
Article:
Li, Degui orcid.org/0000-0001-6802-308X, Tosasukul, Jiraroj and Zhang, Wenyang orcid.org/0000-0001-8391-1122 (2020) Nonlinear Factor-Augmented Predictive Regression Models with Functional Coefficients. Journal of Time Series Analysis. pp. 367-386. ISSN 1467-9892

https://doi.org/10.1111/jtsa.12511

Nonlinear Factor-Augmented Predictive Regression Models with Functional Coefficients

Degui Li[1], Jiraroj Tosasukul[1,2], Wenyang Zhang[1]

[1]*Department of Mathematics, University of York, United Kingdom*

[2]*Department of Mathematics, Faculty of Science, Naresuan University, Thailand*

28 September 2019

Abstract

This paper introduces a new class of functional-coefficient predictive regression models, where the regressors consist of auto-regressors and latent factor regressors, and the coefficients vary with certain index variable. The unobservable factor regressors are estimated through imposing an approximate factor model on high dimensional exogenous 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 (with appropriate rotation) 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. Under regularity conditions, the asymptotic properties of the proposed methods are derived, showing 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 developed model and methodology are further generalised to the factor-augmented vector predictive regression with functional coefficients. Finally, some extensive simulation studies and an empirical application to forecast the UK inflation are given to examine the finite-sample performance of the proposed model and methodology.

Keywords: Bootstrap procedure, Factor models, Functional-coefficient models, Local linear smoothing, Nonlinear forecast, PCA, Vector auto-regression.

1 Introduction

The functional-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 functional-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, Fan and Yao, 2000; Xia, Zhang and Tong, 2004; Kai, Li and Zou, 2011; Jiang *et al*, 2013). Fan and Zhang (2008) and Park *et al* (2015) provide an extensive review of some recent developments in the field. In the dynamic time series context, we start with an introduction of the following functional-coefficient predictive regression model:

$$\begin{aligned} y_{t+1} &= \sum_{i=1}^{q_n} \alpha_{1i}(u_t) z_{ti} + \sum_{j=1}^{d_0} \alpha_{2j}(u_t) y_{t+1-j} + \varepsilon_{t+1} \\ &= \mathbf{Z}_t^\tau \boldsymbol{\alpha}_1(u_t) + \mathbf{Y}_t^\tau \boldsymbol{\alpha}_2(u_t) + \varepsilon_{t+1}, \quad t = 1, \cdots, n, \end{aligned} \tag{1.1}$$

where y_{t+1} is a response variable, $\mathbf{Z}_t = (z_{t1}, \cdots, z_{tq_n})^\tau$ is a q_n-dimensional column vector of random covariates, $\mathbf{Y}_t = (y_t, \cdots, y_{t-d_0+1})^\tau$ is a column vector of d_0 lags of the response, u_t is a univariate index variable, $\boldsymbol{\alpha}_1(\cdot) = [\alpha_{11}(\cdot), \cdots, \alpha_{1q_n}(\cdot)]^\tau$ and $\boldsymbol{\alpha}_2(\cdot) = [\alpha_{21}(\cdot), \cdots, \alpha_{2d_0}(\cdot)]^\tau$ are two column vectors of coefficient functions, and ε_t is the model error. If $u_t = t/n$ with n as the time series length, we may call model (1.1) 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 the present paper, we assume that q_n, the number of exogenous covariates, diverges to infinity as the sample size n grows, but d_0, the number of lags, is fixed. 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 functional-coefficient models, enhancing the nonlinear model prediction accuracy (c.f., Wang and Xia, 2009; Lian, 2012; Fan, Ma and Dai, 2014; Liu, Li and Wu, 2014; Li, Ke and Zhang, 2015). However, as 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. 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* (2018a) 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 (1.1).

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

$$\mathbf{Z}_t = \mathbb{B}_n \mathbf{F}_t + \mathbf{V}_t, \tag{1.2}$$

where $\mathbb{B}_n$ is a $q_n \times k$ matrix of factor loadings, $\mathbf{F}_t = (F_{t1}, \cdots, F_{tk})^\tau$ is a k-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. The number k is usually unknown and can be determined via some data-driven criteria. In this paper, we allow k to increase slowly with the sample size n. In recent years, there has been increasing interest on studying the approximate factor model (1.2), see, for example, Chamberlain and Rothschild (1983), Fama and French (1992), Bai and Ng (2002), Fan, Liao and Mincheva (2013) and the references therein. Through the factor model (1.2), the latent factor time series process may carry a large proportion of the "dynamic information" contained in the high-dimensional observable time series vector $\mathbf{Z}_t$. In addition, by assuming that $\mathbf{F}_t$ and $\mathbf{V}_t$ are uncorrelated, we readily have

$$\boldsymbol{\Sigma}_Z = \mathbb{B}_n \boldsymbol{\Sigma}_F \mathbb{B}_n^\tau + \boldsymbol{\Sigma}_V, \tag{1.3}$$

where $\boldsymbol{\Sigma}_Z$, $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_V$ denote the covariance matrices for $\mathbf{Z}_t$, $\mathbf{F}_t$ and $\mathbf{V}_t$, respectively. In the high-dimensional setting with $q_n > n$, some existing literature such as Fan, Liao and Mincheva (2013) usually assumes that the idiosyncratic error covariance matrix $\boldsymbol{\Sigma}_V$ is sparse with all the eigenvalues bounded (e.g., Bickel and Levina, 2008), indicating low correlation among the components of $\mathbf{V}_t$. On the other hand, the k eigenvalues of $\mathbb{B}_n^\tau \mathbb{B}_n$ are typically assumed to be divergent at the rate of $O(q_n)$, see Assumption 2(ii) in Section 3 below, indicating that the latent factors in (1.2) are pervasive. Hence, the high-dimensional random covariates $\mathbf{Z}_t$ satisfying the approximate factor structure may have high correlation among their components.

In this paper, 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:

$$y_{t+1} = \mathbf{F}_t^\tau \boldsymbol{\beta}_1(u_t) + \mathbf{Y}_t^\tau \boldsymbol{\beta}_2(u_t) + \epsilon_{t+1}, \tag{1.4}$$

where $\boldsymbol{\beta}_1(\cdot) = [\beta_{11}(\cdot), \cdots, \beta_{1k}(\cdot)]^\tau$ and $\boldsymbol{\beta}_2(\cdot) = [\beta_{21}(\cdot), \cdots, \beta_{2d_0}(\cdot)]^\tau$ are two column vectors of coefficient functions, and ϵ_t is the error term. A significant difference between models (1.1) and (1.4) is that the factor regressors $\mathbf{F}_t$ in the latter are unobservable, while all of the regressors in the former are observable. Furthermore, the number of regressors in (1.4) is $k + d_0$, which is much smaller than that in model (1.1). Through a combination of (1.2) and (1.4), we obtain the functional-coefficient predictive regression model with latent factor regressors and call it as the *Factor-Augmented Functional-Coefficient Model (FA-FCM)*. The FA-FCM can be viewed as a generalisation of the linear factor-augmented regression or auto-regression models (c.f., Stock and Watson, 2002; Bernanke, Boivin and Eliasz, 2005; Bai and Ng, 2006; Pesaran, Pick and Timmermann, 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 (1.4) could be more accurate than that using the conventional functional-coefficient time series model (1.1).

The main contribution of this paper is two-fold. First, we introduce a two-stage estimation procedure to estimate the coefficient functions (subject to appropriate rotation) in model (1.4): the *Principal Component Analysis (PCA)* technique is used in stage one to estimate the rotated factor regressors, and then the local linear smoothing method is used in stage two to estimate the rotated coefficient functions with the rotation matrix defined in Section 2 below. Second, we introduce a one-step ahead nonlinear forecasting approach by using the estimates of the rotated coefficient functions and then construct the prediction interval by a wild bootstrap procedure as in Zhang and Peng (2010) and Chen *et al* (2018b). Under some mild conditions, we derive the asymptotic properties of the developed estimation and forecasting methods, from which we find that the local linear estimator and nonlinear forecast using the estimated factor regressors are asymptotically equivalent to the infeasible counterpart using the true latent factor regressors. Furthermore, we also extend the methodology and theory to the factor-augmented vector auto-regression with functional coefficients, substantially generalising the parametric linear factor-augmented vector auto-regression models which have been extensively studied in the literature. Some simulation studies and an empirical application to predict the UK inflation are given to illustrate our model and methodology in finite samples. In particular, our empirical result shows that the proposed nonlinear forecasting method outperforms some commonly-used parametric forecasting methods.

