Factor-Augmented Regressions and their Applications to Financial Markets: A Selective Review[1]

YONGHUI ZHANG *Renmin University of China*

This chapter provides a selective review on the factor-augmented regression (FAR) models, where the factors are usually estimated from a large set of observed data, and then as "generated regressors" enter into the next stage regression. It begins with an introduction to the large dimensional factor models and the widely-used principal component analysis (PCA) estimator. Then we review FAR models with time series data, the extensions of FAR to some nonlinear models and the factor-augemented panel regressions. Lastly, we briefly introduce some applications of FAR to financial markets.

[1]Useful comments from an anonymous referee led to a considerably improved version of the paper. Zhang gratefully acknowledges the financial support from National Natural Science Foundation of China (Project No.71973141). All errors are the authors' sole responsibilities. Yonghui Zhang, School of Economics, Renmin University of China. *E-mail address*: yonghui.zhang@hotmail.com.

3.1 Introduction

Due to rapid technological advances in data collection and the accumulation of massive amounts of information, economists now have the luxury of working in a data-rich environment. Common factor models, as an efficient way to summarize and extract information from large datasets, have received much attention in this revolutionary era of data science and have been widely used in empirical studies. Partial previous applications of factor models in economics and finance include the arbitrage pricing theory (APT) of Ross (1976), the disaggregate business cycle model in Gregory and Head (1999) and Forni and Reichlin (1998), monitoring and forecasting with diffusion indices in Bai and Ng (2006) and Stock et al. (1998); Stock and Watson (2002a,b), the demand system in Gorman (1981), and the identification of common factors for exchange rate in Greenaway-McGrevy et al. (2018), etc. See Bai and Ng (2008), Bai and Wang (2016), Xu et al. (2016), Swanson and Xiong (2018), Karabiyik et al. (2019) and Fan et al. (2021) for selective reviews on factor models from different perspectives.

This chapter provides a selective review of factor-augmented regression (FAR) models and their applications in finance. In the framework of FAR, the factors are usually estimated from a large set of observed data and then, as "generated regressors" that enter into other models such as linear predictive models, quantile regression models, or panel regression models, etc. Since the seminal paper of Stock and Watson (2002b), FAR models have been broadly used for many purposes. Once the factor components have been extracted, they can be used to date the recession Stock and Watson (2016), compare forecasts (Boivin and Ng, 2005), measure uncertainty (Jurado et al., 2015), and evaluate monetary policy (Bernanke and Boivin, 2003).

In the literature on factor models, different methods, such as a principal component analysis (PCA) based on an eigen-analysis of the sample covariance matrix of the data, a maximum likelihood estimation (MLE), or some low-rank regularization methods that use the low-rank structure of factor models, have been proposed to estimate factor models. Among all the estimation methods, PCA is the most popular in empirical applications. As a simple way of transforming the information content in a large number of series into a smaller number of manageable series, PCA can provide consistent estimates for some rotation of common factors under certain assumptions. This chapter focuses on the method of PCA.

The rest of the chapter is organized as follows. We provide an introduction to the factor model and its PCA estimators in Section 2. Section 3 reviews the FAR with time series data, which mainly includes the diffusion index model and the issue of model selection. Section 4 focuses on the nonlinear FAR models covering factor-augmented quantile regression models, factor-augmented nonlinear models, and FARs with structural breaks and threshold effects. Section 5 surveys a FAR with a panel data set. Two applications of FARs to financial markets are mentioned in Section 6. Section 7 presents concluding remarks.

3.2 Factor Models and Principal Component Analysis

In this section, we first introduce the large-dimensional factor models in Bai (2003) and then review the main theories of PCA estimators.

A typical factor model has the following representation:

$$X_{it} = \lambda_i' F_t + e_{it}, \tag{3.2.1}$$

$i = 1, \ldots, N$, $t = 1, \ldots, T$, where X_{it} is the observed data for the i-th cross-section unit or variable at time t, F_t is a vector $(r \times 1)$ of common latent factors, λ_i is a vector $(r \times 1)$ of factor loadings, and e_{it} is the idiosyncratic error. Both F_t's and λ_i's are not observable. We are usually interested in the estimation of common factors $F = (F_1, \ldots, F_T)'$, factor loadings $\Lambda = (\lambda_1, \ldots, \lambda_N)'$, and the common components $C_{it} \equiv \lambda_i' F_t$ when both N and T tend to infinity simultaneously.

In this chapter, we focus on stationary factor models where both $\{F_t\}_{t=1}^T$ and $\{e_{it}\}_{t=1}^T$ are stationary. For an excellent review of nonstationary factor models, see Chapter 6 of Bai and Ng (2008). In the literature, there are two general approaches to estimate the factor model (3.2.1). One is the PCA method, and the other is low-rank regularization. In this review, we restrict our interest to the PCA estimators, which is much simpler and more popular than other methods in applications. For the low-rank regularization of factor models, see Section 3.2 of Fan et al. (2020).

Let $X_i = (X_{i1}, \ldots, X_{iT})'$, $X = (X_1, \ldots, X_N)$, $e_i = (e_{i1}, \ldots, e_{iT})'$, and $e = (e_1, \ldots, e_N)$. Then, the model (3.2.1) can be written in T-vector form $X_i = F\lambda_i + e_i$ or matrix form

$$X = F\Lambda' + e.$$

Note that $F\Lambda' = FHH^{-1}\Lambda'$ for any invertible $r \times r$ matrix H, F and Λ are not separately identifiable. Bai and Ng (2002) impose the following identification restrictions:

$$F'F/T = I_r \text{ and } \Lambda'\Lambda \text{ being diagonal.}$$

The r^2 restrictions given by the normalization of F and the diagonalization of $\Lambda'\Lambda$ fix the r^2 free parameters in matrix H. However, the column signs of F and Λ are still not fixed. Alternatively, one can use the normalization of factor loadings $\Lambda'\Lambda/N = I_r$ and the diagonalization that $F'F$ is diagonal. For other identification restrictions, see Bai and Ng (2013). For any $k(k \leq \min[T, N])$, the PCA estimator with k factors is given by

$$\left(\hat{F}^k, \hat{\Lambda}^k\right) = \arg\min_{F^k, \Lambda^k} S(k) \text{ with}$$

$$S(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \lambda_i^{k\prime}F_t^k\right)^2 = \frac{1}{NT}\mathtt{tr}\left[\left(X - F\Lambda'\right)\left(X - F\Lambda'\right)'\right],$$

subject to

$$\text{(PC1) } F^{k\prime}F^k/T = I_k \text{ and } \Lambda^{k\prime}\Lambda^k \text{ being diagonal; or}$$
$$\text{(PC2) } \Lambda^{k\prime}\Lambda^k/N = I_k \text{ and } F^{k\prime}F^k \text{ being diagonal.}$$

Under restriction (PC1), we can cancel Λ^k by using $\Lambda^{k\prime} = \left(F^{k\prime}F^k\right)^{-1}F^{k\prime}X = F^{k\prime}X/T$ and obtain

$$\hat{F}^k = \arg\min_{F^k}\mathtt{tr}\left(X'M_{F^k}X\right) = \arg\max_{F^k}\mathtt{tr}\left(X'P_{F^k}X\right) = \arg\max_{F^k}\mathtt{tr}\left(F^{k\prime}XX'F^k\right),$$

where $M_{F^k} = I_T - F^k\left(F^{k\prime}F^k\right)^{-1}F^{k\prime} = I_T - F^kF^{k\prime}/T$ and $P_{F^k} = F^k\left(F^{k\prime}F^k\right)^{-1}F^{k\prime} = F^kF^{k\prime}/T$. The estimated factor matrix $\hat{F}^k$ is given by $\sqrt{T}$ times the eigenvectors that correspond to the k largest eigenvalues of the $T \times T$ matrix XX', and the factor loadings are estimated by $\hat{\Lambda}^{k\prime} = \hat{F}^{k\prime}X/T$ by using restriction (PC1). Now, the restriction $\Lambda^{k\prime}\Lambda^k$ being diagonal is satisfied automatically. Alternatively, one can use restriction (PC2) and concentrate out F^k to obtain another set of estimators $\left(\tilde{F}^k, \tilde{\Lambda}^k\right)$: $\tilde{\Lambda}^k$ is $\sqrt{N}$ times the eigenvectors that correspond to the k largest eigenvalues of the $N \times N$ matrix $X'X$, and $\tilde{F}^{k\prime} = X\tilde{\Lambda}^k/N$. Let $\tilde{V}_k$ be the $k \times k$ diagonal matrix that consists of the first k largest eigenvalues of the matrix $X'X/(TN)$ (or $XX'/(TN)$)[2], arranged in decreasing order. It can be shown easily that

$$\frac{\hat{\Lambda}^{k\prime}\hat{\Lambda}^k}{N} = \frac{\tilde{F}^{k\prime}\tilde{F}^k}{T} = \tilde{V}_k,\ \tilde{F}^k = \hat{F}^k\left(\tilde{V}_k\right)^{1/2}, \text{ and } \hat{\Lambda}^{k\prime} = \tilde{\Lambda}^k\left(\tilde{V}_k\right)^{1/2}.$$

[2]Note that $X'X$ and XX' have the same nonzero eigenvalues.

For the determination of the number of factors k, Bai and Ng (2002) propose estimating it by minimizing the following information criteria:

$$\hat{k}_{IC} = \arg \min_{0 \leq k \leq k_{\max}} IC(k) \text{ with } IC(k) = \log S(k) + kg(N,T),$$

or

$$\hat{k}_{PC} = \arg \min_{0 \leq k \leq k_{\max}} PC(k) \text{ with } PC(k) = S(k) + \bar{\sigma}^2 kg(N,T),$$

where $k_{\max} \geq r$ is a prespecified large number, $S(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \hat{\lambda}_i^{k\prime}\hat{F}_t^k\right)^2$, $\bar{\sigma}^2$ is used as the proper scaling penalty term in $PC(k)$, and $g(N,T)$ is a penalty function that satisfies (i) $g(N,T) \to 0$ and (ii) $\delta_{NT} \cdot g(N,T) \to \infty$ as $(N,T) \to \infty$, where $\delta_{NT} \equiv \min[\sqrt{N}, \sqrt{T}]$. The typical function $g(N,T)$ includes $g_1 = \frac{N+T}{NT}\ln\left(\frac{NT}{N+T}\right)$, $g_2 = \frac{N+T}{NT}\ln\left(\delta_{NT}^2\right)$, and $g_3 = \frac{\ln \delta_{NT}^2}{\delta_{NT}^2}$. For $PC(k)$, $\bar{\sigma}^2$ can be estimated by $S(k_{\max}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \hat{\lambda}_i^{k_{\max}\prime}\hat{F}_t^{k_{\max}})^2$. Under standard conditions on factor models, Bai and Ng (2002) show that

$$\Pr\left(\hat{k}_{IC} = r\right) \to 1 \text{ and } \Pr\left(\hat{k}_{PC} = r\right) \to 1$$

as $(N,T) \to \infty$. As shown by Bai and Ng (2002), all criteria (IC, PC) with g_1, g_2 and g_3 perform very well in simulation when both N and T are large. When either N or T is small and errors are uncorrelated across i and t, the ICs with g_1 and g_2 have better finite sample performance. In this chapter, we assume a known r because the large sample theories of the estimated factors and factor loadings are not affected when the number of factors is estimated using information criteria. There are some other criteria on the determination of the number of factors; for example, see Onatski (2009) for a testing approach based on large random matrix theory, Ahn and Horenstein (2013) for the eigenvalue ratio (ER) and the growth ratio (GR) estimator based on the calculation of eigenvalues, Alessi et al. (2010) for the modified information criterion of Bai and Ng (2002) with an additional tuning multiplicative constant in the penalty, and Li et al. (2017) for the information criteria for the factor model with a diverging number of factors.

Since the review focuses on the FARs, we are interested in the large sample properties of the factors estimated by PCA. Therefore, we only state the limiting theories for the estimated factors in Bai (2003). For the limiting distributions of the estimated factor loadings and common components, see Theorems 2-3 in Bai (2003). Let F_t^0 and λ_i^0 denote the true factors and factor loadings, respectively. Under the standard conditions on factor models, Bai (2003) shows that

(i) If $\sqrt{N}/T \to 0$, then for each t,

$$\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) = V_{NT}^{-1}\left(\frac{\hat{F}'F^0}{T}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} + O_p\left(\frac{\sqrt{N}}{T}\right) \tag{3.2.2}$$

$$\xrightarrow{d} N\left(0, V^{-1}Q\Gamma_t Q'V^{-1}\right), \tag{3.2.3}$$

where $H = V_{NT}^{-1}\left(\hat{F}'F^0/T\right)(\Lambda'\Lambda/N)$ is an asymptotically invertible $r \times r$ matrix, V_{NT} is a diagonal matrix that consists of the first r eigenvalues of $(NT)^{-1}XX'$ in decreasing order, $V = \mathtt{diag}(v_1, \ldots, v_r)$ and $Q = \text{plim}_{T\to\infty}\hat{F}'F^0/T = V^{1/2}\Upsilon\Sigma_\Lambda^{-1/2}$, $v_1 > \cdots > v_r$ are the eigenvalues of $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$, and Υ is the corresponding eigenvector matrix such that $\Upsilon'\Upsilon = I_r$ with $\Sigma_\Lambda = \text{plim}_{N\to\infty}\Lambda^{0\prime}\Lambda^0/N$ and $\Sigma_F = \text{plim}_{T\to\infty}F^{0\prime}F^0/T$, and $\Gamma_t = \lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^0\lambda_j^{0\prime}E(e_{it}e_{jt})$.

