

Periodic dynamic factor models: estimation approaches and applications*

Changryong Baek[†]

Dept. of Statistics, Sungkyunkwan University, Korea
e-mail: crbaek@skku.edu

Richard A. Davis

Dept. of Statistics, Columbia University, USA
e-mail: rdavis@stat.columbia.edu

Vladas Pipiras[‡]

Dept. of Statistics and Operations Research, UNC at Chapel Hill, USA
e-mail: pipiras@email.unc.edu

Abstract: A periodic dynamic factor model (PDFM) is introduced as a dynamic factor modeling approach to multivariate time series data exhibiting cyclical behavior and, in particular, periodic dependence structure. In the PDFM, the loading matrices are allowed to depend on the “season” and the factors are assumed to follow a periodic vector autoregressive (PVAR) model. Estimation of the loading matrices and the underlying PVAR model is studied. A simulation study is presented to assess the performance of the introduced estimation procedures, and applications to several real data sets are provided.

MSC 2010 subject classifications: Primary 62M10, 62H12; secondary 62H20.

Keywords and phrases: Dynamic factor model, periodic vector autoregressive (PVAR) model, dimension reduction, adaptive lasso.

Received April 2018.

Contents

1	Introduction	4378
2	Estimation of PDFMs	4381
2.1	Exact model formulation and EM estimation	4381
2.2	Approximate model formulation and PCA estimation	4382
2.3	Asymptotic results for approximate model formulation	4384
2.4	Other approaches based on sparsity and DFMs	4387
2.5	Varying number of factors	4389

*We would like to thank the Associate Editor and anonymous Referees for careful reading of the paper and many useful comments and suggestions.

[†]The work of the first author was supported in part by the Basic Science Research Program from the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (NRF-2017R1A1A1A05000831).

[‡]The third author was supported in part by the NSF grant DMS-1712966.

3	Simulation study	4390
3.1	Approximate model formulation with PCA	4390
3.2	Exact model formulation with EM	4394
4	Applications	4395
4.1	Fraser river water flow	4395
4.2	US quarterly macroeconomic data	4401
5	Conclusions	4402
A	Proofs for approximate model formulation	4404
B	Adaptive lasso in Fourier-PVAR	4408
	References	4409

1. Introduction

Dynamic factor modeling of multivariate (vector) time series is a common and popular approach, especially when the dimension of a time series is large, possibly larger than the sample size. See, for example, Bai and Ng (2008), Stock and Watson (2011), Doz, Giannone and Reichlin (2011, 2012). A dynamic factor model (DFM, for short) of a q -vector time series $\{X_n\}_{n \in \mathbb{Z}}$ is expressed as

$$X_n - \mu = \Lambda_0 F_n + \varepsilon_n, \quad (1.1)$$

where μ is the mean of X_n , Λ_0 is a loading matrix of dimension $q \times r$, $\{F_n\}$ is a r -vector time series consisting of r factors with r typically much smaller than q and $\{\varepsilon_n\}$ is a q -vector time series uncorrelated with $\{F_n\}$, sometimes assumed to be white noise (i.e. $\{\varepsilon_n\} \sim \text{WN}(0, \Sigma_\varepsilon)$) though it can also be correlated in time. The factors $\{F_n\}$ are commonly assumed to follow a vector autoregressive model of order p (VAR(p) model),

$$\Phi(B)F_n = \zeta_n, \quad (1.2)$$

where $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$ is a VAR matrix polynomial with the usual backward shift operator B , each matrix Φ_i of dimension $r \times r$, and $\{\zeta_n\} \sim \text{WN}(0, \Sigma_\zeta)$. In fact, the DFM (1.1)–(1.2) is also called static; more generally, DFM is defined by replacing Λ_0 in (1.1) by a loading matrix polynomial $\Lambda(B)$. Depending on the assumptions and the estimation method used, additional restrictions on the model parameters are also made (e.g. Λ_0 is lower triangular, Σ_ε is diagonal and $\Sigma_\zeta = I_r$, the $r \times r$ identity matrix). The model (1.2) is also commonly assumed to be strictly stationary and causal, though nonstationary DFMs have also been considered.

In this work, we are interested broadly in dynamic factor models for multivariate series exhibiting cyclical behavior with period s . We use the term “cyclical behavior” loosely and rather mean by it a collection of associated models. Two main modeling approaches for univariate series exhibiting cyclical behavior are based on either (multiplicative) seasonal ARIMA models (e.g. Brockwell and Davis (2009), Ghysels and Osborn (2001)) or periodic ARMA models (e.g. Franses and Paap (2004), Hurd and Miamee (2007)). Dynamic factor modeling based on seasonal vector ARIMA models has been studied by several authors.

Alonso, García-Martos, Rodríguez and Sánchez (2011) assume that common factors follow a seasonal VARIMA(p, d, q) $\times$ (P, D, Q) $_s$ model

$$(I - B)^d(I - B^s)^D\Phi(B)\Phi_s(B^s)F_n = c + \Theta(B)\Theta_s(B^s)\zeta_n, \quad (1.3)$$

where s is the period, $\Phi(B)$ is as in (1.2), $\Phi_s(B^s) = I - \Phi_{s,1}B^s - \dots - \Phi_{s,P}B^{sP}$ is a seasonal VAR matrix polynomial, $\Theta(B)$ and $\Theta_s(B^s)$ are, similarly, MA matrix polynomials of respective orders q and Q ,¹ and d, D are non-negative integers (that allow for factors to be non-stationary when $d \geq 1$ or $D \geq 1$). Using a state space formulation of the model (1.1)–(1.2), and imposing flexible restrictions on model parameters, the authors developed an EM algorithm for model estimation. Furthermore, bootstrap strategies were proposed to quantify statistical uncertainty associated with parameter estimates and forecasts. Nieto, Pena and Saboyá (2016) proposed a similar model but identifying separately the factors that are (i) non-stationary and non-seasonal, (ii) (non-stationary) seasonal, and (iii) stationary. A notable procedure to estimate the dimension of the three different types of factors was proposed by the authors, based solely on suitable covariance matrices.

In contrast to seasonal models, there has seemingly been no work on dynamic factor modeling of multivariate series for which periodic models are better or more natural alternatives than the seasonal models above. The goal of this paper is to introduce and to study several approaches to such periodic dynamic factor modeling. The most general form might be to assume, in the “stationary” and “static” form, that

$$X_n - \mu_m = \Lambda_m F_n + \varepsilon_n, \quad (1.4)$$

where $m = m(n) = 1, \dots, s$, refers to the season that n belongs to, the $q \times r$ loading factors Λ_m possibly depend on the season m , $\{\varepsilon_n\}$ is a WN series whose covariance matrix $\mathbb{E}\varepsilon_n\varepsilon_n' = \Sigma_{\varepsilon,m}$ possibly depends on the season m (denoted $\{\varepsilon_n\} \sim \text{WN}(0, \Sigma_{\varepsilon,m})$) and the $r \times 1$ factors $\{F_n\}$, which are independent of $\{\varepsilon_n\}$, are assumed to follow a periodic VAR model of order p (PVAR(p), for short),

$$\Phi_m(B)F_n = \zeta_n \quad (1.5)$$

with the VAR matrix polynomial $\Phi_m(B) = I - \Phi_{m,1}B - \dots - \Phi_{m,p}B^p$ depending on the season m and $\{\zeta_n\} \sim \text{WN}(0, \Sigma_{\zeta,m})$. (The order p may depend on the season as well but we consider the case of a single p for simplicity.) We shall refer to (1.4)–(1.5) as a *periodic dynamic factor model* (PDFM). As with (1.2), we also assume that (1.5) is in the stationary and causal regime.

We note that as defined above, the PDFM is quite general. It includes several potentially interesting special cases including:

- periodic Λ_m , $\Sigma_{\varepsilon,m}$, non-periodic stationary F_n ;
- constant $\Lambda_m = \Lambda$, $\Sigma_{\varepsilon,m} = \Sigma_{\varepsilon}$, periodic F_n ;
- periodic $\Sigma_{\varepsilon,m}$, constant $\Lambda_m = \Lambda$, non-periodic stationary F_n ;

¹ The MA orders q and Q are used only in this part of the paper and should not be confused with the dimension of the series X .

and others (but see also Remark 2.1 below). In all these special and general cases, the component series of the PDFM (1.4)–(1.5) are periodically correlated in the sense that the mean $\mu(n) = \mathbb{E}X_n$ and the autocovariance function $\gamma(n_1, n_2) = \text{Cov}(X_{n_1}, X_{n_2})$ are periodic with period s , that is,

$$\mu(n) = \mu(n + s), \quad \gamma(n_1, n_2) = \gamma(n_1 + s, n_2 + s), \quad n, n_1, n_2 \in \mathbb{Z}. \quad (1.7)$$

Hence, PDFMs should be used primarily when periodic correlations are observed for univariate component series. This will be illustrated on real time series below.

In this work, we shall study estimation methods for the PDFM (1.4)–(1.5) and some of its special cases as in (1.6). Following the approaches taken for DFM in the literature, we shall consider two estimation frameworks for the so-called exact and approximate model formulations. (The distinction is made and discussed nicely in Doz et al. (2012).) For the exact DFM formulation, it is assumed for identification purposes that Σ_ζ is the identity matrix and Λ_0 is lower triangular, and without sacrificing much model flexibility, it is further assumed that Σ_ε is diagonal (see, e.g. Peña and Poncela (2006), Alonso et al. (2011)). Sometimes, the polynomials $\Phi(B)$ are assumed to be diagonal as well (e.g. Nieto et al. (2016)). Other identifiable model formulations are also used (e.g. Bai and Wang (2015)). For the exact model formulation, estimation is typically carried out using the EM algorithm. We shall also make similar assumptions on our PDFM in the exact case and employ the EM algorithm. The exact model though is usually fitted for smaller dimensions q , which we shall also do here.

In higher dimension q , assuming that both q and T , the sample size, converge to infinity, it has been well recognized that weaker assumptions on the model factor structure can be imposed, under the term “approximate DFM”, which lead to consistency results subject to properly rotated factors. See Doz et al. (2011, 2012), Bai and Ng (2007), Forni, Hallin, Lippi and Reichlin (2000). Moreover, in this case, the model parameters in approximate DFMs can be estimated using principal components, though the effects of additional iteration steps involving smoothing as in the EM algorithm have also been studied. In this work, we shall extend the approximate DFM formulation and convergence results to PDFMs, relying heavily on the work by Doz et al. (2011, 2012). An additional important feature of this work is that we shall also be using sparse models on multiple occasions. This should come as no surprise since allowing DFM coefficients to depend on the season increases their number considerably.

The rest of the paper is organized as follows. Estimation of PDFMs is considered in Section 2, including the statements of the main results of this work. A competing approach where standard DFM is applied to the multivariate series obtained by “stacking” X_n ’s across the s seasons, is also discussed in Section 2. A simulation study is presented in Section 3, and applications to several real data sets can be found in Section 4. Section 5 concludes while Appendix A includes the proofs of the results mentioned in Section 2, and Appendix B discusses the adaptive lasso estimation procedure for the so-called Fourier PVAR model.

2. Estimation of PDFMs

We present here estimation methods for PDFMs in the two formulations, exact (Section 2.1) and approximate (Section 2.2), as discussed in the context of DFMs in Section 1. Some asymptotic results for the approximate formulation are also given (Section 2.3) along with estimation methods based on sparsity and the Fourier representation (Section 2.4). The case of a different number of factors across seasons in (1.4) is also discussed (Section 2.5).

2.1. Exact model formulation and EM estimation

In the exact formation, we assume that the PDFM parameters in (1.4)–(1.5) satisfy:

$$\Lambda_m = \begin{pmatrix} I_r \\ B_m \end{pmatrix}, \quad (2.1)$$

where B_m has dimension $(q - r) \times r$. For the usual DFMs, the identifiability of model parameters under the assumption analogous to (2.1) was proved in Bai and Wang (2015), and their arguments extend straightforwardly to PDFMs under (2.1). In fact, a basic reason for (2.1) leading to identifiability is that essentially any $q \times r$ loading matrix Λ_m can be reduced to the form (2.1) as

$$\Lambda_m C_m = \begin{pmatrix} \Lambda_m^{(1)} \\ \Lambda_m^{(2)} \end{pmatrix} C_m = \begin{pmatrix} \Lambda_m^{(1)} C_m \\ \Lambda_m^{(2)} C_m \end{pmatrix} = \begin{pmatrix} I_r \\ B_m \end{pmatrix}, \quad (2.2)$$

where $\Lambda_m^{(1)}$ and $\Lambda_m^{(2)}$ are $r \times r$ and $(q - r) \times r$, respectively, and the $r \times r$ matrix C_m is such that $\Lambda_m^{(1)} C_m = I_r$ (which requires that $\Lambda_m^{(1)}$ is non-singular). Since $\Lambda_m F_n = \Lambda_m C_m C_m^{-1} F_n$, the use of C_m to achieve the identifiable form (2.2) just means that the factors F_n similarly need to be adjusted through C_m^{-1} .

