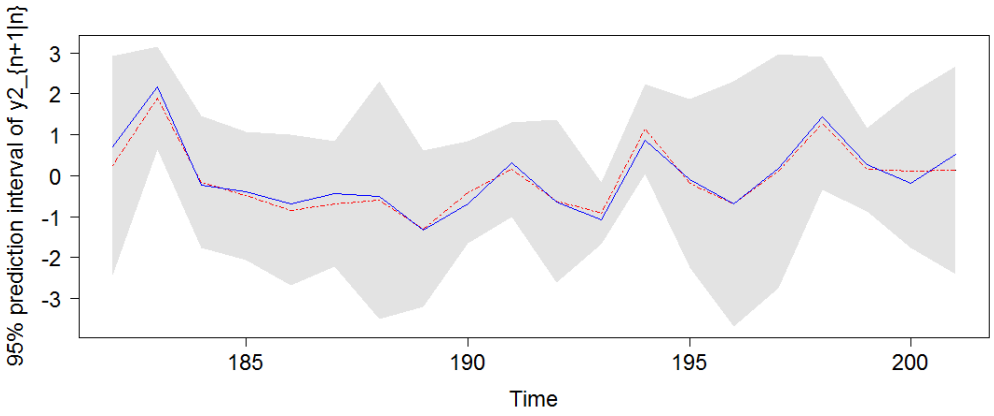


(b) $n = 500$

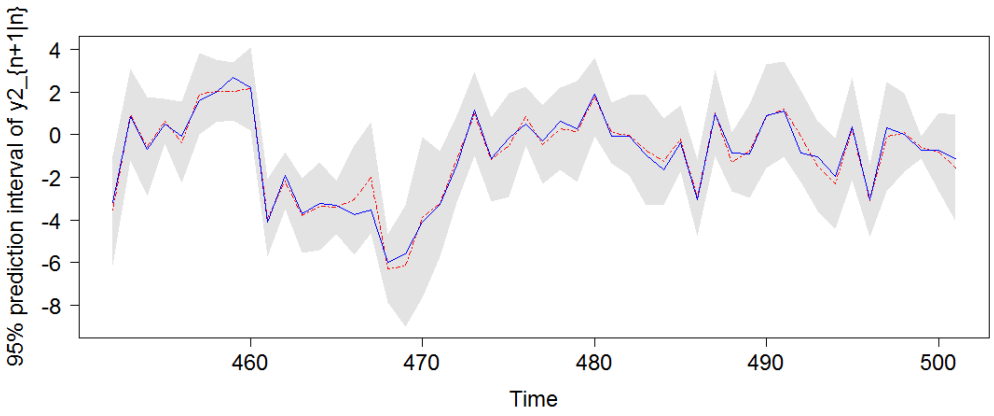


(c) $n = 1000$

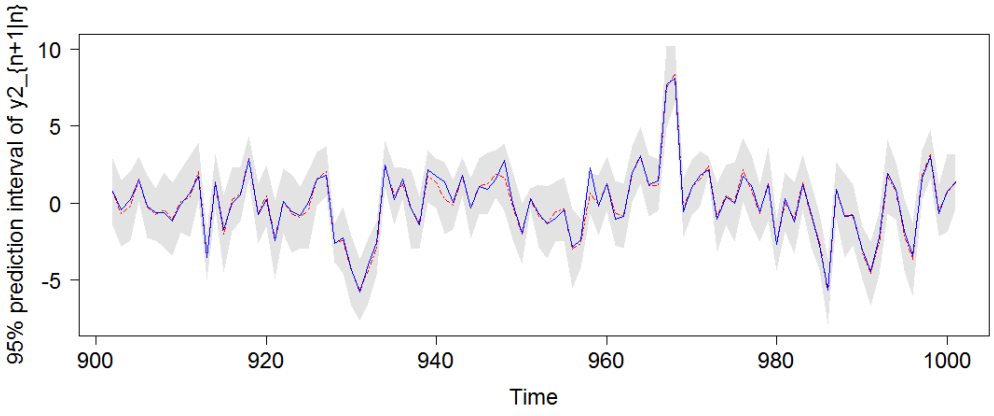
Figure 4.8: The solid curve denotes the true values of the response $y_{2,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 150$ in Example 2.



(a) $n = 200$



(b) $n = 500$



(c) $n = 1000$

Figure 4.9: The solid curve denotes the true values of the response $y_{2,t+1}$, the dotted curve denotes the one-step ahead nonlinear forecasted values, and the grey area denotes the 95% prediction interval of the response in the out-sample when $q_n = 500$ in Example 2.

4.2 An empirical example

In this section, we apply the developed predictive model and forecasting methodology to predict the UK inflation change. The data set were downloaded from the Office for National Statistics (ONS) and the Bank of England (BoE) websites, and covers the time period from the first quarter (Q1) of 1997 to the fourth quarter (Q4) of 2013. This data set has been analysed in Chen et al. (2018) which explore the nonlinear dynamic relationship between the response y_t and the exogenous regressors $\mathbf{Z}_t$ as well as the lags y_{t-j}. This is different from the predictive regression structure considered in this thesis. In the following empirical analysis, the response y_t is defined as the UK consumer price index (CPI), and the exogenous variables $\mathbf{Z}_t$ are the 53 series of measuring the real activity and other economic indicators to forecast CPI.

As in Chen et al. (2018), we divided the dataset into 2 parts used for estimation and prediction, respectively. The first part of the training set covers the time period from Q1/1997 to Q4/2012, and the second part of the forecasting set covers the time period from Q1/2013 to Q4/2013. All of the quarterly observations have been seasonally adjusted. Furthermore, as in Stock and Watson (1998, 1999) and Chen et al. (2018), we considered one of the following 4 transformations on the time series variables (depending on their nature): (i) no transformation, (ii) first difference, (iii) logarithm, and (iv) first difference of logarithms. The transformed CPI and the 53 predictor series were further normalised to have zero mean and unit variance with the training data set of size $n = 62$. The original CPI and the normalised first-difference of CPI are plotted in Figures 4.10 and 4.11.

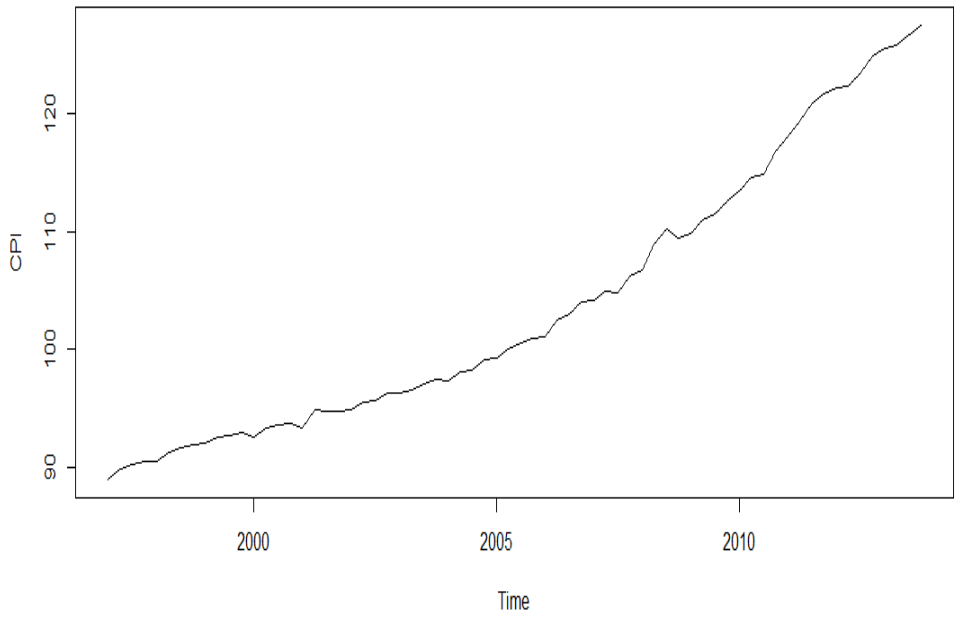


Figure 4.10: Plot of the original CPI series from Q1/1997 to Q4/2013.