The rest of the paper is organised as follows. Section 2 introduces the nonlinear estimation and forecasting methodology. Section 3 gives the asymptotic results of the proposed methods. Section 4 discusses extension of the developed model to nonlinear vector auto-regression with multivariate response. Section 5 reports the simulation studies. Section 6 applies the models and methods to analyse a real data set. Section 7 concludes the paper. All the proofs of the asymptotic results are given in a supplemental document.

2 Methodology

We next introduce nonparametric estimation of the coefficient functions (with appropriate rotation) and one-step ahead nonlinear prediction of the response, and give a data-driven selection criterion to determine the numbers of lags and factor regressors in the predictive regression model (1.4).

2.1 Estimation of the rotated coefficient functions

As the factor regressors in model (1.4) are unobservable, in order to develop a feasible nonparametric estimation and forecasting approach, we next introduce a two-stage estimation procedure. For the time being, we assume that the numbers k and d_0 are known, and Section 2.3 below will discuss how to determine these numbers in practice.

Stage 1: We obtain the estimated factor regressors by using the PCA approach. Specifically, letting $\mathbb{Z}_n = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^\tau$, an $n \times q_n$ matrix of observations, we conduct an eigenanalysis on the $n \times n$ (normalised) matrix $\mathbb{Z}_n\mathbb{Z}_n^\tau/(nq_n)$, and obtain $\hat{\mathbb{F}}_n = \big(\hat{\mathbf{F}}_1, \ldots, \hat{\mathbf{F}}_n\big)^\tau$, an $n \times k$ matrix which consists of the k eigenvectors (multiplied by $\sqrt{n}$) associated with the k largest eigenvalues of the matrix $\mathbb{Z}_n\mathbb{Z}_n^\tau/(nq_n)$ (ranked in the descending order). In addition, we may further construct the estimation of factor loading matrix (with rotation) by $\hat{\mathbb{B}}_n = \mathbb{Z}_n^\tau\hat{\mathbb{F}}_n/n$, where we have used the fact of $\hat{\mathbb{F}}_n^\tau\hat{\mathbb{F}}_n/n = \mathbf{I}_k$ with $\mathbf{I}_k$ being a $k \times k$ identity matrix. This PCA estimation method has been extensively studied in the literature (c.f., Bai and Ng, 2002; Stock and Watson, 2002), which shows that, under some mild conditions, $\hat{\mathbf{F}}_t$ *is a consistent estimate of the rotated common factor* $\mathbf{H}\mathbf{F}_t$ (see also the proof of Lemma B.2 in the supplemental document), where

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

and $\mathbf{Q}_n$ is a $k \times k$ diagonal matrix of the first k largest eigenvalues of $\mathbb{Z}_n\mathbb{Z}_n^\tau/(nq_n)$ 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 the local linear smoothing method (Fan and Gijbels, 1996). Letting $\mathbf{X}_t = \left[(\mathbf{H}\mathbf{F}_t)^\tau, \mathbf{Y}_t^\tau\right]^\tau$, $\boldsymbol{\beta}_H(\cdot) = \left[\boldsymbol{\beta}_1^\tau(\cdot)\mathbf{H}^{-1}, \boldsymbol{\beta}_2^\tau(\cdot)\right]^\tau$ and noting that $\mathbf{H}^{-1}\mathbf{H} = \mathbf{I}_k$, we may rewrite model (1.4) as

$$y_{t+1} = \boldsymbol{\beta}_H^\tau(u_t)\mathbf{X}_t + \epsilon_{t+1}. \tag{2.1}$$

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 Section 3), implying that $\boldsymbol{\beta}_H(\cdot)$ has 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 $\boldsymbol{\beta}_H(\cdot)$. Let $\mathbb{Y}_n = (y_2, \cdots, y_{n+1})^\tau$,

$$\mathbb{X}_n(u) = \begin{pmatrix} \mathbf{X}_1^\tau & \mathbf{X}_1^\tau(u_1 - u) \\ \vdots & \vdots \\ \mathbf{X}_n^\tau & \mathbf{X}_n^\tau(u_n - u) \end{pmatrix}, \quad \mathbb{W}_n(u) = \mathsf{diag}\left\{K_b(u_1, u), \cdots, K_b(u_n, u)\right\}$$

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

$$\tilde{\boldsymbol{\beta}}_H(u) = (\mathbf{I}_{k+d_0}, \mathbf{O}_{k+d_0})\left[\mathbb{X}_n^\tau(u)\mathbb{W}_n(u)\mathbb{X}_n(u)\right]^{-1}\left[\mathbb{X}_n^\tau(u)\mathbb{W}_n(u)\mathbb{Y}_n\right], \tag{2.2}$$

where u is on the support of the index variable u_t and $\mathbf{O}_p$ is a $p \times p$ null matrix. However, the local linear estimation in (2.2) is infeasible and cannot be implemented directly as the factor regressors involved in $\mathbf{X}_t$ are unobservable. In practice, we have to replace $\mathbf{X}_t$ by $\hat{\mathbf{X}}_t = \left(\hat{\mathbf{F}}_t^\tau, \mathbf{Y}_t^\tau\right)^\tau$ with $\hat{\mathbf{F}}_t$ obtained in Stage 1, and let $\hat{\mathbb{X}}_n(u)$ be defined as $\mathbb{X}_n(u)$ but with $\mathbf{X}_t$ replaced by $\hat{\mathbf{X}}_t$. Then, we obtain the following feasible local linear estimate of $\boldsymbol{\beta}_H(u)$:

$$\hat{\boldsymbol{\beta}}_H(u) = (\mathbf{I}_{k+d_0}, \mathbf{O}_{k+d_0})\left[\hat{\mathbb{X}}_n^\tau(u)\mathbb{W}_n(u)\hat{\mathbb{X}}_n(u)\right]^{-1}\left[\hat{\mathbb{X}}_n^\tau(u)\mathbb{W}_n(u)\mathbb{Y}_n\right]. \tag{2.3}$$

In Section 3 below, we will show that the feasible local linear estimator $\hat{\boldsymbol{\beta}}_H(u)$ has the same asymptotic distribution as the infeasible one $\tilde{\boldsymbol{\beta}}_H(u)$ when both n and q_n are sufficiently large.

2.2 One-step ahead nonlinear forecast

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 constructed in Section 2.1, we may obtain the one-step ahead prediction of y_{n+1}:

$$\hat{y}_{n+1|n} = \hat{\boldsymbol{\beta}}^{\tau}_{H,n-1}(u_n)\hat{\mathbf{X}}_n, \tag{2.4}$$

where $\hat{\boldsymbol{\beta}}_{H,n-1}(\cdot)$ is the local linear estimate as in (2.3) using the sample $(y_{t+1}, u_t, \mathbf{Z}_t)$, $t = 1, \cdots, n-1$. In fact, $\hat{y}_{n+1|n}$ in (2.4) can be regarded as a natural estimate of

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

The asymptotic property of $\hat{y}_{n+1|n}$ will be given in Theorem 2 in Section 3 below. In practice, it is often of interest to further construct the confidence interval of $y_{n+1|n}$. For given $0 < \alpha < 1$, the $(1-\alpha)$ confidence interval of $y_{n+1|n}$ can be defined by

$$\left[\hat{y}_{n+1|n} - c_{\alpha/2} \times \sqrt{\widehat{\mathsf{var}}(\hat{y}_{n+1|n})},\ \hat{y}_{n+1|n} + c_{\alpha/2} \times \sqrt{\widehat{\mathsf{var}}(\hat{y}_{n+1|n})}\right], \tag{2.5}$$

where $c_{\alpha/2}$ is the upper $\alpha/2$-percentile of $\left(\hat{y}_{n+1|n} - y_{n+1|n}\right)/\sqrt{\widehat{\mathsf{var}}(\hat{y}_{n+1|n})}$ and $\widehat{\mathsf{var}}(\hat{y}_{n+1|n})$ is the estimate of the variance of $\hat{y}_{n+1|n}$.