As in the usual exact DFMs, the parameters of the models (1.4)–(1.5) can be estimated with the EM algorithm. For this purpose and e.g. when $p = 1$, we have adapted here the arguments in the R code for the EM algorithm found in Shumway and Stoffer (2006), pp. 342–344, with the following changes: in the M-step of the algorithm, we also need to minimize with respect to the loading matrix (A_t in Shumway and Stoffer (2006)); we allow both the loading matrix A_t and the PVAR matrix Φ_t to depend on the season, and, finally, in the M-step, we also work and perform minimization under the parametrization (2.1) of the loadings. The R code implementing this will be available online upon publication of the article. We also note that the EM algorithm requires initial values of the model parameters. For these, we use the values obtained from the principal component approach discussed in Section 2.2 below, after transforming the latter to the used identifiable parametrization through (2.2). As with all likelihood calculations, the EM algorithm can also yield confidence intervals for the parameters of interest, through inverting the numeric Hessian.

2.2. Approximate model formulation and PCA estimation

We discuss here estimation of factors and the underlying PVAR model in the PDFM defined by (1.4)–(1.5) in its approximate formulation. As noted in Section 1, the latter refers to model assumptions and the fact that large q is considered.

One of the basic observations is that PDFM can be viewed as consisting of s separate DFMs. That is, let $X_t^{(m)} = X_{m+(t-1)s}$, where s is the period and $m = 1, \dots, s$ is one of the seasons. Define similarly $\varepsilon_t^{(m)}$, $F_t^{(m)}$ and $\zeta_t^{(m)}$. Then, (1.4) is equivalent to s “static” DFMs: for $m = 1, \dots, s$,

$$X_t^{(m)} - \mu_m = \Lambda_m F_t^{(m)} + \varepsilon_t^{(m)}. \quad (2.3)$$

The factors $\{F_t^{(m)}\}$ can then be estimated as in the usual DFM. The exact procedure is adapted from Doz et al. (2011), and is recalled below. Estimation of the underlying PVAR model in (1.5), which is also discussed below, requires a more delicate extension. As in Doz et al. (2011), we assume without loss of generality that $\mathbb{E}F_t^{(m)}(F_t^{(m)})' = I_r$, where I_r is the $r \times r$ identity matrix.

Suppose that we observe a q -vector time series $X_1, \dots, X_N$ with the sample size

$$N = Ts, \quad (2.4)$$

so that, for simplicity, each season has exactly T observations (or, equivalently, there are exactly T cycles of length s in the sample of size N). Dynamic factors $F_t^{(m)}$, loading matrices Λ_m and PVAR matrices $\Phi_{m,i}$ and $\Sigma_{\zeta,m}$ are estimated through the following steps:

1. Deseasonalize the series $X_1^{(m)}, \dots, X_T^{(m)}$, $m = 1, \dots, s$, to obtain

$$Y_t^{(m)} = X_t^{(m)} - \hat{\mu}_m, \quad t = 1, \dots, T,$$

with $\hat{\mu}_m = \frac{1}{T} \sum_{t=1}^T X_t^{(m)}$.

2. For each season $m = 1, \dots, s$, diagonalize the sample covariance matrix $\hat{\Sigma}_Y^{(m)}$ of $Y_t^{(m)}$, $t = 1, \dots, T$, as

$$\hat{\Sigma}_Y^{(m)} = \frac{1}{T} \sum_{t=1}^T Y_t^{(m)}(Y_t^{(m)})' = \hat{U}_m \hat{D}_m \hat{U}_m', \quad (2.5)$$

where $\hat{D}_m = \text{diag}(\hat{d}_{m,1}, \dots, \hat{d}_{m,q})$ with $\hat{d}_{m,1} \geq \dots \geq \hat{d}_{m,q}$ and $\hat{U}_m = (\hat{U}_{m,1}, \dots, \hat{U}_{m,q})$ is the orthogonal eigenvector matrix such that $\hat{U}_m \hat{U}_m' = \hat{U}_m' \hat{U}_m = I_q$ and $\hat{\Sigma}_Y^{(m)} \hat{U}_{m,i} = \hat{d}_{m,i} \hat{U}_{m,i}$.

3. The principal component estimators of $F_t^{(m)}$ and Λ_m for season $m = 1, \dots, s$, are given by

$$\hat{F}_t^{(m)} = \frac{1}{q} \hat{\Lambda}_m' Y_t^{(m)}, \quad \hat{\Lambda}_m = \sqrt{q}(\hat{U}_1, \dots, \hat{U}_r), \quad t = 1, \dots, T, \quad (2.6)$$

where r is the number of factors used.

4. Use the least squares method to estimate the parameters of the PVAR(p) model in (1.5) based on $\hat{F}_n$, $n = 1, \dots, Ts$, to obtain the estimates $\hat{\Phi}_{m,1}, \dots, \hat{\Phi}_{m,p}$ and $\hat{\Sigma}_{\zeta,m}$, $m = 1, \dots, s$. The covariance matrix $\Sigma_{\varepsilon,m} = \mathbb{E}\varepsilon_t^{(m)}(\varepsilon_t^{(m)})'$ is estimated as $\hat{\Sigma}_{\varepsilon,m} = \hat{\Sigma}_Y^{(m)} - \hat{\Lambda}_m \hat{\Lambda}_m'$.
5. Run one iteration of the Kalman smoother based on (1.4)–(1.5) by replacing μ_m , Λ_m , $\Sigma_{\varepsilon,m}$, $\Phi_{m,1}, \dots, \Phi_{m,p}$ and $\Sigma_{\zeta,m}$ by $\hat{\mu}_m$, $\hat{\Lambda}_m$, $\text{diag}(\hat{\Sigma}_Y^{(m)} - \hat{\Lambda}_m \hat{\Lambda}_m')$, $\hat{\Phi}_{m,1}, \dots, \hat{\Phi}_{m,p}$ and $\hat{\Sigma}_{\zeta,m}$, respectively. The smoother yields the final estimate $\tilde{F}_n$ of the factors.

Several comments regarding the steps above are in order. The estimators (2.6) are defined as they appear in the literature of principal component analysis, e.g., Bai and Ng (2008), Stock and Watson (2011). The principal component estimators in (2.6) of Step 3 are the solutions $F_t^{(m)}$ and Λ_m minimizing the sum of squares

$$\text{SSE}^{(m)}(r) = \frac{1}{qT} \sum_{t=1}^T \left(Y_t^{(m)} - \Lambda_m F_t^{(m)} \right)' \left(Y_t^{(m)} - \Lambda_m F_t^{(m)} \right), \quad (2.7)$$

subject to the normalization $q^{-1} \Lambda_m' \Lambda_m = I_r$. The last step, Step 5, follows Doz et al. (2011). We also note that iterating Steps 4 and 5 is possible and is closely related to employing the EM algorithm. That is, based on the final estimate $\tilde{F}_n$ of the factors, we can get new estimates of PVAR(p) parameters and minimizing (2.7) for given $\tilde{F}_n$ subject to normalization on loadings gives updated estimates of loadings, and so on. This is essentially the same as running the EM algorithm of Section 2.1, as used by Engle and Watson (1981) in the case of DFMs. Doz et al. (2011) also mention iteration procedures but their simulation study shows that just one iteration performs reasonably well in practice. Our simulation study comparing the exact and approximate model approaches in Section 3.2 also shows that just one iteration seems to work well in practice. The expression in (2.5) follows the convention used by Stock and Watson (2011).

Finally, we should also note that the “decoupling” of the original PDFM (1.4)–(1.5) into s separate DFMs should not be too surprising. For example, estimation of time-varying loadings by means of a time-localizing kernel function can be found in Su and Wang (2017). As noted above, it is a bit more delicate to show that the estimated factors can be “pieced” back together into a PVAR model (1.5), assuming that the latter holds at the PDFM level.

Remark 2.1. *Related to the last point, we also make the following important, even if seemingly trivial, observation. The factors F_n in (1.4)–(1.5) can be identified up to rotations only, which further can depend on m . In this regard, note that the PVAR model (1.5) is “closed” under the multiplication of F_n by such rotation matrices, in the sense that this would only change the parameters of the PVAR model. Indeed, with rotations C_m and e.g. when $p = 1$, we have that*

$$\begin{aligned} C_{m(n)} F_n &= \left(C_{m(n)} \Phi_{m,1} C_{m(n-1)}^{-1} \right) C_{m(n-1)} F_{n-1} + C_{m(n)} \zeta_n \\ &=: \tilde{\Phi}_{m,1} (C_{m(n-1)} F_{n-1}) + \tilde{\zeta}_n. \end{aligned}$$

Thus, the new rotated factor process $C_{m(n)}F_n$ is necessarily a PVAR. This precludes the first case in (1.6) as a possibility in the model specification (1.4)–(1.5), and this case should not be considered.

2.3. Asymptotic results for approximate model formulation

The results proved in Doz et al. (2011) can also be used directly to describe the asymptotics of the estimated factors and the PVAR model parameters in the approximate model formulation. This shall require $q, T \rightarrow \infty$; otherwise, no consistency can be expected. We shall suppose that the number of factors r and the underlying PVAR model order p are known – their estimation is discussed below. For identifiability purposes, we also need to fix a rotation up to which the factors and the PVAR model parameters will be identified. Thus, for $m = 1, \dots, s$, let

$$\Lambda'_m \Lambda_m = Q_m D_m Q'_m \quad (2.8)$$

be the eigenvalue decomposition of $\Lambda'_m \Lambda_m$, where D_m is a diagonal matrix consisting of the eigenvalues of $\Lambda'_m \Lambda_m$ and Q_m is the orthogonal matrix consisting of the corresponding eigenvectors. Set

$$G_t^{(m)} = Q'_m F_t^{(m)}, \quad \Pi_m = \Lambda_m Q_m, \quad \Psi_m(B) = Q'_m \Phi_m(B) Q_m, \quad (2.9)$$

so that the PDFM can be expressed as

$$X_t^{(m)} - \mu_m = \Pi_m G_t^{(m)} + \varepsilon_t^{(m)}, \quad (2.10)$$

with the underlying PVAR(p) model for $\{G_t^{(m)}\}$ given by

$$\Psi_n(B) G_n = \eta_{m,n}, \quad (2.11)$$

where $G_{m+(t-1)s} = G_t^{(m)}$ and $\{\eta_{m,n}\} \sim \text{WN}(0, \Sigma_{\eta,m})$ with $\Sigma_{\eta,m} = Q'_m \Sigma_{\zeta,m} Q_m$. The key feature of the model (2.10) is that $\Pi'_m \Pi_m$ is now diagonal (and $\mathbb{E} G_t^{(m)} (G_t^{(m)})' = I_r$ holds as for $F_t^{(m)}$). Let also $\hat{D}_{m,r} = \text{diag}(\hat{d}_{m,1}, \dots, \hat{d}_{m,r})$ and $\hat{P}_{m,r} = (\hat{U}_{m,1}, \dots, \hat{U}_{m,r})$ with $\hat{d}_{m,i}$ and $\hat{U}_{m,i}$ defined following (2.5) and set

$$\hat{G}_t^{(m)} = \hat{D}_{m,r}^{-1/2} \hat{P}_{m,r}' Y_t^{(m)}, \quad \hat{\Pi}_m = \hat{P}_{m,r} \hat{D}_{m,r}^{1/2}. \quad (2.12)$$

The estimates of $\Psi_m(B)$ based on $\hat{G}_n$ will be denoted $\hat{\Psi}_m(B)$. Note that $\hat{F}_t^{(m)}$ and $\hat{\Lambda}_m$ in (2.6) are equal to $\hat{G}_t^{(m)}$ and $\hat{\Pi}_m$ in (2.12) up to the diagonal matrices $\hat{D}_{m,r}^{1/2}/q^{1/2}$ and $\hat{D}_{m,r}^{-1/2} q^{1/2}$, respectively, that is,

$$\hat{F}_t^{(m)} = \frac{\hat{D}_{m,r}^{1/2}}{q^{1/2}} \hat{G}_t^{(m)}, \quad \hat{\Lambda}_m = \hat{\Pi}_m \hat{D}_{m,r}^{-1/2} q^{1/2}. \quad (2.13)$$

As in Step 5 above, we let $\tilde{G}_n$ be the smoothed estimate of $\hat{G}_n$ obtained through the Kalman smoother.

The following assumptions will be made for the PDFM (1.4)–(1.5) following Doz et al. (2011).

- (A1) For any season $m = 1, \dots, s$, $\{X_t^{(m)}\} = \{(X_{1t}^{(m)}, \dots, X_{qt}^{(m)})'\}$ is a stationary time series with finite variance.
- (A2) For each $m = 1, \dots, s$, there is $M > 0$ such that $\text{Var}(X_{it}^{(m)}) \leq M$ for all i and t .
- (A3) For all $m_1, m_2 \in \{1, \dots, s\}$, $\{F_t^{(m_1)}\}$ and $\{\varepsilon_t^{(m_2)}\}$ are independent time series. Furthermore, $\{F_t^{(m)}\}$ has a Wold representation:

$$F_t^{(m)} = \sum_{k=0}^{\infty} C_k^{(m)} Z_{t-k}^{(m)}$$

with $\sum_{k=0}^{\infty} \|C_k^{(m)}\| < +\infty$ and $\{Z_t^{(m)}\}$ is stationary up to order four. The series $\{\varepsilon_t^{(m)}\}$ also admits a Wold representation:

$$\varepsilon_t^{(m)} = \sum_{k=0}^{\infty} D_k^{(m)} \nu_{t-k}^{(m)}$$

with $\sum_{k=0}^{\infty} \|D_k^{(m)}\| < +\infty$ and $\{\nu_t^{(m)}\}$ consists of IID random variables such that $\mathbb{E}(\nu_{it}^{(m)})^4 \leq M$ for some $M > 0$.