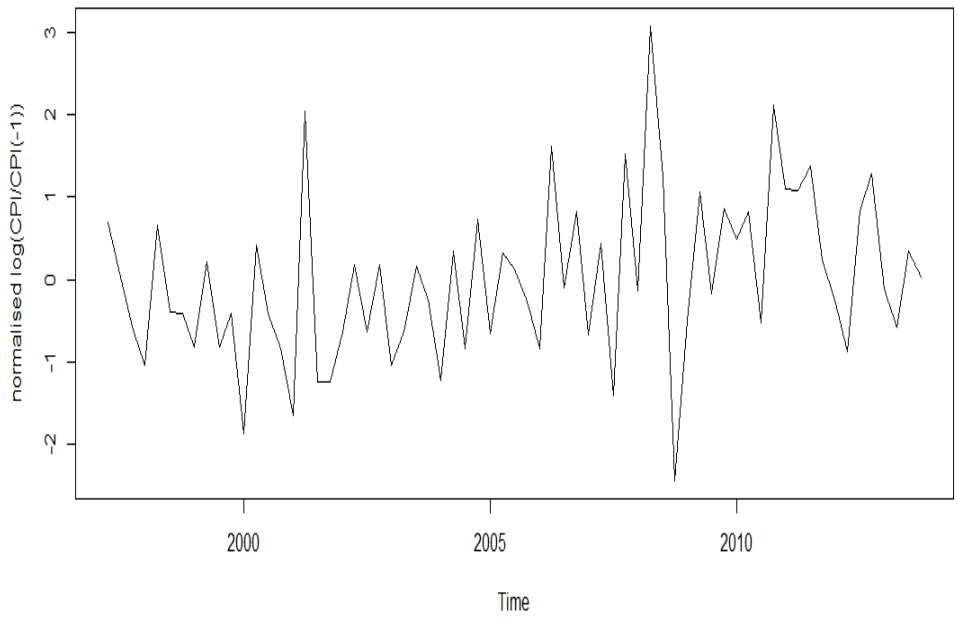


Figure 4.11: Plot of the transformed CPI series from Q1/1997 to Q4/2013.

Due to the high-dimension of the exogenous regressors, we imposed the approximate factor model structure (3.2) on $\mathbf{Z}_t$. As in the simulated examples, we chose four PCA estimated common factors which account for 82.34% of the total variation. Consider the following FA-FCM:

$$y_{t+1} = \sum_{k=1}^{4} F_{tk}\beta_{1k}(u_t) + \sum_{k=1}^{d_0} y_{t+1-k}\beta_{2k}(u_t) + \epsilon_{t+1}, \tag{4.5}$$

where the index variable u_t is y_t, the UK consumer price index.

We first applied the feasible local linear method to estimate the rotated coefficient functions in (4.5), where the Epanechnikov kernel was used and the optimal bandwidth is 0.03 determined by the multi-fold cross-validation criterion introduced in Section 3.4.1. For calculating the AMS values in the multi-fold cross-validation criterion, we set the potential lags as 8. In this analysis, we also tried the different values of the number of lags, and the results are not reported here.

To implement the forecasting model in (4.5), we need to select a number of lags. Consequently, the number of lags, d_0, was determined via the forward selection procedure with the BIC stopping rule introduced in Section 3.4.2. As seen from Figure 4.12, the lag number d_0 is estimated as 6.

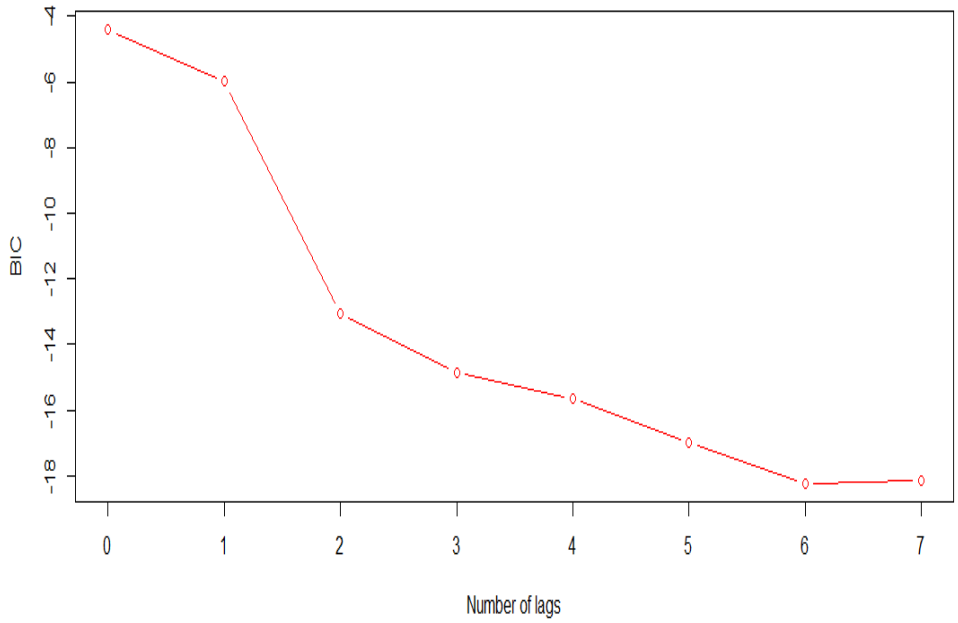


Figure 4.12: The change of the BIC function.

Finally, we constructed the one-step ahead nonlinear forecast as in (3.7). In order to measure the prediction accuracy, we computed the MSPE and the mean absolute prediction error (MAPE) defined as

$$\text{MSPE} = \frac{1}{4}\sum_{t=1}^{4}\left(y_{64+t} - \hat{y}_{64+t|63+t}\right)^2, \quad \text{MAPE} = \frac{1}{4}\sum_{t=1}^{4}\left|y_{64+t} - \hat{y}_{64+t|63+t}\right|.$$

The MPSE and MAPE values for the out-sample prediction using the proposed FA-FCM are 0.0634 and 0.2114, respectively. For the aim of comparison, we also consider using the traditional autoregressive (AR) model, the vector autoregressive (VAR) model consisting of CPI, oil price, effective sterling exchange rate and BoE's base interest rate and the unemployment rate Phillips curve in the out-sample prediction as in Chen et al. (2018). In AR and VAR models, the order of the autoregressions are chosen by the Akaike information criterion (AIC). Let L be the lag operator, and Δ be the first difference operator. The Phillips curve specification

is defined by

$$I_{t+1} - I_t = \alpha + \beta(L)u_t + \gamma(L)\Delta I_t + \epsilon_{t+1},$$

where I_t is the t-th quarter of CPI, u_t is the unemployment rate, $\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3$ and $\gamma(L) = \gamma_0 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3$ are lag polynomials. The relevant MPSE values are 0.0767 (AR), 0.1027 (VAR) and 1.1900 (Phillips) and the relevant MAPE values are 0.2338 (AR), 0.2456 (VAR) and 1.0170 (Phillips), respectively. The MPSE and MAPE results for the AR and VAR models and the unemployment rate Phillips curve are directly quoted from Table 5.3 in Chen et al. (2018). Through the comparison, we find that the developed FA-FCM and the one-step ahead nonlinear forecasting approach have the most accurate performance in predicting the UK inflation change. In addition, Figure 4.13 gives the 95% prediction interval of the transformed CPI, y_{t+1}, from Q1/2013 to Q4/2013 by using the wild bootstrap procedure introduced in Section 3.3.