However, the confidence interval in (2.5) cannot be directly used as neither $c_{\alpha/2}$ nor $\widehat{\mathsf{var}}(\hat{y}_{n+1|n})$ is known. These quantities may be estimated using the asymptotic result (e.g., Theorem 3), and then we could construct a feasible confidence interval. However, such an interval construction based on the asymptotic theory usually does not perform well in small or medium samples. Hence, we next use a wild bootstrap procedure to estimate $c_{\alpha/2}$ and $\widehat{\mathsf{var}}(\hat{y}_{n+1|n})$, and then proceed to construct the prediction interval. The following bootstrap procedure is similar to those in Zhang and Peng (2010) and Chen *et al* (2018b) which construct the point-wise or simultaneous confidence bands in the functional-coefficient models. It can also be seen as a nonparametric generalisation of the bootstrap prediction interval introduced by Gonçalves, Perron and Djogbenou (2017) for the parametric linear factor-augmented regression model.

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 (2.3), 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 2.1, 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 (2.4).

STEP 2: Generate the bootstrap sample:

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

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 $N(0,1)$, 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 $\hat{\text{var}}^*(\hat{y}_{n+1|n})$.

STEP 5: For each $i = 1, \cdots, M$, use the sequence $\hat{y}_{n+1|n}(i)$ and $\hat{\text{var}}^*(\hat{y}_{n+1|n})$ to compute $q^*_n(i) = \left[\hat{y}_{n+1|n}(i) - \hat{y}_{n+1|n}\right] / \sqrt{\hat{\text{var}}^*(\hat{y}_{n+1|n})}$, 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 $\text{var}^*(\hat{y}_{n+1|n})$ obtained in the above bootstrap procedure, we can construct the feasible $(1-\alpha)$ confidence interval of $y_{n+1|n}$ by

$$\left[\hat{y}_{n+1|n} - c^*_{\alpha/2} \times \sqrt{\hat{\text{var}}^*(\hat{y}_{n+1|n})}, \; \hat{y}_{n+1|n} + c^*_{\alpha/2} \times \sqrt{\hat{\text{var}}^*(\hat{y}_{n+1|n})}\right]. \tag{2.6}$$

As in Teräsvirta, Tjøstheim and Granger (2010), here we let the auto-regressors in the time series model be invariant when generating the bootstrap samples, although other bootstrap methods may also be applicable in our setting. In addition, the bias of local linear estimation is ignored in our construction of prediction interval to simplify the methodology. Following the arguments in Zhang and Peng (2010) and Chen *et al* (2018b), we can give theoretical justification of the above bootstrap method. Details are omitted to save the space.

2.3 The forward selection criterion

In order to implement the estimation and forecasting method introduced in Sections 2.1 and 2.2, we need to determine d_0, the number of lags, and k, the number of latent factor regressors. The model selection for parametric linear factor-augmented models is recently studied by Djogbenou (2016). However, his selection criterion is not applicable to our nonparametric model setting. Meanwhile, estimating the number of factors is an important issue in factor analysis and has been extensively studied in the literature (c.f., Bai and Ng, 2002; Onatski, 2009; Ahn and Horenstein, 2013; Li, Li and Shi, 2017). However, the estimated number of factors by some existing approaches such as the information criterion (Bai and Ng, 2002) or ratio criterion (Ahn and Horenstein, 2013) may not be "optimal" in nonlinear forecasting. We next use the forward selection criterion as a screening tool and employ the *Bayesian Information Criteria (BIC)* as the stopping rule to estimate d_0 and k, motivated by the forward selection method introduced by Wang (2009) and Cheng, Honda and Zhang (2016) for the high-dimensional regression models without any latent factor regressor.

Let $\tilde{k}$ be an initial estimate of k whose construction will be discussed in Section 5 below. If the number of lags is assumed to be d, a positive integer, we let $\hat{\boldsymbol{\beta}}_H(u|\tilde{k}, d)$ be the feasible local linear estimated coefficient function (with rotation), similar to the definition of $\hat{\boldsymbol{\beta}}_H(u)$ given in (2.3). Let $\hat{\mathbf{X}}_t(l, d)$ be defined as $\hat{\mathbf{X}}_t$ when the number of lags is d and the number of estimated factors is l. In our data-driven selection procedure, we examine the change of residual sum of squares defined by

$$\hat{\sigma}^2_{n1}(d) = \frac{1}{n}\sum_{t=1}^{n}\left[y_{t+1} - \hat{\boldsymbol{\beta}}_H^{\tau}(u_t|\tilde{k}, d)\hat{\mathbf{X}}_t(\tilde{k}, d)\right]^2,$$

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

$$\mathsf{BIC}_1(d) = \log \hat{\sigma}^2_{n1}(d) + d \cdot \frac{\log(nb)}{nb} \tag{2.7}$$

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 $\mathsf{BIC}_1(0) = \log \hat{\sigma}^2_{n1}(0)$. In the second step, we add the first lag to the model and compute $\mathsf{BIC}_1(1)$. 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 $\mathsf{BIC}_1(\hat{d}+1) > \mathsf{BIC}_1(\hat{d})$.

After obtaining $\hat{d}$, we next re-estimate the number of factor regressors which is optimal for

one-step ahead forecasting. Let

$$\mathcal{K} = \left\{ \lfloor \tilde{k}/2 \rfloor, \lfloor \tilde{k}/2 \rfloor + 1, \cdots, \tilde{k} + \lfloor \tilde{k}/2 \rfloor \right\},$$

where $\lfloor \cdot \rfloor$ denotes the floor function and $\tilde{k}$ is the initial estimate of k. When the number of factor regressors is $l \in \mathcal{K}$, similar to $\hat{\sigma}^2_{n1}(\cdot)$ and $\mathsf{BIC}_1(\cdot)$, we define the residual sum of squares:

$$\hat{\sigma}^2_{n2}(l) = \frac{1}{n} \sum_{t=1}^{n} \left[y_{t+1} - \hat{\boldsymbol{\beta}}^{\intercal}_{H}(u_t | l, \hat{d}) \hat{\mathbf{X}}_t(l, \hat{d}) \right]^2,$$

and the BIC value:

$$\mathsf{BIC}_2(l) = \log \hat{\sigma}^2_{n2}(l) + l \cdot \frac{\log(nb)}{nb}. \tag{2.8}$$

Then, the final estimate of k is obtained by

$$\hat{k} = \arg\min_{l \in \mathcal{K}} \mathsf{BIC}_2(l). \tag{2.9}$$

3 Asymptotic theorems

In this section, we give the asymptotic results for the methodologies developed in Sections 2.1 and 2.2. Throughout the paper, we let $\|\cdot\|$ denote the Euclidean norm of a vector, $\|\cdot\|_F$ and $\|\cdot\|_O$ denote the Frobenius norm and operator norm of a matrix, respectively, 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. We start with some regularity conditions which are needed to prove our 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* $\alpha_j \sim c_\alpha \rho^j$ *as* j *is sufficiently large, 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* $\mathcal{C}$.