(ii) If $\liminf \sqrt{N}/T \geq \tau > 0$, then

$$T\left(\hat{F}_t - H'F_t^0\right) = O_p(1). \tag{3.2.4}$$

From the above results, we conclude that the convergence rate for each estimated factor is $\min[N^{1/2}, T]$, and the asymptotic normality generally holds when $\sqrt{N}/T \to 0$. The rate

$\min[N^{1/2}, T]$ reflects the fact that the factor loadings are estimated. If λ_i^0's are all known, then the convergence rate for the cross-sectional least square estimator for F_t^0 should be $\sqrt{N}$.

To make inference about the estimated factors when $\sqrt{N}/T \to 0$, a consistent estimator for the asymptotic variance $\texttt{Avar}(\hat{F}_t) = V^{-1}Q\Gamma_t Q'V^{-1}$ is needed. Bai (2003) proposes the following estimator

$$\widehat{\texttt{Avar}}(\hat{F}_t) = V_{NT}^{-1}\hat{\Gamma}_t V_{NT}^{-1}, \tag{3.2.5}$$

where the $r \times r$ covariance matrix $\hat{\Gamma}_t$ can be estimated by

$$\text{(a) } \hat{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N}\hat{e}_{it}^2\hat{\lambda}_i\hat{\lambda}_i'; \text{ (b) } \hat{\Gamma}_t = \hat{\sigma}_e^2\frac{1}{N}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\lambda}_i'; \text{ or (c) } \hat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\lambda}_i\hat{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}\hat{e}_{jt},$$

where $\frac{n}{\min[N,T]} \to 0$ in (c) and $\hat{e}_{it} = X_{it} - \hat{\lambda}_i'\hat{F}_t$. (a) works when e_{it} are uncorrelated with e_{jt} for $i \neq j$, and (b) works when e_{it} are uncorrelated with e_{jt} for $i \neq j$ and $E\left(e_{it}^2\right) = \sigma_e^2$ for all i's and t's. As argued by Bai (2003), when a small degree of cross-section correlation in the error terms exists, sometimes restricting them to be 0 could be desirable because the sampling variability from estimating them could generate nontrivial efficiency loss. (c) is termed the cross-section and heteroskedastic autocorrelation consistent (CS-HAC) estimator by Bai (2003) and is appropriate for cross-sectional dependent errors with an additional covariance stationarity assumption that $E\left(e_{it}e_{jt}\right) = \sigma_{ij}$ for all t's. As for the choice of n in (c), one simple rule is to use $n = \min\left\{\lfloor N^{1/2}\rfloor, \lfloor T^{1/2}\rfloor\right\}$ as suggested by Bai (2003), where $\lfloor a\rfloor$ denotes the largest integer less than a.

Since we focus on the cases where the estimated factors are used as regressors in the next stage, the following rough uniform consistency result for the estimated factors $\hat{F}_t$ is given by

$$\max_{1\le t\le T}\left\|\hat{F}_t - H'F_t^0\right\| = O_p\left(T^{-1/2}\right) + O_p\left((T/N)^{1/2}\right).$$

Bai and Ng (2008) provide a sharper bound. If there exists $\ell \geq 4$ such that $E\left\|F_t^0\right\|^\ell \leq M$ and $E\left\|N^{-1/2}\sum_{i=1}^N \lambda_i^0 e_{it}\right\|^\ell \leq M$ for all t, then

$$\max_{1\le t\le T}\left\|\hat{F}_t - H'F_t^0\right\| = O_p\left(T^{-1+1/\ell}\right) + O_p\left(T^{1/\ell}/N^{1/2}\right).$$

When $\ell = 4$, $\max_{1\le t\le T}\left\|\hat{F}_t - H'F_t^0\right\| = O_p\left(T^{-3/4}\right) + O_p\left(T^{1/4}/N^{1/2}\right) \to 0$ given that $T/N^2 \to 0$. In addition, the following results are relevant for the FAR:

(i) $T^{-1}\sum_{t=1}^T\left\|\hat{F}_t - H'F_t^0\right\|^2 = O_p\left(\delta_{NT}^{-2}\right)$;

(ii) Assume that ξ_t is uncorrelated with e_{it} for all i and t and $E\left|\xi_t\right|^2 \leq M$ for all t; then, $T^{-1}\sum_{t=1}^T\left(\hat{F}_t - H'F_t^0\right)\xi_t = O_p\left(\delta_{NT}^{-2}\right)$.

3.3 FARs with Time Series Data

In this section, we mainly review the forecast model using the FARs in Stock and Watson (2002a,b) and Bai and Ng (2006).

3.3.1 Diffusion index forecasting model

When a FAR model is applied to time series data, it is typically for the aim of forecasting some target variables in a data-rich environment. Consider the following the h-step ahead forecasting model for y_t with $h \geq 1$:

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}, t = 1, \ldots, T, \tag{3.3.1}$$

where F_t is an $r \times 1$ vector of unobservable factors, and W_t is a vector of a small number of observables, which typically includes the intercept and the lags of y_t or seasonal dummies, and ε_{t+h} is the error term. Clearly, given $E(\varepsilon_{t+h}|\Omega_T) = 0$, where $\Omega_T = [F_T, W_T, F_{T-1}, W_{T-1}, \cdots]$, the optimal predictor for y_{T+h} at the Tth period is

$$y_{T+h|T} \equiv E(y_{T+h}|\Omega_T) = \alpha' F_t + \beta' W_t. \tag{3.3.2}$$

However, based on the observed data set $\{W_t, y_t\}_{t=1}^T$, it is impossible to obtain the conventional mean-squared optimal prediction of $y_{T+h|T}$ because F_t are unobservable.

In a data-rich environment, suppose that there is a large number of series $\underline{X}_t = (X_{1t}, \ldots, X_{Nt})'$ at the tth period that has a factor structure as follows:

$$X_{it} = \lambda_i' F_t + e_{it}. \tag{3.3.3}$$

Clearly, X_{it} is correlated with the latent factors F_t and is a noisy predictor for y_{t+h}. However, one cannot directly regress y_{t+h} on $\underline{X}_t$ and W_t to construct the forecast due to the large dimension of $\underline{X}_t$ (N) without imposing some special structure. Instead, one can extract the factors from the observables X_{it}'s through PCA and then provide a feasible forecast by using the PC estimates for the latent factors.

When the target variable y_{t+h} is a scalar, the framework in (3.3.1) and (3.3.3) is the famous *diffusion index* (DI) forecasting model proposed by Stock and Watson (2002b). The DI can reduce the dimension of the predictors from N to a much smaller number r, namely, the dimension of F_t. That is, it can exploit the information underlying a large data set in a parsimonious way. Due to its advantage in modeling large data sets, the DI framework has been widely used by government agencies in different countries and many academic researchers.

Now, we turn to the construction of the DI forecast. Consider the PCA estimator $\hat{F}_t$ in the model (3.3.3) based on data $\{X_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$, which is a consistent estimator of $H' F_t$ for some invertible rotation matrix H according to (3.2.2)-(3.2.4) in Section 2. Then, regress y_{t+h} on $\hat{F}_t$ and W_t to obtain $\hat{\delta} \equiv \left(\hat{\alpha}', \hat{\beta}'\right)'$ by using the usual least square method, namely,

$$\hat{\delta} = \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t'\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t y_{t+h},$$

where $\hat{z}_t \equiv \left(\hat{F}_t', W_t'\right)'$. Let $\delta \equiv (\alpha' H^{-1}, \beta')'$ and $z_t \equiv (F_t', W_t')'$. Rewriting

$$\begin{aligned}
y_{t+h} &= \alpha' F_t + \beta' W_t + \varepsilon_{t+h} \\
&= \alpha' H^{-1} \hat{F}_t + \beta' W_t + \varepsilon_{t+h} + \alpha' H^{-1}\left(H F_t - \hat{F}_t\right) \\
&= \delta' \hat{z}_t + \varepsilon_{t+h} + \alpha' H^{-1}\left(H F_t - \hat{F}_t\right)
\end{aligned}$$

leads to the following decomposition:

$$\begin{aligned}
\sqrt{T}\left(\hat{\delta} - \delta\right) &= \left(\frac{1}{T}\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t'\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_t \varepsilon_{t+h} \\
&+ \left(\frac{1}{T}\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t'\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_t \left(H F_t - \hat{F}_t\right)' H^{-1} \alpha,
\end{aligned}$$

where the first term is $O_p(1)$ and becomes the leading term when $\sqrt{T}/N \to 0$, and the second term is $O_p\left(T^{1/2}/N + T^{-1/2}\right)$ according to Lemma A.1 (iii) in Bai and Ng (2006). Under some regular conditions, Bai and Ng (2006) show that when $\sqrt{T}/N \to 0$, $\hat{\delta}$ is $\sqrt{T}$-consistent as the estimator of δ and is asymptotically normally distributed. Specifically, when $\sqrt{T}/N \to 0$, Bai and Ng (2006) establish that

$$\sqrt{T}\left(\hat{\delta}-\delta\right) \xrightarrow{d} N\left(0, \Sigma_\delta\right), \tag{3.3.4}$$

where $\Sigma_\delta \equiv \Phi_0^{-1}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}\Phi_0^{-1}$ with $\Phi_0 \equiv \mathtt{diag}\left(V^{-1}Q\Sigma_\Lambda, I\right)$, $\Sigma_{zz} \equiv \operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} z_t z_t'$ and $\Sigma_{zz,\varepsilon} \equiv \operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} z_t z_t' \varepsilon_{t+h}^2$. Let $\hat{\varepsilon}_{t+h} \equiv y_{t+h} - \hat{y}_{t+h|t}$. A heteroskedasticity consistent (HC) estimator for Σ_δ is given by

$$\hat{\Sigma}_\delta = \left(\frac{1}{T_h}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}\left(\frac{1}{T_h}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\hat{\varepsilon}_{t+h}^2\right)\left(\frac{1}{T_h}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}, \tag{3.3.5}$$

where $T_h = T-h$.

Given the consistency of $\hat{\delta}$, a feasible prediction for y_{T+h} is given by

$$\hat{y}_{T+h|T} = \hat{\alpha}'\hat{F}_T + \hat{\beta}'W_T = \hat{z}_T'\hat{\delta}. \tag{3.3.6}$$

To make inference about the feasible predictor, write

$$\hat{y}_{T+h|T} - y_{T+h|T} = \hat{z}_T'\hat{\delta} - z_T'\delta = \alpha' H^{-1}\left(\hat{F}_t - HF_t\right) + \hat{z}_T'\left(\hat{\delta}-\delta\right), \tag{3.3.7}$$

which has two components due to the estimation of F_t and δ. Recall that as $\sqrt{N}/T \to 0$, $\sqrt{N}\left(\hat{F}_T - H'F_T\right) \xrightarrow{d} N\left(0, \mathtt{Avar}\left(\hat{F}_T\right)\right)$, where $\mathtt{Avar}\left(\hat{F}_T\right) \equiv V^{-1}Q\Gamma_t Q'V^{-1}$. This result combined with (3.3.4) gives the following:

$$\frac{\hat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{Var\left(\hat{y}_{T+h|T}\right)}} \xrightarrow{d} N(0,1),$$

where

$$Var\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\hat{z}_T'\mathtt{Avar}\left(\hat{\delta}\right)\hat{z}_T + \frac{1}{N}\hat{\alpha}'\mathtt{Avar}\left(\hat{F}_T\right)\hat{\alpha}.$$

The forecast error is given by

$$\hat{\varepsilon}_{T+h|T} \equiv \hat{y}_{T+h|T} - y_{T+h} = \left(\hat{y}_{T+h|T} - y_{T+h|T}\right) + \varepsilon_{T+h}$$

and its variance is given by $Var\left(\hat{\varepsilon}_{T+h|T}\right) = Var\left(\hat{y}_{T+h|T}\right) + Var\left(\varepsilon_{T+h}\right)$. If one imposes a further assumption that ε_t is normal with variance σ_ε^2, then the forecasting error also becomes approximately normal:

$$\hat{\varepsilon}_{T+h|T} \overset{a}{\sim} N\left(0, \sigma_\varepsilon^2 + Var\left(\hat{y}_{T+h|T}\right)\right),$$

so that the confidence intervals (CIs) can be constructed for the forecasts.