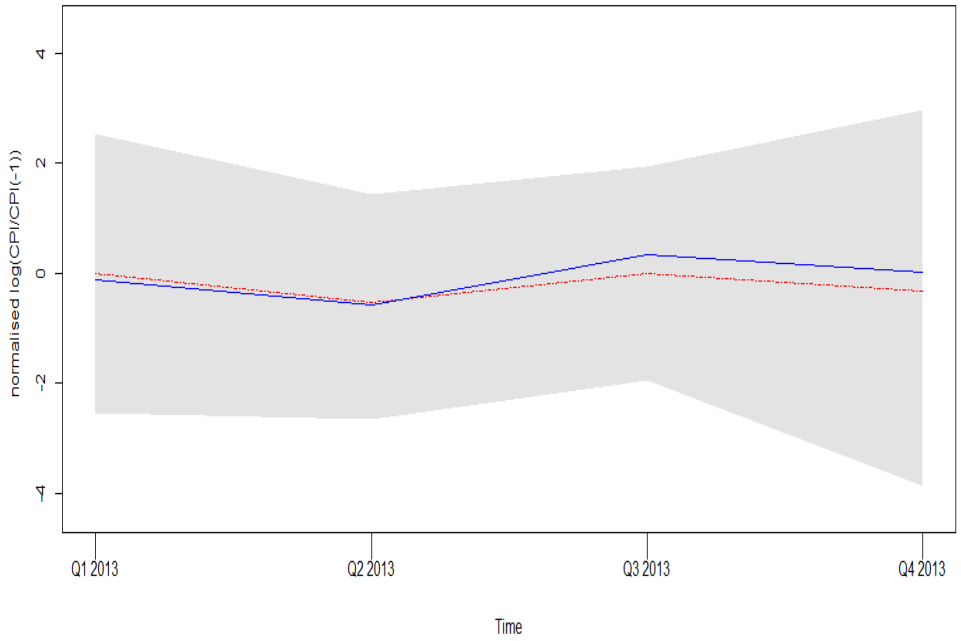


Figure 4.13: The solid curve denotes the true values of the transformed CPI y_{t+1}, the dotted curve denotes the one-step ahead nonlinear forecasted values $\hat{y}_{t+1|t}$, and the grey area denotes the 95% prediction interval of the transformed CPI from Q1/2013 to Q4/2013.

Chapter 5

Main Asymptotic Theorems

In this chapter, we give the asymptotic results for the methodologies developed in Chapter 3. In particular, we show that the local linear estimator and the nonlinear forecast using the estimated factor regressors are asymptotically equivalent to those using the true latent factor regressors under some regularity conditions. In this chapter, we only derive asymptotic theorems for FA-FCM with univariate response. Similar properties also hold for FA-FCM with multivariate response. Throughout this chapter, we let $\|\cdot\|$ denote the Euclidean norm of a vector and $\|\cdot\|_F$ denote the Frobenius norm of a matrix.

5.1 Technical Assumptions

We next list the regularity conditions which are needed for proving the asymptotic theorems.

ASSUMPTION 1. *(i) The process* $\{(y_t, u_t, \mathbf{F}_t, \mathbf{V}_t)\}$ *is stationary and* α*-mixing dependent with the mixing coefficient satisfying* $\alpha_k \sim c_\alpha \rho^k$*, where* $0 < c_\alpha < \infty$ *and* $0 < \rho < 1$.

(ii) The index variable u_t *has a compact support* $\mathcal{C} = [0, 1]$*, and its density function* $f(\cdot)$ *has continuous second-order derivatives and is bounded away from zero and infinity over* $[0, 1]$.

ASSUMPTION 2. *(i) The latent factor regressors satisfy the conditions that* $\mathsf{E}(\mathbf{F}_t) =$

$\mathbf{0}_{k_0}$, *the* $k_0 \times k_0$ *matrix* $\mathbf{\Lambda}_F(u) = \mathsf{E}\left[\mathbf{F}_t\mathbf{F}_t^\intercal | u_t = u\right]$ *is continuous and positive definite over* $u \in \mathcal{C}$, *and* $\mathsf{E}\left[\|\mathbf{F}_t\|^{2(2+\delta)}\right] < \infty$ *for some* $0 < \delta < \infty$, *where* $\mathbf{0}_k$ *is a* k*-dimensional null vector.*

(ii) There exists a $k_0 \times k_0$ *matrix* $\mathbf{\Lambda}_B$ *such that* $\mathbf{\Lambda}_B = \lim_{n\to\infty} \mathbb{B}_n^\intercal \mathbb{B}_n/q_n$. *The matrix* $\mathbf{\Lambda}_B$ *is positive definite with the smallest eigenvalue bounded away from zero.*

(iii) There exist two $k_0 \times k_0$ *non-singular matrices* $\mathbf{Q}_0$ *and* $\hat{\mathbf{\Lambda}}_F$ *such that* $\mathbf{Q}_0 = \lim_{n\to\infty} \mathbf{Q}_n$ *and* $\hat{\mathbf{\Lambda}}_F = \lim_{n\to\infty} \hat{\mathbb{F}}_n^\intercal \mathbb{F}_n/n$ *with probability approaching one, where* $\hat{\mathbb{F}}_n$ *and* $\mathbb{F}_n$ *are defined in Section 3.2.*

ASSUMPTION 3. *Letting* $\mathbf{X}_t^\star = \left(\mathbf{F}_t^\intercal, \mathbf{Y}_t^\intercal\right)^\intercal$, *the* $(k_0 + d_0) \times (k_0 + d_0)$ *matrix* $\mathbf{\Lambda}_X(u) = \mathsf{E}\left[\mathbf{X}_t^\star \mathbf{X}_t^{\star\intercal} | u_t = u\right]$ *is continuous and positive definite over* $u \in \mathcal{C}$. *Moreover,* $\mathsf{E}\left[|y_t|^{2(2+\delta)}\right] < \infty$.

ASSUMPTION 4. *The kernel function* $K(\cdot)$ *is positive and Lipschitz continuous with a compact support.*

ASSUMPTION 5. *The bandwidth* b *satisfies* $b \to 0$ *and* $n^{1-\tau-1/(2+\delta)}b \to \infty$, *where* $\tau > 0$ *can be arbitrarily small and* δ *is defined in Assumption 2(i).*

ASSUMPTION 6. *(i) The idiosyncratic errors* v_{tk} *satisfy* $\mathsf{E}[v_{tk}] = 0$ *and* $\max_{1\leqslant k\leqslant q_n} \mathsf{E}[|v_{tk}|^{2\delta_1}] < \infty$ *with* $\delta_1 > 2$, *and there exists a positive constant* m_0 *such that*

$$\max_{1\leqslant t\leqslant n} \mathsf{E}\left[\left\|\sum_{k=1}^{q_n} \mathbf{B}_k v_{tk}\right\|^{\delta_1}\right] \leqslant m_0 q_n^{\delta_1/2} \tag{5.1}$$

and

$$\max_{1\leqslant t,s\leqslant n} \mathsf{E}\left[\left|\sum_{k=1}^{q_n} \left(v_{tk}v_{sk} - \mathsf{E}[v_{tk}v_{sk}]\right)\right|^{\delta_1}\right] \leqslant m_0 q_n^{\delta_1/2}. \tag{5.2}$$

(ii) Let $q_n/(nb) \to \infty$ *and* $n = o\left([q_n/(nb)]^{\delta_*/2}\right)$, *where* $\delta_*^{-1} = \max\{1/[2(2+\delta)] + 1/\delta_1, 2/\delta_1\}$.

ASSUMPTION 7. *(i) The sequence* $\{\epsilon_t\}$ *is i.i.d. with* $\mathsf{E}[\epsilon_t] = 0$, $0 < \sigma_\epsilon^2 := \mathsf{E}[\epsilon_t^2] < \infty$ *and* $\mathsf{E}[|\epsilon_t|^{2+\delta}] < \infty$, *where* δ *is defined in Assumption*

2(i). Furthermore, ϵ_{t+1} is independent of $(u_s, \mathbf{F}_s, \mathbf{V}_s)$, $s \leqslant t$.

(ii) Both $\boldsymbol{\beta}_1(\cdot)$ and $\boldsymbol{\beta}_2(\cdot)$ have continuous second-order derivatives.