- (A3') The factors $\{F_n\}$ follow PVAR(p) model:

$$\Phi_m(B)F_n = \zeta_{m,n},$$

where the VAR matrix polynomial $\Phi_m(B) = I - \Phi_{m,1}B - \dots - \Phi_{m,p}B^p$ may depend on the season m of the time point n , and $\{\zeta_{m,n}\} \sim \text{IID}(0, \Sigma_{\zeta,m})$ with finite fourth moment. Furthermore, the corresponding VAR model (2.24) of the series $\tilde{F}_t = \text{vec}(F_t^{(1)}, \dots, F_t^{(s)})$, formed by stacking the factors over one full period, is stationary and causal. (The conditions for the latter assumption are well known in the VAR setting but cannot be expressed succinctly in terms of the matrix polynomials $\Phi_m(z)$, $m = 1, \dots, s$.)

Basically, the assumptions (A1)–(A3) impose a factor structure characterized by stationarity and weak temporal dependence of the common and idiosyncratic components. The summability of the coefficient matrices in the Wold decompositions in (A3), in particular, excludes long-memory series. We note that the moment conditions across component indices i in (A2) and (A3) are assumed to hold irrespective of $q \rightarrow \infty$. The condition (A3') posits a parametric PVAR(p) model of the common factors.

We shall further assume weak correlations on idiosyncratic components and relatively pronounced eigenvalues of the underlying factor structure in the sense of approximate factor model of Chamberlain and Rothschild (1983). Let $\|A\| = (\lambda_{\max}(A'A))^{1/2}$ be the spectral norm of a matrix A with $\lambda_{\max}(A'A)$ being the largest eigenvalue of a matrix $A'A$. Let also $\lambda_{\min}(A'A)$ be the smallest eigenvalue. For any $m = 1, \dots, s$, assume:

$$(\text{CR1}) \quad \liminf_{q \rightarrow \infty} \frac{1}{q} \lambda_{\min}(\Lambda'_m \Lambda_m) > 0.$$

- (CR2) $\limsup_{q \rightarrow \infty} \frac{1}{q} \lambda_{\max}(\Lambda'_m \Lambda_m)$ is finite.
- (CR3) $\limsup_{q \rightarrow \infty} \sum_{h \in \mathbb{Z}} \|\mathbb{E}(\varepsilon_t^{(m)} \varepsilon_{t+h}^{(m)'})\|$ is finite.
- (CR4) $\inf_q \lambda_{\min}(\mathbb{E}(\varepsilon_t^{(m)} \varepsilon_t^{(m)'})) > 0$.
- (A4) $\Lambda'_m \Lambda_m$ has distinct eigenvalues.

Remark 2.2. *The summability of the coefficient matrices in the Wold decompositions in (A3) excludes long-memory series for factors, in particular. Whether long memory could be assumed in large DFM analysis remains an open question. Similarly, the Wold decompositions, especially with i.i.d. innovations, preclude many nonlinear time series of potential interest, for example, heteroscedastic models. As noted above, the assumptions (CR1) and (CR2) correspond to strong factors. What happens for weaker factors but for independent observations across time, is considered in Wang and Fan (2017). It remains to be seen whether their results can be extended to the dynamic context.*

The first result concerns the asymptotics of the estimated quantities in the relation (2.10).

Proposition 2.1. *With the notation above and under the assumptions (CR1)–(CR4), (A1)–(A4), (A3'), we have: for each $m = 1, \dots, s$, and fixed t ,*

$$\widehat{G}_t^{(m)} - G_t^{(m)} = O_p\left(\frac{1}{\sqrt{q}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \quad (2.14)$$

$$\widehat{\Pi}_m - \Pi_m = O_p\left(\frac{1}{\sqrt{q}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \quad (2.15)$$

$$\widehat{\Sigma}_{\varepsilon, m} - \Sigma_{\varepsilon, m} = O_p\left(\frac{1}{\sqrt{q}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \quad (2.16)$$

as $q, T \rightarrow \infty$.

The convergence of $\widehat{F}_t^{(m)}$ and $\widehat{\Lambda}_m$ in (2.13) is discussed in Remark A.1 of Appendix A.

The next result concerns the asymptotics of the estimated quantities in the underlying PVAR model.

Proposition 2.2. *With the notation above and under the assumptions (CR1)–(CR4), (A1)–(A4), (A3'), we have: for any $i = 1, \dots, p$, $m = 1, \dots, s$,*

$$\widehat{\Psi}_{m,i} - \Psi_{m,i} = O_p\left(\frac{1}{q}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \quad (2.17)$$

as $q, T \rightarrow \infty$.

Remark 2.3. *We expect that the result analogous to (2.14) holds for the smoothed estimated factors $\widetilde{G}_n$ as in Doz et al. (2011). For brevity, we shall not pursue this result here – we found the effect of smoothing to be of little practical significance.*

Several methods have been proposed to estimate the number of dynamic factors r . For a DFM, one popular approach is based on the information criteria studied by Bai and Ng (2007). We adapt this information criteria to find the number of factors for PDFM by setting

$$\hat{r} = \underset{k}{\operatorname{argmin}} \{ \log (\operatorname{SSE}(k)) + kg(q, N) \}, \quad (2.18)$$

where $\operatorname{SSE}(k)$ is the sum of squared errors assuming k factors calculated by

$$\operatorname{SSE}(k) = \frac{1}{qN} \sum_{n=1}^N \left(X_n - \hat{\mu}_m - \hat{\Lambda}_m \hat{F}_n \right)' (X_n - \hat{\mu}_m - \hat{\Lambda}_m \hat{F}_n) \quad (2.19)$$

and $g(q, N)$ is one of the following four functions:

$$\begin{aligned} g_1(q, N) &= \frac{q+N}{qN} \log \left(\frac{qN}{q+N} \right), & g_2(q, N) &= \frac{q+N}{qN} \log(q \wedge N), \\ g_3(q, N) &= \frac{\log(q \wedge N)}{(q \wedge N)}, & g_4(q, N) &= \frac{q+N-k}{qN} \log(qN), \end{aligned} \quad (2.20)$$

where $q \wedge N = \min(q, N)$. We also mention a popular graphical method of selecting the number of factors through the so-called scree plot. Adapting it to our context, the eigenvalues of the covariance matrix $\hat{\Sigma}_Y$ from the deseasonalized series $X_n - \hat{\mu}_m$ are plotted on the y-axis from the largest to the smallest, and the leveling-off point is selected as the number of factors r . See Figure 8 below for an example. A further discussion related to the choice of r can be found in Section 2.5 below.

The order p of the underlying PVAR model is also estimated using an information criteria, namely, the BIC,

$$\begin{aligned} \hat{p} = \underset{k}{\operatorname{argmin}} \left\{ \frac{1}{s} \sum_{m=1}^s \log \left(|\hat{\Sigma}_{\zeta, m}(k)| \right) \right. \\ \left. + \frac{\log N}{N} (\# \text{ of freely estimated parameters}) \right\}, \end{aligned} \quad (2.21)$$

where $\hat{\Sigma}_{\zeta, m}(k)$ is the covariance matrix estimate of PVAR(k) model residuals at season m .

2.4. Other approaches based on sparsity and DFMs

In the simulation study presented in Section 3 and the applications presented in Section 4, we also estimate the underlying PVAR model by using several other estimation methods, than just a simple least squares regression. On the one hand, we shall use sparse estimation, following our work in Baek, Davis and Pipiras (2017). Sparsity is natural and convenient to use with PVAR models, which have a large number of parameters to estimate even for moderate values of dimension r . On the other hand, we shall also use sparse estimation but in the Fourier representation of PVAR coefficients. Representing the univari-

ate PAR coefficients in the Fourier basis is used in PAR model estimation in Jones and Brelsford (1967), Dudek, Hurd and Wójtcowicz (2016a) (see also references therein), with the idea that only several of the Fourier coefficients will be needed assuming the PAR coefficients change slowly over a period. Extending the Fourier representation approach to PVAR models, the (j, k) element of $\Phi_{m,i}$ is represented as

$$\Phi_{m,i}(j, k) = a_{i,1}^{(j,k)} + \sum_{h=1}^H a_{i,2h}^{(j,k)} \cos\left(\frac{2\pi hm}{s}\right) + a_{i,2h+1}^{(j,k)} \sin\left(\frac{2\pi hm}{s}\right), \quad i = 1, \dots, p, \quad (2.22)$$

where H is the order of the Fourier approximation. For example, if $r = 5$, $s = 24$ and $p = 2$, then the number of parameters for PVAR(2) model is given by $r^2 ps = 1200$. While in the Fourier approximation with $H = 2$, the number of parameters to be estimated is reduced to $r^2 p(2H + 1) = 250$. We can further impose sparsity on the Fourier coefficients to estimate them using adaptive lasso as described in Appendix B. We shall refer to the modeling approach based on (2.22) and adaptive lasso as Fourier-PVAR or FPVAR, for short.

Further insight into PDFM can be given through the following discussion. PVAR time series $\{F_n\}$ satisfying (1.5) are nonstationary in general but the multivariate time series $\tilde{F}_t = \text{vec}(F_t^{(1)}, \dots, F_t^{(s)})$ obtained by stacking F_n 's over one full period is stationary (see also the notation around (2.3)). In fact, the series $\{\tilde{F}_t\}$ has a VAR representation, and this is why additional suitable assumptions on the stationary “block” series $\{F_t^{(m)}\}$ can be made (as in Section 2.3). These observations also naturally suggest the following “competing” dynamic factor approach to multivariate time series with periodic dependence structure. As with the PVAR model above, a natural possibility is to consider the “block” vector series $\tilde{X}_t = \text{vec}(X_t^{(1)}, \dots, X_t^{(s)})$ obtained by stacking the original series X_n over s seasons of the full period. Then, the usual DFM could be postulated for $\{\tilde{X}_t\}$ as in (1.1)–(1.2), namely,

$$\tilde{X}_t - \tilde{\mu} = \tilde{\Lambda}_0 \tilde{F}_t + \tilde{\varepsilon}_t \quad (2.23)$$

and

$$\tilde{\Phi}(B) \tilde{F}_t = \tilde{\zeta}_t, \quad (2.24)$$

where the various components of the model are interpreted as in (1.1)–(1.2). This “stacking” approach was used by e.g., Dordonnat, Koopman and Ooms (2012), in the case of univariate series X_n .

For later use, we shall refer to the approach based on (2.23)–(2.24) as block-DFM (where “block” refers to the blocks $X_t^{(m)}$ used to construct $\tilde{X}_t$). Note that block-DFM becomes our PDFM when $\tilde{\Lambda}_0$ and $\tilde{\Phi}(B)$ have block-diagonal structures, namely,

$$\tilde{\Lambda}_0 = \text{diag}(\Lambda_1, \dots, \Lambda_s), \quad \tilde{\Phi}(B) = \text{diag}(\Phi_1(B), \dots, \Phi_s(B)). \quad (2.25)$$

That is, PDFM can be viewed as block-DFM with restrictions (2.25). This means, in particular, that a larger number of parameters needs to be estimated

in block-DFM: the loading matrix $\tilde{\Lambda}_0$ has dimension $sq \times sr$ and hence s^2qr parameters, and the polynomial $\tilde{\Phi}(B)$ has $p(sr)^2$ parameters, as opposed to the respective numbers sqr and spr^2 for PDFM. As a consequence, if the data actually follows PDFM, we can naturally expect better estimation performance when going the PDFM rather than the block-DFM route. A smaller number of parameters should also lead to better forecasts – incidentally, this is also what we generally observe in data applications (see Section 4 below).

The number of parameters aside, the option of using PDFMs versus block-DFMs is similar to that of using PVAR models versus VARs (or larger dimensional embedded VARs). Discussions for the latter case can be found in Hurd and Miamee (2007) p. 6, Franses and Paap (2004), Section 1.3. Among advantages of periodic models, their flexibility and convenient interpretation are often mentioned. In some applications, periodic dependence structures are also naturally expected: e.g., in environmental applications, this structure may be determined by the annual seasons, or in engineering applications, as in vibrations of heavy machinery, they may reflect the periods associated with machinery rotating parts.

2.5. Varying number of factors

We assumed in the preceding sections that the number of factors is the same across the s seasons. But the approaches described in Sections 2.1 and 2.2 extend straightforwardly to the case of varying number of factors, but with several caveats that are mentioned below. We do not consider this more general case throughout the paper for two reasons: simplicity and also since the real data considered in this work do not present strong evidence in favor of using varying number of factors.

To be more specific, by the case of varying number of factors, we mean the model (1.4)–(1.5) but with

$$\Lambda_m = (\Lambda_{0,m} \ 0_{q \times (r-r_m)}), \quad (2.26)$$

where $\Lambda_{0,m}$ has dimension $q \times r_m$, $0_{q \times (r-r_m)}$ is the $q \times (r-r_m)$ matrix of zeros, $r = \max_m r_m$ and at least one of r_m is smaller than r . In other words, only the first r_m components of the r -vector factors F_n load onto X_n . Consider first such a model in the framework of Sections 2.2–2.3, where q is assumed to be large. Since Steps 1–3 of the estimation algorithm in Section 2.2 apply for fixed season m , they can also be used for the model (1.4)–(1.5) with (2.26). Moreover, a different number of factors might be suggested by using the criterion (2.18) but replacing $SSE(k)$ in (2.19) by $SSE_m(k)$ in (2.7), or also by examining the scree plots of the eigenvalues of the matrices $\hat{\Sigma}_Y^{(m)}$. The factors that load in season m can still be estimated consistently by the results analogous to those in Section 2.3.