Given the consistent estimators for $\mathtt{Avar}(\hat{\delta})$ in (3.3.5) and $\mathtt{Avar}(\hat{F}_T)$ in (3.2.5), the prediction intervals can be straightforwardly constructed. For example, the 95% CI for the forecasting variable y_{T+h} can be given by

$$\left[\hat{y}_{T+h|T} - 1.96\sqrt{\hat{\sigma}_\varepsilon^2 + \widehat{Var}\left(\hat{y}_{T+h|T}\right)}, \hat{y}_{T+h|T} + 1.96\sqrt{\hat{\sigma}_\varepsilon^2 + \widehat{Var}\left(\hat{y}_{T+h|T}\right)}\right]$$

where $\widehat{Var}\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\hat{z}_T'\widehat{\mathtt{Avar}}\left(\hat{\delta}\right)\hat{z}_T + \frac{1}{N}\hat{\alpha}'\widehat{\mathtt{Avar}}\left(\hat{F}_T\right)\hat{\alpha}$ and $\hat{\sigma}_\varepsilon^2 = T^{-1}\sum_{t=h+1}^{T}\hat{\varepsilon}_t^2$ with $\hat{\varepsilon}_{t+h} = y_{t+h} - \hat{z}_t'\hat{\delta}$. For the bootstrap prediction intervals for factor models, see Gonçalves et al. (2017). The bootstrap can relax the assumption of the Gaussianity of the innovations and construct valid prediction intervals under more general conditions. Moreover, even under the Gaussianity assumption, the bootstrap procedure leads to more accurate intervals when N is relatively small.

The estimated factors $\hat{F}_t$ can be combined with an m-vector of W_t to form a factor-augmented vector autoregression (FAVAR) model, which was originally proposed by Bernanke et al. (2005) to measure monetary policy. A FAVAR model assumes that a large number of economic variables are driven by a small vector autoregression (VAR), which includes both latent and observed variables. To be specific, in a FAVAR, both the unobserved factors F_t and the observed factors W_t affect a large number of observed variables X_{it},

$$X_{it} = \lambda_i' F_t + \gamma_i' W_t + e_{it},$$

and the random vector $G_t \equiv (F_t', W_t')'$ follows a VAR of finite order,

$$\Phi(L) G_t = u_t.$$

where $\{u_t\}$ are i.i.d. (0, σ_u^2). For the above model, two ways have been proposed to analyze the system by Bernanke et al. (2005). The first one is a two-step method: Step 1. Form estimates of the space spanned by both F_t and W_t with PCA, where various identification approaches can be used to obtain the estimates $\hat{F}_t$; Step 2. The estimated factors are treated as observed in conducting a VAR analysis of $\hat{G}_t = \left(\hat{F}_t', W_t'\right)'$. Due to the problem of "generated regressors", the two-step estimator has a different inferential theory from the standard large factor models; see Bai et al. (2016). The second method considers the joint estimation of the latent factors and impulse responses with a one-step likelihood approach that uses Gibbs sampling. These two methods can complement one another, with the first one being computationally simple and the second providing possibly better finite sample inference but at a heavier computational cost. See Bernanke et al. (2005), Stock and Watson (2012), and Bai and Wang (2016) for more discussion.

3.3.2 Model selection of the diffusion index forecasting model

In the DI forecasting model, the predictor constructed in (3.3.6) relies on the fact that both the target variable y_{t+h} and the large set of variables X_{it}, $i = 1, \ldots, N$, have the same set of factors. In practice, some factors extracted from the observable X_{it}'s may provide useless information for predicting the target variable. In the literature, different approaches to select the significant predictors exist (the estimated factors and observable predictors), such as model selection criteria, a generalized cross-validation, a shrinkage estimation with sparsity assumptions, and model averaging techniques.

In the following section, we review three commonly used model selection criteria in DI forecasting models.

Groen and Kapetanios (2013) model selection criteria. Let $\hat{F} = \left(\hat{F}_1, \ldots, \hat{F}_T\right)'$ be the $T \times r$ matrix of the estimated factors by applying PCA to model (3.3.3), where the number of factors r can be determined by the criteria in Bai and Ng (2002). Let

$$\hat{\mathcal{F}} = \left\{ \left\{ \hat{F}_t^{(1)} \right\}_{t=1}^{T}, \left\{ \hat{F}_t^{(2)} \right\}_{t=1}^{T}, \ldots, \left\{ \hat{F}_t^{(s)} \right\}_{t=1}^{T} \right\}$$

denote the set of estimated factor variables, where $\hat{F}_t^{(l)}$ indicates the lth candidate vector of the estimated factors at time t and includes the d_l elements of $\hat{F}_{t,}$; the number of all candidate sets is s. Let $\hat{F}^{(l)} = \left(\hat{F}_1^{(l)}, \ldots, \hat{F}_T^{(l)}\right)'$ be the $T \times d_l$ matrix of the estimated factors for the lth candidate set. Clearly, $s \leq 2^r$ because there are 2^r distinct forecasting models when there are r estimated factors. When r is not too large, all 2^r subsets can be considered; but when the number of combinations is too large, a much smaller subset of them can be considered. Furthermore, let

$$\mathcal{F} = \left\{ \left\{ F_t^{(1)} \right\}_{t=1}^{\infty}, \left\{ F_t^{(2)} \right\}_{t=1}^{\infty}, \ldots, \left\{ F_t^{(s)} \right\}_{t=1}^{\infty} \right\},$$

where $F_t^{(l)} \equiv \text{plim}_{N,T\to\infty} \hat{F}_t^{(l)}$ is the probability limit of $\hat{F}_t^{(l)}$ for $l = 1, \ldots, s$. To select the correct specification of factors in the predictive regression, Groen and Kapetanios (2013) propose the following information criteria (IC):

$$IC_l = \frac{T}{2} \ln\left(\hat{\sigma}_\varepsilon^2\right) + d_l \tilde{C}_{N,T},$$

where $\hat{\sigma}_\varepsilon^2 \equiv T^{-1} \sum_{t=h+1}^{T} \hat{\varepsilon}_t^2$, with $\hat{\varepsilon}_t$'s being the ordinary least squares (OLS) residuals from the linear regression

$$y_{t+h} = \hat{z}_t^{(l)\prime} \delta^{(l)} + \varepsilon_{t+h}$$

with $\hat{z}_t^{(l)} \equiv \left(\hat{F}_t^{(l)\prime}, W_t'\right)'$, d_l is the number of factors included in $\hat{F}_t^{(l)}$, and $\tilde{C}_{N,T}$ denotes a penalty term that satisfies $\tilde{C}_{NT} = o(T)$ and $\lim_{N,T\to\infty} T^{-1} \min(N,T)\tilde{C}_{N,T} = \infty$. Let F_t^0 be the true factors in the target predictor variable y_{t+h}. Under the standard assumptions in Bai and Ng (2006), Groen and Kapetanios (2013) establish the following factor selection consistency result:

(i) If there exists a matrix A such that $F_t^0 = AF_t^{(l)}$ for $\forall t$, but no such matrix exists for $F_t^{(j)}$ when $l \neq j$, then

$$\lim_{N,T\to\infty} \Pr\left(IC_l < IC_j\right) = 1; \tag{3.3.8}$$

(ii) If there exist matrices $A^{(l)}$ and $A^{(m)}$ such that $F_t^0 = A^{(l)} F_t^{(l)}$ and $F_t^0 = A^{(m)} F_t^{(m)}$ for $\forall t$ and $d_l < d_m$, then (3.3.8) holds.

The model selection consistency results include two parts: (i) any set of estimated factors whose probability limits span the true factors will be chosen over any set of estimated factors that do not span the true factors given the penalty term being $o(T)$; and (ii) if two sets of estimated factors both span the set of true factors, then the one with the smaller dimension will be chosen.

Note that the conditions on $\tilde{C}_{N,T}$ allow for all popular ICs such as the Bayesian IC (BIC) and the HQIC. To gain a better finite sample performance, Groen and Kapetanios (2013) propose the following modified BIC and HQIC:

$$BICM = \frac{T}{2} \ln\left(\hat{\sigma}_\varepsilon^2\right) + d \ln(T) \left(1 + \frac{T}{N}\right) \text{ and}$$
$$HQICM = \frac{T}{2} \ln\left(\hat{\sigma}_\varepsilon^2\right) + 2d \ln(\ln T) \left(1 + \frac{T}{N}\right).$$

When other variables W_t such as the lags of the target variable in the forecasting regression also need selection, Groen and Kapetanios (2013) provide the modified criteria to select the estimated factors and other predictors jointly:

$$BICM = \frac{T}{2} \ln\left(\hat{\sigma}_\varepsilon^2\right) + k \ln T + d \ln(T) \left(1 + \frac{T}{N}\right) \text{ and}$$
$$HQICM = \frac{T}{2} \ln\left(\hat{\sigma}_\varepsilon^2\right) + 2k \ln(\ln T) + 2d \ln(\ln T) \left(1 + \frac{T}{N}\right),$$

where k is the number of variables in the subset of W_t used in the forecasting regression.

Ando and Tsay (2014) PMSE criterion. Let $u_{1+h}, \ldots, u_{T+h}$ be replicates of the target variables $y_{1+h}, \ldots, y_{T+h}$ given the true values of factors $F_1, \ldots, F_T$. To assess the predictive ability of the estimated model, the predictive mean squared error (PMSE) is considered:

$$\eta \equiv \frac{1}{T} \int \left[\sum_{t=1}^{T} \left(u_{t+h} - \hat{z}_t' \hat{\delta}\right)^2\right] dG(\mathbf{u}), \tag{3.3.9}$$

where $dG(\mathbf{u})$ is the Lebesgue measure with respect to the probability density $g(\mathbf{u})$ of the joint distribution of $\mathbf{u} = (u_{1+h}, \ldots, u_{T+h})'$. The PMSE is positive unless $u_{t+h} = \hat{z}_t' \hat{\delta}$ almost surely holds. The best statistical model is chosen by minimizing the PMSE. However, the PMSE depends on (a) the unobserved replicates $u_{1+h}, \ldots, u_{T+h}$, (b) the observed data (through the estimate $\hat{\delta}$), and (c) the use of proxies $\hat{z}_1, \ldots, \hat{z}_T$ because the true factors in z_t are unobservable. A natural estimator for η in (3.3.9) is given by

$$\hat{\eta} = \frac{1}{T} \sum_{t=1}^{T} \left(y_{t+h} - \hat{z}_t' \hat{\delta} \right)^2,$$

which replaces the unknown distribution $G(\mathbf{u})$ with an empirical distribution with a mass of $1/T$ on each observation. However, $\hat{\eta}$ is biased due to the estimation error. The bias is given by

$$b = \int \left\{ \frac{1}{T} \int \left[\sum_{t=1}^{T} \left(u_{t+h} - \hat{z}_t' \hat{\delta} \right)^2 \right] dG(\mathbf{u}) - \frac{1}{T} \sum_{t=1}^{T} \left(y_{t+h} - \hat{z}_t' \hat{\delta} \right)^2 \right\} dG(\mathbf{y}).$$

Assuming that $\hat{b}$ estimates b consistently by using some procedure, the following bias-corrected PMSE is considered:

$$PMSE = \hat{\eta} - \hat{b}. \tag{3.3.10}$$

See Remark 4 in Ando and Tsay (2014) for a detailed discussion of b and its consistent estimator $\hat{b}$. Ando and Tsay (2014) choose the best diffusion-index model that minimizes this PMSE score in (3.3.10). Based on the bias-corrected PMSE, the criterion is a natural extension of the traditional Akaike information criterion (AIC), but it relaxes the restrictive distributional assumptions for the likelihood.

Djogbenou (2021) selection procedures for FAR models. The first procedure is a generalization of leave-d-out cross-validation (CV), which can select the smallest basis for the space spanned by the true factors. Given the b random draws of d indexes S in $\{1, \ldots, T\}$ called validation samples, for each draw $S = \{s(1), \ldots, s(d)\}$, define

$$\mathbf{y}_S = \begin{pmatrix} y_{s(1)} \\ y_{s(2)} \\ \vdots \\ y_{s(d)} \end{pmatrix} \text{ and } \hat{\mathbf{Z}}_S(m) = \begin{pmatrix} \hat{F}_{s(1)}(m) & W_{s(1)} \\ \hat{F}_{s(2)}(m) & W_{s(2)} \\ \vdots & \vdots \\ \hat{F}_{s(d)}(m) & W_{s(d)} \end{pmatrix},$$

where $\hat{F}_t(m)$ is a subset from the $r \times 1$ estimated factors $\hat{F}_t$ by using PCA, and m is any of the 2^r subsets of indices in $\{1, \ldots, r\}$ denoted $\mathcal{M}$ including the empty set. The corresponding construction sample is indexed by $S^c = \{1, \ldots, T\} \backslash S$ with $\mathbf{y}_{S^c}$ being the complement of $\mathbf{y}_S$ in $\mathbf{y}$, and $\hat{\mathbf{Z}}_{S^c}$ being the complement of $\hat{\mathbf{Z}}_S$ in $\hat{\mathbf{Z}} = (\hat{z}_1, \ldots, \hat{z}_T)'$. Denote $\hat{\mathbf{y}}_S(m) = \hat{\mathbf{Z}}_S(m)\hat{\delta}_S(m)$, $\hat{\delta}_S(m) = \left(\hat{\mathbf{Z}}_{S^c}'\hat{\mathbf{Z}}_{S^c}\right)^{-1} \hat{\mathbf{Z}}_{S^c}'\mathbf{y}_{S^c}$. The Monte Carlo leave-$d$-out CV estimated model is obtained by minimizing

$$CV_d(m) = \frac{1}{d \cdot b} \sum_{S \in \mathcal{R}} \|\mathbf{y}_S - \hat{\mathbf{y}}_S(m)\|^2,$$

where $\mathcal{R}$ represents a collection of b subsets of size d randomly drawn from $\{1, \ldots, T\}$. When $d = 1$, $S = \{t\}$, and $\mathcal{R} = \{\{1\}, \ldots, \{T\}\}$, $CV_1(m)$ is the usual leave-one-out CV objective function. The consistency of the leave-d-out CV for the diffusion index model is established in Djogbenou (2021). In addition, Djogbenou (2021) shows that the usual CV fails to give consistent model selections.