REMARK 1. The α-mixing dependence condition on the stationary process in Assumption 1(i) is mild and has been widely used on analysing nonlinear time series (c.f., Bosq, 1998). The geometric decaying rate on the mixing coefficient and the compact support restriction on the index variable in Assumption 1 are imposed to facilitate our proofs and can be relaxed at the cost of more lengthy arguments in the poofs. The conditions in Assumption 2 are common in PCA estimation of the approximate factor models (c.f., Bai and Ng, 2002, 2006). Assumptions 3–5 and 7(ii) are needed as the local linear smoothing method is used in our estimation and forecasting procedures. In particular, the strong moment condition and bandwidth restriction can ensure the validity of uniform consistency results in Theorem 1 and Corollary 1 in Section 5.2. Assumption 6(i) is similar to the condition B4 in Chen et al. (2018), indicating that for any t, v_{it} is allowed to be weakly dependent over i. Assumption 6(ii) shows that both n and q_n diverge to infinity simultaneously in our asymptotic results and their relationship relies on the moment conditions. In particular, the dimension of exogenous regressors is allowed to be much larger than the time series length.

5.2 Asymptotic Properties

In this section, we establish the asymptotic properties for the estimation of rotated coefficient functions in Section 3.2 and the one-step ahead nonlinear forecasting technique in Section 3.3. The following theorem shows that the feasible local linear estimator $\hat{\boldsymbol{\beta}}_H(u)$ defined in (3.6) is asymptotically equivalent to the infeasible one $\tilde{\boldsymbol{\beta}}_H(u)$ uniformly over $u \in \mathcal{C}$.

THEOREM 1. *Suppose that Assumptions 1–6 in Section 5.1 are satisfied. Then,*

$$\sup_{u \in \mathcal{C}} \left\| \hat{\boldsymbol{\beta}}_H(u) - \tilde{\boldsymbol{\beta}}_H(u) \right\| = o_P\left((nb)^{-1/2}\right). \tag{5.3}$$

Let $\mu_j = \int u^j K(u)du$, $\nu_j = \int u^j K^2(u)du$, $\mathbf{H}_0 = \mathbf{Q}_0^{-1}\hat{\mathbf{\Lambda}}_F\mathbf{\Lambda}_B$, $\bar{\mathbf{H}}_0 = \mathsf{diag}\{\mathbf{H}_0, \mathbf{I}_{d_0}\}$,

$\mathbf{\Lambda}(u) = \bar{\mathbf{H}}_0 \left\{ \mathsf{E}\left[\mathbf{X}_t^\star \mathbf{X}_t^{\star\intercal} | u_t = u \right] \right\} \bar{\mathbf{H}}_0^\intercal$, and define $\boldsymbol{\beta}_H''(\cdot)$ as the second-order derivative of $\boldsymbol{\beta}_H(\cdot)$. Then, by (3.4) and the standard argument in the local linear estimation (c.f., Fan and Gijbels, 1996), we have

$$(nb)^{1/2} \left[\tilde{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u) - \frac{1}{2}\mu_2 \boldsymbol{\beta}_H''(u) b^2 \right] \xrightarrow{d} \mathsf{N}\left[\mathbf{0}_{k_0+d_0}, \frac{\sigma_\epsilon^2 \nu_0}{f(u)} \mathbf{\Lambda}^{-1}(u) \right]. \tag{5.4}$$

Furthermore, following the uniform consistency results in Hansen (2008), we may show that

$$\sup_{u \in \mathcal{C}} \left\| \tilde{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u) \right\| = O_P\left(b^2 + [\log n/(nb)]^{1/2} \right). \tag{5.5}$$

Let $a_n \propto b_n$ denote that $0 < \underline{c} \leqslant a_n/b_n \leqslant \bar{c} < \infty$ when n tends to infinity. Combining (5.4) and (5.5) with Theorem 1, we readily have the following corollary.

COROLLARY 1. *Suppose Assumptions 1–7 in Section 5.1 are satisfied and* $b \propto n^{-1/5}$. *Then we have*

$$(nb)^{1/2} \left[\hat{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u) - \frac{1}{2}\mu_2 \boldsymbol{\beta}_H''(u) b^2 \right] \xrightarrow{d} \mathsf{N}\left[\mathbf{0}_{k_0+d_0}, \frac{\sigma_\epsilon^2 \nu_0}{f(u)} \mathbf{\Lambda}^{-1}(u) \right], \tag{5.6}$$

and

$$\sup_{u \in \mathcal{C}} \left\| \hat{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u) \right\| = O_P\left(b^2 + [\log n/(nb)]^{1/2} \right). \tag{5.7}$$

In Section 3.3, we construct the one-step ahead nonlinear forecast $\hat{y}_{n+1|n}$ using the local linear estimates. The following theorem describes the accuracy of the one-step ahead forecast.

THEOREM 2. *Suppose that the conditions in Corollary 1 are satisfied. Then,*

$$\hat{y}_{n+1|n} - y_{n+1} = \Delta_n - \epsilon_{n+1} + o_P\left(1/(nb)^{1/2} \right), \tag{5.8}$$

where $\Delta_n = \left[\hat{\boldsymbol{\beta}}_{H,n-1}(u_n) - \boldsymbol{\beta}_H(u_n) \right]^\intercal \mathbf{X}_n$. *Furthermore, conditional on* $u_n = u_*$ *and* $\mathbf{X}_n = \mathbf{X}_*$, *we have*

$$(nb)^{1/2} \left[\Delta_n - \frac{1}{2}\mu_2 b^2 \mathbf{X}_*^\intercal \boldsymbol{\beta}_H''(u_*) \right] \xrightarrow{d} \mathsf{N}\left[\mathbf{0}_{k_0+d_0}, \frac{\sigma_\epsilon^2 \nu_0}{f(u_*)} \mathbf{X}_*^\intercal \mathbf{\Lambda}^{-1}(u_*) \mathbf{X}_* \right]. \tag{5.9}$$

REMARK 2. By the definition of $y_{n+1|n}$ in Section 3.3, we may show that

$$\hat{y}_{n+1|n} - y_{n+1|n} = \Delta_n + o_P\left(1/(nb)^{1/2}\right),$$

indicating that $(nb)^{1/2}\left(\hat{y}_{n+1|n} - y_{n+1|n}\right)$ has the same asymptotic normal distribution as that in (5.9).

5.3 Proofs of Theorems

This section contains the proofs of the main asymptotic theorems and some technical lemmas together with their proofs. Throughout the proofs, we let C be a generic positive constant whose value may change from line to line.

5.3.1 Proofs of the Asymptotic Theorems

In this section, we give the detailed proofs of the asymptotic theorems in Section 5.2.

PROOF OF THEOREM 1. Note that

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_H(u) - \tilde{\boldsymbol{\beta}}_H(u) &= (\mathbf{I}_{k_0+d_0}, \mathbf{O}_{k_0+d_0}) \left\{ \left[\hat{\mathbb{X}}_n^{\intercal}(u)\mathbb{W}_n(u)\hat{\mathbb{X}}_n(u)\right]^{-1} \right. \\
&\quad \left. - \left[\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{X}_n(u)\right]^{-1} \right\} \left[\hat{\mathbb{X}}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{Y}_n\right] \\
&\quad + (\mathbf{I}_{k_0+d_0}, \mathbf{O}_{k_0+d_0}) \left[\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{X}_n(u)\right]^{-1} \\
&\quad \left[\hat{\mathbb{X}}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{Y}_n - \mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{Y}_n\right] \\
&=: \boldsymbol{\Pi}_{n1}(u) + \boldsymbol{\Pi}_{n2}(u).
\end{aligned}$$

Letting

$$\begin{aligned}
\mathbb{M}_n(u) &= \frac{1}{nb}\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{X}_n(u), \;\; \hat{\mathbb{M}}_n(u) = \frac{1}{nb}\hat{\mathbb{X}}_n^{\intercal}(u)\mathbb{W}_n(u)\hat{\mathbb{X}}_n(u), \\
\mathbb{N}_n(u) &= \frac{1}{nb}\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{Y}_n, \;\; \hat{\mathbb{N}}_n(u) = \frac{1}{nb}\hat{\mathbb{X}}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{Y}_n,
\end{aligned}$$

we have

$$\boldsymbol{\Pi}_{n1}(u) = (\mathbf{I}_{k_0+d_0}, \mathbf{O}_{k_0+d_0}) \left[\hat{\mathbb{M}}_n^{-1}(u) - \mathbb{M}_n^{-1}(u)\right] \hat{\mathbb{N}}_n(u), \tag{5.10}$$

and

$$\mathbf{\Pi}_{n2}(u) = \left(\mathbf{I}_{k_0+d_0)}, \mathbf{O}_{k_0+d_0}\right) \mathbb{M}_n^{-1}(u) \left[\hat{\mathbb{N}}_n(u) - \mathbb{N}_n(u)\right]. \tag{5.11}$$

