Economic Growth Centre
Economics, School of Social Sciences
Nanyang Technological University
14 Nanyang Drive
Singapore 637332

# TIME-VARYING STRUCTURAL APPROXIMATE DYNAMIC FACTOR MODEL

**Ziyan Zhao**
**Qingfeng Liu**

*September 2024*

EGC Report No: 2024/01

HSS-04-90A
Tel: +65 67906073
Email: D-EGC@ntu.edu.sg

*Economic Growth Centre Working Paper Series*

The author(s) bear sole responsibility for this paper.

Views expressed in this paper are those of the author(s) and not necessarily those of the Economic Growth Centre, NTU.

# Time-Varying Structural Approximate Dynamic Factor Model

Ziyan Zhao*
Economic Growth Centre, School of Social Sciences, Nanyang Technological University
and
Qingfeng Liu †
Department of Industrial and Systems Engineering, Hosei University

January 14, 2024

**Abstract**

This study proposes a time-varying structural approximate dynamic factor (TVS-ADF) model by extending the ADF model in state-space form. The TVS-ADF model considers time-varying coefficients and a time-varying variance–covariance matrix of its innovation terms, so that it can capture complex dynamic economic characteristics. We propose the identification scheme of the common factors in the TVS-ADF and derive the identification theory. We also propose an effective Markov chain Monte Carlo (MCMC) algorithm to estimate the TVS-ADF. To avoid the overparameterization caused by the time-varying characteristics of the TVS-ADF, we include the shrinkage and sparsification approaches in the MCMC algorithm. Additionally, we propose several effective information criteria for the determination of the number of factors in the TVS-ADF. Extensive artificial simulations demonstrate that the TVS-ADF has better forecast performance than the ADF in almost all settings for different numbers of explained variables, numbers of explanatory variables, sparsity levels, and sample sizes. An empirical application to macroeconomic forecasting also indicates that our model can substantially improve predictive accuracy and capture the dynamic features of an economic system better than the ADF.

*Keywords:* MCMC, shrinkage, sparsification, overparameterization, algorithms

---

*E-mail: sss-zyzhao@ntu.edu.sg. The authors would like to thank Florian Huber, Yasumasa Matsuda, Qihui Chen, and the audience at the Asian meeting of the econometric society in China, 2022, for valuable comments. The authors also gratefully acknowledge the financial supports of the Japan Society for the Promotion of Science through KAKENHI (B) No. 22H00833 (Liu) and the Nomura Foundation for Social Science Grant No. N21-3-E30-010 (Liu). The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

†Corresponding author. E-mail: qliu@hosei.ac.jp.



Electronic copy available at: https://ssrn.com/abstract=3996594

# 1 Introduction

The approximate dynamic factor (ADF) model in state-space form proposed by Forni et al. (2009) has become increasingly popular in various economics and finance applications over the past decade or so. However, the ADF has a limitation, in that it does not consider the time-varying characteristics of coefficients and the variance covariance matrix of the innovations, although this type of time-varying characteristics exists in many macroeconomic variables and financial time series as the relationships between economic variables are not always time invariant in an economic system. Hence, capturing these time-varying relationships between economic variabels is a crucial task in economic analysis. There are many studies devoted to this topic (e.g., Cogley 2005; Primiceri 2005; Karakatsani and Bunn 2008; Galí and Gambetti 2015; Antolin-Diaz et al. 2017; Aharon and Demir 2022; Gao et al. 2023).

Recently, vector autoregressive (VAR) models with time-varying parameters have enjoyed significant popularity in time series analysis (e.g., Cogley and Sargent 2005; Primiceri 2005; Koop et al. 2009; Nakajima et al. 2011; Korobilis 2013; Baumeister and Peersman 2013; Koop et al. 2019; Huber et al. 2020; Chan 2023). Among these studies, the time-varying structural VAR (TVS-VAR) model proposed by Primiceri (2005) considers time-varying coefficients and covariances of the innovations so that it can characterize the non-linearities and time variation of both the relationships between variables and innovations. The time-varying structure of the TVS-VAR may be successfully applied to the ADF to address its time-invariant limitation.

In this study, we propose a new model—a time-varying structural approximate dynamic factor (TVS-ADF) model—by extending the ADF. The main contributions of this study are fourfold.

First, we introduce a time-varying structure similar to that of the TVS-VAR into the



Electronic copy available at: https://ssrn.com/abstract=3996594

ADF to form the TVS-ADF, which considers observed explanatory variables, time-varying coefficients, and a time-varying variance–covariance matrix of the innovations. Additionally, we provide the identification scheme of the common factors in the TVS-ADF and derive the identification theory.

Second, we propose several effective information criteria for the determination of the number of factors in the TVS-ADF. These information criteria are straightforward, hence one can easily carry out them in practical applications.

Third, we provide a Markov chain Monte Carlo (MCMC) algorithm for estimating the TVS-ADF. Although maximum likelihood estimation could be considered an alternative, the maximization of the likelihood function would be extremely difficult, if not impossible, when the dimensions of the TVS-ADF parameters are very high. The MCMC algorithm is thus a realizable choice for high-dimensional situations.

Lastly, we provide solutions for shrinkage and sparsification to avoid overparameterization. The flexibility of the TVS-ADF arising from its time-varying characteristics comes at the cost of overparameterization, which can lead to perfect in-sample fit but poor out-of-sample forecast performance. To deal with this issue, we propose shrinkage and sparsification solutions. To shrink the TVS-ADF, we use the continuous shrinkage prior (Dirichlet–Laplace prior) proposed by Bhattacharya et al. (2015), which can be expressed as global–local scale mixtures of Gaussians and facilitate computation for high-dimensional situations. As Bhattacharya et al. (2015) pointed out, under the Bayesian paradigm, sparsity is routinely induced through two-component mixture priors with a probability mass of zero; however, such priors encounter daunting computational problems in high dimensions. Hence, we do not consider the sparsification-only approach for our TVS-ADF. As another solution, Huber et al. (2020) showed that carrying out sparsification after shrinkage can yield better predictive performance in some empirical applications. In their algorithm, the shrinkage procedure, which was conducted first, enabled them to adopt a sparsification



Electronic copy available at: https://ssrn.com/abstract=3996594

procedure with low computation cost. We also adopt the approach of Huber et al. (2020) with both shrinkage and sparsification for our TVS-ADF.

It is worth mentioning that, although there are some studies that have developed dynamic factor models with time-varying parameters, they are different from our model. The models in these studies can be classified into two categories: parametric and semiparametric models. As for parametric models,[1] Del Negro and Otrok (2008) allowed the time-varying factor loadings and stochastic volatility of the innovations, but the variance–covariance matrix of the innovations is diagonal. Mumtaz and Surico (2012) allowed time-varying coefficients in the state equation of the factors, but the factor loadings and the variance of the innovations are time invariant. Combining the different characteristics of these two models, Bjørnland and Thorsrud (2019) proposed a new factor model in which more time-varying parameters are allowed, but the variance–covariance matrix of the innovations is still diagonal. Marcellino et al. (2016) proposed a mixed frequency dynamic factor model in which the disturbances of both the factor and innovations have time-varying stochastic volatilities, but the factor loadings are time invariant. Mikkelsen et al. (2019) considered time-varying factor loadings, but the variance matrix of the innovations is both time invariant and diagonal. Based on Bjørnland and Thorsrud (2019) and Marcellino et al. (2016), Thorsrud (2020) introduced time-varying factor loadings with some restrictions, but the variance matrix of the innovations is still assumed to be both time invariant and diagonal. It is evident that all the aforementioned models assume the variance–covariance matrix of the innovations to be diagonal; in other words, innovations are not allowed to have cross-sectional dependence. As Barigozzi (2018) pointed out, the dynamic factor models in which innovations are allowed to have cross-sectional dependence are the most realistic. By contrast, our TVS-ADF not only considers the cross-sectional dependence of the innovations but also allows the variance–covariance matrix of the innovations to be time

[1] Note that we only focus on dynamic factor models in which the parameters are modeled as evolving stochastically. For positing a break in the parameters, see Stock and Watson (2016) for an overview.