The second procedure proposed by Djogbenou (2021) is a generalization of the bootstrap approximation of the squared error of prediction of Shao (1996). Djogbenou (2021) proves the validity of the described bootstrap scheme under some conditions. The algorithm for bootstrapping the squared error of prediction is as follows:

Step 1. Obtain the estimates $\hat{F}$ and $\hat{\Lambda}$ for F and Λ from X with the PCA method .

Step 2. For each model m :

1. Compute $\hat{\delta}(m)$ by regressing $\mathbf{y}$ on $\hat{\mathbf{Z}}(m) = \left(\hat{F}(m), W\right)$.
2. Generate B bootstrap samples $X_{it}^* = \hat{F}_t'\hat{\lambda}_i + e_{it}^*$, $\mathbf{y}^*(m) = \hat{\mathbf{Z}}(m)\hat{\delta}(m) + \varepsilon^*$, where $\{e_{it}^*\}$ and $\{\varepsilon_t^*\}$ are the bootstrap errors based on $\{\hat{e}_{it}\}$ and $\{\hat{\varepsilon}_t\}$, respectively, with $\hat{\varepsilon}_t$ being the residual when all the estimated factors are used.
3. $\{e_{it}^*\}$ are obtained by multiplying $\{\hat{e}_{it}\}$ i.i.d. external draws η_{it} with $E\eta_{it} = 0$ and $Var(\eta_{it}) = 1$ for $i = 1, \ldots, N$, and $t = 1, \ldots T$.
4. $\{\varepsilon_t^*\}_{t=1,\ldots,T}$ are the i.i.d. draws of $\left\{\frac{\sqrt{T/\kappa}}{\sqrt{1-(r+q)/T}}\left(\hat{\varepsilon}_t - T^{-1}\sum_{t=1}^T \hat{\varepsilon}_t\right)\right\}$, with $\kappa \to \infty$ and $\frac{\kappa}{\min[N,T]} \to 0$, and q is the dimension of W_t.
5. For each bootstrap sample, extract $\hat{F}^*$ from X^* and estimate $\hat{\delta}_\kappa^*(m)$ based on $\hat{\mathbf{Z}}^*(m) = \left(\hat{F}^*(m), W\right)$ and $y^*(m)$.

Step 3. Obtain $\hat{m}$ as the model that minimizes the average of $\hat{\Gamma}_\kappa^j(m) = T^{-1}\left\|\mathbf{y} - \hat{\mathbf{Z}}^{*j}(m)\hat{\delta}_\kappa^{*j}(m)\right\|^2$ over the B samples indexed by j, as given by $\hat{\Gamma}_\kappa(m) = B^{-1}\sum_{j=1}^B \hat{\Gamma}_\kappa^j(m)$.

Other recent developments. For a slightly different forecasting model, where both latent factors and idiosyncratic components from a large set of predictors may enter the predictive regression, Fosten (2017) proposes some model selection criteria that considers the uncertainty in estimating both components. The criteria can jointly select the estimated factors and idiosyncratic components consistently. For other approaches to determine the forecasting model with FAR, see Bai and Ng (2009) for the boosting approach to diffusion index, Cheng and Hansen (2015) for the forecast combination with FAR based on frequentist model averaging criteria, Kelly and Pruitt (2015) for the three-pass filter regression approach, and Carrasco and Rossi (2016) for the regularization methods for in-sample inference and forecasting in misspecified factor models.

3.4 Nonlinear FARs

The FAR framework is also useful in nonlinear or semiparametric regressions. Recent developments for FAR in nonlinear models include a factor-augmented functional-coefficient predictive regression, factor-augmented quantile prediction regression, factor-augmented forecasting model with structural changes or threshold effects, etc. In this section, we give a selective review on nonlinear FAR models.

3.4.1 Factor-augmented functional-coefficient predictive regression models

Li et al. (2020) introduce a new class of functional-coefficient predictive regression models, where the regressors consist of lagged target variables and latent factor regressors, and the coefficients vary with certain index variables. Specifically, they consider the following one-step ahead predictive regression model:

$$\begin{aligned} y_{t+1} &= \sum_{l=1}^{q_n} \alpha_{1l}(u_t) z_{tl} + \sum_{j=1}^{d_0} \alpha_{2j}(u_t) y_{t+1-j} + \varepsilon_{t+1} \\ &= Z_t'\alpha_1(u_t) + Y_t'\alpha_2(u_t) + \varepsilon_{t+1}, \end{aligned} \tag{3.4.1}$$

$t = 1, \ldots, T$, where y_{t+1} is the dependent variable, $Z_t = (z_{t1}, \ldots, z_{tq_T})$ is a $q_T \times 1$ vector of exogenous covariates with $q_T \to \infty$ as $T \to \infty$, $Y_t = (y_t, \ldots, y_{t-d_0+1})'$ includes fixed

d_0 lags of the response, u_t is a univariate index variable, $\alpha_1(\cdot) = [\alpha_{11}(\cdot), \ldots, \alpha_{1q_T}(\cdot)]'$ and $\alpha_2(\cdot) = [\alpha_{21}(\cdot), \ldots, \alpha_{2d_0}(\cdot)]'$ are two vectors of coefficient functions, and ε_{t+1} is the error. When $u_t \equiv t/T$ is the time regressor, the model (3.4.1) becomes the time-varying coefficient (TVC) model. In the above predictive model, Z_t are determined by some latent factors outside of the models, and Y_t are determined within the model.

When the dimension of regressors is extremely high or moderately large, some commonly used approaches such as a shrinkage estimation or screening method are used to remove the insignificant regressors or reduce the model dimension. However, as Fan and Lv (2008) points out, when irrelevant regressors are highly correlated with some relevant ones, both methods tend to select these irrelevant regressors into the model with higher priority than some other relevant regressors, which then causes high false positive rates and low true positive rates. In time series models, the problem may be more severe since regressors often contain the lags of the dependent variable and cause strong correlations among the regressors.

To address this problem, Li et al. (2020) develop an alternative dimension-reduction technique for the high-dimensional functional-coefficient predictive regression model by imposing a factor structure on Z_t:

$$Z_t = \mathbb{B}_T F_t + V_t \tag{3.4.2}$$

where $\mathbb{B}_T$ is a $q_T \times r$ matrix of factor loadings, F_t is a $r \times 1$ vector of latent common factors that is stationary and weakly dependent, and V_t is a q_n-dimensional vector of idiosyncratic errors. The number of factors r is usually unknown and may increase slowly as the sample size increases; see Li et al. (2017). Using the structure model in (3.4.2), Li et al. (2020) consider the following functional-coefficient predictive model by using the factor regressors:

$$y_{t+1} = F_t'\beta_1(u_t) + Y_t'\beta_2(u_t) + \varepsilon_{t+1} \tag{3.4.3}$$

where $\beta_1(\cdot) = (\beta_{11}(\cdot), \ldots, \beta_{1r}(\cdot))' = \mathbb{B}_T'\alpha_1(\cdot)$, $\beta_2(\cdot) = \alpha_2(\cdot)$, and $\varepsilon_{t+1} = V_t'\alpha_1(u_t) + \varepsilon_{t+1}$. The model in (3.4.2) and (3.4.3) is called the *factor-augmented functional-coefficient model (FA-FCM)*. Noting that the factor regressors are unobservable in (3.4.3), Li et al. (2020) propose to replace them by their PC estimators with data Z_t's. Specifically, they introduce a two-stage estimation procedure.

Step 1. Estimate the factor regressors by using PCA. Let $\hat{\mathbb{F}}_T \equiv \left(\hat{F}_1, \ldots, \hat{F}_T\right)'$ be the PC estimators, which are an $T \times r$ matrix that consists of the r eigenvectors (multiplied by $\sqrt{n}$) associated with the r largest eigenvalues of $\mathbb{Z}_T\mathbb{Z}_T'/(Tq_T)$ (ranked in descending order), where $\mathbb{Z}_T = (Z_1, \ldots, Z_T)'$.

Step 2. Estimate the rotated coefficient functions by the local linear (LL) smoothing method. Let $X_t = \left[(HF_t)', Y_t'\right]'$ and $\beta_H(\cdot) = [\beta_1'(\cdot)H, \beta_2'(\cdot)]'$. Write the model (3.4.3) as follows:

$$y_{t+1} = X_t'\beta_H(u_t) + \varepsilon_{t+1}.$$

Assuming that the coefficient functions $\beta_H(u_t)$ have continuous second-order derivatives, one can estimate the above model with the LL smoothing method by using the estimated factors in Step 1. Let $\hat{X}_t = \left[\hat{F}_t', \hat{Y}_t'\right]'$, $\mathbb{Y}_T = (y_2, \ldots, y_T)'$,

$$\hat{\mathbb{X}}_T(u) = \begin{pmatrix} \hat{X}_1' & \hat{X}_1'(u_1 - u) \\ \vdots & \vdots \\ \hat{X}_n' & \hat{X}_n'(u_1 - u) \end{pmatrix}, \text{ and } \mathbb{W}_T(u) = \texttt{diag}\{K_b(u_1, u), \ldots, K_b(u_T, u)\}$$

where $K_b(u_i, u) \equiv K((u_i - u)/b)$, $K(\cdot)$ is a kernel function, and b is the bandwidth. The local linear estimate of $\beta_H(u)$ is given by

$$\hat{\beta}_H(u) = \left[I_{r+d_0}, 0_{(r+d_0)\times(r+d_0)}\right]\left[\hat{\mathbb{X}}_T'(u)\,\mathbb{W}_T(u)\hat{\mathbb{X}}_T(u)\right]^{-1}\hat{\mathbb{X}}_T'(u)\,\mathbb{W}_T(u)\mathbb{Y}_T \tag{3.4.4}$$

where $u \in \mathcal{U}$ with $\mathcal{U}$ is the support of the index variable u_t.

Given the feasible LL estimate in (3.4.4), the one-step ahead prediction of y_{T+1} is given by

$$\hat{y}_{T+1|T} = \hat{\beta}_{H,T-1}(u_T)' \hat{X}_T \tag{3.4.5}$$

where $\hat{\beta}_{H,T-1}$ is the LL estimate as in (3.4.4) that uses the sample (y_{t+1}, u_t, Z_t), $t = 1, \ldots, T$. Clearly, $\hat{y}_{T+1|T}$ in (3.4.5) is an estimate for

$$y_{T+1|T} = F_T' \beta_1(u_T) + Y_T' \beta_T(u_t) = X_T' \beta_H(u_T).$$

For a given $0 < \alpha < 1$, the $(1-\alpha)$ CI of $y_{T+1|T}$ can be defined by

$$\left[\hat{y}_{T+1|T} - c_{\alpha/2}\sqrt{\widehat{Var}(\hat{y}_{T+1|T})}, \hat{y}_{T+1|T} + c_{\alpha/2}\sqrt{\widehat{Var}(\hat{y}_{T+1|T})}\right]$$

where $c_{\alpha/2}$ is the upper $\alpha/2$-percentile of $\left(\hat{y}_{T+1|T} - y_{T+1|T}\right)/\sqrt{\widehat{Var}(\hat{y}_{T+1|T})}$, and $\widehat{Var}(\hat{y}_{T+1|T})$ is the estimate of the variance of $\hat{y}_{T+1|T}$. To construct the feasible CI, these above quantities have to be estimated by using the asymptotic result (e.g., Theorem 3 in Li et al. (2020)). However, due to the slow convergence rate caused by the nonparametric nature, such CI construction based on asymptotic theory usually does not perform well in finite samples. Li et al. (2020) propose the use of a wild bootstrap procedure to estimate $c_{\alpha/2}$ and $\widehat{Var}(\hat{y}_{T+1|T})$, and we then proceed to construct the prediction interval.

Under some regularity conditions on factor models, Li et al. (2020) establish the asymptotic properties of the proposed methods:

(i) $\sup_{u\in\mathcal{U}} \left\|\hat{\beta}_H(u) - \tilde{\beta}_H(u)\right\| = o_p\left((Tb)^{-1/2}\right)$, where $\tilde{\beta}_H(u)$ is the infeasible estimator that uses the true factors.

(ii) $\sqrt{Tb} D_T \left(\hat{\beta}_H(u) - \beta_H(u) - \frac{1}{2}\mu_2 \beta_H^{(2)}(u) b^2\right) \xrightarrow{d} N\left(0_{r+d_0}, \Xi(u)\right)$, where $\mu_2 = \int u^2 K(u)\, du$.