We first consider $\mathbf{\Pi}_{n2}(u)$. Observe that

$$\begin{aligned}
\mathbb{M}_n(u) &= \frac{1}{nb}\mathbb{X}_n^{\intercal}(u)\mathbb{W}_n(u)\mathbb{X}_n(u) \\
&= \begin{bmatrix} \frac{1}{nb}\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^{\intercal} K_b(u_t,u) & \frac{1}{nb}\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^{\intercal}(\frac{u_t-u}{b})K_b(u_t,u) \\ \frac{1}{nb}\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^{\intercal}(\frac{u_t-u}{b})K_b(u_t,u) & \frac{1}{nb}\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t^{\intercal}(\frac{u_t-u}{b})^2K_b(u_t,u) \end{bmatrix} \\
&=: \begin{bmatrix} \mathbb{M}_{n0}(u) & \mathbb{M}_{n1}(u) \\ \mathbb{M}_{n1}(u) & \mathbb{M}_{n2}(u) \end{bmatrix}.
\end{aligned}$$

Letting $\bar{\mathbf{H}} = \mathsf{diag}\{\mathbf{H}, \mathbf{I}_{d_0}\}$ and $\mathbf{X}_t^\star = \left(\mathbf{F}_t^{\intercal}, \mathbf{Y}_t^{\intercal}\right)^{\intercal}$, we readily have

$$\mathbb{M}_{nj}(u) = \bar{\mathbf{H}}\left[\frac{1}{nb}\sum_{t=1}^n \mathbf{X}_t^\star\mathbf{X}_t^{\star\intercal}(\frac{u_t-u}{b})^j K_b(u_t,u)\right]\bar{\mathbf{H}}^{\intercal}.$$

By Lemma B.1 in Section 5.3.2, we have

$$\mathbb{M}_n(u) = f(u)\cdot \mathsf{diag}\{\mu_0,\mu_2\} \otimes \left\{\bar{\mathbf{H}}_0 \mathsf{E}\left[\mathbf{X}_t^\star\mathbf{X}_t^{\star\intercal}|u_t=u\right]\bar{\mathbf{H}}_0^{\intercal}\right\} + o_P(1) \tag{5.12}$$

uniformly for $u \in \mathcal{C}$, where $\bar{\mathbf{H}}_0$ is defined in Section 5.2 of the main document. By Assumptions 2(ii)(iii) and 3, the limit matrix on the right hand side of (5.12) is invertible, which together with Lemma B.2 in Section 5.3.2, implies that

$$(nb)^{1/2}\sup_{u\in\mathcal{C}}\|\mathbf{\Pi}_{n2}(u)\| = o_P(1). \tag{5.13}$$

We next consider $\mathbf{\Pi}_{n1}(u)$. By Lemma B.2 and following the proof of Lemma B.1 in Section 5.3.2, we readily have

$$\begin{aligned}
\sup_{u\in\mathcal{C}}\left\|\hat{\mathbb{N}}_n(u)\right\| &= \sup_{u\in\mathcal{C}}\left\|\hat{\mathbb{N}}_n(u) - \mathbb{N}_n(u)\right\| + \sup_{u\in\mathcal{C}}\|\mathbb{N}_n(u)\| \\
&= o_P\left((nb)^{-1/2}\right) + O_P(1) = O_P(1).
\end{aligned} \tag{5.14}$$

This, together with Lemma B.3, leads to

$$(nb)^{1/2} \sup_{u \in \mathcal{C}} \|\mathbf{\Pi}_{n1}(u)\| = o_P(1). \tag{5.15}$$

In view of (5.13) and (5.15), we complete the proof of Theorem 1. □

PROOF OF THEOREM 2. Note that

$$\begin{aligned} \hat{y}_{n+1|n} - y_{n+1} &= \hat{\boldsymbol{\beta}}_{H,n-1}^{\intercal}(u_n)\hat{\mathbf{X}}_n - \boldsymbol{\beta}_H^{\intercal}(u_n)\mathbf{X}_n - \epsilon_{n+1} \\ &= \hat{\boldsymbol{\beta}}_{H,n-1}^{\intercal}(u_n)\left(\hat{\mathbf{X}}_n - \mathbf{X}_n\right) + \left[\hat{\boldsymbol{\beta}}_{H,n-1}(u_n) - \boldsymbol{\beta}_H(u_n)\right]^{\intercal}\mathbf{X}_n - \epsilon_{n+1} \\ &=: \Delta_{n1} + \Delta_n - \epsilon_{n+1}. \end{aligned} \tag{5.16}$$

By (3.7) in Corollary 1 and the arguments in Lemma B.2, we may show that

$$\begin{aligned} \Delta_{n1} &= \left[\hat{\boldsymbol{\beta}}_{H,n-1}(u_n) - \boldsymbol{\beta}_H(u_n)\right]^{\intercal}\left(\hat{\mathbf{X}}_n - \mathbf{X}_n\right) + \boldsymbol{\beta}_H^{\intercal}(u_n)\left(\hat{\mathbf{X}}_n - \mathbf{X}_n\right) \\ &= O_P\left(b^2 + [\log n/(nb)]^{1/2}\right) \cdot o_P\left(1/(nb)^{1/2}\right) + O_P(1) \cdot o_P\left(1/(nb)^{1/2}\right) \\ &= o_P\left(1/(nb)^{1/2}\right). \end{aligned} \tag{5.17}$$

By (5.16) and (5.17), we readily have (3.8). The proof of (3.9) follows directly from (3.6) in Corollary 1. □

5.3.2 Some Technical Lemmas and Their Proofs

In this section, we state some technical lemmas which have been used in the proofs of the main asymptotic theorems and give their proofs.

LEMMA B.1. *Suppose that Assumptions 1–5 in Section 5.1 are satisfied. Then,*

$$\mathbb{M}_n(u) = f(u) \cdot \mathsf{diag}\left\{\mu_0, \mu_2\right\} \otimes \left\{\bar{\mathbf{H}}_0 \mathsf{E}\left[\mathbf{X}_t^{\star}\mathbf{X}_t^{\star\intercal} | u_t = u\right]\bar{\mathbf{H}}_0^{\intercal}\right\} + o_P(1) \tag{5.18}$$

uniformly for $u \in \mathcal{C}$.

PROOF. By Assumption 2(ii)(iii), we readily show that $\mathbf{H}_0$ is the probability limit

of the rotation matrix $\mathbf{H}$ when n tends to infinity. Hence, it suffices to prove

$$\frac{1}{nb}\sum_{t=1}^{n}\mathbf{X}_t^{\star}\mathbf{X}_t^{\star\intercal}\left(\frac{u_t-u}{b}\right)^j K_b(u_t,u)=\mathbf{\Sigma}_j(u)+o_P(1) \tag{5.19}$$

uniformly for $u\in\mathcal{C}$, where $\mathbf{\Sigma}_j(u)=\mu_j f(u)\mathsf{E}\left[\mathbf{X}_t^{\star}\mathbf{X}_t^{\star\intercal}|u_t=u\right]$, $j=0,1,2$. The proof of the uniform consistency result (5.19) is standard. Let $m_n=n^{1/(2+\delta)+(\tau/2)}$ with δ and τ defined in Assumptions 2(i) and 5, respectively, and $\mathbf{V}_t=\mathbf{X}_t^{\star}\mathbf{X}_t^{\star\intercal}$ to simplify notation. We define

$$\bar{\mathbf{V}}_t=\mathbf{V}_t\cdot I\big(\|\mathbf{V}_t\|_F\leqslant m_n\big),\ \ \tilde{\mathbf{V}}_t=\mathbf{V}_t\cdot I\big(\|\mathbf{V}_t\|_F>m_n\big),$$

where $I(\cdot)$ is an indicator function. It is easy to show that

$$\begin{aligned}&\frac{1}{nb}\sum_{t=1}^{n}\mathbf{X}_t^{\star}\mathbf{X}_t^{\star\intercal}\left(\frac{u_t-u}{b}\right)^j K_b(u_t,u)\\ =\ &\frac{1}{nb}\sum_{t=1}^{n}\bar{\mathbf{V}}_t\left(\frac{u_t-u}{b}\right)^j K_b(u_t,u)+\frac{1}{nb}\sum_{t=1}^{n}\tilde{\mathbf{V}}_t\left(\frac{u_t-u}{b}\right)^j K_b(u_t,u).\end{aligned}$$