ASSUMPTION 2. *(i) The latent factor regressors satisfy that* $\mathsf{E}(\mathbf{F}_t) = \mathbf{0}_k$ *and* $\max_{1 \leqslant s \leqslant k} \mathsf{E}\left[|F_{ts}|^{2(2+\delta)}\right] < \infty$*, where* $0 < \delta < \infty$*,* $\mathbf{0}_k$ *is a* k*-dimensional null vector and* F_{ts} *is the* s*-th element of* $\mathbf{F}_t$*. In addition, the* $k \times k$ *matrix* $\boldsymbol{\Lambda}_F = \mathsf{E}\left[\mathbf{F}_t \mathbf{F}_t^{\intercal}\right]$ *is positive definite.*

(ii) There exists a $k \times k$ *matrix* $\mathbf{\Lambda}_B$ *such that*

$$\left\| \frac{\mathbb{B}_n^\intercal \mathbb{B}_n}{q_n} - \mathbf{\Lambda}_B \right\|_F = o(1).$$

The matrix $\mathbf{\Lambda}_B$ *is positive definite with eigenvalues bounded from zero and infinity.*

(iii) There exist two $k \times k$ *positive definite matrices* $\mathbf{Q}_0$ *and* $\hat{\mathbf{\Lambda}}_F$ *with eigenvalues bounded from zero and infinity such that*

$$\|\mathbf{Q}_n - \mathbf{Q}_0\|_F = o_P(1) \text{ and } \left\| \frac{\hat{\mathbb{F}}_n^\intercal \mathbb{F}_n}{n} - \hat{\mathbf{\Lambda}}_F \right\|_F = o_P(1),$$

where $\mathbf{Q}_n$, $\hat{\mathbb{F}}_n$ *and* $\mathbb{F}_n$ *are defined in Section 2.1.*

(iv) The eigenvalues of the $k \times k$ *matrix* $\mathbf{\Lambda}_B^{1/2} \mathbf{\Lambda}_F \mathbf{\Lambda}_B^{1/2}$ *are distinct.*

ASSUMPTION 3. *Letting* $\mathbf{X}_t^\star = \left(\mathbf{F}_t^\intercal, \mathbf{Y}_t^\intercal\right)^\intercal$, *the* $(k + d_0) \times (k + d_0)$ *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* $u \in \mathcal{C}$. *The smallest eigenvalue of* $\mathbf{\Lambda}_X(u)$ *is larger than a positive constant uniformly 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* $kb \to 0$ *and* $n^{1-\tau-1/(2+\delta)} b/k \to \infty$, *where* $\tau > 0$ *can be arbitrarily small and* δ *is defined in Assumption 2(i).*

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

$$\max_{1 \leqslant s \leqslant k} \mathsf{E}\left[\left|\sum_{i=1}^{q_n} b_{is} v_{ti}\right|^{\delta_1}\right] \leqslant m_0 q_n^{\delta_1/2} \tag{3.1}$$

and

$$\mathsf{E}\left[\left|\sum_{i=1}^{q_n} (v_{t_1 i} v_{t_2 i} - \mathsf{E}[v_{t_1 i} v_{t_2 i}])\right|^{\delta_1}\right] \leqslant m_0 q_n^{\delta_1/2}, \tag{3.2}$$

where v_{ti} *is the* i*-th random element in* $\mathbf{V}_t$ *and* b_{is} *is the* (i, s)*-entry of* $\mathbb{B}_n$.

(ii) Let $q_n/(nb) \to \infty$,

$$n = o\left(\min\left\{[q_n/(nb)]^{\delta_1/2}k^{-(1+5\delta_1/2)}, [q_n/(nb)]^{\delta_*/2}k^{-(2\delta_*+1)}, [q_n/(nb)]^{\delta_1/4}k^{-\delta_1/4}\right\}\right),$$

where $\delta_*^{-1} = 1/[2(2+\delta)] + 1/\delta_1$. *In addition,* $k = o\left(n^{(2+\delta)\tau/4} \wedge n^{1/(2+\delta)+5\tau/6}\right)$.

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 $\beta_1(\cdot)$ *and* $\beta_2(\cdot)$ *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 Remark 2 below. Assumption 6(i) is similar to the condition B4 in Chen *et al* (2018a), indicating that for any t, v_{ti} is allowed to be weakly dependent over i. For the special case of cross-sectional independent v_{ti} (over i), if $\max_i \mathsf{E}[|v_{ti}|^{\delta_1}] < \infty$ and b_{is} is bounded uniformly over i and s, by the Rosenthal inequality for sum of independent random variables (e.g., Rosenthal, 1970), we can easily verify (3.1). Similarly (3.2) can be verified if, in addition, $\max_i \mathsf{E}[|v_{ti}|^{2\delta_1}] < \infty$. Assumption 6(ii) shows that both n and q_n diverge to infinity simultaneously in our asymptotic results and their relationship is relevant to the moment conditions. In particular, the dimension of exogenous regressors is allowed to be much larger than the time series length. Assumptions 5 and 6(ii) also impose some mild restriction on k, the number of factors. This number is allowed to be divergent at a slow polynomial rate of n.

The following theorem shows that the feasible local linear estimator $\hat{\boldsymbol{\beta}}_H(u)$ defined in (2.3) 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 are satisfied. Then,*

$$\sup_{u\in\mathcal{C}} \left\|\hat{\boldsymbol{\beta}}_H(u) - \tilde{\boldsymbol{\beta}}_H(u)\right\| = o_P\left((nb)^{-1/2}\right). \tag{3.3}$$

We next turn to the point-wise asymptotic distribution theory for the feasible local linear estimator $\hat{\boldsymbol{\beta}}_H(u)$. Let $\mu_j = \int u^j K(u)du$, $\nu_j = \int u^j K^2(u)du$, $\mathbf{H}_0 = \mathbf{Q}_0^{-1}\hat{\boldsymbol{\Lambda}}_F\boldsymbol{\Lambda}_B$, $\bar{\mathbf{H}}_0 = \mathsf{diag}\{\mathbf{H}_0, \mathbf{I}_{d_0}\}$,

$$\boldsymbol{\Lambda}(u) = \bar{\mathbf{H}}_0\boldsymbol{\Lambda}_X(u)\bar{\mathbf{H}}_0^\tau,$$

and define $\boldsymbol{\beta}''_H(\cdot)$ as the second-order derivative of $\boldsymbol{\beta}_H(\cdot)$. Let $\mathbf{D}_n$ be a $k_0 \times (k + d_0)$ matrix (which might depend on u) such that $\|\mathbf{D}_n\|_O$ is bounded and

$$\mathbf{D}_n\boldsymbol{\Lambda}^{-1}(u)\mathbf{D}_n^\tau \to \mathbf{D}_0(u) \text{ as } n \to \infty, \tag{3.4}$$

where $\mathbf{D}_0(u)$ is a $k_0 \times k_0$ symmetric and nonnegative definite matrix and k_0 is a fixed positive integer.

THEOREM 2. *Suppose Assumptions 1–7 are satisfied and* $b \propto n^{-1/5}$. *Then we have*

$$(nb)^{1/2}\mathbf{D}_n\left[\hat{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u) - \frac{1}{2}\mu_2\boldsymbol{\beta}''_H(u)b^2\right] \xrightarrow{d} N\left[\mathbf{0}_{k_0}, \frac{\sigma_\epsilon^2\nu_0}{f(u)}\mathbf{D}_0(u)\right], \quad n \to \infty. \tag{3.5}$$

REMARK 2. If the number of latent factor regressors is fixed, $k + d_0$ would be a fixed positive integer. In this special case, we may choose $k_0 = k + d_0$ and $\mathbf{D}_n = \mathbf{I}_{k+d_0}$, and consequently the asymptotic normal distribution in (3.5) becomes

$$(nb)^{1/2}\left[\hat{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u) - \frac{1}{2}\mu_2\boldsymbol{\beta}''_H(u)b^2\right] \xrightarrow{d} N\left[\mathbf{0}_{k+d_0}, \frac{\sigma_\epsilon^2\nu_0}{f(u)}\boldsymbol{\Lambda}^{-1}(u)\right], \quad n \to \infty. \tag{3.6}$$

Furthermore, when k is fixed, following the uniform consistency results in Hansen (2008) and Li, Lu and Linton (2012), we may show that

$$\sup_{u\in[\gamma,1-\gamma]}\left\|\tilde{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u)\right\| = O_P\left(b^2 + [\log n/(nb)]^{1/2}\right), \tag{3.7}$$

where γ is a small positive constant between 0 and 1/2. Combining (3.3) in Theorem 1 and (3.7), we readily have that

$$\sup_{u\in[\gamma,1-\gamma]}\left\|\hat{\boldsymbol{\beta}}_H(u) - \boldsymbol{\beta}_H(u)\right\| = O_P\left(b^2 + [\log n/(nb)]^{1/2}\right). \tag{3.8}$$

In Section 2.2, we construct the one-step ahead nonlinear forecast $\hat{y}_{n+1|n}$ using the local linear estimates. The following theorem describes the asymptotic prediction accuracy. For simplicity, we let the number of factors k be fixed.