Step 4 of that algorithm, however, does not apply. One is presented with an interesting problem of estimating the parameters of a PVAR model but where for some season(s) m , only r_m -vector parts of the r -vector series F_n are actually

observed. This issue of “missing” data could be dealt with naturally through the EM algorithm as e.g. in Shumway and Stoffer (2006), Section 6.4. But one question that remains to be addressed (and that we currently do not know how to answer) concerns identifiability issues for such PVAR models with “missing” data. What we mean can be explained through the following simple example. Suppose F_n is one-dimensional and satisfies a PAR(1) model with period $s = 4$ and the AR parameters ϕ_1, ϕ_2, ϕ_3 and ϕ_4 for the four seasons. If, for example, F_n is “missing” in the periods 2 and 3, it is unrealistic that ϕ_2 and ϕ_3 can be identified since there is no data to deduce the chronological order of seasons 2 and 3 (that is, the model parameters ϕ_2 and ϕ_3 may as well be interchanged).

A similar discussion also applies to the model (1.4)–(1.5) with (2.26), in the framework of Section 2.1 where q is thought to be small. That is, modulo the identifiability question raised above, there are a priori no major issues of running the EM algorithm for the model (1.4)–(1.5) with (2.26) in its state space formulation. One smaller issue to mention is that in the M-step of the algorithm, we do not have a closed-form solution to Λ_m and use direct optimization. The EM algorithm was tested successfully on simple models with varying number of factors, and the associated code is available upon request.

3. Simulation study

In this section, we study the finite sample performance of the proposed estimation methods for PDFMs by Monte Carlo simulations. We focus on the PCA method of Sections 2.2–2.3 and then present briefly the results for the EM algorithm of Section 2.1.

3.1. Approximate model formulation with PCA

For brevity, we consider loadings and idiosyncratic error terms that do not depend on the season (but naturally the factors follow a PVAR model). The univariate components $X_{j,n}$ of X_n in the PDFM (1.4) can then be written as

$$X_{jn} = \sum_{k=1}^r \lambda_{jk} F_{kn} + \varepsilon_{jn}, \quad j = 1, \dots, q.$$

The factor loadings λ_{jk} of the matrix $\Lambda = (\lambda_{j,k})$ are generated as the (j, k) components of the QR decomposition of $q \times r$ matrix consisting of i.i.d. $\mathcal{N}(0, 1)$ random variables and multiplied by $\sqrt{q}$ so that

$$\frac{1}{q} \Lambda' \Lambda = I_r$$

(see, in particular, Assumptions (CR1) and (CR2) in Appendix A). The idiosyncratic component is assumed to follow the AR(1) model

$$\varepsilon_{jn} = \rho \varepsilon_{j,n-1} + \xi_{jn} \quad (3.1)$$

with ξ_{jn} being i.i.d. $\mathcal{N}(0, 1 - \rho^2)$ so that this yields serial correlations with zero mean and unit variance. Regarding factors F_{kn} , we considered two PVAR(1) models. The first model (DGP1) parameters are given by: for $s = 4$ and $r = 3$,

$$\Phi_{1,1} = \begin{pmatrix} .1 & .5 & .5 \\ 0 & 0 & .3 \\ 0 & .25 & .5 \end{pmatrix}, \quad \Phi_{2,1} = \begin{pmatrix} .8 & 0 & 0 \\ .1 & 0 & 0 \\ 0 & 0 & .3 \end{pmatrix},$$

$$\Phi_{3,1} = \begin{pmatrix} 0 & .6 & 0 \\ 0 & 0 & .2 \\ 0 & .1 & 0 \end{pmatrix}, \quad \Phi_{4,1} = \begin{pmatrix} .7 & 0 & 0 \\ 0 & -.5 & 0 \\ 0 & -.2 & .8 \end{pmatrix}$$

and

$$\Sigma_{\zeta,1} = \Sigma_{\zeta,2} = \Sigma_{\zeta,3} = \Sigma_{\zeta,4} = \begin{pmatrix} 1 & 1/4 & 1/6 \\ 1/4 & 1 & 0 \\ 1/4 & 0 & 1 \end{pmatrix}.$$

The second model (DGP2) parameters are: for $s = 2$ and $r = 4$,

$$\Phi_{1,1} = \begin{pmatrix} .2 & 0 & .1 & 0 \\ 0 & .2 & 0 & .1 \\ .1 & 0 & 0 & 0 \\ 0 & .1 & 0 & .3 \end{pmatrix}, \quad \Phi_{2,1} = \begin{pmatrix} -.2 & 0 & .1 & 0 \\ 0 & .2 & .1 & .5 \\ 0 & 0 & .7 & 0 \\ .1 & .1 & 0 & -.3 \end{pmatrix},$$

$$\Sigma_{\zeta,1} = \Sigma_{\zeta,2} = \begin{pmatrix} 1 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1 & 0 & 0 \\ 1/4 & 0 & 1 & 0 \\ 1/4 & 0 & 0 & 1 \end{pmatrix}.$$

Since factors are identifiable up to a rotation (see Section 2 and also Remark A.1 in Appendix A), we need to reestimate them and PVAR coefficients by fixing a particular rotation. We estimate the rotation matrix Q in (2.8) (which does not depend on m by our assumptions) as in Doz et al. (2011), p. 195,

$$\hat{Q} = \sum_{n=1}^N F_n \tilde{G}'_n \left(\sum_{n=1}^N \tilde{G}_n \tilde{G}'_n \right)^{-1}.$$

Then, the estimates of the PDFM factors and loadings are defined as $\bar{F}_n = \hat{Q} \tilde{G}_n$, $\bar{\Lambda} = \hat{\Pi}_m \hat{Q}'$, and the PVAR model parameters $\Phi_{m,i}$ are re-estimated based on $\bar{F}_n$, denoted $\bar{\Phi}_{m,1}$, $m = 1, \dots, s$.

We measure the performance of the estimation procedure through the following statistics. The first statistic measures the discrepancy of the estimated factors through a trace R^2 of the multivariate regression of $\bar{F}$ onto F . Introduced in Stock and Watson (2002), it is defined as

$$R^2 = \frac{\widehat{\mathbb{E}} [\text{tr} (\bar{F}' F (F' F)^{-1} F' \bar{F})]}{\widehat{\mathbb{E}} [\text{tr} (\bar{F}' \bar{F})]},$$

TABLE 1
The performance of the PDFM for DGP1 measured by R^2 and $\chi^2(\bar{F}, F)$.

	R^2				$\chi^2(\bar{F}, F)$			
	$\rho = .3$		$\rho = .7$		$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$q = 20$	.956	.959	.955	.96	$q = 20$	.046	.041	.047
$q = 50$	.976	.981	.975	.981	$q = 50$	.025	.019	.026
$q = 100$	.983	.988	.983	.987	$q = 100$	.018	.012	.019

TABLE 2
The MSEs of the PDFM model for DGP1.

	$\text{MSE}(\bar{\Lambda}, \Lambda)$				$\text{MSE}(\bar{\Phi}, \Phi)\text{-PVAR}$			
	$\rho = .3$		$\rho = .7$		$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$q = 20$	.146	.086	.175	.088	$q = 20$	.352	.171	.363
$q = 50$	.340	.168	.386	.192	$q = 50$	.335	.163	.334
$q = 100$	.667	.331	.751	.375	$q = 100$	.327	.157	.318

	$\text{MSE}(\bar{\Phi}, \Phi)\text{-sPVAR}$				$\text{MSE}(\bar{\Phi}, \Phi)\text{-FPVAR}$			
	$\rho = .3$		$\rho = .7$		$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$q = 20$	.252	.121	.258	.123	$q = 20$	.345	.177	.356
$q = 50$	.235	.113	.237	.111	$q = 50$	.333	.171	.331
$q = 100$	.236	.111	.235	.111	$q = 100$	.329	.160	.331

TABLE 3
The performance of the PDFM for DGP2 measured by R^2 and $\chi^2(\bar{F}, F)$.

	R^2				$\chi^2(\bar{F}, F)$			
	$\rho = .3$		$\rho = .7$		$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$q = 20$	.949	.953	.949	.952	$q = 20$	.056	.05	.057
$q = 50$	.973	.977	.973	.978	$q = 50$	.031	.024	.032
$q = 100$	.981	.986	.981	.986	$q = 100$	.023	.016	.023

where $\hat{\mathbb{E}}$ denotes the empirical expectation calculated by taking average over Monte Carlo replications. The second summary statistic is a chi-square type statistic for common component used in Forni et al. (2000). It is given by

$$\chi^2(\bar{F}, F) = \frac{(\bar{\Lambda}\bar{F} - \Lambda F)'(\bar{\Lambda}\bar{F} - \Lambda F)}{(\Lambda F)'(\Lambda F)}.$$

The last summary statistic is a mean squared error of the loading matrices and PVAR parameters given by

$$\begin{aligned} \text{MSE}(\bar{\Lambda}, \Lambda) &= \text{tr}(\bar{\Lambda} - \Lambda)(\bar{\Lambda} - \Lambda)', \\ \text{MSE}(\bar{\Phi}, \Phi) &= \sum_{m=1}^s \sum_{i=1}^p \text{tr}(\bar{\Phi}_{m,i} - \Phi_{m,i})(\bar{\Phi}_{m,i} - \Phi_{m,i})'. \end{aligned}$$

We considered three different dimensions $q = 20, 50$ and 100 with sample sizes $T = 100$ and $T = 200$. For the idiosyncratic errors following AR(1) model, we considered $\rho = .3$ and $.7$. All results are based on 500 Monte Carlo replications. Tables 1 and 2 report on the performance of PDFM for DGP1. It is observed that R^2 approaches 1 and $\chi^2(\bar{F}, F)$ diminishes as the sample size T

TABLE 4
The MSEs of the PDFM model for DGP2.

MSE($\bar{\Lambda}, \Lambda$)					MSE($\bar{\Phi}, \Phi$)–PVAR				
	$\rho = .3$		$\rho = .7$			$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$		$T = 100$	$T = 200$	$T = 100$	$T = 200$
$q = 20$	.467	.255	.525	.288	$q = 20$	.367	.181	.377	.191
$q = 50$	1.070	.530	1.179	.585	$q = 50$	.352	.168	.362	.173
$q = 100$	2.098	1.045	2.289	1.138	$q = 100$	.340	.168	.346	.164

MSE($\bar{\Phi}, \Phi$)–sPVAR					MSE($\bar{\Phi}, \Phi$)–FPVAR				
	$\rho = .3$		$\rho = .7$			$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$		$T = 100$	$T = 200$	$T = 100$	$T = 200$
$q = 20$	.285	.148	.287	.156	$q = 20$	.249	.120	.243	.128
$q = 50$	.278	.139	.284	.140	$q = 50$	.239	.119	.240	.118
$q = 100$	.280	.140	.271	.136	$q = 100$	.237	.119	.239	.115

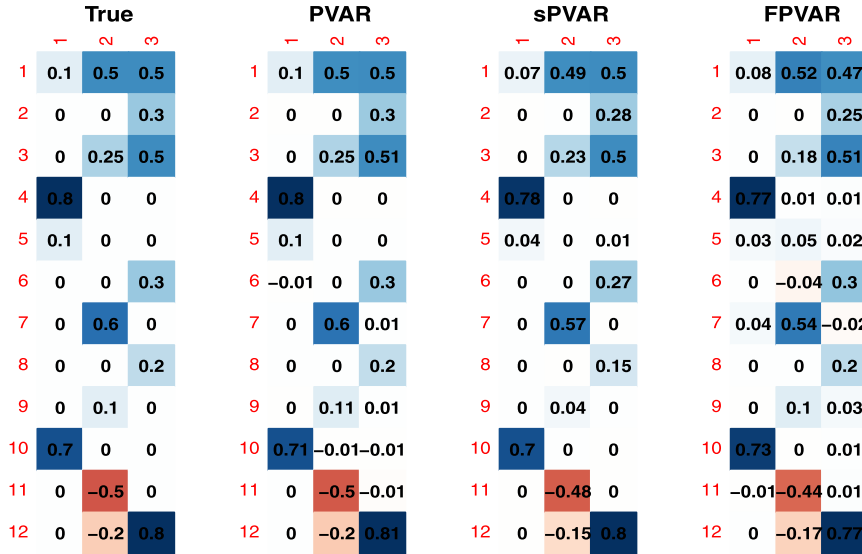


FIG 1. The coefficients of stacked PVAR matrices $\Phi_{1,1}, \Phi_{2,1}, \Phi_{3,1}, \Phi_{4,1}$ for DGP1 (left) and the averages of estimated coefficients for three different models, namely, PVAR, sPVAR and FPVAR with $q = 100, T = 100$ and $\rho = .3$.

increases and the dimension q increases, as in Doz et al. (2011). This means that the precision of the factor estimation increases as the data dimension increases in both directions. However, the MSE of loading matrices are increasing as the dimension q increases while the sample size T is fixed. The MSEs for PVAR coefficients are all decreasing as the sample size T increases, and sparse PVAR (sPVAR) method shows the smallest MSEs as we expected since the true model is sparse. Note also that the performance of PDFM is getting worse under stronger correlations for idiosyncratic errors.