By the moment conditions in Assumptions 2(i) and 3 and using the Markov inequality, for any $\xi>0$,

$$\begin{aligned}&\mathsf{P}\left(\sup_{u\in\mathcal{C}}\left\|\frac{1}{nb}\sum_{t=1}^{n}\tilde{\mathbf{V}}_t\left(\frac{u_t-u}{b}\right)^j K_b(u_t,u)\right\|_F>\xi\right)\\ \leqslant\ &\mathsf{P}\left(\max_{1\leqslant t\leqslant n}\|\mathbf{V}_t\|_F>m_n\right)\leqslant\sum_{t=1}^{n}\mathsf{P}\left(\|\mathbf{V}_t\|_F>m_n\right)\\ \leqslant\ &n\mathsf{E}\left[\|\mathbf{V}_t\|_F^{2+\delta}\right]/m_n^{2+\delta}=O(n^{-(2+\delta)\tau/2})=o(1).\end{aligned} \tag{5.20}$$

Let $\bar{\mathbf{V}}_{t,j}(u)=\bar{\mathbf{V}}_t\left(\frac{u_t-u}{b}\right)^j K_b(u_t,u)/b$. As $m_n\to\infty$, by Assumptions 3, 4 and 5 as well as some standard arguments,

$$\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(u)\right]=\mu_j f(u)\mathsf{E}\left[\mathbf{X}_t^{\star}\mathbf{X}_t^{\star\intercal}|u_t=u\right]+o_P(1)$$

uniformly for $u\in\mathcal{C}$. We next consider covering $\mathcal{C}$ by a finite number of disjoint intervals $\mathcal{C}_k$ with centre s_k and radius $r=o(b^2)$, where the total number of $\mathcal{C}_k$ is

$N = O(r^{-1})$. Observe that

$$\begin{aligned}
&\sup_{u\in\mathcal{C}}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\bar{\mathbf{V}}_{t,j}(u)-\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(u)\right]\right\}\right\|_F \\
\leqslant\ & C\left(\max_{1\leqslant k\leqslant N}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\bar{\mathbf{V}}_{t,j}(s_k)-\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(s_k)\right]\right\}\right\|_F\right. \\
&+\max_{1\leqslant k\leqslant N}\sup_{u\in\mathcal{C}_k}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\bar{\mathbf{V}}_{t,j}(u)-\bar{\mathbf{V}}_{t,j}(s_k)\right\}\right\|_F \\
&\left.+\max_{1\leqslant k\leqslant N}\sup_{u\in\mathcal{C}_k}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(u)\right]-\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(s_k)\right]\right\}\right\|_F\right) \\
\leqslant\ & C\max_{1\leqslant k\leqslant N}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\bar{\mathbf{V}}_{t,j}(s_k)-\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(s_k)\right]\right\}\right\|_F+o_P(1). \qquad (5.21)
\end{aligned}$$

Furthermore, by Assumptions 1 and 5 in Section 5.1, and the exponential type inequality for the α-mixing sequence (c.f., Theorem 2.18 in Fan and Yao, 2003), we have for any $\xi > 0$

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant k\leqslant N}\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\bar{\mathbf{V}}_{t,j}(s_k)-\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(s_k)\right]\right\}\right\|_F>\xi\right) \\
\leqslant\ & \sum_{k=1}^{N}\mathsf{P}\left(\left\|\frac{1}{n}\sum_{t=1}^{n}\left\{\bar{\mathbf{V}}_{t,j}(s_k)-\mathsf{E}\left[\bar{\mathbf{V}}_{t,j}(s_k)\right]\right\}\right\|_F>\xi\right) \\
=\ & O\left(N\exp\left\{-c_1 b m_n^{-1}n^{1-(\tau/3)}\right\}\right)+O\left(Nm_n^{1/2}b^{-1/2}n^{1-(\tau/3)}\rho^{n^{(\tau/3)}}\right) \\
=\ & O\left(N\exp\left\{-c_1 n^{1-1/(2+\delta)-(5\tau/6)}b\right\}\right)+O_P\left(Nb^{-1/2}n^{1+1/(4+2\delta)-(\tau/12)}\rho^{n^{(\tau/3)}}\right) \\
=\ & o(1), \qquad (5.22)
\end{aligned}$$

where c_1 is a positive constant. Then, we can prove (5.19) by using (5.20)–(5.22), and thus the proof of the lemma is completed. □

LEMMA B.2. *Suppose that Assumptions 1–6 in Section 5.1 are satisfied. Then we have*

$$\sup_{u\in\mathcal{C}}\left\|\hat{\mathbb{N}}_n(u)-\mathbb{N}_n(u)\right\|=o_P\left((nb)^{-1/2}\right),$$

where $\mathcal{C}$ is the compact support of u_t.

PROOF. By the definitions of $\hat{\mathbf{F}}_t$ and $\mathbf{H}$, we readily have

$$\begin{aligned}
\mathbf{Q}_n\left(\hat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right) &= \frac{1}{nq_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathbf{F}_s^{\intercal}\mathbf{B}_k v_{tk} + \sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathbf{F}_t^{\intercal}\mathbf{B}_k v_{sk}\right. \\
&\quad \left. + \sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s v_{sk}v_{tk}\right) \\
&= \frac{1}{nq_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathbf{F}_s^{\intercal}\mathbf{B}_k v_{tk} + \sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathbf{F}_t^{\intercal}\mathbf{B}_k v_{sk}\right. \\
&\quad \left. + \sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathsf{E}\left[v_{sk}v_{tk}\right]\sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\left\{v_{sk}v_{tk} - \mathsf{E}\left[v_{sk}v_{tk}\right]\right\}\right) \\
&=: \frac{1}{nq_n}\left(\mathbf{W}_{nt,1} + \mathbf{W}_{nt,2} + \mathbf{W}_{nt,3} + \mathbf{W}_{nt,4}\right) \qquad (5.23)
\end{aligned}$$

for any $1 \leqslant t \leqslant n$, where $\mathbf{B}_k$ is the k-th row of of $\mathbb{B}_n$ and v_{tk} is the k-th component of $\mathbf{V}_t$.