Electronic copy available at: https://ssrn.com/abstract=3996594

varying. Furthermore, in the models above, either the factor loadings or the variances of the innovations are time invariant. By contrast, our TVS-ADF allows both the factor loadings and the variance–covariance matrix of the innovations to be time-varying.

For semiparametric models, the main idea for capturing the dynamics of economic data is to model factor loadings as a function of time or observed variables (e.g., Motta et al., 2011; Eichler et al., 2011; Su and Wang, 2017; Ma et al., 2020; Cataño et al., 2021; Barigozzi et al., 2021; Pelger and Xiong, 2021). The main difference from our approach is that the variance–covariance matrix of the innovations in these models is specified as either diagonal or time invariant. Another drawback of these models is that semiparametric factor loadings are not easy to interpret in empirical applications.

The TVS-ADF may be related to the class of factor augmented VAR (FAVAR) models with time-varying parameters, but the models in this class have a different structure from the TVS-ADF and consist of two equations: the factor regression equation and the VAR equation. Addtionally, the factor regression equations of most models cannot capture all the time-varying characteristics that can be captured by our TVS-ADF. For example, the model proposed by Bianchi et al. (2009) includes the time-varying parameters in the VAR equation, but the factor loadings and the variances of innovations in the factor regression equation are time invariant. Liu et al. (2011) and Abbate et al. (2016) built different FAVAR models with time-varying parameters, in which the factor loadings are allowed to be time varying, but the variances of the innovations in the factor regression equation are time invariant. Korobilis (2013) proposed the time-varying parameter factor augmented VAR (TVP-FAVAR) model, in which the variances of the innovations in the factor regression equation are allowed to be time varying, but the factor loadings are constant. Subsequently, based on Korobilis (2013), Koop and Korobilis (2014) allowed the factor loadings to be time varying in the TVP-FAVAR; but, as mentioned below, this is different from our TVS-ADF, as are the other variants. All these models are different from our TVS-ADF



Electronic copy available at: https://ssrn.com/abstract=3996594

as follows. Their variance–covariance matrices of the innovations in the factor regression equation are assumed to be diagonal. Hence, they cannot allow innovations to have cross-sectional dependence. By contrast, our TVS-ADF allows the cross-sectional dependence of the innovations and their time-varying variance–covariance matrix. Additionally, in the factor regression equations of some models above, either the factor loadings are time invariant or the innovations are time invariant. However, we allow them both to be time varying in the TVS-ADF.

The rest of the paper is organized as follows. Section 2 describes the proposed TVS-ADF model, identification of factors, information criteria for the determination of the number of factors, the estimation methodology with shrinkage and sparsification, and ordering issue of variables. Section 3 conducts extensive artificial simulations. Section 4 carries out an empirical application of macroeconomic forecasting. Section 5 concludes. All the technical proofs, details of the algorithm, and additional simulations are provided in the supplementary material.

# 2 Model

We construct the TVS-ADF model with $n$ units and $T$ sample periods as follows:

$$Y_t = B_t X_t + \Lambda_t F_t + \xi_t, \qquad \xi_t \sim N(0, \Gamma_t^{\xi}), \tag{1}$$

$$F_t = C F_{t-1} + \eta_t, \qquad \eta_t \sim N(0, I), \tag{2}$$

where $Y_t$ is an $n \times 1$ vector of explained variables ($t = 1, \cdots, T$), $B_t$ is an $n \times m$ matrix of time-varying coefficients, $X_t$ is an $m \times 1$ vector of observed explanatory variables, $\Lambda_t$ is an $n \times r$ matrix of time-varying factor loadings, $F_t$ is an $r \times 1$ vector of unobserved common factors, $\xi_t$ is an $n \times 1$ vector of idiosyncratic errors, $\Gamma_t^{\xi}$ is an $n \times n$ positive definite matrix,



Electronic copy available at: https://ssrn.com/abstract=3996594

$C$ is an $r \times r$ matrix of coefficients, $\eta_t$ is an $r \times 1$ vector of innovations, $I$ is an $r \times r$ identity matrix. Specifically,

$$Y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix}_{n\times 1}, \quad B_t = \begin{pmatrix} \beta'_{1t} & 0 & \cdots & 0 \\ 0 & \beta'_{2t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta'_{nt} \end{pmatrix}_{n\times m}, \quad X_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{pmatrix}_{m\times 1},$$

where $y_{it}$ is a scalar, $\beta_{it}$ is an $\tilde{m} \times 1$ vector, $x_{it}$ is an $\tilde{m} \times 1$ vector, $i = 1, \cdots, n$, and $n\tilde{m} = m$. $\Lambda_t = (\lambda_{1t}, \cdots \lambda_{nt})'$, where $\lambda_{it}$ is an $r \times 1$ vector for $i = 1, \cdots, n$. $F_t = (F_{1t}, F_{2t}, \cdots, F_{rt})'$, where $F_{st}$ denotes the $s$-th element of $F_t$ for $s = 1, \cdots, r$. As $\Gamma_t^\xi$ is positive definite, it can be factorized with Cholesky decomposition: $\Gamma_t^\xi = A_t^{-1} H_t A_t^{-1'}$, where

$$A_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1,t} & a_{n2,t} & \cdots & 1 \end{pmatrix}, \quad H_t = \begin{pmatrix} h_{1t} & & & \\ & h_{2t} & & \\ & & \ddots & \\ & & & h_{nt} \end{pmatrix}.$$

This decomposition of the variance–covariance matrix above is a common technique used in time-varying models (see Cogley and Sargent, 2005, Cogley, 2005, Primiceri, 2005, Korobilis, 2013 and Koop and Korobilis, 2014). Similar to Primiceri (2005), we have the following transformed formula of (1), which is convenient for calculating the distributions of $A_t$ and $H_t$ conditional on the other parameters:

$$Y_t = B_t X_t + \Lambda_t F_t + A_t^{-1} H_t^{1/2} e_t, \qquad e_t \sim N(0, I). \tag{3}$$

Let $\beta_t = (\beta'_{1t}, \cdots, \beta'_{nt})'$ and $\lambda_t = (\lambda'_{1t}, \cdots, \lambda'_{nt})'$. Let $a_t$ be the vector of the non-zero and non-one elements of $A_t$ (stacked by rows) and $h_t$ be the vector of the diagonal elements of



Electronic copy available at: https://ssrn.com/abstract=3996594

$H_t$. The dynamics of the time-varying parameters are specified as follows:

$$\beta_t = \beta_{t-1} + \bar{d}_t, \qquad \bar{d}_t \sim N(0, \bar{D}), \tag{4}$$

$$\lambda_t = \lambda_{t-1} + \tilde{d}_t, \qquad \tilde{d}_t \sim N(0, \tilde{D}), \tag{5}$$

$$a_t = \rho a_{t-1} + u_t, \qquad u_t \sim N(0, S), \tag{6}$$

$$log(h_t) = \varphi log(h_{t-1}) + \gamma_t, \qquad \gamma_t \sim N(0, \Sigma), \tag{7}$$

where the diagonal matrices $\bar{D} = \text{diag}\left(\bar{D}_1, \cdots, \bar{D}_n\right)_{m\times m}$, $\tilde{D} = \text{diag}\left(\tilde{D}_1, \cdots, \tilde{D}_n\right)_{nr\times nr}$, $\bar{D}_i$ is an $\tilde{m}\times\tilde{m}$ diagonal matrix $(i = 1, \cdots, n)$ that denotes the variance of $\beta_{it}$ conditional on $t-1$, $\tilde{D}_i$ is an $r \times r$ diagonal matrix $(i = 1, \cdots, n)$ that denotes the conditional variance of $\lambda_{it}$; $\rho$ and $\varphi$ are scalar parameters; the diagonal matrix $S = \text{diag}\left(S_2, \cdots, S_n\right)_{\sum_{i=2}^{n}(i-1)\times\sum_{i=2}^{n}(i-1)}$, and for $i = 2, \cdots, n$, $S_i$ is a $(i-1)\times(i-1)$ diagonal matrix that denotes the conditional variance of $(a_{i1,t}, \cdots, a_{ii-1,t})$. Note that $e_t$, $\eta_t$, $\bar{d}_t$, $\tilde{d}_t$, $u_t$, and $\gamma_t$ are mutually independent, and $\bar{D}$, $\tilde{D}$, $S$, and $\Sigma$ are all diagonal matrices.