(iii) Let $\Delta(u_T, X_T) \equiv \left[\hat{\beta}_{H,T-1}(u_T) - \beta_H(u_T)\right]' X_T$. For a fixed r,

$$\hat{y}_{T+1|T} - y_{T+1|T} = \Delta(u_T, X_T) - \varepsilon_{T+1} + o_p\left(1/(Tb)^{1/2}\right),$$

and conditional on $u_T = u^*$ and $X_T = X^*$,

$$\sqrt{Tb}\left[\Delta(u^*, X^*) - \frac{1}{2}\mu_2 b^2 X^{*\prime} \beta_H^{(2)}(u^*)\right] \xrightarrow{d} N(0, \Sigma^*);$$

where D_T, $\Xi(u)$ and Σ^* are well-defined matrices in Li et al. (2020). The result in (i) shows that the LL estimator and the nonlinear forecast that uses the estimated factor regressors are asymptotically equivalent to those that use the true latent factor regressors. Result (ii) gives the pointwise asymptotic distribution of the LL estimator of the functional coefficient, and result (iii) presents the limiting distribution of the feasible predictor. In practice, to implement the estimation and forecasting method, Li et al. (2020) use the forward selection criterion as a screening tool and adopt the BIC as the stopping rule to estimate d_0 and r.

3.4.2 Factor-augmented quantile regressions

Since the seminal work by Koenker and Bassett Jr (1978), quantile regressions (QRs) have been widely used in empirical studies such as labor economics, macroeconomics and risk management in finance, and their theoretical properties have been extensively investigated in the econometric and statistical literature. In a data-rich environment, the factor structure, as an efficient tool that summarizes the information in a large set of predictors or regressors, is incorporated into a QR. For example, Ando and Tsay (2011) study the estimation of factor-augmented conditional QR models and propose an information criterion based on the bias-corrected expected log-likelihood to select the estimated factors and observable predictors; Ohno and Ando (2018) adopt a factor-augmented quantile predictive regression system to predict the stock returns with the shrinkage method to select the useful predictors. Other recent developments include the model averaging approach to forecasting with factor-augmented quantile autoregressions in Phella (2020) and the quantile-based asset pricing model in Ando et al. (2019) and Belloni et al. (2019). In the literature of factor-augmented QR models, another strand of works treat the factors as unknown parameters to be estimated; see Ando and Bai (2020), Chen et al. (2021), and Ma et al. (2021). In this subsection, we focus on the developments of the factor-augmented approach to QRs.

Ando and Tsay (2011) consider the τth conditional quantile of y_t given X_t and W_t:

$$q_\tau\left(y_t|z_t;\gamma\right) = \alpha\left(\tau\right)' F_t + \beta\left(\tau\right)' W_t \equiv \gamma\left(\tau\right)' z_t \tag{3.4.6}$$

where $\gamma\left(\tau\right) = \left(\alpha\left(\tau\right)', \beta\left(\tau\right)'\right)'$ is a vector of the coefficients that depend on the quantile τ, and $z_t = (F_t', W_t')'$. If $\tau = 0.5$, then the model in (3.4.6) reduces to the conditional median regression, which is more robust to outliers than the usual conditional mean regression. In addition, the set of latent factors F_t also enters a panel data set

$$\underline{X}_t = \Lambda F_t + \varepsilon_t, t = 1, \ldots, T,$$

where $\underline{X}_t = (X_{1t}, \ldots, X_{Nt})'$ is an N-dimensional observable random vector.

The estimation procedure consists of two stages. In the first stage, common factors are estimated from the panel data set $\{X_t\}$ with PCA. In the second stage, the estimated factors are used in the standard quantile regression. Putting the estimated factors into the model yields

$$q_\tau\left(y_t|\hat{z}_t;\gamma\right) = \alpha\left(\tau\right)' \hat{F}_t + \beta\left(\tau\right)' W_t \equiv \gamma\left(\tau\right)' \hat{z}_t$$

where $\hat{z}_t = (\hat{F}_t', W_t')'$, and $\hat{F}_t$ is the r-dimensional PC estimator for common factors F_t. The unknown parameters $\gamma\left(\tau\right)$ are estimated by maximizing the log-likelihood function:

$$l_\tau\left(\mathbf{y}_T; \gamma\left(\tau\right), \hat{Z}\right) = \frac{1}{T}\log\left[\tau\left(1-\tau\right)\exp\left\{-\sum_{t=1}^{T}\rho_\tau\left(y_t - \gamma\left(\tau\right)'\hat{z}_t\right)\right\}\right]$$

with $\rho_\tau\left(u\right) = u\left(\tau - \mathbf{1}\left(u<0\right)\right)$, $\mathbf{y}_T = (y_1, \ldots, y_T)'$, and $\hat{Z} = (\hat{z}_1, \ldots, \hat{z}_T)'$. The estimate $\hat{\gamma}\left(\tau\right)$ is obtained as a solution of $\partial l_\tau\left(\gamma\right)/\partial\gamma = 0$, which is a linear optimization problem. Koenker and Bassett Jr (1978) showed that the solution is

$$\hat{\gamma}\left(\tau\right) = \hat{Z}^{-1}\left(h_\tau\right) y\left(h_\tau\right),$$

where h_τ is a p-element index subset from the set $\{1, 2, \ldots, T\}$, $\hat{Z}\left(h_\tau\right)$ refers to indexed rows in $\hat{Z}$, and $y\left(h_\tau\right)$ refers to the elements in y selected by h_τ. The asymptotic normality of $\hat{\gamma}\left(\tau\right)$ is established by Bai and Ng (2008). Under some mild conditions and $T^{5/8}/N \to 0$,

$$\sqrt{T}\left[\hat{\gamma}\left(\tau\right) - \gamma\left(\tau\right)\right] \xrightarrow{d} N\left(0, \Omega_\tau\right),$$

where Ω_τ is the positive asymptotic variance matrix; see Lemma A.2 in Ando and Tsay (2011) for the detailed expression of Ω_τ.

By replacing the unknown parameter $\gamma\left(\tau\right)$ with $\hat{\gamma}\left(\tau\right)$, we obtain the estimated quantile regression model with factor-augmented predictors:

$$q_\tau\left(y_t|\hat{z}_t;\hat{\gamma}\left(\tau\right)\right) = \hat{\gamma}\left(\tau\right)'\hat{z}_t, t = 1, 2, ..., T.$$

The asymptotic properties of $q_\tau\left(y_t|\hat{z}_t;\hat{\gamma}\left(\tau\right)\right)$ can be analyzed similar to the term in (3.3.7). After estimating the model, the goodness of fit can be assessed from a predictive point of view. Let $\mathbf{u}_T = (u_1, \ldots, u_T)'$ be replicates of the dependent variable $\mathbf{y}_T$ drawn from $g\left(u\right)$, which is the target probability that generates the data. The goodness of fit for $q_\tau(y_t|\hat{z}_t;\hat{\gamma})$ can be given by the *expected log-likelihood*

$$\eta_\tau\left(G,\hat{\gamma},\hat{Z}\right)=\int l_\tau\left(\mathbf{u}_T;\hat{\gamma}(\tau),\hat{Z}\right)dG(\mathbf{u}_T) \tag{3.4.7}$$

where $dG(\mathbf{u})$ is the Lebesgue measure with respect to the true probability density $g(u)$. The best candidate model should maximize the expected log-likelihood function in (3.4.7). In this sense, the optimal number of factors is defined as the value of r that maximizes the expected log-likelihood function.

Note that the true distribution $G(\mathbf{u})$ is unknown, which can be replaced by the empirical distribution $\hat{G}(\cdot)$. Thus, a natural estimator of the expected log-likelihood is the sample-based log-likelihood:

$$\eta_\tau\left(\hat{G},\hat{\gamma},\hat{Z}\right)=\int l_\tau\left(\mathbf{y}_T;\hat{\gamma}(\tau),\hat{Z}\right)d\hat{G}(\mathbf{y}_T)=\log\left[\tau(1-\tau)\right]-\frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_t-\gamma(\tau)'\hat{z}_t\right). \tag{3.4.8}$$

However, the estimate in (3.4.8) is positively biased because the same data are used both in the estimation of unknown parameters and the evaluation of the expected log-likelihood. The bias term is given by

$$b_\tau(G)=\int\left[\eta_\tau\left(\hat{G},\hat{\gamma},\hat{Z}\right)-\eta_\tau\left(G,\hat{\gamma},\hat{Z}\right)\right]dG(Y_T).$$

A consistent estimator $\hat{b}_\tau(G)$ is given by Theorem 3.1 of Ando and Tsay (2011). Then, based on the bias-corrected estimator for the expected log-likelihood, a model selection criterion for evaluating the estimate is given by

$$IC=-2l_\tau\left(\mathbf{y}_T;\hat{\gamma}(\tau),\hat{Z}\right)+2T\hat{b}_\tau(G). \tag{3.4.9}$$

Let $B_{NT}=\min\left[N,T^{3/2}\right]$. Ando and Tsay (2011) show that the IC in (3.4.9) is a consistent estimator of the expected log-likelihood with order $O_p\left(B_{NT}^{-1}\right)$ and then can asymptotically select the optimal number of factors that maximizes the expected log-likelihood.

Recently, Ohno and Ando (2018) have used the factor-augmented quantile predictive regression system to predict stock returns, where a shrinkage method is adopted to select the useful estimated factors and observed predictors. To be Specifically, they consider the following quantile predictive model:

$$q_\tau(y_{t+h}|z_t;\gamma)=\alpha(\tau)'F_t+\beta(\tau)'W_t\equiv\gamma(\tau)'z_t,$$
$$X_{it}=\lambda_i'F_t+\varepsilon_{it},t=1,\ldots,T,i=1,\ldots,N.$$

A two-stage estimation procedure is considered. In the first stage, the factors are estimated with PCA from the observables X_{it}'s. In the second stage, the shrinkage method is combined with the usual quantile regression to select the predictors. The unknown parameters $\gamma(\tau)$ are estimated by using a shrinkage method, which is obtained by minimizing

$$\frac{1}{T}\sum_{t=1}^{T-h}\rho_\tau\left(y_{t+h}-\gamma(\tau)'\hat{z}_t\right)+p_\delta(\gamma(\tau)),$$

where $\rho_\tau(\cdot)$ is the check function, and $p_\delta(\cdot)$ is a penalty function of the coefficients indexed by a regularization parameter δ that controls the trade-off between the loss function and the penalty. They propose the use of a smoothly clipped absolute deviation (SCAD) penalty in Fan and Li (2001) $p_\delta(\gamma)=\sum_{k=1}^{p}p_{\kappa,\delta}(\gamma_k)$, where

$$p_{\delta,\kappa}(\gamma_k)=\begin{cases}\kappa|\gamma_k|, & \text{if } |\gamma_k|\le\kappa\\ \frac{\kappa\delta|\gamma_k|-0.5(\gamma_k^2+\kappa^2)}{\delta-1}, & \text{if } \kappa<|\gamma_k|\le\kappa\delta\\ 0.5\kappa^2(\delta+1), & \text{if } \kappa\delta<|\gamma_k|\end{cases}$$

for $\kappa > 0$ and $\delta > 2$. For the algorithm for QR with the SCAD penalty, see Wu and Liu (2009). Ohno and Ando (2018) establish the variable selection consistency and asymptotic normality of the estimator under some mild conditions. The regularization parameter is selected by using the prediction error because the aim is to predict the behavior of the stock market. The FAR quantile predictive model is applied to analyze the Tokyo Stock Exchange, and they show that macroeconomic variables play an important role in predicting the characteristics of the stock market of Japan even after controlling the various asset-pricing factors.

3.4.3 Time-varying DI forecasting model

The standard DI forecasting model assumes *time invariant* factor loadings and coefficients in the predictive regression model. However, considerable empirical evidence of parameter instability and time-varying effects has been found in the literature of macroeconomics and finance. Ignoring such instabilities in the diffusion model may lead to inconsistent estimators of the factors (Bates et al., 2013) or an inconsistent estimation of the coefficients (Clements and Hendry, 1996). Although the rolling window (RW) method is usually used to address the unstable parameters in empirical applications, the choice of windows is arbitrary or just based on past experience. As Rossi and Inoue (2012) and Inoue et al. (2017) point out, the forecasting performance of the RW scheme is sensitive to the choice of the window size since it balances the estimation efficiency and the misspecification bias.