Note that

$$\hat{\mathbf{X}}_t - \mathbf{X}_t = \left((\hat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t)^{\intercal}, \mathbf{0}_{d_0}^{\intercal}\right)^{\intercal}. \qquad (5.24)$$

By (5.24), to prove Lemma B.2, we only need to consider $\mathbf{R}_{nj}(u)$ defined by

$$\begin{aligned}
\mathbf{R}_{nj}(u) &= \frac{1}{nb}\sum_{t=1}^{n}\left(\hat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)y_{t+1}\left(\frac{u_t - u}{b}\right)^j K_b(u_t, u) \\
&= \mathbf{Q}_n^{-1}\frac{1}{nb}\sum_{t=1}^{n}\mathbf{Q}_n\left(\hat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right)y_{t+1}\left(\frac{u_t - u}{b}\right)^j K_b(u_t, u) \\
&= \mathbf{Q}_n^{-1}\frac{1}{n^2bq_n}\sum_{t=1}^{n}\sum_{l=1}^{4}\mathbf{W}_{nt,l}y_{t+1}\left(\frac{u_t - u}{b}\right)^j K_b(u_t, u). \qquad (5.25)
\end{aligned}$$

Following the argument in the proof of Lemma B.1, we may show that

$$\frac{1}{nb}\sum_{t=1}^{n}|y_{t+1}|\left|\frac{u_t - u}{b}\right|^j K_b(u_t, u) = O_P(1) \qquad (5.26)$$

uniformly for $u \in \mathcal{C}$. Using (5.26) and the fact that the limit of $\mathbf{Q}_n^{-1}$ exists (by Assumption 2(iii)), to prove Lemma B.2, it suffices to show

$$\max_{1\leqslant t\leqslant n}\|\mathbf{W}_{nt,l}\| = o_P\left(nq_n/(nb)^{1/2}\right), \quad l = 1, 2, 3, 4. \qquad (5.27)$$

By Assumption 2(iii), we have $\left\|\sum_{s=1}^{n}\hat{\mathbf{F}}_s\mathbf{F}_s^{\intercal}\right\|_F = O_P(n)$. By (5.1) in Assumption 6(i), we have

$$\begin{aligned}&\mathsf{P}\left(\max_{1\leqslant t\leqslant n}\left\|\sum_{k=1}^{q_n}\mathbf{B}_k v_{tk}\right\| > mn^{1/\delta_1}q_n^{1/2}\right) \leqslant \sum_{t=1}^{n}\mathsf{P}\left(\left\|\sum_{k=1}^{q_n}\mathbf{B}_k v_{tk}\right\| > mn^{1/\delta_1}q_n^{1/2}\right)\\ &\leqslant \max_{1\leqslant t\leqslant n}\mathsf{E}\left[\left\|\sum_{k=1}^{q_n}\mathbf{B}_k v_{tk}\right\|^{\delta_1}\right]/(m^{\delta_1}q_n^{\delta_1/2}) < \xi \qquad (5.28)\end{aligned}$$

for any $\xi > 0$, by letting $m > (m_0/\xi)^{1/\delta_1}$, where m_0 is defined in Assumption 6(i). Hence, we readily have

$$\max_{1\leqslant t\leqslant n}\|\mathbf{W}_{nt,1}\| = O_P(n^{1+1/\delta_1}q_n^{1/2}) = o_P\left(nq_n/(nb)^{1/2}\right), \qquad (5.29)$$

as $n = o\left([q_n/(nb)]^{\delta_1/2}\right)$ from Assumption 6(ii).

By Assumption 2(i) and following the proof of (5.28), we have

$$\max_{1\leqslant t\leqslant n}\|\mathbf{F}_t\| = O_P\left(n^{1/[2(2+\delta)]}\right).$$

When $l = 2$, we may show that

$$\begin{aligned}\max_{1\leqslant t\leqslant n}\|\mathbf{W}_{nt,2}\| &= \max_{1\leqslant t\leqslant n}\left\|\sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathbf{F}_t^{\intercal}\mathbf{B}_k v_{sk}\right\|\\ &\leqslant \max_{1\leqslant t\leqslant n}\|\mathbf{F}_t\|\left(\sum_{s=1}^{n}\left\|\hat{\mathbf{F}}_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{n}\left\|\sum_{k=1}^{q_n}\mathbf{B}_k v_{sk}\right\|^2\right)^{1/2}\\ &= O_P\left(n^{1/[2(2+\delta)]}\right)\cdot O_P\left(n^{1/2}\right)\cdot O_P\left(n^{1/2+1/\delta_1}q_n^{1/2}\right)\\ &= O_P\left(n^{1+1/[2(2+\delta)]+1/\delta_1}q_n^{1/2}\right)\\ &= o_P\left(nq_n/(nb)^{1/2}\right), \qquad (5.30)\end{aligned}$$

as $n^{1/[2(2+\delta)]+1/\delta_1} = o\left([q_n/(nb)]^{1/2}\right)$ from Assumption 6(ii).

For $l = 3$, by Assumptions 1(i) and 6(i), we have

$$\begin{aligned}
\max_{1\leqslant t\leqslant n}\|\mathbf{W}_{nt,3}\| &= \max_{1\leqslant t\leqslant n}\left\|\sum_{s=1}^{n}\sum_{k=1}^{q_n}\hat{\mathbf{F}}_s\mathsf{E}\left[v_{sk}v_{tk}\right]\right\| \\
&= \max_{1\leqslant t\leqslant n}\left(\sum_{s=1}^{n}\|\hat{\mathbf{F}}_s\|^2\right)^{1/2}\left(\sum_{s=1}^{n}\left(\sum_{k=1}^{q_n}\mathsf{E}\left[v_{sk}v_{tk}\right]\right)^2\right)^{1/2} \\
&= O_P\left(n^{1/2}\right)\cdot O_P(q_n) \\
&= o_P\left(nq_n/(nb)^{1/2}\right) \qquad (5.31)
\end{aligned}$$

as $b \to 0$.

By (5.2) in Assumption 6(i), similarly to the proof of (5.28), we can prove that

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant s,t\leqslant n}\left|\sum_{k=1}^{q_n}\left(v_{sk}v_{tk}-\mathsf{E}\left[v_{sk}v_{tk}\right]\right)\right| > mn^{2/\delta_1}q_n^{1/2}\right) \\
\leqslant\ &\sum_{s=1}^{n}\sum_{t=1}^{n}\mathsf{P}\left(\left|\sum_{k=1}^{q_n}\left(v_{sk}v_{tk}-\mathsf{E}\left[v_{sk}v_{tk}\right]\right)\right| > mn^{2/\delta_1}q_n^{1/2}\right) \\
\leqslant\ &\sum_{s=1}^{n}\sum_{t=1}^{n}\mathsf{E}\left[\left|\sum_{k=1}^{q_n}\left(v_{sk}v_{tk}-\mathsf{E}\left[v_{sk}v_{tk}\right]\right)\right|^{\delta_1}\right]/(m^{\delta_1}n^2q_n^{\delta_1/2}) \\
\leqslant\ &(m_0/m)^{\delta_1} < \xi,
\end{aligned}$$

where m is chosen as that in (5.28). Then, for $l = 4$, we have

$$\begin{aligned}
\max_{1\leqslant t\leqslant n}\|\mathbf{W}_{nt,4}\| &= \max_{1\leqslant t\leqslant n}\left\|\sum_{s=1}^{n}\hat{\mathbf{F}}_s\sum_{k=1}^{q_n}\left(v_{sk}v_{tk}-\mathsf{E}\left[v_{sk}v_{tk}\right]\right)\right\| \\
&= \max_{1\leqslant t\leqslant n}\left(\sum_{s=1}^{n}\|\hat{\mathbf{F}}_s\|^2\right)^{1/2}\left(\sum_{s=1}^{n}\left(\sum_{k=1}^{q_n}\left(v_{sk}v_{tk}-\mathsf{E}\left[v_{sk}v_{tk}\right]\right)\right)^2\right)^{1/2} \\
&= O_P\left(n^{1/2}\right)\cdot O_P\left(n^{1/2}n^{2/\delta_1}q_n^{1/2}\right) \\
&= o_P\left(nq_n/(nb)^{1/2}\right) \qquad (5.32)
\end{aligned}$$

as $n^{2/\delta_1} = o\left([q_n/(nb)]^{1/2}\right)$ from Assumption 6(ii). We have completed the proof of (5.27). □

LEMMA B.3. *Suppose that Assumptions 1–6 in Section 5.1 are satisfied. Then we*

have

$$\sup_{u\in\mathcal{C}}\left\|\hat{\mathbb{M}}_n^{-1}(u)-\mathbb{M}_n^{-1}(u)\right\|_F=o_P\left((nb)^{-1/2}\right),$$

where $\mathcal{C}$ is the compact support of the index variable u_t.