As previously discussed, in contrast to the ADF, our model considers additional issues related to the observed explanatory variables, time-varying coefficients, and time-varying variance–covariance matrix of the innovations. Incorporating the explanatory variables is intended to capture the impacts of known important economic variables on the explained variables other than the latent factors. Allowing for time variation in both the coefficients and variance–covariance matrix of the innovations can capture additional dynamic characteristics of the economy. Specifically, the drifting coefficients, $B_t$ and $\Lambda_t$, can capture time variation in the parameters or nonlinearities, which reflect the dynamic impact of the explanatory variables on the dependent variables. Furthermore, the time-varying variance–covariance matrix of the innovations, $\Gamma_t^{\xi}$, can capture possible dynamic heteroscedasticity in the innovations and dynamic nonlinearities in the simultaneous relationships between model variables.



Electronic copy available at: https://ssrn.com/abstract=3996594

However, these time-varying parameters can lead to overparameterization. For instance, economic variables $X_t$ may sometimes only have a very small effect on $Y_t$. Therefore, if we do not push $\beta_t$ toward zero, it will cause poor results. Additionally, the impact of some variables on $Y_t$ could be time invariant; hence, in such cases, it is critical to push the variance of $\beta_t$ toward zero. As mentioned before, we take two approaches to shrink and sparsify the TVS-ADF to avoid overparameterization: (i) only shrink the model using the approach of Bhattacharya et al. (2015) and (ii) both shrink and sparsify the model using the approach of Huber et al. (2020). We use notation TVS-ADF(s) for the model that is only shrunk and TVS-ADF(ss) for the model that is both shrunk and sparsified. The methods for shrinkage and sparsification are incorporated into the MCMC algorithm and are described in detail in Appendix B of the supplementary material.

## 2.1 Identification of factors

To ensure that the common factors can be identified, we still need some additional assumptions for the TVS-ADF as follows:

**Assumption 1.** *Let $\lambda_{i0}$ denote the initial value of $\lambda_{it}$ for $i = 1, \cdots, n$. $E(\lambda_{i0}) \neq 0$ and $E(\lambda_{i0}\lambda'_{i0})$ is a positive definite matrix.*

**Assumption 2.** *$C$ can be decomposed as $C = PMP^{-1}$, where $P$ and $M$ denote the eigenvector matrix and eigenvalue matrix of $C$, respectively. All eigenvalues of $C$ are nonzero and the maximum absolute value of the eigenvalues is smaller than 1.*

**Assumption 3.** *Let $F_0$ denote the initial value of $F_t$. $E(F_0F'_0)$ can be factorized as $E(F_0F'_0) = P_0M_0P'_0$, where $P_0$ and $M_0$ denote the eigenvector matix and eigenvalue matrix of $E(F_0F'_0)$, respectively, and $M_0$ is a positive definite matrix.*

**Assumption 4.** *Let $\gamma_t^{\xi,(n)}$ denote the maximum eigenvalue of $\Gamma_t^{\xi}$ satisfying the equations (6) and (7) with $|\rho| < 1$ and $|\varphi| < 1$. For all $t$, $\gamma_t^{\xi,(n)}$ is $o_p(n)$.*



Electronic copy available at: https://ssrn.com/abstract=3996594

Next, we will detailedly elucidate the identification scheme of the common factors in the TVS-ADF.

**Theorem 1.** *Suppose that Assumptions 1, 2, 3, and 4 hold, then the common component* $\Lambda_t F_t$ *and idiosyncratic component* $\xi_t$ *can be seperated in probability as* $n \to \infty$.

In Theorem 1, we allow the cumulative variance of the common component across $n$ units to be $O(n)$ by Assumptions 1, 2, and 3, while we impose some restrictions on $\Gamma_t^{\xi}$ by Assumption 4 to make the cumulative sum of all elements in $\Gamma_t^{\xi}$ is $o_p(n)$ (in other words, only the mild cross-sectional dependence of $\xi_t$ is allowed in $\Gamma_t^{\xi}$ as $n$ increases). Then, in this situation, the factors have an intuitive meaning from the perspective of the aggregation of all units across series, that is, the cumulative sum of the variance of the common component across series can diverge faster as $n \to \infty$ relative to the idiosyncratic component.

Although we can disentangle the common component from the idiosyncratic component, the loadings and factors cannot be identified without restrictions. Since, for an arbitrary $r \times r$ invertible matrix $Q$, it always holds that $\Lambda_t F_t = \Lambda_t Q Q^{-1} F_t$. Obviously, there are $r^2$ free elements in $Q$. Hence, we need $r^2$ restrictions to identify $\Lambda_t$ and $F_t$. Note that, as a matter of fact, the $r^2$ restrictions already exist in our model specification. First, in the TVS-ADF, the variance–covariance matrix of $\eta_t$ is set as an identity matrix, which is a standard normalization assumption for factor models. This provides $r(r+1)/2$ restrictions on the conditional variance–covariance matrix of $F_t$. Second, diagonal $\tilde{D}_i$ for all $i$ (i.e., the loadings of different latent factors are uncorrelated) implies that the variance–covariance matrix of $\Lambda_t$ (i.e., $E\Lambda_t' \Lambda_t - (E\Lambda_t)' E\Lambda_t$) is diagonal, which provides additional $r(r-1)/2$ restrictions. Finally, $\Lambda_t$ and $F_t$ are identified with the $r^2$ restrictions.

**Remark 1** One may concern that $\Gamma_t^{\xi}$ with the dynamics of (6) and (7) can not statisfy Assumption 4. But, actually, (6) and (7) are not contradictory to Assumption 4. For



Electronic copy available at: https://ssrn.com/abstract=3996594

simplicity, let us consider a three-dimensional example:

$$\Gamma_t^{\xi} = \begin{pmatrix} 1 & 0 & 0 \\ a_{21,t} & 1 & 0 \\ a_{31,t} & a_{32,t} & 1 \end{pmatrix}^{-1} \begin{pmatrix} h_{1t} & & \\ & h_{2t} & \\ & & h_{3t} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ a_{21,t} & 1 & 0 \\ a_{31,t} & a_{32,t} & 1 \end{pmatrix}^{-1'}$$

$$= \begin{pmatrix} h_{1t} & -a_{21,t}h_{1t} & (a_{21,t}a_{32,t} - a_{31,t})h_{1t} \\ -a_{21,t}h_{1t} & a_{21,t}^2 h_{1t} + h_{2t} & (a_{31,t} - a_{21,t}a_{32,t})h_{1t} - a_{32,t}h_{2t} \\ (a_{21,t}a_{32,t} - a_{31,t})h_{1t} & (a_{31,t} - a_{21,t}a_{32,t})h_{1t} - a_{32,t}h_{2t} & (a_{21,t}a_{32,t} - a_{31,t})^2 h_{1t} + a_{32,t}^2 h_{2t} + h_{3t} \end{pmatrix}$$

If $a_{31,t-1} = 0$, $a_{32,t-1} = 0$, $S_{31} = 0$ (the conditional variance of $a_{31,t}$), and $S_{32} = 0$ (the conditional variance of $a_{32,t}$), then

$$\Gamma_t^{\xi} = \begin{pmatrix} h_{1t} & -a_{21,t}h_{1t} & 0 \\ -a_{21,t}h_{1t} & a_{21,t}^2 h_{1t} + h_{2t} & 0 \\ 0 & 0 & h_{3t} \end{pmatrix}$$

according to (6), which implies that (6) and (7) can make $\Gamma_t^{\xi}$ have enough zeros to satisfy Assumption 4. Hence, (6), (7), and Assumption 4 can be assumed to hold together.