Similar results are observed for DGP2 as shown in Tables 3–4. It is interesting that Fourier-PVAR (FPVAR) model gives the smallest MSEs amongst all the

	True				PVAR				sPVAR				FPVAR			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	0.2	0	0.1	0	0.21	-0.01	0.1	-0.01	0.13	-0.01	0.05	0	0.13	0	0.04	0
2	0	0.2	0	0.1	0	0.21	0	0.1	0	0.13	-0.01	0.05	0	0.15	0	0.12
3	0.1	0	0	0	0.09	0	0	0	0.03	0	0	0	0.02	0	0.02	0
4	0	0.1	0	0.3	0	0.1	0	0.32	0	0.05	0	0.24	0.01	0.04	0	0.25
5	-0.2	0	0.1	0	-0.18	0	0.09	-0.01	-0.12	-0.01	0.03	-0.01	-0.13	0	0.03	0
6	0	0.2	0.1	0.5	-0.01	0.22	0.1	0.5	0.02	0.16	0.04	0.47	0.01	0.15	0.02	0.39
7	0	0	0.7	0	0	0.01	0.72	-0.01	0.01	0	0.69	0	0.01	0	0.64	0
8	0.1	0.1	0	-0.3	0.1	0.1	0	-0.3	0.05	0.06	0.01	-0.23	0.02	0.04	0	-0.25

FIG 2. The coefficients of stacked PVAR coefficient matrices $\Phi_{1,1}, \Phi_{2,1}$ for DGP2 (left) and the averages of estimated coefficients for three different models, namely, PVAR, sPVAR and FPVAR with $q = 100, T = 100$ and $\rho = .7$.

methods considered in this particular simulation setting. This is because, as can be checked, the number of zero coefficients in the Fourier representation is even larger than that in the Φ_m -representation of the DGP2 model. The average estimated PVAR coefficients are illustrated in Figures 1–2 with the coloring and its lightness associated with the displayed numerical values. In summary, the proposed methods for estimating PDFM perform well in finite sample.

3.2. Exact model formulation with EM

We report here briefly on the performance of the estimation procedure in the exact model formulation with the EM algorithm introduced in Section 2.1. The data generating process (DGP3) uses the identifiable model assuming (2.1). We set $q = 4$ with $r = 2$ and $s = 4$ with common loadings

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ .15 & -.4 \\ 1.5 & -2 \end{pmatrix}$$

over the seasons. The underlying factors follows the PVAR(1) model with

$$\Phi_{1,1} = \begin{pmatrix} .2 & 0 \\ 0 & .2 \end{pmatrix}, \quad \Phi_{2,1} = \begin{pmatrix} .1 & .3 \\ .3 & .1 \end{pmatrix}, \quad \Phi_{3,1} = \begin{pmatrix} -.2 & 0 \\ 0 & .2 \end{pmatrix}, \quad \Phi_{4,1} = \begin{pmatrix} .5 & 0 \\ .1 & .1 \end{pmatrix},$$

and idiosyncratic component follows the AR(1) model (3.1) with $\rho = .3$ or $\rho = .7$. We considered $T = 100$ and $T = 200$ and the performance measures are $\text{MSE}(\hat{\Lambda}, \Lambda)$ and $\text{MSE}(\hat{\Phi}, \Phi)$. The initial estimates for the EM algorithm are based on PCA, that is, the principal component estimator (2.6) is rotated as in the relation (2.2).

TABLE 5
The performance of PDFM under the exact model formulation with EM for DGP3.

	MSE($\hat{\Lambda}, \Lambda$)				MSE($\hat{\Phi}, \Phi$)			
	$\rho = .3$		$\rho = .7$		$\rho = .3$		$\rho = .7$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
Two-stage	.637	.280	1.470	.388	.433	.272	.448	.276
EM	.596	.265	1.290	.388	.374	.226	.373	.241

Table 5 summarizes the results. We compared the EM results with the two-stage PCA estimator. For fair comparison, we rotated the estimates from the two-stage estimation by multiplying C based on the relation (2.2) and re-estimated the PVAR parameters from the rotated factors $C^{-1}F_n$. Observe first that the MSE decreases as the sample size T increases and the autocorrelation in idiosyncratic component is getting weaker. In short, the EM algorithm successfully estimates the underlying parameters. But note also that our two-stage PCA estimator is comparable to EM even if the dimension q of the data is 4, that is, quite low.

4. Applications

We apply here PDFMs to several real data sets. As noted in Section 1, PDFM should be entertained only in the case when univariate component series exhibit periodic correlations. In the applications below, the latter is the first thing examined and discussed. We also note that the detection tools not only reveal periodic correlations but also suggest what the period s is. When fitting a PDFM, the number of factors needs to be estimated as discussed at the end of Section 2.3. This is also one of the first tasks carried out in the applications below. The rest of PDFM analysis involves estimation of the loadings Λ_m and the parametric model of the factors F_n . Here, we shall consider a number of possibilities, discussed in more detail below, with these model components being either periodic or non-periodic, with estimation being either sparse or not, etc. Furthermore, the results will be compared to a number of alternative approaches, for example, the most elementary one where just the univariate series are being modeled.

4.1. Fraser river water flow

The first data set considered here is a monthly water flow from the Fraser River in British Columbia (BC), Canada. The Fraser River is the longest river in BC and the largest river by volume flowing into the Pacific seaside. It runs almost 1400 km and the annual discharge at its mouth is 112 cubic kilometers. The analyzed data are the monthly averages of the daily discharge measurements, in cubic kilometers. Six locations are selected: the Fraser River above Texas Creek, at Hansard, at McBride, at Hope, at Red Pass and at Shelly from 1959 to 2010; hence 624 observations are used in the analysis. We also reserved the last 24 observations for out-of-sample forecasts. The data set is obtained from the Wateroffice of Canada – see <http://wateroffice.ec.gc.ca> for more information.

Fraser Water Flow

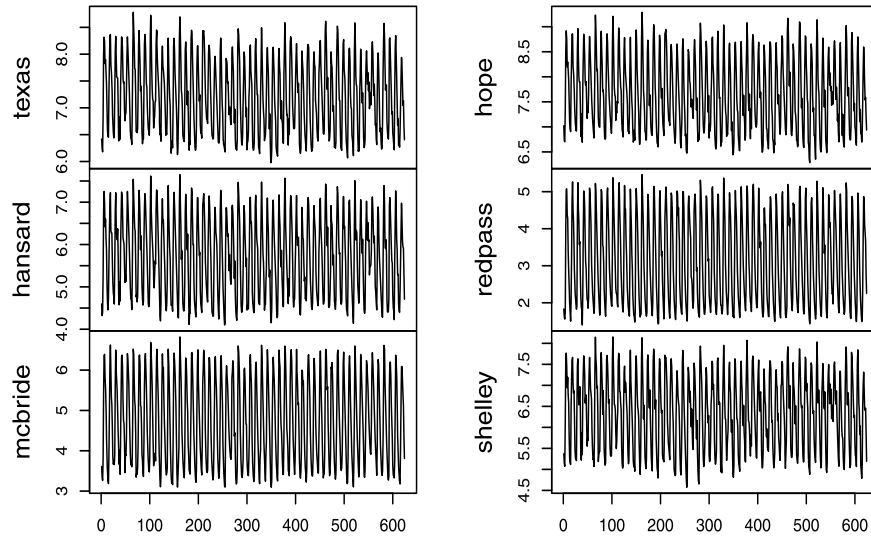


FIG 3. Time plots of log-transformed monthly water flow of Fraser river.

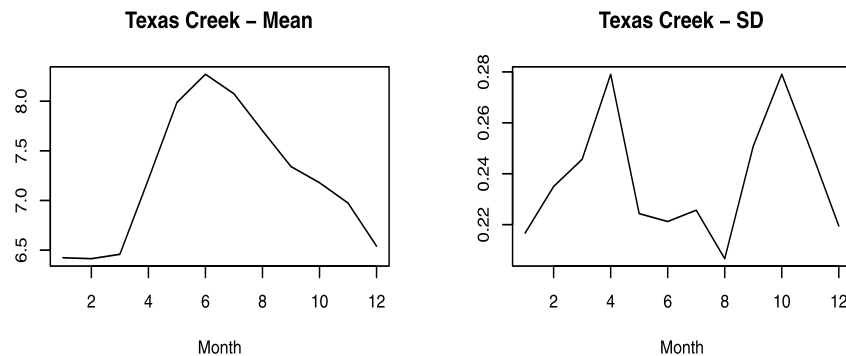


FIG 4. Seasonal mean and standard deviation at Texas Creek.

The data is log-transformed and Figure 3 shows time plots of log-transformed monthly water flow of Fraser river at the six locations. It is clearly observed that the data show cyclic variations. Figure 4 shows seasonal mean and standard deviation at the Fraser above Texas creek. The water flow seems low during the winter season, from December to May, then increasing rapidly from April, achieving the highest flow in July, and then slowly decreasing afterwards. The standard deviation is “bimodal” having peaks at April and October, so the water flow dynamics changes around those times. That is, the water flow dynamics enters a higher water flow period from May to September, and then changes to

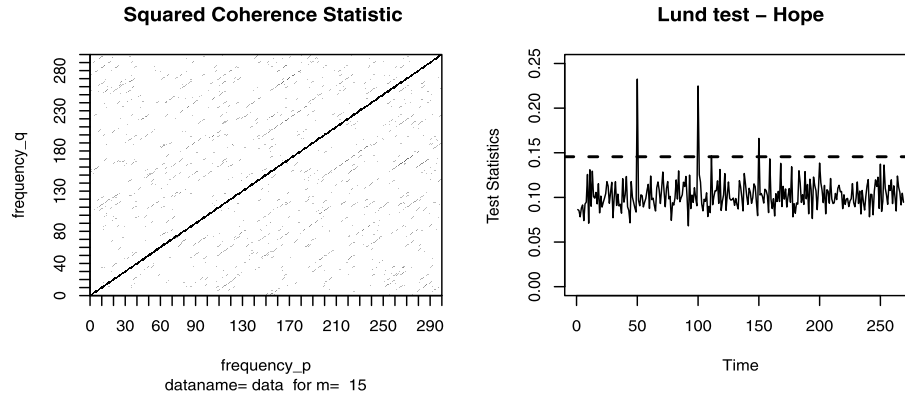


FIG 5. Spectral coherence plot and Lund test plot for Hope.

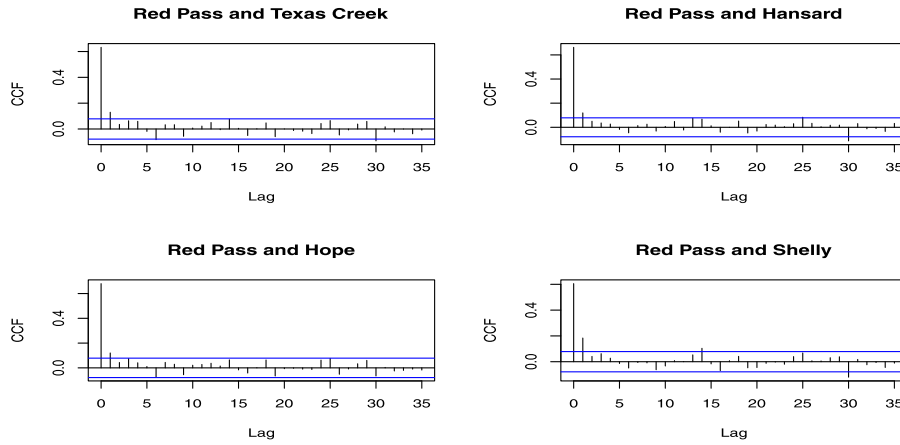
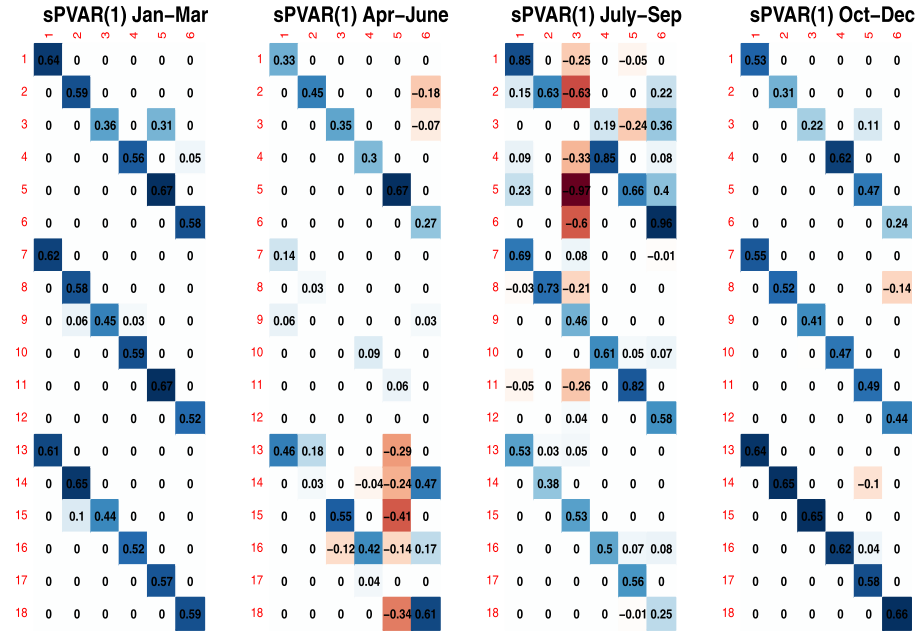


FIG 6. The cross-correlation plots at some selected locations.