THEOREM 3. *Suppose that the conditions in Theorem 2 are satisfied and* k *is a fixed positive integer. Then,*

$$\hat{y}_{n+1|n} - y_{n+1} = \Delta(u_n, \mathbf{X}_n) - \epsilon_{n+1} + o_P\left(1/(nb)^{1/2}\right), \tag{3.9}$$

where

$$\Delta(u_n, \mathbf{X}_n) = \left[\hat{\boldsymbol{\beta}}_{H,n-1}(u_n) - \boldsymbol{\beta}_H(u_n)\right]^\tau \mathbf{X}_n.$$

Furthermore, conditional on $u_n = u_*$ *and* $\mathbf{X}_n = \mathbf{X}_*$*, we have*

$$(nb)^{1/2}\left[\Delta(u_*, \mathbf{X}_*) - \frac{1}{2}\mu_2 b^2 \mathbf{X}_*^\tau \boldsymbol{\beta}_H''(u_*)\right] \xrightarrow{d} N\left[0, \frac{\sigma_\epsilon^2 \nu_0}{f(u_*)} \mathbf{X}_*^\tau \boldsymbol{\Lambda}^{-1}(u_*) \mathbf{X}_*\right]. \tag{3.10}$$

REMARK 3. By the definition of $y_{n+1|n}$ in Section 2.2, we may show that

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

indicating that $(nb)^{1/2}\left(\hat{y}_{n+1|n} - y_{n+1|n}\right)$ has the same (conditional) asymptotic normal distribution as that in (3.10).

REMARK 4. In this section, we assume that the number of factor regressors and lags of response is correctly specified when deriving the asymptotic theory. Under-estimation and over-estimation of this number have different impacts on the asymptotic results. For example, if fewer than the true number of factors is selected in the PCA estimation, the model is under-fitted, leading to inconsistent functional coefficient estimation and inaccurate one-step ahead prediction. On the other hand, if more than true number of factors is selected, the model is over-fitted, which would not affect the estimation consistency but may reduce the estimation efficiency.

4 Extension to 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, 2006). 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 with linear vector auto-regression, and studying the

so-called factor-augmented vector auto-regressive models. Such a modelling approach is first introduced by Bernanke, Boivin and Eliasz (2005), and has been extensively studied in the literature (c.f., Bai and Ng, 2006; Bai, Li and Lu, 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 Section 2 to the case of multiple response variables. Specifically, suppose (1.2) and

$$\bar{\boldsymbol{y}}_{t+1} = \mathbf{B}_0^\intercal(u_t)\mathbf{F}_t + \sum_{j=1}^{d_1} \mathbf{B}_j^\intercal(u_t)\bar{\boldsymbol{y}}_{t+1-j} + \bar{\boldsymbol{\epsilon}}_{t+1}, \tag{4.1}$$

where $\bar{\boldsymbol{y}}_{t+1}$ is p_0-dimensional column vector of response variables, $\mathbf{B}_0(\cdot)$ is a $k \times p_0$ matrix of coefficient functions and $\mathbf{B}_j(\cdot)$ is a $p_0 \times p_0$ matrix of coefficient functions, $j = 1, \cdots, d_1$, and $\bar{\boldsymbol{\epsilon}}_t$ is a p_0-dimensional column vector of errors. Model (4.1) 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 2.1. Let

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

and then rewrite (4.1) as

$$\bar{\boldsymbol{y}}_{t+1} = \bar{\mathbf{B}}_H(u_t)\bar{\mathbf{X}}_t + \bar{\boldsymbol{\epsilon}}_{t+1}. \tag{4.2}$$

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 2.1. Let $\tilde{\mathbb{Y}}_n = (\bar{\boldsymbol{y}}_2, \cdots, \bar{\boldsymbol{y}}_{n+1})$ 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 (2.3), we can obtain the

following feasible local linear estimates of $\bar{\mathbf{B}}_H(u)$ and its derivative $\bar{\mathbf{B}}'_H(u)$:

$$\left[\tilde{\mathbf{B}}_H(u), \tilde{\mathbf{B}}'_H(u)\right] = \left[\tilde{\mathbb{Y}}_n \mathbb{W}_n(u) \tilde{\mathbb{X}}_n(u)\right] \left[\tilde{\mathbb{X}}_n^\intercal(u) \mathbb{W}_n(u) \tilde{\mathbb{X}}_n(u)\right]^{-1}, \tag{4.3}$$

where $\mathbb{W}_n(u)$ is defined in Section 2.1. As in Theorem 1 given in Section 3, under some regularity conditions, we may analogously prove that the above local linear estimators are asymptotically equivalent to those directly using the unobservable rotated factor regressors $\mathbf{HF}_t$. Details are omitted here to save space. Finally, the one-step ahead nonlinear forecast of $\bar{\boldsymbol{y}}_{n+1}$ can be constructed by following that in Section 2.2, i.e.,

$$\tilde{\boldsymbol{y}}_{n+1|n} = \tilde{\mathbf{B}}_{H,n-1}(u_n) \tilde{\mathbf{X}}_n, \tag{4.4}$$

where $\tilde{\mathbf{B}}_{H,n-1}(\cdot)$ is the local linear estimate as constructed in (4.3) using the sample $(\bar{\boldsymbol{y}}_{t+1}, u_t, \mathbf{Z}_t)$, $t = 1, \cdots, n-1$.

In this section, we limit our attention to the case of fixed p_0, i.e., the number of response variables is fixed. In practical applications, it may be not uncommon that the dimension of $\bar{\boldsymbol{y}}_t$ grows with n, resulting in high-dimensional vector FA-FCM. For the latter case, the methodology and theory developed in Sections 2 and 3 need to be substantially extended. For example, a penalised version of local linear smoothing (e.g., Wang and Xia, 2009) may be applied to estimate the high-dimensional coefficient matrices in (4.1) and the sparsity assumption is usually needed to develop sensible asymptotic theory. This will be explored in our future study.

5 Simulation studies

An important issue in the local linear estimation and one-step ahead nonlinear forecasting approaches is the choice of bandwidth b. As the underlying process is assumed to be stationary and weakly dependent (see Assumption 1 in Section 3), the classical cross-validation method is not applicable to our setting. We next use a modified multi-fold cross-validation criterion proposed by Cai, Fan and Yao (2000) to determine an appropriate bandwidth in the simulation studies. 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

define the following average mean squared error:

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

where

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

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 (2.3) with the bandwidth re-scaled to be $b \cdot [n/(n-qm)]^{1/5}$. In the simulation studies, as suggested by Cai, Fan and Yao (2000), we use $m = \lfloor 0.1n \rfloor$ and $Q = 4$. To save computational time, the above modified cross-validation bandwidth selection is only applied to the model determined by the data-driven selection criterion in 2.3, and the bandwidth b in (2.7) and (2.8) is simply chosen by the rule of thumb.

The criterion of selecting the factor number proposed in Section 2.3 requires an initial estimate $\tilde{k}$. In our numerical studies, the initial estimated number of factors is determined by choosing the first few eigenvectors of $\mathbb{Z}_n\mathbb{Z}_n^{\tau}/(nq_n)$ (corresponding to the first few largest eigenvalues) such that at least 80% of the total variation is accounted for.

We next present two simulated examples: one with univariate response variable and the other with bivariate response vector, corresponding to the methodologies developed in Sections 2 and 4, respectively.

EXAMPLE 1. Consider the following univariate FA-FCM:

$$y_{t+1} = \mathbf{F}_t^{\tau}\boldsymbol{\beta}_1(u_t) + \mathbf{Y}_t^{\tau}\boldsymbol{\beta}_2(u_t) + \sigma \cdot \epsilon_{t+1}, \quad t = 1, \cdots, n, \tag{5.1}$$

where $\boldsymbol{\beta}_1(u) = [\beta_{11}(u), \cdots, \beta_{1k}(u)]^{\tau}$ and $\boldsymbol{\beta}_2(u) = [\beta_{21}(u), \cdots, \beta_{2d_0}(u)]^{\tau}$ are two column vectors of coefficient functions, $k = 4$, $d_0 = 3$, $\beta_{11}(u) = \sin(u)$, $\beta_{12}(u) = \cos(u)$, $\beta_{13}(u) = \sqrt{u}$, $\beta_{14}(u) = \log(1+u)$, $\beta_{21}(u) = \frac{1}{4}\sin(u)$, $\beta_{22}(u) = \frac{1}{4}\cos(u)$ 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 $\mathsf{N}(0,1)$, $\sigma = 0.2$, $\mathbf{F}_t = (F_{t1}, \cdots, F_{t4})^{\tau}$ is a four-dimensional latent common factor vector with each factor component generated from the following $\mathsf{AR}(1)$ process:

$$F_{tj} = 0.5 \cdot F_{t-1,j} + z_{tj}, \quad j = 1, 2, 3, 4, \tag{5.2}$$

where z_{tj} are *i.i.d.* and follow the standard normal distribution. In addition, we use the factor model structure (1.2) to generate the exogenous observations $\mathbf{Z}_t$, where each row of the $q_n \times k$ factor loading matrix $\mathbb{B}_n$ is independently generated from $N(\mathbf{0}_4, \mathbf{I}_4)$ and the idiosyncratic error vector $\mathbf{V}_t \sim N(\mathbf{0}_{q_n}, \mathbf{I}_{q_n})$. The sample size n is set to be 200, 500 and 1000, whereas the dimension q_n is set to be 20, 150 and 500. The replication number is 200.