**Remark 2** Regarding time variation in the coefficients and factors, one may concern that $Q$ can be time varying instead of being constant. However, this is impossible under our model setting because rescaled factors $Q_t^{-1}F_t$ and rescaled loadings $\Lambda_t Q_t$ cannot satisfy (2) and (5).

**Remark 3** Although this study focuses on the identification method of Theorem 1, we discuss two other identification schemes to reveal that the common factors can also be identified from the perspective of the aggregation of all units across time.

The first one is that we allow the common factors $F_t$ to have autocorrelations and some



Electronic copy available at: https://ssrn.com/abstract=3996594

factors in $F_t$ are non-stationary (which means that the maximum absolute value of the eigenvalues of $C$ is equal to or greater than 1), while the idiosyncratic errors are not allowed to have autocorrelations and their covariances are bounded in probability; addtionally, we need to suppose that $n$ is fixed or $n \to \infty$. The factors have an intuitive meaning in terms of the dynamic aggregation of all units across time, that is, the cumulative variances of the common component $\Lambda_t F_t$ across time can diverge as $T$ increases, whereas the cumulative sum of the variance of the idiosyncratic component $\xi_t$ is bounded in probability as $T \to \infty$. Hence, the common component can be separated from the idiosyncratic component, which means that $\lambda_t F_t$ can capture the main information of all units (i.e., the cumulative dynamic variances and covariances of the unstationary factors during the whole period and stationary factors over infinite finite-time intervals) as $T \to \infty$. For the assumption about $n$, we use an expamle to explain. For instance, there are 10 variables having latent common factors in an economic system; if we only consider one of these 10 variables and omit the other variables, then the common factors identified by this one variable may not be the same as those obtained by the 10 variables (even though the common component of this one variable can be separated from the idiosyncratic component). Hence, one either assumes a fixed $n$, or assumes $n \to \infty$ in a practical application.

The second scheme is, on the basis of the first method mentioned above, to impose Assumption 4 of Theorem 1 on $\Gamma_t^{\xi}$ to deal with the identifiation of factors. Then, it is easy to obtain that the common component can be disentangled from the idiosyncratic component as $n \to \infty$ or $T \to \infty$ in probability.

## 2.2 Number of factors

Let $r_0$ denote the number of the common factors and suppose that $r_0 \in \{1, 2, \cdots, R\}$ where $R$ is a positive integer. Drawing on the idea of Bai and Ng (2002), we propose four



Electronic copy available at: https://ssrn.com/abstract=3996594

information criteria ($IC$) to determine the number of factors as follows [2]:

$$
\begin{aligned}
IC_1(r) &= \frac{1}{nT}\sum_{t=1}^{T} \mathrm{tr}\left[\hat{\Gamma}_t^{\xi,r}\right] + r\frac{\log(n)}{n+T},\\
IC_2(r) &= \frac{1}{nT}\sum_{t=1}^{T} \mathrm{tr}\left[\hat{\Gamma}_t^{\xi,r}\right] + r\frac{\log(n+T)}{n+T},\\
IC_3(r) &= \frac{1}{nT}\sum_{t=1}^{T} \mathrm{tr}\left[\hat{\Gamma}_t^{\xi,r}\right] + r\frac{\log(T)}{n+T},\\
IC_4(r) &= \frac{1}{nT}\sum_{t=1}^{T} \mathrm{tr}\left[\hat{\Gamma}_t^{\xi,r}\right] + r\frac{\log(nT)}{n+T},
\end{aligned}
$$

where $\hat{\Gamma}_t^{\xi,r}$ denotes the estimate of $\Gamma_t^{\xi}$ with $r$ factors. Then we can obtain an estimate of the number of factors using an information criterion, that is,

$$
\hat{r}_j = \operatorname*{arg\,min}_{r\in\{1,\cdots,R\}} IC_j(r), \quad j = 1, \cdots, 4.
$$

The first terms of the criteria above are motivated by a bootstrap method. To describe this motivation better, we need to introduce additional notations. We rewrite $B_t$, $\Lambda_t$, $F_t$, and $\Gamma_t^{\xi}$ as $B_t^r$, $\Lambda_t^r$, $F_t^r$, and $\Gamma_t^{\xi,r}$, respectively, to signify that there are $r$ common factors in the TVS-ADF. For each $r$, we generate $N$ (sufficiently large) sequences, which are denoted by $Y^{r,i} = (Y_1^{r,i}, \cdots, Y_T^{r,i})$ for $i = 1, \cdots, N$, following

$$
Y_t^{r,i} = \hat{B}_t^r X_t + \hat{\Lambda}_t^r \hat{F}_t^r + \varepsilon_t^i, \quad \varepsilon_t^i \sim N(0, \hat{\Gamma}_t^{\xi,r}), \quad t = 1, \cdots, T,
$$

where $\hat{B}_t^r$, $\hat{\Lambda}_t^r$, $\hat{F}_t^r$, and $\hat{\Gamma}_t^{\xi,r}$ denote the estimates of $B_t^r$, $\Lambda_t^r$, $F_t^r$, and $\Gamma_t^{\xi,r}$; additionally, we

[2] We also consider log forms of these four information criteria. Details see Appendix C.2 of the supplementary material.



Electronic copy available at: https://ssrn.com/abstract=3996594

draw $\varepsilon_t^i$ from $N(0, \hat{\Gamma}_t^{\xi,r})$. Then we can choose a number of factors, $r^*$, by

$$r^* = \underset{r\in\{1,\cdots,R\}}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \|Y_t^{r,i} - \hat{B}_t^r X_t - \hat{\Lambda}_t^r \hat{F}_t^r\|^2 = \underset{r\in\{1,\cdots,R\}}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \|\varepsilon_t^i\|^2.$$

Let $\varepsilon_{t,j}^i$ denote $j$-th row of $\varepsilon_t^i$. Since as $N \to \infty$, $\frac{1}{N}\sum_{i=1}^{N}\left(\varepsilon_{t,j}^i\right)^2 \to E[\left(\varepsilon_{t,j}^i\right)^2] = \hat{\Gamma}_{t,jj}^{\xi,r}$, where $\hat{\Gamma}_{t,jj}^{\xi,r}$ denotes the $j$-th diagonal element of $\hat{\Gamma}_t^{\xi,r}$, we have $1/N\sum_{i=1}^{N}\sum_{t=1}^{T}\|\varepsilon_t^i\|^2 \to \sum_{t=1}^{T}\text{tr}\left[\hat{\Gamma}_t^{\xi,r}\right]$ as $N \to \infty$.

The second term of each criterion above is a penalty term, which is used to ensure that under and overparameterized models are not be chosen. Since the cardinality of $Y_t$, $n$, is a crucial element for the identification of factors in the TVS-ADF according to Theorem 1; additionally, the sample size $T$ can affect the estimation results of parameters, we propose four different penalties both depending $n$ and $T$. Note that these four penalty terms can vanish gradually as $n \to \infty$ and $T \to \infty$.

## 2.3 MCMC algorithm for estimation

We use the MCMC algorithm to estimate the model. As it is difficult and complex to obtain the closed form of the joint posterior distribution of all parameters, we simulate the joint posterior distribution using Gibbs sampling, sequentially drawing the parameters of the TVS-ADF from the conditional posterior distributions. Our proposed algorithm including shrinkage and sparsification for the Gibbs sampling comprises seven steps (for details, see Appendix B of the supplementary material): drawing $B_t$ and $\Lambda_t$; drawing $F_t$; darwing $A_t$; drawing $H_t$; drawing $C$, $\rho$ and $\phi$; drawing some important parameters in the prior distributions we use; drawing some hyperparameters. In our algorithm, we will conduct shrinkage and sparsification for the time-varying coeffeicients and covariances of the innovations. First the shrinkage is carried out using the Dirichlet–Laplace prior, then the sparsification is done by a method proposed by Huber et al. (2020).