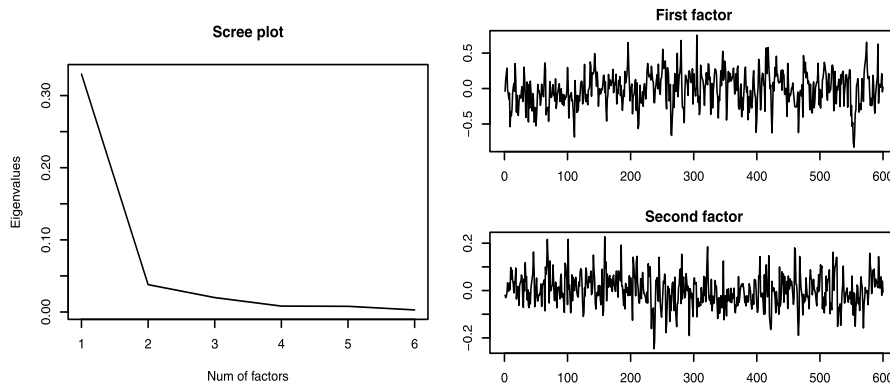
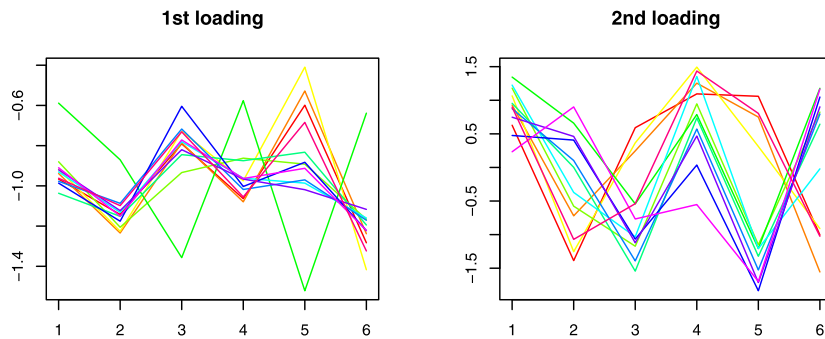
a slow period from November to March.

Figure 5 presents two commonly used plots to detect periodic correlations in univariate series (after removing its seasonal mean): the spectral coherence plot according to Hurd and Gerr (2008), and a related test statistic with a critical value line from Lund (2011). The spectral coherence is plotted using the R package perARMA (Dudek, Hurd and Wojtowicz (2016b)). The two plots are for the univariate series at Hope. If a series exhibits periodic correlations at period s (after removing the seasonal mean), the spectral coherence plot should have diagonal lines emerging at multiples of the index N/s , here $N/s = 600/12 = 50$. The plot in Figure 5 indeed suggests that the first major diagonal line is around the index 50. This could be determined easier from the Lund test statistic plot, which essentially averages the spectral coherence statistic at different indices along the corresponding diagonals, and also provides a critical values as the

FIG 7. Coefficient plots of $sPVAR(1;110)$.

dashed line. Note that the Lund test statistic value exceed the critical value at the indices 50 and 100, suggesting that periodic correlations are present in this univariate series. Figure 6 presents the sample cross-correlations between some selected locations prewhitened by fitting univariate AR(1) models. All four panels show significant lag 1 (cross-) correlation and suggest that the multivariate modeling is also potentially necessary to account for between-location dependence.

To account for such periodically correlated observations, we first applied sPVAR (sparse PVAR) models as in Baek et al. (2017). The best model is sPVAR(1;110), where 110 indicates the number of non-zero coefficients, by minimizing the BIC and using adaptive lasso to estimate coefficients. Figure 7 shows the autoregressive coefficient matrix stacked by 12 months. For example, the first 6×6 matrix on the left panel shows the autoregressive coefficient matrix for January, the 6×6 submatrix below corresponds to February, and so on. It is observed that coefficients from December to March look similar by having larger coefficients on the diagonal and little-to-none correlation for the off-diagonal entries. The coefficients for April are similar to December-March, but having smaller coefficients and negative coefficients appear first at the Shelley location. In May, the diagonal entries shrink; then negative correlations amongst different locations are prominent in June and July. Then, cross-correlations are dramatically weakened in August and within series correlations become dominant with a few exceptions through September to November. This shows that the water

FIG 8. Scree plot and estimated factors with $r = 2$.FIG 9. Estimated loadings with $r = 2$.

flow in the Fraser goes through several changes over a year. That is, during the winter period, December to March, the water flows appear almost decorrelated across locations but having strong within-series dependence. From April, within-series dependence is getting weaker, but between-locations dependence is getting stronger and peaks during June and July, and then rapidly disappears. This is consistent with the observations from Figure 4.

The results for PDFM are the following. Using the scree plot and the information criterion, the number of factors is selected as $r = 2$; so the PDFM reduces dimension from 6 to 2. Figure 8 shows the scree plot and time plots of the two factors found in the PDFM (two factors are also commonly selected when we considered scree plot for each season).

Twelve loading coefficients corresponding to each season are overlaid in Figure 9. Observe that there are some similarities in the shapes of loading coefficients, and if these are grouped by taking averages, Figure 10 is obtained. The grouping follows three patterns: one from December to March, then April to November except May, and May. Similar to the earlier interpretations, there is

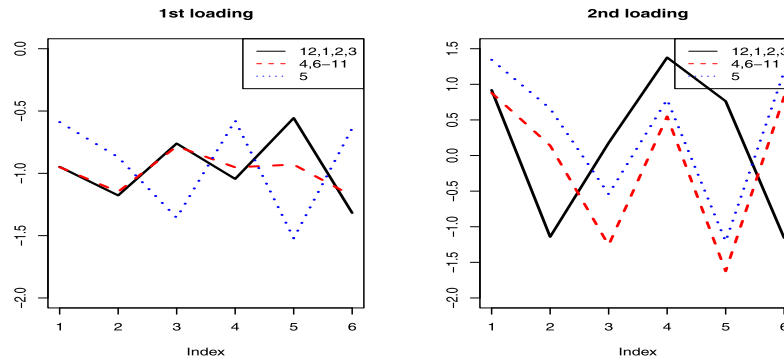


FIG 10. Estimated loadings grouped by three similar patterns.

TABLE 6
h-step-ahead out-of-sample forecasting errors ($10 \times \text{MSE}$) in Fraser river data.

Models	# of non-zero parameters	$10 \times \text{MSE}$		
		$h = 1$	$h = 2$	$h = 4$
PAR(1)	72	.398	.509	.645
PVAR(1)	432	.398	.504	.636
sPVAR(1)	123	.382	.491	.599
FPVAR(1)	42	.367	.504	.639
DFM-VAR(1)	4	.389	.536	.632
DFM-sVAR(1)	3	.388	.526	.621
DFM-PVAR(1)	48	.393	.521	.642
DFM-sPVAR(1)	29	.393	.522	.618
DFM-FPVAR(1)	36	.377	.507	.612
DFMS-PVAR(1)	48	.400	.524	.635
DFMS-sPVAR(1)	30	.390	.502	.616
DFMS-FPVAR(1)	36	.391	.487	.590
DFMG-sPVAR(1)	34	.388	.502	.620
EM-DFM-PVAR(1)	48	.388	.526	.621

a seasonal shift in water flow in the Fraser, mainly during the winter season and the rest of time. However, a further finding is that the water flow dynamics in May really differs from that for other months. It almost moves in the opposite direction and can be interpreted in that the negative between-locations correlations are dominant at that time.

To evaluate the forecasting ability, we also obtained the h -step-ahead out-of-sample forecasts and computed the empirical mean squared error (MSE),

$$\text{MSE} = \frac{1}{q(N_r - h + 1)} \sum_{n=N-N_r}^{N-h} \left(X_{n+h} - \hat{X}_{n+h} \right)' \left(X_{n+h} - \hat{X}_{n+h} \right), \quad (4.1)$$

where $\hat{X}_{n+h}$ is the h -step-ahead best linear predictor of X_{n+h} based on the training sample size $N - N_r$ and the test sample size $N_r - h + 1$. Table 6 shows the h -step-ahead out-of-sample forecast error together with the number

of non-zero parameters for $h = 1, 2$ and 4 . In the column “Models”, DFM corresponds to a dynamic factor model with common loadings across seasons while DFMS stands for DFM with seasonal loadings. For example, DFM-PVAR is the model where the common loadings are assumed for the dynamic factors and PVAR model is applied for the estimated factors. The DFMG represents DFM with grouped loadings as in Figure 10. As before, sPVAR corresponds to sparse PVAR obtained using adaptive lasso, FPVAR to Fourier-PVAR, and PAR to univariate periodic autoregressive model applied to each component series. EM-DFM-PVAR refers to DFM-PVAR estimated using the EM algorithm as discussed in Section 2.1; all other dynamic factor models in the table use PCA estimation.

The results are mixed and no method outperforms in all cases. For the 1-step-ahead forecasting, FPVAR(1) model shows the smallest MSE while DFMS-FPVAR(1) performs best for $h = 2$ and 4 , and DFM-FPVAR(1) is also quite comparable, while using 6 fewer parameters than FPVAR(1). In general, it is observed that the PDFM with FPVAR show better forecasting performance than full models, while using fewer parameters. As seen from Figures 8–10 and the discussion above, the PDFMs also have nice interpretations. EM-PDFM-PVAR(1), which is based on the estimation method suggested in Section 2.1, does not assume sparsity and generally performs worse than its sparse analogues; but note also that it outperforms (in almost all cases) the non-sparse DFM-VAR(1), DFM-PVAR(1), DFMS-PVAR(1) and is comparable to the sparse DFM-sVAR(1) (their performance measures only differ in the fourth decimal places).

We also report here briefly on comparison to forecasting using block-DFM (see Section 2.4). Since the block-DFM takes the full period as one block, it is not fair to compare one-step-ahead out-of-sample forecasts. Instead, we calculated the sum of 12-step-ahead forecasting errors, where the sum of m -step-ahead forecasting errors is given by

$$\sum_{h=1}^m \left(X_{N-m+h} - \hat{X}_{N-m}(h) \right)' \left(X_{N-m+h} - \hat{X}_{N-m}(h) \right) \quad (4.2)$$

and $\hat{X}_{N-m}(h)$ is the best linear predictor of X_{N-m+h} from the training sample $\{X_1, \dots, X_{N-m}\}$. The block-DFM performs worse than the PDFMs, for example, the forecasting error is 153.37 versus 131.70 for DFM-sPVAR.

4.2. US quarterly macroeconomic data

We also considered the quarterly macroeconomic time series of e.g., GDP, GNP and industrial product index for the United States analyzed in Stock and Watson (2009). Stock and Watson (2009) argued that a single break is plausible in 1984; so we only considered the post change data from 1984:I to 2006:IV. We excluded subaggregate series and applied the transformations detailed in Table A.1 of Stock and Watson (2009). The final data set has dimension $q \times N = 109 \times 92$,

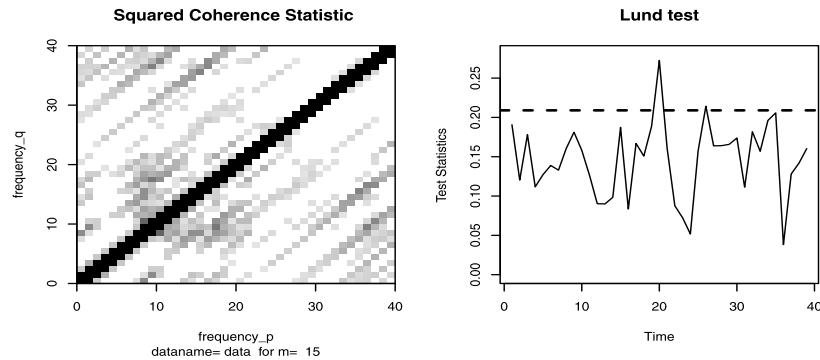


FIG 11. Spectral coherence plot and Lund test plot for HSWST series in US quarterly data.

reserving the last 12 observations for the evaluation of MSE. The dimension leads to an infeasible model fit when VAR models are considered, and even a univariate PAR(1) model requires $109 \times 4 = 436$ parameters to estimate. PDFM seems a natural model to consider for such high-dimensional data.

We checked for periodic correlations through the squared coherence statistic and Lund test discussed in Section 4.1. For example, Figure 11 shows the plots for HSWST (housing start west) variable. Here, periodic correlations should manifest themselves at the index $T/s = 80/4 = 20$, which is seen more clearly from the Lund test statistic plot. The scree plot is presented in the left panel of Figure 12, but it is not very informative to decide on a number of factors. The information criterion with the penalty function g_1 in (2.20) suggested two factors, which we use here. The other penalty functions g_2 and g_4 suggested only one factor and g_3 suggested four factors. Based on two factors, the estimated factors and loadings are depicted in Figures 12 and 13. Similar overall pattern of loadings amongst quarters suggests DFM or PDFM with common loadings.