PROOF. Note that

$$\begin{aligned}\left\|\hat{\mathbb{M}}_n^{-1}(u)-\mathbb{M}_n^{-1}(u)\right\|_F &= \left\|\hat{\mathbb{M}}_n^{-1}(u)\left[\mathbb{M}_n(u)-\hat{\mathbb{M}}_n(u)\right]\mathbb{M}_n^{-1}(u)\right\|_F\\ &\leqslant \left[\left\|\hat{\mathbb{M}}_n^{-1}(u)-\mathbb{M}_n^{-1}(u)\right\|_F+\left\|\mathbb{M}_n^{-1}(u)\right\|_F\right]\\ &\quad \left\|\mathbb{M}_n(u)-\hat{\mathbb{M}}_n(u)\right\|_F\cdot\left\|\mathbb{M}_n^{-1}(u)\right\|_F,\end{aligned}$$

leading to

$$\left\|\hat{\mathbb{M}}_n^{-1}(u)-\mathbb{M}_n^{-1}(u)\right\|_F\leqslant\frac{\left\|\mathbb{M}_n(u)-\hat{\mathbb{M}}_n(u)\right\|_F\cdot\left\|\mathbb{M}_n^{-1}(u)\right\|_F^2}{1-\left\|\mathbb{M}_n(u)-\hat{\mathbb{M}}_n(u)\right\|_F\cdot\left\|\mathbb{M}_n^{-1}(u)\right\|_F}. \tag{5.33}$$

By Assumptions 2(ii)(iii) and 3, and Lemma B.1, we have

$$\sup_{u\in\mathcal{C}}\left\|\mathbb{M}_n^{-1}(u)\right\|_F=O_P(1). \tag{5.34}$$

On the other hand, following the proof Lemma B.2,

$$\sup_{u\in\mathcal{C}}\left\|\mathbb{M}_n(u)-\hat{\mathbb{M}}_n(u)\right\|_F=o_P\left((nb)^{-1/2}\right). \tag{5.35}$$

Then, by (5.33)–(5.35), we can complete the proof of Lemma B.3.

Chapter 6

Conclusion and Future Research

In this thesis, we have introduced a new nonlinear factor-augmented predictive regression model with functional coefficients and developed a feasible local linear smoothing method to estimate the coefficient functions (with appropriate rotation), where the latent (rotated) factor regressors are estimated by the principal component analysis approach and the number of auto-regressors is determined by the forward selection procedure with the BIC stopping rule. Furthermore, we have extended the proposed methodologies to the more general case of multiple response variables. The developed model is also generalised to the factor-augmented vector auto-regression model with functional coefficients. Moreover, the one-step ahead nonlinear forecast of the response is obtained by using the local linear estimated functional coefficients and the prediction interval is constructed via the wild bootstrap procedure as introduced in Chapter 3. In addition, the asymptotic theory in Chapter 5 shows that the proposed local linear estimator and nonlinear forecast using the estimated factor regressors are asymptotically equivalent to those assuming that the true latent factor regressors were observable. The asymptotic property is supported by the simulation studies in finite samples in Section 4.1. The proposed methods perform reasonably well in the simulation. Furthermore, the developed predictive model and forecasting methodology are applied to predict the UK inflation change, providing satisfactory forecasting performance. In particular, our empirical result shows that our proposed nonlinear forecasting method outperforms some commonly-used parametric forecasting methods.

There are some interesting topics that are related to this thesis and will be considered in future studies. In Section 2.3.2, different methods for determining the number of common factors are reviewed. It might be interesting to connect the selection of factor number to nonlinear forecasting and choose the optimal factor number which results in good forecasting performance. Another interesting topic is on the multi-step ahead nonlinear forecast of the proposed model and construction of prediction intervals.

References

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., K. Li and L. Lu (2016). Estimation and inference of FAVAR models. *Journal of Business and Economic Statistics 34*(4), 620–641.

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

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

Bernanke, B. S., J. Boivin and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*(1), 387–422.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer-Verlag.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.

Bühlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics 136*(1), 163–188.

Cai, Z., J. Fan and R. Li (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association 95*(451), 888–902.

Cai, Z., J. Fan and Q. Yao (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association 95*(451), 941–956.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica 51*(5), 1305–1323.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets.

Chen, J., D. Li, O. Linton and Z. Lu (2018). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association 113*(522), 919–932.

Chen, R. and R. S. Tsay (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association 88*(421), 298–308.

Chen, X. B., J. Gao, D. Li and P. Silvapulle (2018). Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business and Economic Statistics 36*(1), 88–100.

Cheng, M.-Y., T. Honda and J.-T. Zhang (2016). Forward variable selection for sparse ultra-high dimensional varying coefficient models. *Journal of the American Statistical Association 111*(515), 1209–1221.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*, 280–293.

Cleveland, W., E. Grosse and W. Shyu (1992). Local regression models. In J. M. Chambers and T. J. Hastie (Eds.), *Statistical Models in S*. New York: Chapman & Hall.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

De Boor, C. (2001). *A practical Guide to Splines*. New York: Springer-Verlag.

Djogbenou, A. (2017). Model selection in factor-augmented regressions with estimated factors. Technical report. Working paper.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American statistical Association 87*(420), 998–1004.

Fan, J., M. Farmen and I. Gijbels (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(3), 591–608.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

Fan, J. and T. Huang (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli 11*(6), 1031–1057.

Fan, J., Y. Liao and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society: Series B 75*(4), 603–680.

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911.

Fan, J. and J. Lv (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica 20*(1), 101.

Fan, J., Y. Ma and W. Dai (2014). Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models. *Journal of the American Statistical Association 109*(507), 1270–1284.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer-Verlag.

Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *Annals of Statistics 27*(5), 1491–1518.

Fan, J. and W. Zhang (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface 1*(1), 179–195.

Gasser, T. and H.-G. Müller (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics 11*(3), 171–185.

Gonçalves, S., B. Perron and A. Djogbenou (2017). Bootstrap prediction intervals for factor models. *Journal of Business and Economic Statistics 35*(1), 53–69.

Hamilton, J. D. (1994). *Time Series Analysis*, Volume 2. Princeton University Press Princeton, NJ.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*, 726–748.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B 55*(4), 757–796.

Hastie, T., R. Tibshirani and M. Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC press.

Jiang, J. (2014). Multivariate functional-coefficient regression models for nonlinear vector time series data. *Biometrika 101*(3), 689–702.

Jiang, Q., H. Wang, Y. Xia and G. Jiang (2013). On a principal varying coefficient model. *Journal of the American Statistical Association 108*(501), 228–236.

Jolliffe, I. (2002). *Principal Component Analysis*. New York: Springer-Verlag.

Kai, B., R. Li and H. Zou (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of Statistics 39*(1), 305–332.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*(1), 141–151.

Li, D., Y. Ke and W. Zhang (2015). Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. *The Annals of Statistics 43*(6), 2676–2705.

Lian, H. (2012). Variable selection for high-dimensional generalized varying-coefficient models. *Statistica Sinica 22*(4), 1563–1588.

Liu, J., R. Li and R. Wu (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association 109*(505), 266–274.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications 9*(1), 141–142.

Park, B. U., E. Mammen, Y. K. Lee and E. R. Lee (2015). Varying coefficient regression models: a review and new developments. *International Statistical Review 83*(1), 36–64.

Pesaran, M. H., A. Pick and A. Timmermann (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics 164*, 173–187.

Robinson, P. M. (1989). Nonparametric estimation of time-varying paramete. In P. Hackl (Ed.), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253–264. Berlin: Springer-Verlag.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 341–360.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

Stock, J. H. and M. W. Watson (1998). Diffusion indexes. Technical report, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

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

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological) 36*(2), 111–147.

Teräsvirta, T., D. Tjostheim and C. Granger (2010). *Modelling Nonlinear Economic Time Series*. Oxford: Oxford University Press.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3 rd ed.). John Wiley & Sons.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

Wand, M. P. and M. C. Jones (1994). *Kernel smoothing*. Chapman and Hall/CRC.

Wang, H. (2009). Forward regression for ultra-high dimensional variable screening. *Journal of the American Statistical Association 104*(488), 1512–1524.

Wang, H. and Y. Xia (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association 104*(486), 747–757.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics: Series A 26*(4), 359–372.

Xia, Y., W. Zhang and H. Tong (2004). Efficient estimation for semivarying-coefficient models. *Biometrika 91*(3), 661–681.

Zhang, W. and H. Peng (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis 101*, 1656–1680.