Electronic copy available at: https://ssrn.com/abstract=3996594

## 2.4 The ordering issue

In empirical application, we could face the ordering issue of the variables in $Y_t$ (i.e., $y_{1t}, \cdots, y_{nt}$) due to Cholesky decomposition. First, the order of these variables affects their distributions conditional on $t-1$. For simplicity, let us consider a two-variable example:

$$\Gamma_t^{\xi} = A_t^{-1} H_t A_t^{-1'} = \begin{pmatrix} 1 & \\ a_{21,t} & 1 \end{pmatrix} \begin{pmatrix} h_{1t} & \\ & h_{2t} \end{pmatrix} \begin{pmatrix} 1 & a_{21,t} \\ & 1 \end{pmatrix} = \begin{pmatrix} h_{1t} & a_{21,t} h_{1t} \\ a_{21,t} h_{1t} & a_{21,t}^2 h_{1t} + h_{2t} \end{pmatrix}$$
$$= \begin{pmatrix} e^{log(h_{1t-1})+\gamma_{1t}} & a_{21,t} h_{1t} \\ a_{21,t} h_{1t} & (a_{21,t-1} + u_{21,t}) e^{log(h_{1t-1})+\gamma_{1t}} + e^{log(h_{2t-1})+\gamma_{2t}} \end{pmatrix}.$$

The expression above clarifies that, conditional on the information at $t-1$, the distribution of the first diagonal element of $\Gamma_t^{\xi}$ is a log-normal distribution, whereas the second diagonal element is not. Hence, different orderings will imply different distributions for the variables in $Y_t$, which could affect the model's predictive results.

Second, the order of $(y_{1t}, \cdots, y_{nt})$ essentially determines the contemporaneous reaction relationships between different variables. To describe this clearly, let us consider a three-demensional $Y_t$. According to (B.10) in Appendix B of the supplementary material, we have

$$\xi_{1t} = \sqrt{h_{1t}} e_{1t} \tag{8}$$

$$\xi_{2t} = \sqrt{h_{2t}} e_{2t} - a_{21,t} \xi_{1t} \tag{9}$$

$$\xi_{3t} = \sqrt{h_{3t}} e_{3t} - a_{31,t} \xi_{1t} - a_{32,t} \xi_{2t}. \tag{10}$$

Obviously, if $\xi_{1t}$ increases one unit at $t$ (thus $y_{1t}$ will increase one unit), then $\xi_{1t}$ will have a contemporaneous effect on $\xi_{2t}$ and $\xi_{3t}$ (thus $y_{2t}$ and $y_{3t}$ will be affected) according to (9) and (10). However, if $\xi_{2t}$ increases one unit at $t$, then it will only have a contemporaneous



Electronic copy available at: https://ssrn.com/abstract=3996594

effect on $\xi_{3t}$ but not have a contemporaneous effect on $\xi_{1t}$ according to (10). Similarly for $\xi_{3t}$, it can not have a contemporaneous effect on $\xi_{1t}$ and $\xi_{2t}$. In other words, $y_{1t}$ can contemporaneously affect $y_{2t}$ and $y_{3t}$, but not conversely; $y_{2t}$ can contemporaneously affect $y_{3t}$, but not conversely.

The second characteristic above gives us an important implication that one can determine the ordering of variables according to the contemporaneous reaction relationships between different variables, which can make empirical analysis be meaningful. For instance, in monetary policy analysis, inflation react to the policy instrument (e.g., interest rate) with at least one period of lag (in other words, the policy instrument does not contemporaneously affect inflation), then we can put inflation in the front of the policy instrument.

# 3 Artificial simulation

We first present evidence on the performance of our model based on simulation experiments using artificial data generated from the TVS-ADF in Section 3.1; then, in Section 3.2, we show the the performance of the proposed information criteria for the determination of the number of factors in the TVS-ADF.

## 3.1 Point forecasts

To assess how well the different models perform across different numbers of explained variables, numbers of explanatory variables, degrees of sparsity, and lengths of time series, we set $n = 10, 20, 30$; for each $n$, we consider three sparsity levels, labeled as dense (with 10% zeros in $\alpha_i$ in (B.4)), moderate (with 40% zeros), and sparse (with 70% zeros); For each sparsity level, we consider 4 explanatory variables (consisting of $\tilde{m} = 2$ obeserved variables and $r = 2$ unobserved common factors) and 8 explanatory variables (consisting of $\tilde{m} = 4$ obeserved variables and $r = 4$ unobserved common factors); we set sample sizes



Electronic copy available at: https://ssrn.com/abstract=3996594

$T = 100, 200$. We randomly generate $N = 100$ simulated datasets for each variant. We set $x_{it} \sim N(0, I_x)$, $S = 0.1^2 I_S, \Sigma = 0.1^2 I_\Sigma, \bar{D} = 0.1^2 I_{\bar{D}}, \tilde{D} = 0.1^2 I_{\tilde{D}}$, where $I_x$, $I_S$, $I_\Sigma$, $I_{\bar{D}}$ and $I_{\tilde{D}}$ are identity matrices of dimensions $\tilde{m} \times \tilde{m}$, $[n \times (n-1)/2] \times [n \times (n-1)/2]$, $n \times n$, $n\tilde{m} \times n\tilde{m}$ and $nr \times nr$, respectively. Moreover, for $r = 2$, we set

$$C = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.3 \end{pmatrix};$$

for $r = 4$, we set

$$C = \begin{pmatrix} 0.5 & 0.1 & 0 & 0 \\ 0.1 & 0.3 & 0 & 0 \\ 0 & 0 & 0.6 & -0.1 \\ 0 & 0 & -0.1 & 0.4 \end{pmatrix}.$$

The initial value of $F_t$ is set to zero. The initial value of $\Gamma_t^\xi$, $\Gamma_1^\xi$, is generated by $1/2(a + a')$, where $a$ is an $n \times n$ matrix and each element of $a$ is generated from a normal distribution, $N(0, 0.1)$. Each diagonal element of $\Gamma_1^\xi$ is replaced by one to ensure $\Gamma_1^\xi$ is a positive definite matrix. Then, we can easily obtain the initial values of $A_t$ and $H_t$ using the Cholesky decomposition for $\Gamma_1^\xi$, respectively. Next, we set 70%, 90%, and 95% zeros in $A_1$ and $S$ for $n = 10$, 20, and 30, respectively, to ensure $\gamma_t^{\xi,(n)} = o_p(n)$.

We use the two-step method of Forni et al. (2009) to estimate the ADF and the MCMC algorithm described in Section 3 to estimate the TVS-ADF. We use the first $T$ observations to estimate the models, and then the resulting estimates to predict the $T+1$-th observation. The precision of point forecasts is measured by the following mean-squared error: $\text{MSE}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \text{MSE}_i$, $i = 1, \cdots, n$, where $\text{MSE}_i = \frac{1}{N}\sum_{j=1}^{N}(y_{i,T+1}^j - \hat{y}_{i,T+1}^j)^2$, $y_{i,T+1}^j$ refers to the $T+1$-th observation of the $i$-th unit (i.e., $i$-th explained variable) in the $j$-th simulated dataset and $\hat{y}_{i,T+1}^j$ denotes its fitted value. Note that $\text{MSE}^{(n)}$ measures the predictive precision of all explained variables.