The best model is DFM-sPVAR(1;6) showing the smallest h -step-ahead forecasting errors for all $h = 1, 2$ and 4 (see Table 7). It is remarkable that it only requires 6 non-zero parameters – just 1.37% of a univariate PAR(1) model while giving the smaller h -step-ahead MSE. The estimated sPVAR(1;6) parameters are presented in Figure 14. It shows that most of periodic nature is due to lag-1 dependence of the first factor, and that the second factor makes a significant effect cross-sectionally for the third quarter. The sum of 12-step-ahead forecasting errors of block-DFM as in (4.2) is 838.31 while that of DFMS-sPVAR is 697.56, so PDFM performs better than the block-DFM approach.

5. Conclusions

In this work, we introduced a periodic dynamic factor model (PDFM) that generalizes the usual dynamic factor model (DFM) by allowing its loadings to depend on “season,” and its factors to follow a periodic VAR model with

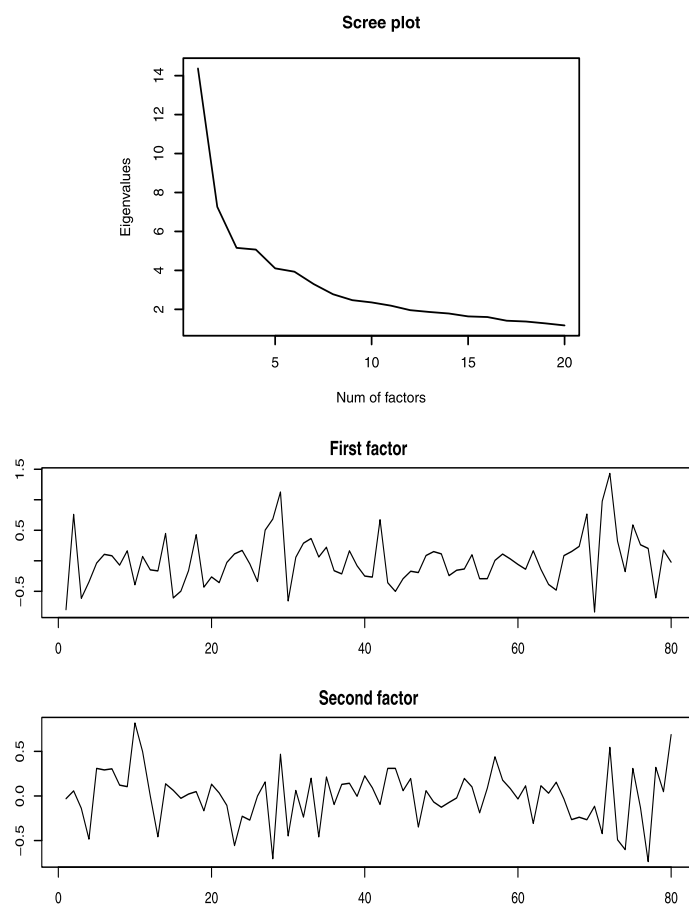
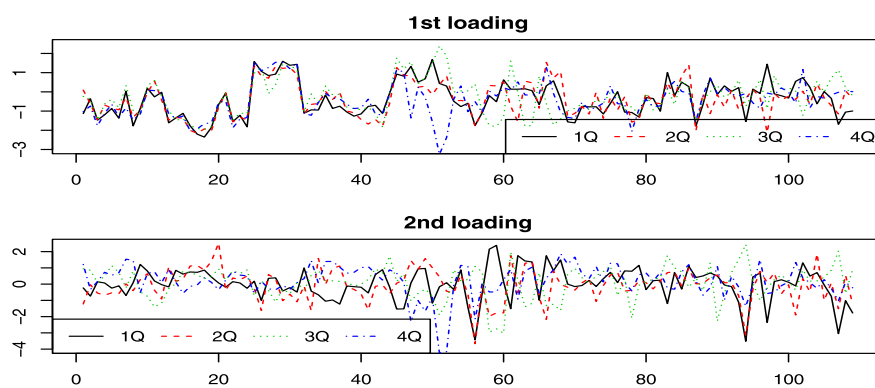
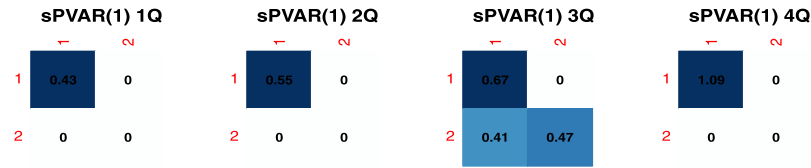
FIG 12. Scree plot and estimated factors ($r = 2$) for US quarterly data.

FIG 13. Estimated loadings for US quarterly data.

FIG 14. Estimated coefficients of $sPVAR(1;6)$ for US quarterly data.TABLE 7
1-step-ahead out-of-sample forecasting errors in US quarterly data.

Models	# of non-zero parameters	MSE		
		$h = 1$	$h = 2$	$h = 4$
PAR(1)	436	1.136	1.126	1.165
DFM-VAR(1)	4	.978	.890	.783
DFM-sVAR(1)	3	.979	.883	.771
DFM-PVAR(1)	16	1.037	.967	.849
DFM-sPVAR(1)	6	.908	.783	.693
DFM-FPVAR(1)	8	.941	.843	.745
DFMS-PVAR(1)	16	1.106	.970	.830
DFMS-sPVAR(1)	8	1.025	.860	.729
DFMS-FPVAR(1)	8	.976	.840	.722

coefficients depending on “season” (the so-called PVAR model). Two estimation approaches for PDFM were considered: the EM estimation for the so-called exact formulation and the PCA estimation for the approximate formulation. Some asymptotic results were proved for the approximate formulation, by adapting the work on DFM by Doz et al. (2011, 2012). The simulation study showed satisfactory performance of the proposed estimation procedures and two real data sets were considered to illustrate the use of PDFMs and related models.

A number of open questions related to PDFMs and the standard DFMs were already raised in the text, including extensions to long-memory and nonlinear models, and to weak factors (Remark 2.2), and identifiability issues related to the varying number of factors (Section 2.5). Another interesting direction would be to develop detection and testing methods for the various special cases of PDFMs, as in (1.6). In this work, these various special cases were assessed just based on the out-of-sample forecasting performance.

Appendix A: Proofs for approximate model formulation

Proof of Proposition 2.1. The relations (2.14)–(2.16) follow immediately by applying Proposition 2 of Doz et al. (2011) for each fixed season $X_t^{(m)}$, $m = 1, \dots, s$. $\square$

Remark A.1. By (2.13), (2.14) above and Lemma 2 in Doz et al. (2011) (applied for fixed season m), the principal component estimator $\hat{F}_t^{(m)}$ in (2.13)

behaves asymptotically as

$$\frac{D_{m,r}^{1/2}}{q^{1/2}} G_t^{(m)}.$$

However, it is not clear whether the assumptions in Doz et al. (2011) stated above imply the convergence of $D_{m,r}^{1/2}/q^{1/2}$. (The conditions (CR1) and (CR2), in particular, do not even imply the convergence of the smallest and largest diagonal elements of $D_{m,r}^{1/2}/q^{1/2}$.) If the following stronger assumption is made, namely, that

$$\frac{1}{q} \Lambda'_m \Lambda_m \rightarrow A_m \quad (\text{A.1})$$

for some positive definite matrix A_m , then one would have the convergence

$$\widehat{F}_t^{(m)} \xrightarrow{p} B_m^{1/2} G_t^{(m)} = B_m^{1/2} Q'_m F_t^{(m)},$$

where B_m is a diagonal matrix consisting of the eigenvalues of A_m . That is, $\widehat{F}_t^{(m)}$ would converge to a (scaled) rotation of $F_t^{(m)}$. Similar conclusion applies for $\widehat{\Lambda}_m$: under (A.1), one would have

$$\widehat{\Lambda}_m \xrightarrow{p} \Lambda_m Q_m B_m^{-1/2}.$$

In the dynamic factor literature, the assumption (A.1) is made in e.g., Bai and Ng (2008).

Proof of Proposition 2.2. We shall adapt the approach used in the proof of Proposition 3 of Doz et al. (2011). Suppose for simplicity that $p < m$; appropriate changes in other cases are indicated below. Note that the model (2.11) can be expressed as

$$G_t^{(m)} = \Psi_1^{(m)} G_t^{(m,p)} + \eta_t^{(m)}, \quad (\text{A.2})$$

where $\eta_t^{(m)} = \eta_{m,m+(t-1)s}$ and

$$\Psi_1^{(m)} = (\Psi_{m,1}, \Psi_{m,2}, \dots, \Psi_{m,p}), \quad G_t^{(m,p)} = \begin{pmatrix} G_t^{(m-1)} \\ G_t^{(m-2)} \\ \vdots \\ G_t^{(m-p)} \end{pmatrix}. \quad (\text{A.3})$$

(For example, if p and m are such that $-s < m - p \leq 0$, $G_t^{(m,p)}$ would have to be defined based on $G_t^{(m-1)}, \dots, G_t^{(1)}, G_{t-1}^{(s)}, \dots, G_{t-1}^{(m-p+s+1)}$.) In particular,

$$\Psi_1^{(m)} = \left(\mathbb{E} G_t^{(m)} G_t^{(m,p)'} \right) \left(\mathbb{E} G_t^{(m,p)} G_t^{(m,p)'} \right)^{-1} =: \Upsilon_1^{(m,p)} \left(\Upsilon_0^{(m,p)} \right)^{-1}. \quad (\text{A.4})$$

Placing the least squares estimators $\widehat{\Psi}_{m,1}, \dots, \widehat{\Psi}_{m,p}$ into an analogous matrix $\widehat{\Psi}_1^{(m)} = (\widehat{\Psi}_{m,1}, \dots, \widehat{\Psi}_{m,p})$, one can write similarly

$$\widehat{\Psi}_1^{(m)} = \widehat{\Upsilon}_1^{(m,p)} \left(\widehat{\Upsilon}_0^{(m,p)} \right)^{-1}, \quad (\text{A.5})$$

where

$$\hat{\Upsilon}_1^{(m,p)} = \frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m)} \left(\hat{G}_t^{(m,p)} \right)' = \left(\frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m)} (\hat{G}_t^{(m_1)})' \right)_{m_1=m-1, \dots, m-p}, \quad (\text{A.6})$$

$$\hat{\Upsilon}_0^{(m,p)} = \frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m,p)} \left(\hat{G}_t^{(m,p)} \right)' = \left(\frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m_1)} (\hat{G}_t^{(m_2)})' \right)_{m_1, m_2=m-1, \dots, m-p}. \quad (\text{A.7})$$

As at the end of the proof of Proposition 2 of Doz et al. (2011), it is enough to show that, for any $m_1, m_2 = 1, \dots, s$,

$$\frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m_1)} (\hat{G}_t^{(m_2)})' - \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) = O_p \left(\frac{1}{q} \right) + O_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{A.8})$$

The relation (A.8) can be proved by adapting the proof of part (i) of Proposition 3 in Doz et al. (2011). We assume for simplicity that $\mu_m = 0$, so that $Y_t^{(m)}$ can be taken as $X_t^{(m)}$. Write

$$\frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m_1)} \left(\hat{G}_t^{(m_2)} \right)' = \hat{D}_{m_1, r}^{-1/2} \hat{P}_{m_1, r}' S_{m_1, m_2} \hat{P}_{m_2, r} \hat{D}_{m_2, r}^{-1/2}, \quad (\text{A.9})$$

where

$$S_{m_1, m_2} = \frac{1}{T} \sum_{t=1}^T Y_t^{(m_1)} \left(Y_t^{(m_2)} \right)'.$$

Now, decompose the quantity in (A.9) as

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \hat{G}_t^{(m_1)} \left(\hat{G}_t^{(m_2)} \right)' &= \hat{D}_{m_1, r}^{-1/2} \hat{P}_{m_1, r}' \Pi_{m_1} \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \Pi_{m_2}' \hat{P}_{m_2, r} \hat{D}_{m_2, r}^{-1/2} \\ &\quad + \hat{D}_{m_1, r}^{-1/2} \hat{P}_{m_1, r}' \left(S_{m_1, m_2} - \Pi_{m_1} \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \Pi_{m_2}' \right) \hat{P}_{m_2, r} \hat{D}_{m_2, r}^{-1/2}. \end{aligned} \quad (\text{A.10})$$

By using the notation in the middle of (2.9), the first term in (A.10) can be written as

$$\hat{D}_{m_1, r}^{-1/2} \hat{P}_{m_1, r}' P_{m_1, r} D_{m_1, r}^{1/2} \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) D_{m_2, r}^{1/2} P_{m_2, r}' \hat{P}_{m_2, r} \hat{D}_{m_2, r}^{-1/2}.$$

It follows from Lemma 2, (iv), and Lemma 4, (i), in Doz et al. (2011) applied to the series for seasons m_1 and m_2 , and the fact $\mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) = \mathbb{E} \left(G_0^{(m_1)} (G_0^{(m_2)})' \right)$ that the first term in (A.10) satisfies

$$\hat{D}_{m_1, r}^{-1/2} \hat{P}_{m_1, r}' \Pi_{m_1} \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \Pi_{m_2}' \hat{P}_{m_2, r} \hat{D}_{m_2, r}^{-1/2}$$

$$= \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) + O_p \left(\frac{1}{q} \right) + O_p \left(\frac{1}{\sqrt{T}} \right).$$