Table 1:
Percentages of correctly estimating k in Example 1

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	91.0%	95.5%	98.5%
$q_n = 150$	91.5%	96.5%	97.0%
$q_n = 500$	91.5%	95.0%	97.0%

Table 2:
Percentages of correctly estimating d_0 in Example 1.

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	90.0%	94.5%	95.0%
$q_n = 150$	97.0%	99.5%	100.0%
$q_n = 500$	96.0%	99.0%	100.0%

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 in the predictive model (5.1), and subsequently obtain the one-step ahead nonlinear forecast of the response as in Section 2.2. The common factor vector $\mathbf{F}_t$ and its dimension are usually unknown in practice. In the simulation, the latent factors are estimated by the PCA technique. The number of factors as well as the number of lags in the predictive model are determined by the selection criterion in Section 2.3. Tables 1 and 2 report the frequency of correctly estimating k and d_0, showing that the proposed selection method works well in finite samples. The bandwidth in the local linear smoothing method (for the selected model) is determined by the modified multi-fold cross-validation method introduced at the beginning of this section. 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 20 replications, and then use this average value as the optimal bandwidth in our simulation.

Table 3: MSPE of the infeasible one-step ahead forecast in Example 1

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	0.050247 (0.018)	0.043212 (0.009)	0.041964 (0.007)
$q_n = 150$	0.048863 (0.017)	0.042863 (0.009)	0.042404 (0.006)
$q_n = 500$	0.052585 (0.017)	0.042962 (0.008)	0.041039 (0.006)

Table 4: MSPE of the feasible one-step ahead forecast in Example 1

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	0.102878 (0.054)	0.090175 (0.040)	0.089434 (0.039)
$q_n = 150$	0.062720 (0.019)	0.055809 (0.012)	0.053106 (0.008)
$q_n = 500$	0.052585 (0.017)	0.046164 (0.009)	0.044172 (0.006)

We next examine the finite-sample performance of the one-step ahead nonlinear forecast constructed in Section 2.2. 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):

$$\mathsf{MSPE} = \frac{1}{\lfloor 0.1n \rfloor} \sum_{t=n-\lfloor 0.1n \rfloor}^{n} \left(\hat{y}_{t+1|t} - y_{t+1} \right)^2, \tag{5.3}$$

where $\hat{y}_{t+1|t}$ is defined as in (2.4). For the aim of comparison, we also consider the infeasible local linear estimation defined in (2.2) and use it to construct the infeasible one-step ahead nonlinear forecast. Tables 3 and 4 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, Figure 1 gives the 95% prediction interval in the out-sample forecasting by using the wild bootstrap procedure introduced in Section 2.2.

EXAMPLE 2. We next consider the bivariate FA-FCM with the following form:

$$\bar{\boldsymbol{y}}_{t+1} = \mathbf{B}_0^\tau(u_t)\mathbf{F}_t + \sum_{j=1}^{2} \mathbf{B}_j^\tau(u_t)\bar{\boldsymbol{y}}_{t+1-j} + \bar{\boldsymbol{\epsilon}}_{t+1}, \tag{5.4}$$

where $\bar{\boldsymbol{y}}_t = (y_{1t}, y_{2t})^\tau$ is a bivariate response vector, $\mathbf{F}_t$ is a four-dimensional latent factor vector generated in the same way as in Example 1, $\bar{\boldsymbol{\epsilon}}_t = (\epsilon_{1t}, \epsilon_{2t})^\tau$ 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 $\mathbf{B}_j(\cdot)$ is a 2×2 diagonal

matrix of coefficient functions, $j = 1, 2$. Specifically,

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

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 $\bar{\boldsymbol{y}}_t$ 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.

Table 5:

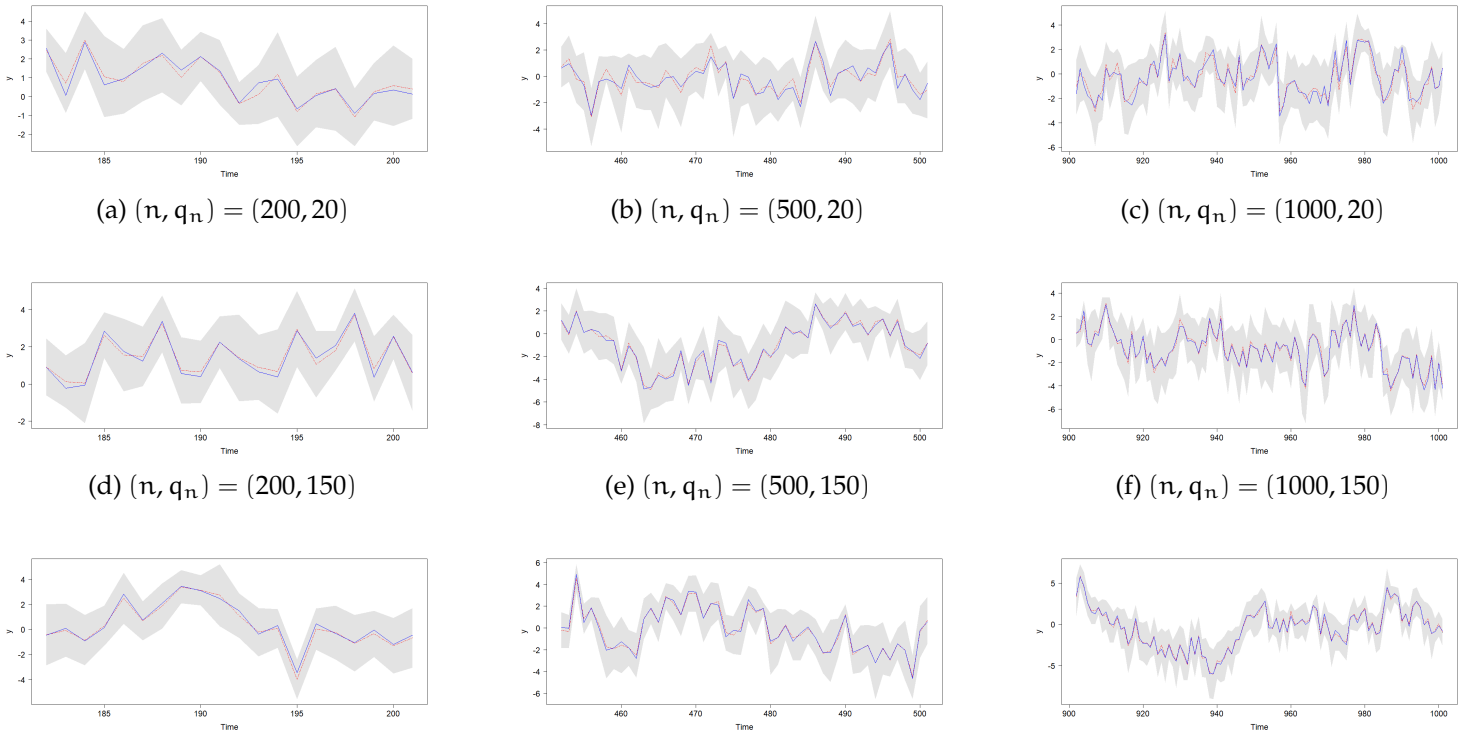
Percentages of correctly estimating k in Example 2

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	92.5%	94.5%	96.5%
$q_n = 150$	90.5%	95.5%	98.0%
$q_n = 500$	93.5%	96.0%	98.5%

Table 6:

Percentages of correctly estimating d_0 in Example 2.