Electronic copy available at: https://ssrn.com/abstract=3996594

We tabulate $\text{MSE}^{(n)}$ for three models with different numbers of explained variables, numbers of explanatory variables, sample sizes, and sparsity levels in Table 1. We can make the following observations. First, in the most cases, the mean-squared errors of the three models become smaller as the sparsity level gradually increases. Considering the time-varying characteristic tends to become weaker as the sparsity level rises and more parameters become constants, we think that this could cause $\text{MSE}^{(n)}$ for the ADF to decline gradually. Additionally, for both TVS-ADF(s) and TVS-ADF(ss), the increase in the sparsity level could be conducive to an improvement in the shrinakge and sparsification.

Second, $\text{MSE}^{(n)}$ of the TVS-ADF(s) and TVS-ADF(ss) increases as sample size $T$ becomes large. This may be because although increasing the sample size benefits estimation accuracy, it also increases the number of unknown parameters; thus, an increase in the number of unknown parameters can offset the benefit of a larger sample. Moreover, the $\text{MSE}^{(n)}$ for the ADF has a similar tendency. A possible reason is that, in the random walk process, increasing time $T$ can make parameters have larger aggregate movements, which can offset the benefit of a larger sample.

Third, TVS-ADF(s) and TVS-ADF(ss) show substantial improvements with respect to predictive accuracy relative to the ADF in all cases, which indicates that our models can better capture time-varying information. Moreover, the TVS-ADF(s) and TVS-ADF(ss) yield similar outcomes in most cases, but the TVS-ADF(s) performs slightly better than the TVS-ADF(ss).

Overall, the TVS-ADF(s) and TVS-ADF(ss) almost always perform better than the ADF for different numbers of explained variables, numbers of explanatory variables, sparsity levels, and sample sizes. Furthermore, the TVS-ADF(s) and TVS-ADF(ss) display similar results.



Electronic copy available at: https://ssrn.com/abstract=3996594

Table 1: $\text{MSE}^{(n)}$ of out-of-sample point forecasts

| | | ADF | TVS-ADF(s) | TVS-ADF(ss) | ADF | TVS-ADF(s) | TVS-ADF(ss) |
|---|---|---|---|---|---|---|---|
| | | | $\tilde{m}=2$, $r=2$ | | | $\tilde{m}=4$, $r=4$ | |
| $n=10$ | | | | | | | |
| | | | $T=100$ | | | $T=100$ | |
| | dense(10%) | 5.34644 | 3.12305 | 3.38085 | 8.74673 | 7.15022 | 7.43124 |
| | moderate(40%) | 4.21234 | 2.81839 | 2.95329 | 5.10891 | 3.44360 | 3.64375 |
| | sparse(70%) | 2.43962 | 2.41033 | 2.43900 | 3.99359 | 1.92644 | 2.10490 |
| | | | $T=200$ | | | $T=200$ | |
| | dense(10%) | 8.82200 | 4.46352 | 4.57566 | 17.34683 | 12.37428 | 12.66816 |
| | moderate(40%) | 4.58688 | 3.38852 | 3.55007 | 11.10494 | 7.69730 | 8.15528 |
| | sparse(70%) | 5.53746 | 5.36589 | 5.42170 | 5.37767 | 2.85125 | 3.10249 |
| $n=20$ | | | | | | | |
| | | | $T=100$ | | | $T=100$ | |
| | dense(10%) | 4.64027 | 3.42310 | 3.61690 | 8.41308 | 6.71999 | 6.98837 |
| | moderate(40%) | 3.45378 | 2.63137 | 2.75044 | 5.30972 | 4.03870 | 4.26704 |
| | sparse(70%) | 2.33500 | 2.08634 | 2.16436 | 3.60206 | 2.15200 | 2.28437 |
| | | | $T=200$ | | | $T=200$ | |
| | dense(10%) | 8.93815 | 4.90091 | 5.03344 | 17.55887 | 11.44914 | 11.87710 |
| | moderate(40%) | 4.44473 | 3.46787 | 3.63615 | 10.74631 | 8.25037 | 8.57664 |
| | sparse(70%) | 4.41872 | 3.97067 | 4.08648 | 4.43319 | 2.33699 | 2.51044 |
| $n=30$ | | | | | | | |
| | | | $T=100$ | | | $T=100$ | |
| | dense(10%) | 4.58403 | 3.53453 | 3.72323 | 8.80589 | 7.33162 | 7.58540 |
| | moderate(40%) | 3.50203 | 2.71417 | 2.83595 | 5.82228 | 4.35927 | 4.60039 |
| | sparse(70%) | 2.21307 | 2.05951 | 2.12199 | 3.48129 | 2.62252 | 2.72115 |
| | | | $T=200$ | | | $T=200$ | |
| | dense(10%) | 8.93286 | 5.09744 | 5.26160 | 17.47703 | 11.91445 | 12.39145 |
| | moderate(40%) | 5.43414 | 4.10248 | 4.26533 | 11.19039 | 8.17423 | 8.49913 |
| | sparse(70%) | 4.04739 | 3.67277 | 3.75833 | 5.08611 | 3.39573 | 3.60816 |



Electronic copy available at: https://ssrn.com/abstract=3996594

## 3.2 Information criteria

To assess the performance of the proposed information criteria across different numbers of factors, we set the true number of factors $r_0 = 2$ and 4; additionaly, for each $r_0$, we consider $n = 10$, 20 and $T = 100$, 200. The other settings (e.g., the sparsity levels and parameter settings) are the same as that described in section 3.1. For all cases, each information criterion determines the number of factors $\hat{r}$ from the set $\{1, 2, 3, 4, 5\}$. We carry out 100 replications for each variant.

Table 2 reports the averages of $\hat{r}$ determined by different $IC$ over 100 replications for $r_0 = 2$ and 4, respectively [3]. First, for $r_0 = 2$, the performance of all information criteria can improve for the determination of the number of factors in the TVS-ADF(s) with different sparsity levels as $T$ increases. Additionally, $IC_1$ outperforms the other information criterions in most cases. For TVS-ADF(ss), the $IC_2$, $IC_3$, and $IC_4$ can provide more suitable results in most cases relative to the $IC_1$. Moreover, the $IC_2$, $IC_3$, and $IC_4$ have a better perfromance as $n$ and $T$ increases. For $r_0 = 4$, all information criteria have similar performance to $r_0 = 2$ for the TVS-ADF(s) and TVS-ADF(ss). Furthermore, all information criteria are inclined to underestimate the number of factors as the sparsity level increases in some cases regardless of $r_0 = 2$ or $r_0 = 4$. We think that since the high sparsity level can decrease the degree of time-varying characteristics in our model, the information of observations over $n$ units used for capturing common factors can be weakened in comparison with the low sparsity level.

[3] We also provide the results of the log-form information criteria in Appendix C.2.



Electronic copy available at: https://ssrn.com/abstract=3996594

Table 2: The number of factors $\hat{r}$ determined by different information criteria