To accommodate the time-varying factor loadings and time-varying regression coefficients, Wei and Zhang (2020) propose a general time-varying DI forecasting model. Given N observable predictors $\underline{X}_t = (X_{1t}, \cdots, X_{Nt})'$ and a target variable y_t, they consider forecasting $y_{T+\ell}$ for some integer $\ell > 0$ given the information set up according to the Tth period. Assume that X_{it} has a static factor representation with time-varying factor loadings as Su and Wang (2017):

$$X_{it} = \lambda_{it}' F_t + e_{it},\ i = 1, \ldots, N, t = 1, \ldots, T, \tag{3.4.10}$$

where F_t is an $r \times 1$ vector of the latent factors, $\lambda_{it} \equiv \lambda_i(t/T)$ are the time-varying factor loadings with $\lambda_i(\cdot) : [0,1] \to \mathbb{R}^r$ being a vector of unknown piecewise smooth functions, and e_{it} is the idiosyncratic error. Meanwhile, the target variable to be predicted is also driven by the same set of common factors and some other observables:

$$y_{t+\ell} = \alpha_t' F_t + \beta_t' W_t + \varepsilon_{t+\ell}, \tag{3.4.11}$$

where W_t is a $p \times 1$ vector of the observed variables including the constant, the lags of y_t or time trends, $\alpha_t \equiv \alpha(t/T)$ and $\beta_t \equiv \beta(t/T)$ are the TVCs for predictors F_t and W_t, respectively, with $\alpha(\cdot)$ and $\beta(\cdot)$ being unknown smooth functions on $[0,1]$, and $\varepsilon_{t+\ell}$ is the error term. The models (3.4.10) and (3.4.11) are termed the time-varying diffusion index (TVDI) forecasting models. If F_t is observable and the conditional mean of $\varepsilon_{t+\ell}$ is zero given the past information, then the (mean-squared) optimal prediction of $y_{T+\ell}$ is given by

$$y_{T+\ell|T} = E\left(Y_{T+\ell}|F_T, W_T, F_{T-1}, W_{T-1}, \cdots\right) = \alpha'(1)F_T + \beta'(1)W_T.$$

Here, factors F_T and TVCs $\alpha(1)$ and $\beta(1)$ are all unknown. To address this issue, it is tempting to follow the road map of Bai and Ng (2006) to estimate F_t's first from the large number of observed predictors and then use them as generated regressors to estimate $\alpha(1)$ and $\beta(1)$ nonparametrically in (3.4.11). The two-step procedure to construct the time-varying DI forecast is as follows:

Step 1. Apply the local principal component analysis (LPCA) in Su and Wang (2017) to estimate the locally weighted factors from a large number of predictors. Let $K_h(\cdot) \equiv h^{-1}K(\cdot/h)$ and $K_{h,tr} \equiv K_h(\tau_t - \tau_r)$, where $\tau_s \equiv s/T$ for $1 \leq s \leq T$, and $h \to 0$ is the bandwidth as $(N,T) \to \infty$. At the Tth period, estimate $\{\lambda_{iT}\}_{r=1}^{T}$ and $\{F_t\}_{t=1}^{T}$ by minimizing the following kernel weighted least squares objective function:

$$\min_{\{\lambda_{iT}\}_{i=1}^{N},\{F_t\}_{t=1}^{T}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\lambda_{iT}'F_t\right)^2 K_{h,tT}$$
$$\Leftrightarrow \min_{\Lambda_T,F^{(T)}} \sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}^{(T)}-\lambda_{iT}'F_t^{(T)})^2 = \mathtt{tr}[(X^{(T)}-F^{(T)}\Lambda_T')(X^{(T)}-F^{(T)}\Lambda_T')'] \tag{3.4.12}$$

where $X_{it}^{(T)} \equiv X_{it}K_{h,tT}^{1/2}$, $X_i^{(T)} \equiv (X_{i1}^{(T)},\cdots,X_{iN}^{(T)})'$, $X^{(T)} \equiv (X_1^{(T)},\cdots,X_N^{(T)})$, $F_t^{(T)} \equiv F_tK_{h,tT}^{1/2}$, $F^{(T)} \equiv (F_1^{(T)},\cdots,F_T^{(T)})'$, and $\Lambda_T \equiv (\lambda_{1T},\cdots,\lambda_{NT})'$. To identify $F^{(T)}$ and Λ_T in (3.4.12), the usual restrictions for PCA are used: $F^{(T)\prime}F^{(T)}/T = I_r$ and $\Lambda_T'\Lambda_T$ is a diagonal matrix, where I_r is the $r \times r$ identity matrix. After concentrating out Λ_T, we can estimate $F^{(T)}$ by

$$\hat{F}^{(T)} = \arg\max_{F^{(T)}} \mathtt{tr}\left[F^{(T)\prime}X^{(T)}X^{(T)\prime}F^{(T)}\right] \tag{3.4.13}$$

subject to $F^{(T)\prime}F^{(T)}/T = I_r$. The estimator $\hat{F}^{(T)} = (\hat{F}_1^{(T)},\cdots,\hat{F}_T^{(T)})'$ is $\sqrt{T}$ times the eigenvectors that correspond to the r largest eigenvalues of the matrix $X^{(T)}X^{(T)\prime}$, arranged in descending order, and $\hat{\Lambda}_T = X^{(T)\prime}\hat{F}^{(T)}(\hat{F}^{(T)\prime}\hat{F}^{(T)})^{-1} = X^{(T)\prime}\hat{F}^{(T)}/T$. To mitigate the boundary bias problem since T is at the right boundary, a boundary-corrected kernel function should be used.

Step 2. Augment the forecasting regression with the estimated factors in (3.4.13) to estimate the TVCs with the kernel method. Because t is near T, transform (3.4.11) as follows:

$$\begin{aligned} y_{t+\ell} &= \alpha'(\tau_t)[H^{(T)\prime}]^{-1} \times K_{h,tT}^{-1/2}H^{(T)\prime}F_t^{(T)} + \beta'(\tau_t)W_t + \varepsilon_{t+\ell} \\ &\equiv \gamma^{(T)}(\tau_t)'D_t^{(T)} + \beta'(\tau_t)W_t + \varepsilon_{t+\ell} \\ &= \gamma^{(T)}(\tau_t)'\hat{D}_t^{(T)} + \beta'(\tau_t)W_t + \varepsilon_{t+\ell}^* \\ &= \delta^{(T)\prime}(\tau_t)\hat{G}_t^{(T)} + \varepsilon_{t+\ell}^* \end{aligned}$$

where $\gamma^{(T)}(\tau_t)' \equiv \alpha'(\tau_t)[H^{(T)\prime}]^{-1}$, $D_t^{(T)} \equiv K_{h,tT}^{-1/2}H^{(T)\prime}F_t^{(T)}$, $\hat{D}_t^{(T)} \equiv K_{h,tT}^{-1/2}\hat{F}_t^{(T)}$, $\delta^{(T)}(\tau_t) = \left(\gamma^{(T)}(\tau_t)',\beta(\tau_t)'\right)'$ and $\hat{G}_t^{(T)} = \left(\hat{D}_t^{(T)},W_t\right)$. In the above third equation, the unobservable composite regressors $D_t^{(T)}$ are replaced by their consistent estimators $\hat{D}_t^{(T)} \equiv K_{h,tT}^{-1/2}\hat{F}_t^{(T)}$, which come from the local PCA in Step 1. Note that $\delta^{(T)}(\tau_t)$ are the TVCs. A feasible forecast is proposed by combining the estimated factors and the nonparametric estimators of the coefficients. The proposed feasible forecast for $Y_{T+\ell}$ follows as

$$\hat{y}_{T+\ell|T} = \hat{\delta}^{(T)\prime}(1)\hat{G}_T^{(T)}.$$

where $\hat{\delta}^{(T)\prime}(1) = \left(\sum_{t=1}^{T-\ell}\hat{G}_t^{(T)}\hat{G}_t^{(T)\prime}K_{h^*,Tt}\right)^{-1}\sum_{t=1}^{T-\ell}\hat{G}_t^{(T)}y_{t+\ell}K_{h^*,Tt}$ is the local constant estimator of $\delta^{(T)}(1)$, with h^* being the bandwidth parameter.

For the model in (3.4.10) and (3.4.11) without W_t, Corradi and Swanson (2014) are concerned with testing the following hypotheses:

$$\mathbb{H}_0 : \Lambda_t = (\lambda_{1t},\ldots,\lambda_{Nt}) = \Lambda_0 \text{ and } \alpha_t = \alpha_0 \text{ for all } t,$$

versus

$$\mathbb{H}_1 : \Lambda_t = \begin{cases} \Lambda_{0,1} & \text{for } t/T \leq \tau_\lambda^{(1)} \\ \Lambda_{0,2} & \text{for } \tau_\lambda^{(1)} + 1 \leq t/T \leq \tau_\lambda^{(2)} \\ \vdots & \\ \Lambda_{0,q_\lambda+1} & \text{for } \tau_\lambda^{(q_\lambda)} + 1 \leq t/T \leq 1 \end{cases} \quad \text{and/or}$$

$$\alpha_t = \begin{cases} \alpha_{0,1} & \text{for } t/T \leq \tau_\alpha^{(1)} \\ \alpha_{0,2} & \text{for } \tau_\alpha^{(1)} + 1 \leq t/T \leq \tau_\alpha^{(2)} \\ \vdots & \\ \alpha_{0,q_\alpha+1} & \text{for } \tau_\alpha^{(q_\alpha)} + 1 \leq t/T \leq 1 \end{cases}$$

where the factor loadings Λ_t and slopes α_t are time-invariant under the null hypothesis and may have shifted under the alternative hypothesis. To construct the test, they proceed in three steps. First, the IC of Bai and Ng (2002) is used to determine the number of breaks by using the full sample. Second, the factors are estimated by again using the full sample. The PC estimator is denoted as $\hat{G}_T$. Third, an estimator of the sample covariance between $y_{t+\ell}$ and the estimated factors is constructed by using both the full sample and rolling windows of observations. The difference between the full sample and the rolling estimators of the covariance between $y_{t+\ell}$ and the estimated factors is the key ingredient of their statistic. For $\ell = 1$, the testing statistic is constructed as follows:

$$Z_{P,R} = \sqrt{P}\left[\frac{1}{T}\sum_{t=2}^{T}\hat{G}_{t-1}y_t - \frac{1}{P}\sum_{t=R+1}^{T}\left(\frac{1}{R}\sum_{j=t-1+R}^{t}\hat{G}_{j-1}y_t\right)\right],$$

where R is the window for the rolling estimator, and $P = T - R$. The statistic is more in the spirit of a Hausman-type test than tests for structural stability. Indeed, they compare two estimators that converge to the same probability limit under the null hypothesis but to different probability limits under the alternative hypotheses. Furthermore, under the null hypothesis, the full sample estimator is more efficient. Under some conditions, such as $\sqrt{T}/N \to 0$ and $P/R \to \pi > 0$, then under $\mathbb{H}_0$,

$$Z_{P,R} \xrightarrow{d} N(0, \Omega_0)$$

for some positive matrix Ω_0, and

$$\lim_{P\to\infty} \Pr\left(P^{-1/2}\|Z_{P,R}\| \geq \varepsilon\right) = 1.$$

To implement the test, they propose using the quadratic form $Z'_{P,R}\hat{\Omega}_T^{-1}Z_{P,R}$, where $\hat{\Omega}_T$ is a consistent estimator for Ω_0. Clearly, $Z'_{P,R}\hat{\Omega}_T^{-1}Z_{P,R} \to_d \chi_r^2$ under $\mathbb{H}_0$ and $\Pr\left(P^{-1}Z'_{P,R}\hat{\Omega}_T^{-1}Z_{P,R} \geq \varepsilon\right) \to 1$ under $\mathbb{H}_1$ for any $\varepsilon > 0$.

3.4.4 FAR models with structural breaks or threshold effects

Massacci (2019) considers an unstable diffusion index model as follows

$$\underline{X}_t = 1(t/T \leq \pi_x)\Lambda_1 F_t + 1(t/T > \pi_x)\Lambda_2 F_t + e_t, \tag{3.4.14}$$

$$y_{t+h} = 1(t/T \leq \pi_y)\left(\gamma_1' F_t + \beta_1' W_t\right) + 1(t/T > \pi_y)\left(\gamma_2' F_t + \beta_2' W_t\right) + \varepsilon_{t+h}, \tag{3.4.15}$$

where $t = 1, \ldots, T$, $\underline{X}_t = (X_{1t}, \cdots, X_{Nt})'$, F_t is a $k \times 1$ vector of factors, $\Lambda_l = (\Lambda_{l1}', \cdots \Lambda_{lN}')'$ is a $N \times k$ matrix of factor loadings, $l = 1, 2$, γ_1 and γ_2 are $k \times 1$ vectors of the coefficients for factor F_t, β_1, β_2 are $p \times 1$ vectors of the coefficients for W_t in the predictive model, and $h \geq 0$.

Obviously, there are structural changes in the factor model of X_{it} and the predictive regression model of y_{t+h}.

For the breakpoint factor model in (3.4.14), for a given number R of factors, the joint objective function to be minimized is

$$S\left(\Lambda^R, F^R, \pi_x\right) = \frac{1}{NT}\sum_{t=1}^{T}\left\{\left\|\underline{X}_t - 1_{1t}\left(\pi_x\right)\Lambda_1^R F_t^R - 1_{2t}\left(\pi_x\right)\Lambda_2^R F_t^R\right\|^2\right\}$$

where $1_{1t}\left(\pi_x\right) = 1\left(t/T \le \pi_x\right)$, $\Lambda^R = \left(\Lambda_1^R, \Lambda_2^R\right)$, and $F^R = \left(F_1^R, \ldots, F_T^R\right)$. For the above optimization problem, for a given π_x, the PCA method is used to estimate the factors and factor loadings, and π_x is estimated with the concentrated least squares method. See Massacci (2017) for the estimation of threshold factor models. Once the factors and breakpoints are estimated, we can estimate the model (3.4.15) by minimizing the following feasible objective function:

$$L\left(\theta, \pi_y\right) = T^{-1}\sum_{t=1}^{T-h}\left[y_{t+h} - 1_{1t}\left(\pi_y\right)\left(\gamma_1'\hat{F}_t + \beta_1' W_t\right) - 1_{2t}\left(\pi_y\right)\left(\gamma_2'\hat{F}_t + \beta_2' W_t\right)\right]^2$$

where $\theta = \left(\gamma_1', \gamma_2', \beta_1', \beta_2'\right)'$. As the usual threshold model, we can obtain the estimators $\hat{\theta}$ and $\hat{\pi}_y$ for the unknown coefficients and breakpoint, respectively. The consistency of the estimator $\hat{\pi}_y$ of π and the limiting distribution for the estimated coefficients are established under some mild conditions. A test for structural change in (3.4.15) is also provided by Massacci (2019).