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	73.0%	86.0%	96.0%
$q_n = 150$	71.5%	80.5%	90.5%
$q_n = 500$	74.0%	82.0%	91.5%



(a) $(n, q_n) = (200, 20)$ (b) $(n, q_n) = (500, 20)$ (c) $(n, q_n) = (1000, 20)$

(d) $(n, q_n) = (200, 150)$ (e) $(n, q_n) = (500, 150)$ (f) $(n, q_n) = (1000, 150)$

(g) $(n, q_n) = (200, 500)$ (h) $(n, q_n) = (500, 500)$ (i) $(n, q_n) = (1000, 500)$

Figure 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.

The methods used to choose the optimal bandwidth in the local linear estimation, determine the number of common factors and estimate the number of lags in the bivariate predictive model are the same as those in Example 1. Tables 5 and 6 give the percentages of correctly choosing 4 latent factors and the percentages of accurately identifying 2 lags in the model, respectively. Like in the case of univariate response (see Example 1), the proposed methods have reliable numerical performance in specifying the predictive model structure.

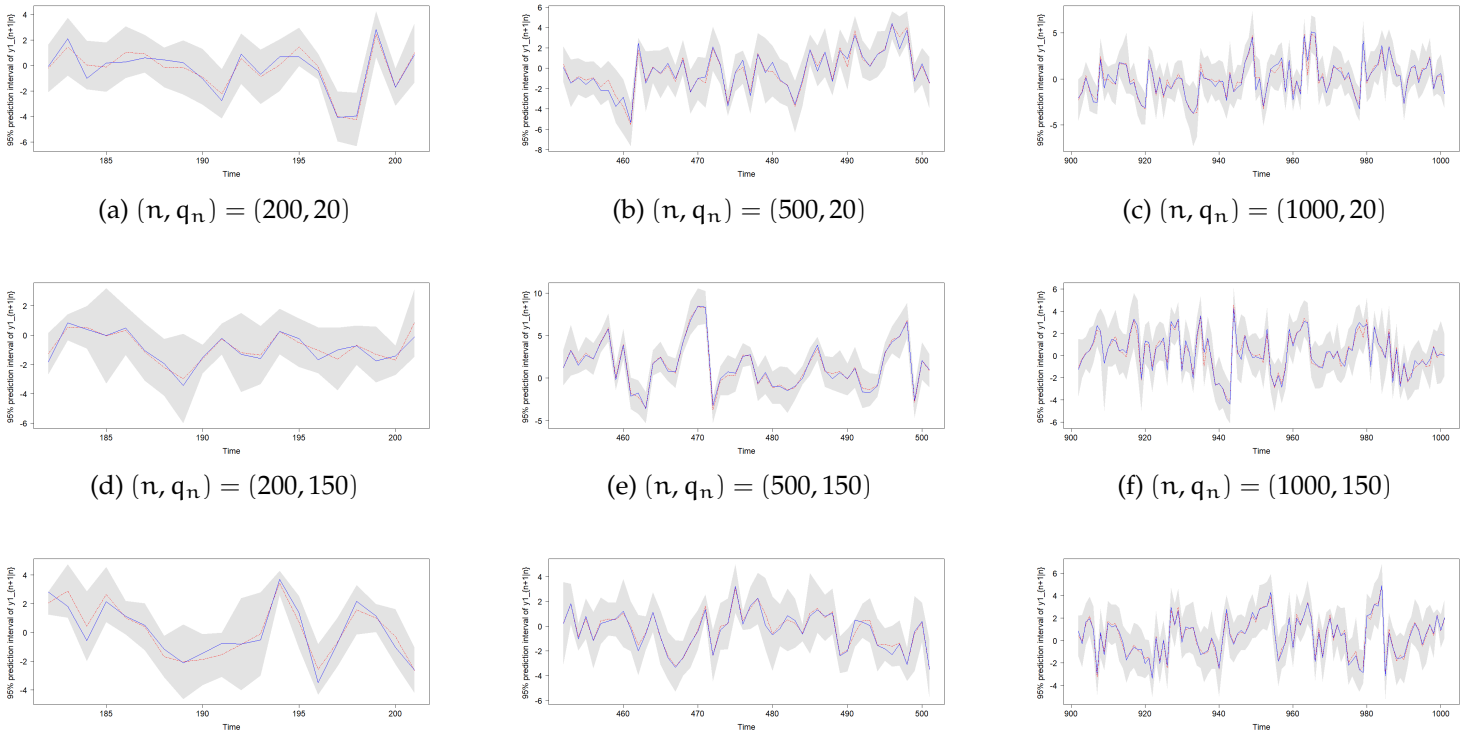
Table 7: MSPE of the infeasible one-step ahead forecast in Example 2

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	0.711482 (0.362)	0.645547 (0.207)	0.622903 (0.130)
$q_n = 150$	0.752796 (0.415)	0.644184 (0.181)	0.625819 (0.123)
$q_n = 500$	0.700678 (0.377)	0.640495 (0.248)	0.590312 (0.129)

Table 8: MSPE of the feasible one-step ahead forecast in Example 2

q_n \ n	$n = 200$	$n = 500$	$n = 1000$
$q_n = 20$	0.750918 (0.305)	0.688301 (0.244)	0.650344 (0.130)
$q_n = 150$	0.793823 (0.426)	0.677696 (0.180)	0.652213 (0.124)
$q_n = 500$	0.745995 (0.380)	0.661221 (0.250)	0.645943 (0.129)

As in Example 1, we 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 7 and 8, from which we may find that the nonlinear forecast using the feasible local linear estimation has a prediction accuracy similar 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 2 and 3 give the 95% prediction interval for $y_{1,t+1}$ and $y_{2,t+1}$, respectively, in the out-sample forecasting, where the bootstrap procedure is used.