| | | TVS-ADF(s) | | | | TVS-ADF(ss) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $IC_1$ | $IC_2$ | $IC_3$ | $IC_4$ | $IC_1$ | $IC_2$ | $IC_3$ | $IC_4$ |
| $r_0 = 2$ | | | | | | | | | |
| | | $n = 10, T = 100$ | | | | $n = 10, T = 100$ | | | |
| | dense(10%) | 1.16 | 1.00 | 1.00 | 1.00 | 2.95 | 1.54 | 1.60 | 1.21 |
| | moderate(40%) | 1.44 | 1.12 | 1.13 | 1.04 | 3.79 | 2.13 | 2.14 | 1.42 |
| | sparse(70%) | 1.08 | 1.01 | 1.01 | 1.00 | 3.89 | 1.58 | 1.65 | 1.11 |
| | | $n = 10, T = 200$ | | | | $n = 10, T = 200$ | | | |
| | dense(10%) | 2.28 | 1.72 | 1.73 | 1.35 | 4.09 | 2.93 | 3.00 | 2.62 |
| | moderate(40%) | 1.60 | 1.08 | 1.08 | 1.02 | 4.01 | 2.60 | 2.60 | 1.96 |
| | sparse(70%) | 2.51 | 2.12 | 2.12 | 1.91 | 4.35 | 3.60 | 3.60 | 2.60 |
| | | $n = 20, T = 100$ | | | | $n = 20, T = 100$ | | | |
| | dense(10%) | 1.19 | 1.00 | 1.03 | 1.00 | 2.27 | 1.74 | 1.78 | 1.31 |
| | moderate(40%) | 1.02 | 1.00 | 1.00 | 1.00 | 2.12 | 1.46 | 1.52 | 1.08 |
| | sparse(70%) | 1.00 | 1.00 | 1.00 | 1.00 | 1.60 | 1.12 | 1.14 | 1.00 |
| | | $n = 20, T = 200$ | | | | $n = 20, T = 200$ | | | |
| | dense(10%) | 1.99 | 1.50 | 1.52 | 1.16 | 3.04 | 2.48 | 2.48 | 2.07 |
| | moderate(40%) | 1.75 | 1.22 | 1.26 | 1.06 | 3.09 | 2.21 | 2.27 | 1.84 |
| | sparse(70%) | 1.97 | 1.67 | 1.67 | 1.37 | 3.55 | 2.49 | 2.52 | 2.09 |
| $r_0 = 4$ | | | | | | | | | |
| | | $n = 10, T = 100$ | | | | $n = 10, T = 100$ | | | |
| | dense(10%) | 2.92 | 1.83 | 1.91 | 1.48 | 4.43 | 3.75 | 3.82 | 2.73 |
| | moderate(40%) | 1.76 | 1.27 | 1.28 | 1.14 | 3.60 | 2.21 | 2.24 | 1.64 |
| | sparse(70%) | 1.08 | 1.00 | 1.00 | 1.00 | 3.35 | 1.49 | 1.55 | 1.11 |
| | | $n = 10, T = 200$ | | | | $n = 10, T = 200$ | | | |
| | dense(10%) | 3.51 | 2.75 | 2.75 | 2.41 | 4.35 | 3.89 | 3.91 | 3.58 |
| | moderate(40%) | 3.02 | 2.21 | 2.21 | 1.94 | 4.32 | 3.69 | 3.72 | 3.21 |
| | sparse(70%) | 2.24 | 1.43 | 1.43 | 1.18 | 3.97 | 2.36 | 2.37 | 1.79 |
| | | $n = 20, T = 100$ | | | | $n = 20, T = 100$ | | | |
| | dense(10%) | 2.14 | 1.48 | 1.50 | 1.16 | 3.70 | 2.52 | 2.71 | 1.78 |
| | moderate(40%) | 1.30 | 1.07 | 1.10 | 1.01 | 2.49 | 1.74 | 1.77 | 1.24 |
| | sparse(70%) | 1.01 | 1.00 | 1.00 | 1.00 | 1.76 | 1.20 | 1.21 | 1.02 |
| | | $n = 20, T = 200$ | | | | $n = 20, T = 200$ | | | |
| | dense(10%) | 3.34 | 2.60 | 2.63 | 2.08 | 4.16 | 3.77 | 3.78 | 3.21 |
| | moderate(40%) | 2.43 | 1.86 | 1.86 | 1.56 | 3.91 | 3.27 | 3.27 | 2.47 |
| | sparse(70%) | 1.30 | 1.07 | 1.07 | 1.01 | 2.82 | 1.71 | 1.73 | 1.16 |



Electronic copy available at: https://ssrn.com/abstract=3996594

# 4 Empirical application

We use the FRED-MD database from McCracken and Ng (2016), which consists of monthly US macroeconomic data. The sample period is from January 1998 to September 2023. The dataset includes economic variables in eight groups: output and income; labor market; housing; consumption, orders, and inventories; money and credit; interest and exchange rates; prices; and stock market. We choose 21 representative variables from different groups by selecting the highest-level indices in each group. Then, these variables are standardized using the transformation codes provided by McCracken and Ng (2016). Subsequently, we divide these variables into six groups following the block method of Belviso and Milani (2006) and Korobilis (2013). Then we use the strategy mentioned in section 2.4 to determine the ordering of these groups as follows: real activity; money, credit, and finance; exchange rate; price; expectations; and monetary policy (interest rate). Table 5 (see Appendix C.1 of the supplementary material) details the variables. We conduct point forecasts for these variables using the ADF and TVS-ADF, respectively. We determine the number of factors in the TVS-ADF using the four information criteria in section 2.2, while we use the three information criteria ($PC_{p1}$, $PC_{p2}$, and $PC_{p3}$) proposed by Bai and Ng (2002) to choose the number of factors for the ADF following Forni et al. (2009). We set $r \in \{1, 2, 3, \cdots, 10\}$. We take the first lags, second lags, and third lags of the 21 economic variables as three observed explanatory variables.

We consider $h$-step-ahead point forecasts, $h = 1, 2, 3, 4$, and two options, $T = 100$ and 200 as the sample sizes. For each variant, we adopt the following rolling window scheme to carry out point forecasts. For example, let us consider $T = 100$ and the 1-step-ahead point forecasts (i.e., $h = 1$). As the starting point of the rolling window, we use the first 100 observations from the sample period, January 1998 to April 2006, to estimate the models, which are then used to predict the outcomes for May 2006. Then, we move the rolling



Electronic copy available at: https://ssrn.com/abstract=3996594

window one step ahead (i.e., the sample period is from February 1998 to May 2006) and use the resulting estimates to predict the outcomes for June 2006. We proceed recursively 100 times in this fashion and obtain a sequence of forecasts from May 2006 to August 2014. Similarly for the other variants.

We measure the precision of the $h$-step-ahead point forecasts for the $i$-th explained variable using the mean-squared error: $\mathrm{MSE}_i^h = \frac{1}{l}\sum_{t=T+h}^{T+h+l}(y_{it} - \hat{y}_{it})^2$, $i = 1, \cdots, n$, where $l = 100$ (i.e., 100 times), $n = 21$ (i.e., 21 variables), and $\hat{y}_{it}$ denotes the predicted value of $y_{it}$. Additionally, we measure the predictive accuracy of all explained variables using $\mathrm{MSE}_{(n)}^h = 1/n \sum_{i=1}^n \mathrm{MSE}_i^h$. To measure the predictive accuracy of all explained variables on the time dimension, we use the cumulative sum of forecasting errors: $\mathrm{CSE}_{(n),\tau}^h = \sum_{t=T+h}^{\tau} \mathrm{SE}_t$, $\tau = T + h, \cdots, T + h + l$, for all explained variables, where $\mathrm{SE}_t = 1/n \sum_{i=1}^n (y_{it} - \hat{y}_{it})^2$.

Tables 3 and 4 present the $\mathrm{MSE}_i^h$ and $\mathrm{MSE}_{(n)}^h$ for the $h$-step-ahead point forecasts of the three models for different sample sizes[4]. The deep gray figures indicate the lowest $\mathrm{MSE}_i^h$ across the three models for a given sample size, while the light gray figures are the second lowest $\mathrm{MSE}_i^h$. The last line of each block with a specific $h$ in the tables gives the $\mathrm{MSE}_{(n)}^h$. The values of $\mathrm{MSE}_i^h$ show that the predictive performances of the TVS-ADF(ss) and TVS-ADF(s) are better than that of the ADF for most economic variables, regardless of sample size and forecast horizon $h$, which could arise from the capacity of our models to capture economic dynamics. Additionally, the results for the TVS-ADF(ss) and TVS-ADF(s) are similar for different cases.