Other recent developments include Wang et al. (2015), where a structural break in the coefficient of W_t is studied, and Yan and Cheng (2022) for factor-augmented forecasting regressions with a threshold effect in the coefficient of F_t in (3.4.15), where the threshold effect is caused by some common random variables.

3.5 FAR with Panel Data

In this section, we review the development of factor augmented panel regression models.

A prototypical panel data regression model with a factor error structure is given by

$$y_{it} = \beta' X_{it} + u_{it}, \tag{3.5.1}$$

$$u_{it} = \lambda_i^{u\prime} F_t^u + \varepsilon_{it} \tag{3.5.2}$$

$i = 1, \ldots, N$, $t = 1, \ldots, T$, where X_{it} is a $p \times 1$ vector of the explanatory variables, F_t^u is an r-vector of the latent common factors in the composite error u_{it}, λ_i^u is a $r \times 1$ vector of factor loadings, and ε_{it} is the idiosyncratic error. The regressor X_{it} is allowed to be correlated with common component $\lambda_i^{u\prime} F_t^u$. This correlation causes the endogeneity bias issue when the least squares (LS) or least-squares dummy variables (LSDV) methods are adopted to estimate the unknown parameters β.

In the literature, several methods have been proposed to estimate unknown parameters in the panel data models with multifactor structural errors. For example, Ahn et al. (2013) estimate the model based on moment restrictions on the error terms, and Bai (2009) considers the joint estimation of the β, factor space $\{F_1^u, \ldots, F_T^u\}$ and factor loadings $\{\lambda_1^u, \ldots, \lambda_N^u\}$ by using the iterated PCA method. Alternatively, one may augment the panel regression directly with the estimated factors or their proxies from the observable data X_{it}'s or (y_{it}, X_{it})'s. For example, Pesaran (2006) proposes the common correlated effects (CCE) estimators that use the cross-sectional averages of $(y_{it}, X_{it}')'$ as the proxies of the unobservable factors. However, the consistency of CCE relies on the rank condition, which is usually difficult to verify in practice. In contrast, Kapetanios and Pesaran (2007) and Greenaway-McGrevy et al. (2012) estimate the latent factors from the observables by using the PCA method, which avoids the rank condition in CCE. For a comprehensive comparison between the CCE and PC approaches, see Westerlund and Urbain (2015) and Reese and Westerlund (2018). For some empirical applications of factor-augmented panel regressions, see Giannone et al. (2010) for the study of the international saving-investment relationship with the global factor being extracted from the observables and Kapetanios and Pesaran (2007) for modeling asset returns. Now, write the model (3.5.1)-(3.5.2) in vector form:

$$y_i = X_i\beta + F^u\lambda_i^u + \varepsilon_i \tag{3.5.3}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$, $X_i = (X_{i1}, \ldots, X_{iT})'$, $F^u = (F_1^u, \ldots, F_T^u)'$, and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$. The correlation between X_{it} and the common component $\lambda_i^{u\prime}F_t^u$ is modeled through a latent factor structure in X_{it}, namely,

$$X_{it} = F_t^{X\prime}\lambda_i^X + V_{it}, \tag{3.5.4}$$

where F_t^X is a vector of the latent factors underlying X_{it}, λ_i^X is the corresponding vector of factor loadings, and $F_t^{X\prime}\lambda_i^X$ can be correlated with $F_t^{u\prime}\lambda_i^u$. In matrix form, (3.5.4) can be written as $X_i = F^X\lambda_i^X + V_i$, where $F^X = \left(F_1^X, \ldots, F_T^X\right)'$, and $V_i = (V_{i1}, \ldots, V_{iT})'$. Let $F^y = \left(F_1^y, \ldots, F_T^y\right)'$ denote the common factors in y_i and λ_i^y denote the associated factor loadings. Define F to be the $T \times m$ matrix that consists of a subset of the columns of F^X and F^y such that $T^{-1}F'F$ is nonsingular, and

$$F^X = FA^X \text{ and } F^y = FA^y$$

for some selection matrices A^X and A^y. That is, F contains all unique columns in F^X and F^y. When X_i and y_i share the same factor, it is included only once as a column in F. Then,

$$F^u\lambda_i^u = F^y\lambda_i^y - F^X\lambda_i^X\beta = F\left(A^y\lambda_i^y - A^X\lambda_i^X\beta\right) \equiv F\left(\lambda_{y,i} - \lambda_{X,i}\beta\right) \equiv F\lambda_{u,i}.$$

Replacing F^u with the full set of factors F, we can write the model (3.5.3) as

$$y_i = X_i\beta + F\lambda_{u,i} + \varepsilon_i$$

with $\lambda_{u,i}$ satisfying $F\lambda_{u,i} = F^u\lambda_i^u$. Then, an infeasible ordinary factor-augmented estimator is given by

$$\hat{b}_{I,FAE} = \left(\sum_{i=1}^N X_i'M_FX_i\right)^{-1}\sum_{i=1}^N X_i'M_Fy_i,$$

where $M_F = I_T - F\left(F'F\right)^{-1}F'$. When F is replaced with a PC estimate $\hat{F}$ from the observables $\{(y_{it}, X_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$, a feasible estimator is given by

$$\begin{aligned}\hat{b}_{FAE} &= \left(\sum_{i=1}^N X_i'M_{\hat{F}}X_i\right)^{-1}\sum_{i=1}^N X_i'M_{\hat{F}}y_i \\ &= \beta + \left(\sum_{i=1}^N X_i'M_{\hat{F}}X_i\right)^{-1}\sum_{i=1}^N X_i'M_{\hat{F}}\varepsilon_i + \left(\sum_{i=1}^N X_i'M_{\hat{F}}X_i\right)^{-1}\sum_{i=1}^N X_i'M_{\hat{F}}F\lambda_{u,i}\end{aligned}$$

where $M_{\hat{F}} = I_T - \hat{F}\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'$. The properties of $\hat{b}_{FAE}$ depend heavily on the last term. In particular, consistency requires that $(NT)^{-1}\sum_{i=1}^N\sum_{i=1}^N X_i'M_{\hat{F}}X_i(\varepsilon_i + F\lambda_{u,i}) \to_p 0$, while to make $\hat{b}_{FAE}$ have an unbiased limiting distribution, we require $(NT)^{-1/2}\sum_{i=1}^N\sum_{i=1}^N X_i'M_{\hat{F}}X_i(\varepsilon_i + F\lambda_{u,i})$ to be centered at zero. The consistency of $\hat{b}_{FAE}$ is established for the general estimated factors that are not limited to the PC estimator. Given that $T^{-1}\left\|\hat{F} - FH\right\|^2 = o_p(1)$ for some nonsingular matrix H, as $(N, T) \to \infty$,

$$\hat{b}_{FAE} \xrightarrow{p} \beta.$$

This result shows that the consistency of a standard FAE requires only the consistency of the associated factor estimator, given the PC estimator $\hat{F}$ in the sense that

$$T^{-1}\left\|\hat{F}-FH\right\|^{2}=O_{p}(\delta_{NT}^{-2})$$

with $\delta_{NT}=\min[\sqrt{N},\sqrt{T}]$ and for some asymptotic nonsingular H. Denote the $\hat{b}_{FAE}$ by using PCs as $\hat{b}_{PCA}$. With the above convergence rate for the PC estimators, when $T/N\to 0$, $N/T^{3}\to 0$, then

$$\sqrt{NT}\left(\hat{b}_{PCA}-\hat{b}_{I,FAE}\right)=o_{p}\left(1\right).$$

The above result suggests that if N is large compared to T and if T is not too small, then the asymptotic distribution of the PCA is equivalent to that of its infeasible counterpart. Under the requirement on (N,T),

$$\sqrt{NT}\left(\hat{b}_{PCA}-\hat{b}_{I,FAE}\right)\to_{d}N\left(0,\Xi\right),$$

where $\Xi=\Sigma_{V}^{-1}\Omega_{0}\Sigma_{V}^{-1}$ with $\Omega_{0}=\operatorname{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}X_{i}'M_{F}\varepsilon_{i}\varepsilon_{i}'M_{F}X_{i}$ and $\Sigma_{V}=\operatorname{plim}_{N,T\to\infty}$ $\frac{1}{NT}\sum_{i=1}^{N}X_{i}'M_{F}X_{i}$. However, if $T>N$, then the asymptotic distribution of $\sqrt{NT}\left(\hat{b}_{PCA}-\beta\right)$ is biased. If $N/T\to\in(0,\infty)$, then the asymptotic distribution of $\hat{b}_{PCA}$ is biased. A bias correction as proposed in Bai (2009) should be used to carry out a valid statistical inference.

Finally, we mention the factor-augmented panel regression model where the factors may have mixed factor strength. Despite the generated regressor problem caused by the use of estimated rather than known factors, the PC estimators of β are still $\sqrt{NT}$-consistent and asymptotically normal. This requires strong factors. However, in practice, some factors are only weakly influential; see Chudik et al. (2011) and Chudik and Pesaran (2015b) for some motivating examples. Bai and Ng (2008) and Boivin and Ng (2006) find that PC factors can be severely impaired when the factors are not strong. Chudik et al. (2011) investigate the implications of weak, semi-weak and semi-strong factors on the CCE estimator and find that the CCE estimator performs better and has fewer size distortions than the iterated PC approach of Bai (2009). The findings are because the presence of weak or semi-strong factors leads to inconsistent PC estimates of the (rotated) factors but does not affect the CCE estimator since its aim is to address the endogeneity that arises from error cross-section dependence. In addition, Onatski (2012) studies the case when some of the included factors are semi-strong within the context of a PC estimation of a pure common factor model and finds that the presence of such factors causes the PC estimator to be inconsistent.

Recently, Reese and Westerlund (2018) offer an analysis of the effect of weak, semi-weak and semi-strong factors on two of the most popular estimators for FARs, namely, PC and CCE. As a comparison, we present the main findings for two popular estimators that may facilitate the choice of estimators in applications. They set factor loadings that go to zero at the rate of N^{-a} to capture the different weaknesses of the factor, where $a\in[0,1]$ and N are related to T via $T=N^{k}$. Under the standard assumption of homogenous slopes, the main findings are summarized as follows.

(i) To ensure $\sqrt{NT}$-consistency and to alimit normal distributions, both estimators require strong or semi-strong factors ($a<1/2$). The requirement on k generally differs between the estimators. However, both sets shrink toward 1 as the factors become weaker ($a\to 0$).

(ii) Unless $k<1$, both estimators are asymptotically biased.

(iii) Both estimators can be consistent when the factors are semi-weak ($a<1$), but only PC allows for a consistent estimation in the weak factor case ($a=1$) with additional requirements on k. CCE has binding restrictions on k only for semi-weak factors. By contrast, PC restricts k from below for semi-strong factors and requires that $k>3$ when the factors are weak, which is restrictive.

3.6 Two Applications of FAR in Finance

FAR has been widely used in finance. In this section, we present two typical applications of FAR to financial markets: Ludvigson and Ng (2009) use the FAR to study the bond risk premia, and Qiu et al. (2022) use a factor-augmented HAR model to forecast the realized volatility (RV).

3.6.1 Ludvigson and Ng's (2009) factor models for bond risk premia

Ludvigson and Ng (2009) use the methodology of a dynamic factor analysis for large datasets to study the possible linkages between forecastable variation in excess bond returns and macroeconomic fundamentals. They consider the following model for excess bond returns:

$$rx_{t+1}^{(n)} = \beta' Z_t + \gamma' \underline{X}_t + \varepsilon_{t+1},$$

where $rx_{t+1}^{(n)}$ is the continuously compounded (log) excess return on an n-year discount bond in period $t+1$, Z_t is a set of K predetermined conditioning variables at time t and may include the individual forward rates, the single forward factor, or other predictors based on a few macroeconomic series, $\underline{X}_t = (X_{1t}, \ldots, X_{Nt})'$ is a $N \times 1$ vector of macroeconomic Fundamentals, and N is large and possibly larger than the number of time periods (T). To handle the large dimension of the regressors, they assume that X_{it} has a factor structure:

$$X_{it} = \lambda_i' f_t + e_{it}$$

where f_t is an $r \times 1$ vector of latent common factors, λ_i is the corresponding $r \times 1$ vector of factor loadings, and e_{it} is a vector of idiosyncratic errors. One prominent feature for the factor structure of X_{it} is $r << N$, so that substantial dimension reduction can be achieved by considering the regression

$$rx_{t+1}^{(n)} = \alpha' F_t + \beta' Z_t + \varepsilon_t$$

where $F_t \subset f_t$ because the factors that are pervasive for the panel of data X_{it} need not be important for predicting $rx_{t+1}^{(n)}$.