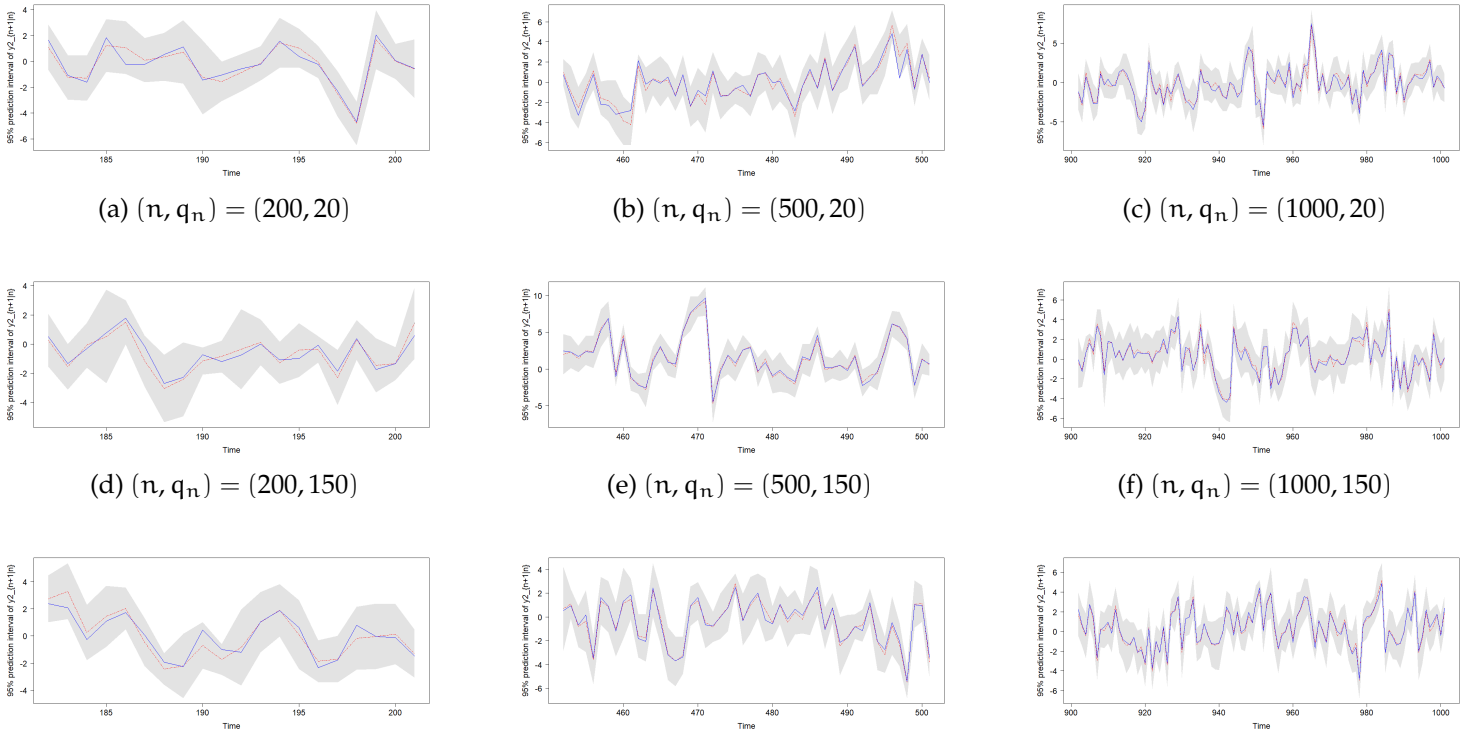


(a) $(n, q_n) = (200, 20)$ (b) $(n, q_n) = (500, 20)$ (c) $(n, q_n) = (1000, 20)$

(d) $(n, q_n) = (200, 150)$ (e) $(n, q_n) = (500, 150)$ (f) $(n, q_n) = (1000, 150)$

(g) $(n, q_n) = (200, 500)$ (h) $(n, q_n) = (500, 500)$ (i) $(n, q_n) = (1000, 500)$

Figure 2: 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.



(a) $(n, q_n) = (200, 20)$ (b) $(n, q_n) = (500, 20)$ (c) $(n, q_n) = (1000, 20)$

(d) $(n, q_n) = (200, 150)$ (e) $(n, q_n) = (500, 150)$ (f) $(n, q_n) = (1000, 150)$

(g) $(n, q_n) = (200, 500)$ (h) $(n, q_n) = (500, 500)$ (i) $(n, q_n) = (1000, 500)$

Figure 3: 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.

6 An empirical example

In this section, we apply the developed predictive model and forecasting methodology to forecast 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* (2018a) which explore the nonlinear dynamic relationship between the response y_t and the exogenous regressors $\mathbf{Z}_t$ as well as the lags $y_{t-j}, j = 1, 2, , \cdots$. This is different from the predictive regression structure considered in the present paper. 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* (2018a), we divided the dataset into two 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* (2018a), 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. Due to the high-dimension of the exogenous regressors, we imposed the approximate factor model structure (1.2) on $\mathbf{Z}_t$. Consider the following FA-FCM:

$$y_{t+1} = \sum_{i=1}^{k} F_{ti}\beta_{1i}(u_t) + \sum_{j=1}^{d_0} y_{t+1-j}\beta_{2j}(u_t) + \epsilon_{t+1}, \;\; u_t = y_t. \tag{6.1}$$

The initial estimate of the factor number is $\tilde{k} = 4$, accounting for 82.34% of the total variation. The number of lags, d_0, was determined via the forward selection procedure with the BIC stopping rule. As seen from Figure 4, the number d_0 is estimated as 6. Furthermore, using the selection criterion in Section 2.3, the final estimate of the factor number is 5.

We first applied the feasible local linear method to estimate the rotated coefficient functions in (6.1), where the Epanechnikov kernel was used and the optimal bandwidth is 0.03 determined by the multi-fold cross-validation criterion introduced in Section 5. Then we constructed the one-step ahead nonlinear forecast as in (2.4). In order to measure the prediction accuracy, we computed the

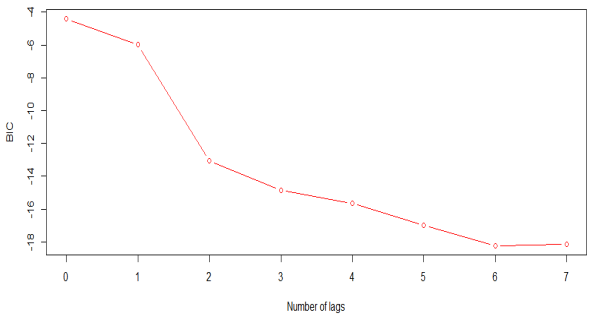


Figure 4: The change of the BIC function.

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 0.0562 and 0.2176, respectively. For the aim of comparison, we also consider using the traditional AR and VAR models and the unemployment rate Phillips curve[1] in the out-sample prediction as in Chen *et al* (2018a). The relevant MSPE 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.[2] In addition, we compare the one-step ahead prediction performance between our modelling method and the linear factor-augmented modelling method and find that our FA-FCM method has a better forecasting performance.

[1]The Phillips curve is named after A.W. Phillips who published some pioneering works in studying the relationship between unemployment rates and wage changes (e.g., Phillips, 1958, 1959). It is also commonly used to forecast the inflation change with the following specification (e.g., Stock and Watson, 1999; Chen *et al*, 2018a):

$$\Delta y_{t+1} = y_{t+1} - y_t = \alpha + \beta(L)u_t^* + \gamma(L)\Delta y_t + e_{t+1},$$

where y_t denotes the CPI, u_t^* denotes 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, L denotes the lag operator, and Δ denotes the first difference operator.

[2]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* (2018a).

7 Conclusion

In this paper 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 PCA approach. The number of factor regressors and the number of auto-regressors are determined by the data-driven selection procedure with the BIC stopping rule. 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. The asymptotic theory in Section 3 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. Such an asymptotic property is supported by the simulation studies in finite samples. Furthermore, the developed predictive model and forecasting methodology are applied to predict the UK inflation change and have satisfactory forecasting performance. In particular, our method outperforms some commonly-used forecasting approaches in the empirical application.

A possible extension of our predictive model framework is to include some lags of $\mathbf{F}_t$ in the factor-augmented functional-coefficient models. These lags may contain some useful dynamic information, which is helpful to improve the model forecasting performance. The estimation and prediction methods developed in the present paper can be easily generalised to deal with this case, but the selection criterion in Section 2.3 needs to be appropriately modified to estimate an "optimal" number of lags of $\mathbf{F}_t$.

Acknowledgements

The authors would like to thank a Co-Editor and two reviewers for the valuable comments, which substantially improve an earlier version of the paper.

Supplemental materials

The online supplemental materials contain the empirical data set, the detailed proofs of the main asymptotic theorems as well as some technical lemmas.

Data availability statement

The empirical data set used in Section 6 is available in the supplemental materials.

References

Ahn, S. and Horenstein, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, **81**, 1203–1227.

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

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

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, **74**, 1135–1150.

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

Bickel, P. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, **36**, 2577–2604.

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

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

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

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, **51**, 1305–1324.

Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**, 759–771.

Chen, J., Li, D., Linton, O. and Lu, Z. (2018a). Semiparametric model averaging of ultrahigh dimensional time series. *Journal of the American Statistical Association*, **113**, 919–932.

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

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

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

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

Djogbenou, A. (2016). Model selection in factor-augmented regressions with estimated factors. *Manuscript*.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, **47**, 427–465.

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

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

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

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

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

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

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

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

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

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

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

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

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

Li, D., Lu, Z. and Linton, O. (2012). Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory*, **28**, 935–958.

Li, H., Li, Q. and Shi, Y. (2017). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics*, **197**, 76–86.

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

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

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, **77**, 1447–1479.

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

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

Phillips, A. W. (1958).The relationship between unemployment and the rate of change of money wages in the United Kingdom 1861–1957. *Economica* **25**, 283–299.

Phillips, A. W. (1959). Wage changes and unemployment in Australia, 1947–1958. *Economic Society of Australia and New Zealand Economic Monograph* **219**.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. *Statistical Analysis and Forecasting of Economic Structural Change* (ed. by P. Hackl). Springer, Berlin, pp. 164–253.

Rosenthal, H. P. (1970). On the subspaces of L_p $(p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics*, **8**, 273–303.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

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

Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. *NBER Working Paper 6702*.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *NBER Working Paper 7023*.

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

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

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

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

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

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