On the other hand, the second term in (A.10) is dealt with as follows. By Lemma 2, (iii), in Doz et al. (2011), applied to season m and assumptions (CR1) and (CR2), we have $q^{1/2} \widehat{D}_{m,r}^{-1/2} = O_p(1)$. It is then enough to show that

$$\frac{1}{q} \left(S_{m_1, m_2} - \Pi_{m_1} \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \Pi_{m_2}' \right) = O_p \left(\frac{1}{q} \right) + O_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{A.11})$$

This can be achieved by writing, in view of (2.10),

$$\begin{aligned} & \frac{1}{q} \left(S_{m_1, m_2} - \Pi_{m_1} \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \Pi_{m_2}' \right) \\ &= \frac{1}{q} \Pi_{m_1} \left(\frac{1}{T} \sum_{t=1}^T G_t^{(m_1)} (G_t^{(m_2)})' - \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \right) \Pi_{m_2}' \\ &+ \frac{1}{qT} \sum_{t=1}^T G_t^{(m_1)} (\varepsilon_t^{(m_2)})' + \frac{1}{qT} \sum_{t=1}^T \varepsilon_t^{(m_1)} (G_t^{(m_2)})' \Pi_{m_2}' \\ &+ \frac{1}{q} \left(\frac{1}{T} \sum_{t=1}^T \varepsilon_t^{(m_1)} (\varepsilon_t^{(m_2)})' - \mathbb{E} \left(\varepsilon_t^{(m_1)} (\varepsilon_t^{(m_2)})' \right) \right) + \frac{1}{q} \mathbb{E} \left(\varepsilon_t^{(m_1)} (\varepsilon_t^{(m_2)})' \right), \end{aligned}$$

and then arguing that, for any $m_1 \neq m_2$,

$$\mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^T G_t^{(m_1)} (G_t^{(m_2)})' - \mathbb{E} \left(G_t^{(m_1)} (G_t^{(m_2)})' \right) \right\|^2 = O \left(\frac{1}{T} \right), \quad (\text{A.12})$$

$$\frac{1}{T} \sum_{t=1}^T G_t^{(m_1)} (\varepsilon_t^{(m_2)})' = O_p \left(\frac{\sqrt{q}}{\sqrt{T}} \right), \quad (\text{A.13})$$

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_t^{(m_1)} (\varepsilon_t^{(m_2)})' - \mathbb{E} \left(\varepsilon_t^{(m_1)} (\varepsilon_t^{(m_2)})' \right) = O_p \left(\frac{q}{\sqrt{T}} \right). \quad (\text{A.14})$$

In fact, the relations (A.12) and (A.14) follow from similar relations

$$\mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^T G_t G_t' - \mathbb{E} (G_t G_t') \right\|^2 = O \left(\frac{1}{T} \right), \quad (\text{A.15})$$

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t' - \mathbb{E} (\varepsilon_t \varepsilon_t') = O_p \left(\frac{q}{\sqrt{T}} \right), \quad (\text{A.16})$$

where $G_t' = \left((G_t^{(1)})', (G_t^{(2)})', \dots, (G_t^{(s)})' \right)$ and $\varepsilon_t' = \left((\varepsilon_t^{(1)})', (\varepsilon_t^{(2)})', \dots, (\varepsilon_t^{(s)})' \right)$.

The multivariate series $\{G_t\}$ and $\{\varepsilon_t\}$ are stationary and the relations (A.15) and (A.16) can be proved by adapting the arguments of Brockwell and Davis (2009), pp. 226-227 (as indicted by Doz et al. (2011)). The relation (A.13), on the other hand, can be dealt with similarly by using the argument in the proof of Lemma 2, (i), in Doz et al. (2011). $\square$

Appendix B: Adaptive lasso in Fourier-PVAR

Applying adaptive lasso is immediate once Fourier-PVAR is represented as a linear model. First, for n falling into season m , the PVAR(p) model is written in a matrix form as

$$\begin{aligned} F_n &= \Phi_{m,1}F_{n-1} + \dots + \Phi_{m,p}F_{n-p} + \zeta_{m,n} \\ &= ((F'_{n-1}, \dots, F'_{n-p}) \otimes I_q) \text{vec}(\Phi_{m,1}, \dots, \Phi_{m,p}) + \zeta_n. \end{aligned}$$

By assuming the Fourier representation of the coefficients, the AR coefficients are represented as

$$\text{vec}(\Phi_{m,1}, \dots, \Phi_{m,p}) = (I_{pq^2} \otimes w(n)) \alpha,$$

where

$$\begin{aligned} w(n) &= \left(1, \cos\left(\frac{2\pi n}{s}\right), \sin\left(\frac{2\pi n}{s}\right), \cos\left(\frac{2\pi 2n}{s}\right), \sin\left(\frac{2\pi 2n}{s}\right), \right. \\ &\quad \left. \dots, \cos\left(\frac{2\pi Hn}{s}\right), \sin\left(\frac{2\pi Hn}{s}\right) \right) \end{aligned}$$

is the $1 \times (2H+1)$ vector of Fourier basis functions used in estimation and

$$\alpha = \left(\phi_{1,1}^{(1,1)}, \dots, \phi_{1,2H+1}^{(1,1)}, \phi_{1,1}^{(2,1)}, \dots, \phi_{1,2H+1}^{(2,1)}, \dots, \phi_{p,1}^{(q,q)}, \dots, \phi_{p,2K+1}^{(q,q)} \right)'$$

is the $pq^2(2H+1) \times 1$ vector of the corresponding Fourier coefficients. Then, we can write F_n as

$$F_n = ((F'_{n-1}, \dots, F'_{n-p}) \otimes I_q) (I_{pq^2} \otimes w(n)) \alpha + \zeta_{m,n} =: b(n)\alpha + \zeta_n.$$

Therefore, the Fourier-PVAR model can be written in a matrix form as

$$\mathbf{F} = \mathbf{B}\alpha + \boldsymbol{\zeta},$$

where

$$\mathbf{F} = \begin{pmatrix} F_{p+1} \\ \vdots \\ F_N \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b(p+1) \\ \vdots \\ b(N) \end{pmatrix}, \quad \boldsymbol{\zeta} = \begin{pmatrix} \zeta_{p+1} \\ \vdots \\ \zeta_N \end{pmatrix}.$$

By incorporating possible correlations of ζ_n through an estimated covariance matrix $\widehat{\Sigma}$ of ζ_n , the adaptive lasso estimator of Fourier-PVAR coefficients is defined iteratively by

$$\begin{aligned} \widehat{\alpha}^{(\ell)} &= \underset{\alpha}{\operatorname{argmin}} \left(\frac{1}{N-p} \|(I_{N-p} \otimes \widehat{\Sigma}_\ell^{-1/2})\mathbf{F} - (I_{N-p} \otimes \widehat{\Sigma}_\ell^{-1/2})\mathbf{B}\alpha\|^2 \right. \\ &\quad \left. + \lambda_\ell \sum_{j=1}^{pq^2(2H+1)} w_j^{(\ell)} |\alpha_j| \right), \end{aligned}$$

where α_j is the j th element of $\boldsymbol{\alpha}$ and the weights are proportional to the inverse of coefficients, namely,

$$w_j^{(\ell)} = \frac{1}{|\hat{\alpha}_j^{(\ell-1)}|}.$$

The covariance matrix at the ℓ th iteration is obtained by

$$\hat{\Sigma}_\ell = \frac{1}{N-p} \sum_{i=p+1}^N \left(F_i - b(i) \hat{\boldsymbol{\alpha}}^{(\ell-1)} \right) \left(F_i - b(i) \hat{\boldsymbol{\alpha}}^{(\ell-1)} \right)'.$$

We have used the 10-fold cross-validation rule to find the tuning penalty parameter λ_ℓ and the initial estimator of the correlation matrix is given by the identity matrix of dimension q .

References

- Alonso, A. M., García-Martos, C., Rodríguez, J. and Sánchez, M. J. (2011), ‘Seasonal dynamic factor analysis and bootstrap inference: Application to electricity market forecasting’, *Technometrics* **53**(2), 137–151. [MR2808340](#)
- Baek, C., Davis, R. A. and Pipiras, V. (2017), ‘Sparse seasonal and periodic vector autoregressive modeling’, *Computational Statistics & Data Analysis* **106**, 103–126. [MR3566078](#)
- Bai, J. and Ng, S. (2007), ‘Determining the number of primitive shocks in factor models’, *Journal of Business & Economic Statistics* **25**(1), 52–60. [MR2338870](#)
- Bai, J. and Ng, S. (2008), *Large Dimensional Factor Analysis*, Now Publishers Inc.
- Bai, J. and Wang, P. (2015), ‘Identification and Bayesian Estimation of Dynamic Factor Models’, *Journal of Business & Economic Statistics* **33**(2), 221–240. [MR3337059](#)
- Brockwell, P. and Davis, R. A. (2009), *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, New York. Reprint of the second (1991) edition. [MR1093459](#)
- Chamberlain, G. and Rothschild, M. (1983), ‘Arbitrage, factor structure, and mean-variance analysis on large asset markets’, *Econometrica* **51**(5), 1281–1304. [MR0736050](#)
- Dordonnat, V., Koopman, S. J. and Ooms, M. (2012), ‘Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling’, *Computational Statistics & Data Analysis* **56**(11), 3134–3152. [MR2943886](#)
- Doz, C., Giannone, D. and Reichlin, L. (2011), ‘A two-step estimator for large approximate dynamic factor models based on Kalman filtering’, *Journal of Econometrics* **164**, 188–205. [MR2821802](#)

- Doz, C., Giannone, D. and Reichlin, L. (2012), ‘A quasi-maximum likelihood approach for large approximate dynamic factor model’, *Review of Economics and Statistics* **94**, 1014–1024.
- Dudek, A., Hurd, H. and Wójtowicz, W. (2016a), ‘PARMA methods based on Fourier representation of periodic coefficients’, *Wiley Interdisciplinary Reviews: Computational Statistics* **8**(3), 130–149. [MR3511662](#)
- Dudek, A., Hurd, H. and Wojtowicz, W. (2016b), *perARMA: Periodic Time Series Analysis*. R package version 1.6.
- Engle, R. and Watson, M. (1981), ‘A one-factor multivariate time series model of metropolitan wage rates’, *Journal of the American Statistical Association* **76**(376), 774–781.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), ‘The generalized dynamic-factor model: identification and estimation’, *The Review of Economics and Statistics* **82**(4), 540–554.
- Franses, P. H. and Paap, R. (2004), *Periodic Time Series Models*, Advanced Texts in Econometrics, Oxford University Press, Oxford. [MR2061576](#)
- Ghysels, E. and Osborn, D. R. (2001), *The Econometric Analysis of Seasonal Time Series*, Themes in Modern Econometrics, Cambridge University Press, Cambridge. [MR1868590](#)
- Hurd, H. and Gerr, N. (2008), ‘Graphical method for determining the presence of periodic correlation’, *Journal of Time Series Analysis* **12**, 337–350. [MR1131006](#)
- Hurd, H. L. and Miamee, A. (2007), *Periodically Correlated Random Sequences*, Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ. [MR2348769](#)
- Jones, R. H. and Brelsford, W. M. (1967), ‘Time series with periodic structure’, *Biometrika* **54**, 403–408. [MR0223041](#)
- Lund, R. (2011), Choosing seasonal autocovariance structures: PARMA or SARMA, in W. Bell, S. Holan and T. McElroy, eds, ‘Economic Time Series: Modelling and Seasonality’, Chapman and Hall, pp. 63–80. [MR3076011](#)
- Nieto, F. H., Pena, D. and Saboyá, D. (2016), ‘Common seasonality in multivariate time series’, *Statistica Sinica* **26**(4), 1389–1410. [MR3586220](#)
- Peña, D. and Poncela, P. (2006), ‘Nonstationary dynamic factor analysis’, *Journal of Statistical Planning and Inference* **136**(4), 1237 – 1257. [MR2253761](#)
- Shumway, R. H. and Stoffer, D. S. (2006), *Time Series Analysis and Its Applications: With R examples*, Springer. [MR2228626](#)
- Stock, J. H. and Watson, M. W. (2002), ‘Forecasting using principal components from a large number of predictors’, *Journal of the American Statistical Association* **97**(460), 1167–1179. [MR1951271](#)
- Stock, J. H. and Watson, M. W. (2009), Forecasting in dynamic factor models subject to structural instability, in ‘The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry’, Oxford University Press, pp. 1–57. [MR3204305](#)
- Stock, J. H. and Watson, M. W. (2011), ‘Dynamic factor models’, *Oxford Handbook of Economic Forecasting*. Oxford University Press, USA pp. 35–59. [MR3204189](#)

- Su, L. and Wang, X. (2017), ‘On time-varying factor models: Estimation and testing’, *Journal of Econometrics* **198**(1), 84 – 101. [MR3628100](#)
- Wang, W. and Fan, J. (2017), ‘Asymptotics of empirical eigenstructure for high dimensional spiked covariance’, *The Annals of Statistics* **45**(3), 1342–1374. [MR3662457](#)