The common factors f_t are estimated by $\hat{f}_t$ with PCA. To determine the composition of $\hat{F}_t$, form different subsets of $\hat{f}_t$ and/or functions of $\hat{f}_t$ (such as $\hat{f}_{1t}^2$), run regression of $rx_{t+1}^{(n)}$ on $\hat{F}_t$ and Z_t, and evaluate the corresponding Bayesian information criterion (BIC) and $\bar{R}^2$. Following Stock and Watson (2002b), minimizing the BIC yields the preferred set of factors $\hat{F}_t$.

The above method is used to study the predictability of excess returns on U.S. government bonds. The empirical results show that there is a strong predictable variation in excess bond returns that is associated with macroeconomic activity. The predictive power of the estimated factors is not only statistically significant but also economically important, with factors explaining between 21% and 26% of one-year-ahead excess bond returns. The selected factors also exhibit stable and strongly statistically significant out-of-sample forecasting power for future returns. In addition, risk premia are found to be substantially higher in recessions when the macroeconomic factors are added to the information already contained in current bond market data.

3.6.2 Qiu et al.'s (2022) panel HAR model to forecast volatility

Qiu et al. (2022) propose a heterogenous autoregressive panel (HARP) model with error cross-sectional dependence to control for possibly unobserved common effects in volatility across a class of assets. Let y_{it} be the RV of the i-th individual asset at time t. Consider the h-period direct forecasting panel data model for y_{it} :

$$y_{i,t+h} = \alpha_i' d_t + \sum_{l \in \mathcal{L}} \phi_i^{(l)} \bar{y}_{it}^{(l)} + \beta_i' X_{it} + u_{i,t+h} \tag{3.6.1}$$

$i = 1, \dots, N$, $t = 1, \dots, T$, where $\bar{y}_{it}^{(l)} = l^{-1} \sum_{s=0}^{l} y_{i,t-l}$, d_t is a vector of the observed common effects that may include the constant, the deterministic trends or seasonal dummies, X_{it} is a $k \times 1$ vector of regressors specific to cross-sectional unit i at time t, and α_i and β_i are parameter vectors specific for the h-period forecasting model. Note that the HAR component $\bar{y}_{it}^{(l)}$ is the average of previous l periods of y_{it}, and $\phi_i^{(l)}$ is the corresponding coefficient. $\mathcal{L}$ is the set of lag index vectors of l. Furthermore, assume that the error term, u_{it}, comprises m unobserved common factors:

$$u_{it} = \lambda_i' F_t + \varepsilon_{it} \tag{3.6.2}$$

where F_t is the $m \times 1$ vector of the unobserved common factors, λ_i is the $m \times 1$ vector of the factor loadings, and ε_{it} are the idiosyncratic errors. With the factor structure error, the RV of individual stocks is correlated beyond what can be explained by the observed determinants. Finally, F_t can be modeled by a VAR or by a more general relationship:

$$F_{t+h} = \Phi F_t + \zeta_{t+h}. \tag{3.6.3}$$

Given the model setup in (3.6.1)-(3.6.3), they are interested in forecasting y_{T+h} at the Tth period. Plugging (3.6.2) into (3.6.1) leads to

$$y_{i,t+h} = \alpha_i' d_t + \sum_{l \in \mathcal{L}} \phi_i^{(l)} \bar{y}_{it}^{(l)} + \beta_i' X_{it} + \lambda_i' F_{t+h} + \varepsilon_{i,t+h}$$

and the optimal forecast for $y_{i,T+h}$ is given by

$$y_{i,T+h|T} = \alpha_i' d_T + \sum_{l \in \mathcal{L}} \phi_i^{(l)} \bar{y}_{iT}^{(l)} + \beta_i' X_{iT} + \lambda_i' F_{T+h|T}.$$

Clearly, an estimator for $y_{i,T+h|T}$ can be constructed as

$$\hat{y}_{i,T+h|T} = \hat{\alpha}_i' d_T + \sum_{l \in \mathcal{L}} \hat{\phi}_i^{(l)} \bar{y}_{iT}^{(l)} + \hat{\beta}_i' X_{iT} + \hat{\lambda}_i' \hat{F}_{T+h|T}$$

where each estimated quantity is obtained as follows.

(i) Obtain $\hat{\alpha}_i$, $\hat{\phi}_i^{(l)}$'s, and $\hat{\beta}_i$ based on the dynamic CCE in Chudik and Pesaran (2015a).

(ii) Given the CCE estimators, define

$$\hat{u}_{it} = y_{it} - \hat{\alpha}_i' d_{t-h} + \sum_{l \in \mathcal{L}} \hat{\phi}_i^{(l)} \bar{y}_{i,t-h}^{(l)} + \hat{\beta}_i' X_{i,t-h}, t = h+1, \dots T,$$

which has an approximate factor structure. By applying the PCA method to $\{\hat{u}_{it}\}$ to obtain the estimators, $\hat{\lambda}_i$ for $i = 1, \dots, N,$, and $\hat{F}_t$ for $t = h+1, \dots, T$.

(iii) Based on $\left\{\hat{F}_t\right\}_{t=h+1}^{T}$, obtain the usual least-squares estimate $\hat{\Phi}$ for Φ and the estimate for $F_{T+h|T}$:

$$\hat{F}_{T+h|T} = \hat{\Phi} \hat{F}_T.$$

They use the model in (3.6.1)-(3.6.3) to study the linkage among the realized volatilities of component stocks and find that the linkage is important to forecast the relevant index volatility. Empirical studies show that the RV models that exploit the linkage effects lead to significantly better out-of-sample forecast performance, for example, an up to 32% increase in the pseudo R^2.

3.7 Concluding Remarks

In this chapter, we give a selective review of FARs and their applications to financial markets. For the estimation of factor models, we focus on the PCA method. We emphasize that many important topics are not covered in the chapter since too many breakthrough methodologies and applications have been developed for factor models in the literature. For example, the factor structure is used in the multiple testing, model selection, and robust estimation of high dimensional models to control the correlation among regressors or errors. More importantly, many learning theories and methods on the low-rank structure of high dimensional models have refreshed the modern understanding of econometric modelling in the last ten years. The low-rank structure is one of the key features of factor models. It provides a new perspective and opportunities to study factor models and leads to many new discoveries and understanding. For a more comprehensive account on this topic, see Chapters 9-11 of Fan et al. (2020).

Bibliography

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Ahn, S. C., Y. H. Lee, and P. Schmidt (2013). Panel data models with multiple time-varying individual effects. *Journal of econometrics 174*(1), 1–14.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters 80*(23-24), 1806–1813.

Ando, T. and J. Bai (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association 115*(529), 266–279.

Ando, T., J. Bai, M. Nishimura, and J. Yu (2019). A quantile-based asset pricing model. *working paper*.

Ando, T. and R. S. Tsay (2011). Quantile regression models with factor-augmented predictors and information criterion. *The Econometrics Journal 14*(1), 1–24.

Ando, T. and R. S. Tsay (2014). A predictive approach for selection of diffusion index models. *Econometric Reviews 33*(1-4), 68–99.

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

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J., K. Li, and L. Lu (2016). Estimation and inference of favar models. *Journal of Business & 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.

Bai, J. and S. Ng (2008). *Large dimensional factor analysis*. Now Publishers Inc.

Bai, J. and S. Ng (2009). Boosting diffusion indices. *Journal of Applied Econometrics 24*(4), 607–629.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of econometrics 176*(1), 18–29.

Bai, J. and P. Wang (2016). Econometric analysis of large factor models. *Annual Review of Economics 8*, 53–80.

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

Belloni, A., M. Chen, O. H. M. Padilla, et al. (2019). High dimensional latent panel quantile regression with an application to asset pricing. *Working paper*.

Bernanke, B., 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.

Bernanke, B. S. and J. Boivin (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics 50*(3), 525–546.

Boivin, J. and S. Ng (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking 1*(3), 117 – 152.

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

Carrasco, M. and B. Rossi (2016). In-sample inference and forecasting in misspecified factor models. *Journal of Business & Economic Statistics 34*(3), 313–338.

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*(2), 280–293.

Chudik, A. and M. H. Pesaran (2015a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics 188*(2), 393–420.

Chudik, A. and M. H. Pesaran (2015b). Large panel data models with cross-sectional dependence: a survey. Technical report, Oxford University Press.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal 1*(14), C45–C90.

Clements, M. P. and D. F. Hendry (1996). Intercept corrections and structural change. *Journal of Applied Econometrics 11*(5), 475–494.

Corradi, V. and N. R. Swanson (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics 182*(1), 100–118.

Djogbenou, A. A. (2021). Model selection in factor-augmented regressions with estimated factors. *Econometric Reviews 40*(5), 470–503.

Fan, J., K. Li, and Y. Liao (2021). Recent developments in factor models and applications in econometric learning. *Annual Review of Financial Economics 13*, 401–430.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association 96*(456), 1348–1360.

Fan, J., R. Li, C.-H. Zhang, and H. Zou (2020). *Statistical foundations of data science*. Chapman and Hall/CRC.

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.

Forni, M. and L. Reichlin (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies 65*(3), 453–473.

Fosten, J. (2017). Model selection with estimated factors and idiosyncratic components. *Journal of Applied Econometrics 32*(6), 1087–1106.

Giannone, D., M. Lenza, and L. Reichlin (2010). Business cycles in the euro area. In A. Alesina and F. Giavazzi (Eds.), *Europe and the Euro*, pp. 141–167. Cambridge, MA: National Bureau of Economic Research: University of Chicago Press.

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

Gorman, W. M. (1981). *Some engel curves*. Cambridge University Press.

Greenaway-McGrevy, R., C. Han, and D. Sul (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics 169*(1), 48–53.

Greenaway-McGrevy, R., N. C. Mark, D. Sul, and J.-L. Wu (2018). Identifying exchange rate common factors. *International Economic Review 59*(4), 2193–2218.

Gregory, A. W. and A. C. Head (1999). Common and country-specific fluctuations in productivity, investment, and the current account. *Journal of Monetary Economics 44*(3), 423–451.

Groen, J. J. and G. Kapetanios (2013). Model selection criteria for factor-augmented regressions. *Oxford bulletin of economics and statistics 75*(1), 37–63.

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics 196*(1), 55–67.

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

Kapetanios, G. and M. H. Pesaran (2007). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In G. Phillips and E. Tzavalis (Eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, pp. 239 – 281. Cambridge: Cambridge University Press.

Karabiyik, H., F. C. Palm, and J.-P. Urbain (2019). Econometric analysis of panel data models with multifactor error structures. *Annual Review of Economics 11*, 495–522.

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

Koenker, R. and G. Bassett Jr (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Li, D., J. Tosasukul, and W. Zhang (2020). Nonlinear factor-augmented predictive regression models with functional coefficients. *Journal of Time Series Analysis 41*(3), 367–386.

Li, H., Q. Li, and Y. Shi (2017). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics 197*(1), 76–86.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

Ma, S., O. Linton, and J. Gao (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics 222*(1), 295–323.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics 197*(1), 101–129.

Massacci, D. (2019). Unstable diffusion indexes: With an application to bond risk premia. *Oxford Bulletin of Economics and Statistics 81*(6), 1376–1400.

Ohno, S. and T. Ando (2018). Stock return predictability: A factor-augmented predictive regression system with shrinkage method. *Econometric Reviews 37*(1), 29–60.

Onatski, A. (2009). A formal statistical test for the number of factors in the approximate factor models. *Econometrica 77*(5), 1447–1480.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics 168*(2), 244–258.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.

Phella, A. (2020). Forecasting with factor-augmented quantile autoregressions: A model averaging approach. *arXiv preprint arXiv:2010.12263*.

Qiu, Y., T. Xie, J. Yu, and Q. Zhou (2022). Forecasting equity index volatility by measuring the linkage among component stocks. *Journal of Financial Econometrics 20*(1), 160–186.

Reese, S. and J. Westerlund (2018). Estimation of factor-augmented panel regressions with weakly influential factors. *Econometric Reviews 37*(5), 401–465.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Rossi, B. and A. Inoue (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics 30*(3), 432–453.

Shao, J. (1996). Bootstrap model selection. *Journal of the American statistical Association 91*(434), 655–665.

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 & Economic Statistics 20*(2), 147–162.

Stock, J. H. and M. W. Watson (2012). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*. USA: Oxford University Press.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, Volume 2, pp. 415–525. Elsevier.

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

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics 198*(1), 84–101.

Swanson, N. R. and W. Xiong (2018). Big data analytics in economics: What have we learned so far, and where should we go from here? *Canadian Journal of Economics/Revue canadienne d'économique 51*(3), 695–746.

Wang, S., G. Cui, and K. Li (2015). Factor-augmented regression models with structural change. *Economics Letters 130*, 124–127.

Wei, J. and Y. Zhang (2020). A time-varying diffusion index forecasting model. *Economics Letters 193*, 109337.

Westerlund, J. and J.-P. Urbain (2015). Cross-sectional averages versus principal components. *Journal of Econometrics 185*(2), 372–377.

Wu, Y. and Y. Liu (2009). Variable selection in quantile regression. *Statistica Sinica 19*(2), 801–817.

Xu, Q.-H., Z.-W. Cai, and Y. Fang (2016). Panel data models with cross-sectional dependence: A selective review. *Applied Mathematics-A Journal of Chinese Universities 31*(2), 127–147.

Yan, Y. and T. Cheng (2022). Factor-augmented forecasting regressions with threshold effects. *The Econometrics Journal 25*(1), 134–154.