Figures 1 and 2 report $\mathrm{CSE}_{(n),\tau}^h$ of $h$-step-ahead point forecasts of the three models for different sample sizes, which illustrates the increasing path of the predictive error. We

[4]Note that, to simplify empirical exercise and present results better, we first calculate the information criteria using the estimation results of each model based on all observations (i.e., the data from January 1998 to September 2023) to select the number of factors; then we conduct 100 rolling windows for each model using a fixed $r$ determined in the first step. All $IC$ of the TVS-ADF report $\hat{r} = 5$, whereas three criteria of the ADF support $\hat{r} = 4$. To convince readers better, we also present the results of the three models for different fixed numbers of factors (i.e., $r = 1, 2, 3, 4, 5$) in Appendix C.3 of the supplementary material. Additionally, we provide the density forecasts of the explained variables in the TVS-ADF and compare the TVS-ADF with other classical models in Appendix C.4 and C.5, respectively.



Electronic copy available at: https://ssrn.com/abstract=3996594

Table 3: $\text{MSE}_i^h$ of $h$-step-ahead point forecasts of three models for different sample sizes

| | | $T=100$ | | | $T=200$ | | |
|---|---|---|---|---|---|---|---|
| | | ADF | TVS-ADF(s) | TVS-ADF(ss) | ADF | TVS-ADF(s) | TVS-ADF(ss) |
| $h=1$ | **Group 1: Real Activity** | | | | | | |
| | CLF | 0.0000450 | 0.0000055 | 0.0000049 | 0.0000825 | 0.0000298 | 0.0000253 |
| | CE | 0.0000915 | 0.0000067 | 0.0000065 | 0.0007973 | 0.0007076 | 0.0002664 |
| | CUR | 0.3965476 | 0.0315664 | 0.0377275 | 3.3851808 | 1.4075778 | 1.2677961 |
| | RPI | 0.0003357 | 0.0000809 | 0.0000841 | 0.0015409 | 0.0011144 | 0.0008798 |
| | HSTNPO | 0.0199556 | 0.0072581 | 0.0117212 | 0.0106306 | 0.0086367 | 0.0119494 |
| | NPHP | 0.0189670 | 0.0046647 | 0.0074855 | 0.0109669 | 0.0038720 | 0.0052756 |
| | RPCE | 0.0002834 | 0.0000110 | 0.0000094 | 0.0008656 | 0.0004306 | 0.0002899 |
| | **Group 2: Money, Credit and Finance** | | | | | | |
| | TRDI | 0.2047704 | 0.0148469 | 0.0154991 | 0.0203103 | 0.0029985 | 0.0033184 |
| | M1MS | 0.0016393 | 0.0002308 | 0.0002550 | 0.0264531 | 0.0249902 | 0.0271451 |
| | M2MS | 0.0002194 | 0.0000272 | 0.0000232 | 0.0000905 | 0.0000582 | 0.0000464 |
| | CIL | 0.0000924 | 0.0000885 | 0.0000603 | 0.0005342 | 0.0003362 | 0.0002723 |
| | **Group 3: Exchange Rate** | | | | | | |
| | EXJPUS | 0.0077644 | 0.0006302 | 0.0005711 | 0.0011294 | 0.0004696 | 0.0004429 |
| | **Group 4: Price** | | | | | | |
| | PPI:CM | 0.0400774 | 0.0020502 | 0.0022087 | 0.0053281 | 0.0024342 | 0.0025617 |
| | PPI:IM | 0.0003436 | 0.0001564 | 0.0001279 | 0.0001901 | 0.0000848 | 0.0000760 |
| | PPI:FG | 0.0002074 | 0.0001105 | 0.0000962 | 0.0001486 | 0.0000986 | 0.0000878 |
| | CPI:AI | 0.0000274 | 0.0000201 | 0.0000117 | 0.0000178 | 0.0000100 | 0.0000077 |
| | PCE:CI | 0.0000312 | 0.0000086 | 0.0000063 | 0.0000086 | 0.0000056 | 0.0000041 |
| | **Group 5: Expectations** | | | | | | |
| | CSI | 19.4868038 | 18.8759556 | 18.8132371 | 21.9949929 | 17.4075369 | 17.3981120 |
| | NOCG | 0.0063040 | 0.0006792 | 0.0005801 | 0.0022776 | 0.0011229 | 0.0010969 |
| | TBI | 0.0008947 | 0.0000699 | 0.0000560 | 0.0001253 | 0.0000455 | 0.0000514 |
| | **Group 6: Monetary Policy (interest rate)** | | | | | | |
| | EFFR | 0.0217789 | 0.0196307 | 0.0213777 | 0.0239301 | 0.0263335 | 0.0313433 |
| | $\text{MSE}_{(n)}^h$ | 0.9622467 | 0.9027666 | 0.9005309 | 1.2136001 | 0.8994711 | 0.8929071 |
| $h=2$ | **Group 1: Real Activity** | | | | | | |
| | CLF | 0.0000462 | 0.0000053 | 0.0000048 | 0.0001014 | 0.0000252 | 0.0000249 |
| | CE | 0.0000911 | 0.0000079 | 0.0000067 | 0.0011149 | 0.0008415 | 0.0002801 |
| | CUR | 0.3898965 | 0.0309646 | 0.0379273 | 4.8598623 | 1.3112633 | 1.2198440 |
| | RPI | 0.0003378 | 0.0000927 | 0.0000826 | 0.0014216 | 0.0011788 | 0.0008798 |
| | HSTNPO | 0.0197571 | 0.0110340 | 0.0198930 | 0.0198257 | 0.0131269 | 0.0194080 |
| | NPHP | 0.0241978 | 0.0084029 | 0.0143740 | 0.0198768 | 0.0065821 | 0.0104900 |
| | RPCE | 0.0002777 | 0.0000101 | 0.0000098 | 0.0010307 | 0.0004464 | 0.0003412 |
| | **Group 2: Money, Credit and Finance** | | | | | | |
| | TRDI | 0.1873265 | 0.0143373 | 0.0148909 | 0.0209176 | 0.0030982 | 0.0034679 |
| | M1MS | 0.0018103 | 0.0002551 | 0.0002538 | 0.0294288 | 0.0289803 | 0.0271637 |
| | M2MS | 0.0002274 | 0.0000219 | 0.0000217 | 0.0001257 | 0.0000955 | 0.0000479 |
| | CIL | 0.0001126 | 0.0000943 | 0.0000578 | 0.0006107 | 0.0003134 | 0.0002675 |
| | **Group 3: Exchange Rate** | | | | | | |
| | EXJPUS | 0.0077150 | 0.0006108 | 0.0005828 | 0.0012101 | 0.0004802 | 0.0004703 |
| | **Group 4: Price** | | | | | | |
| | PPI:CM | 0.0419332 | 0.0021432 | 0.0022102 | 0.0050280 | 0.0025708 | 0.0025400 |
| | PPI:IM | 0.0004297 | 0.0001356 | 0.0001263 | 0.0001909 | 0.0000862 | 0.0000805 |
| | PPI:FG | 0.0002821 | 0.0000940 | 0.0000936 | 0.0001456 | 0.0000976 | 0.0000893 |
| | CPI:AI | 0.0000267 | 0.0000262 | 0.0000138 | 0.0000181 | 0.0000100 | 0.0000073 |
| | PCE:CI | 0.0000314 | 0.0000084 | 0.0000059 | 0.0000085 | 0.0000089 | 0.0000045 |
| | **Group 5: Expectations** | | | | | | |
| | CSI | 18.9694357 | 18.6283004 | 18.6495128 | 25.9140567 | 17.6935390 | 17.6294505 |
| | NOCG | 0.0059201 | 0.0006019 | 0.0005482 | 0.0038068 | 0.0013883 | 0.0012646 |
| | TBI | 0.0008354 | 0.0000793 | 0.0000612 | 0.0001306 | 0.0000415 | 0.0000534 |
| | **Group 6: Monetary Policy (interest rate)** | | | | | | |
| | EFFR | 0.0253043 | 0.0238917 | 0.0256199 | 0.0376003 | 0.0333866 | 0.0354828 |
| | $\text{MSE}_{(n)}^h$ | 0.9369521 | 0.8914818 | 0.8936332 | 1.4722148 | 0.9094077 | 0.9024599 |



Electronic copy available at: https://ssrn.com/abstract=3996594

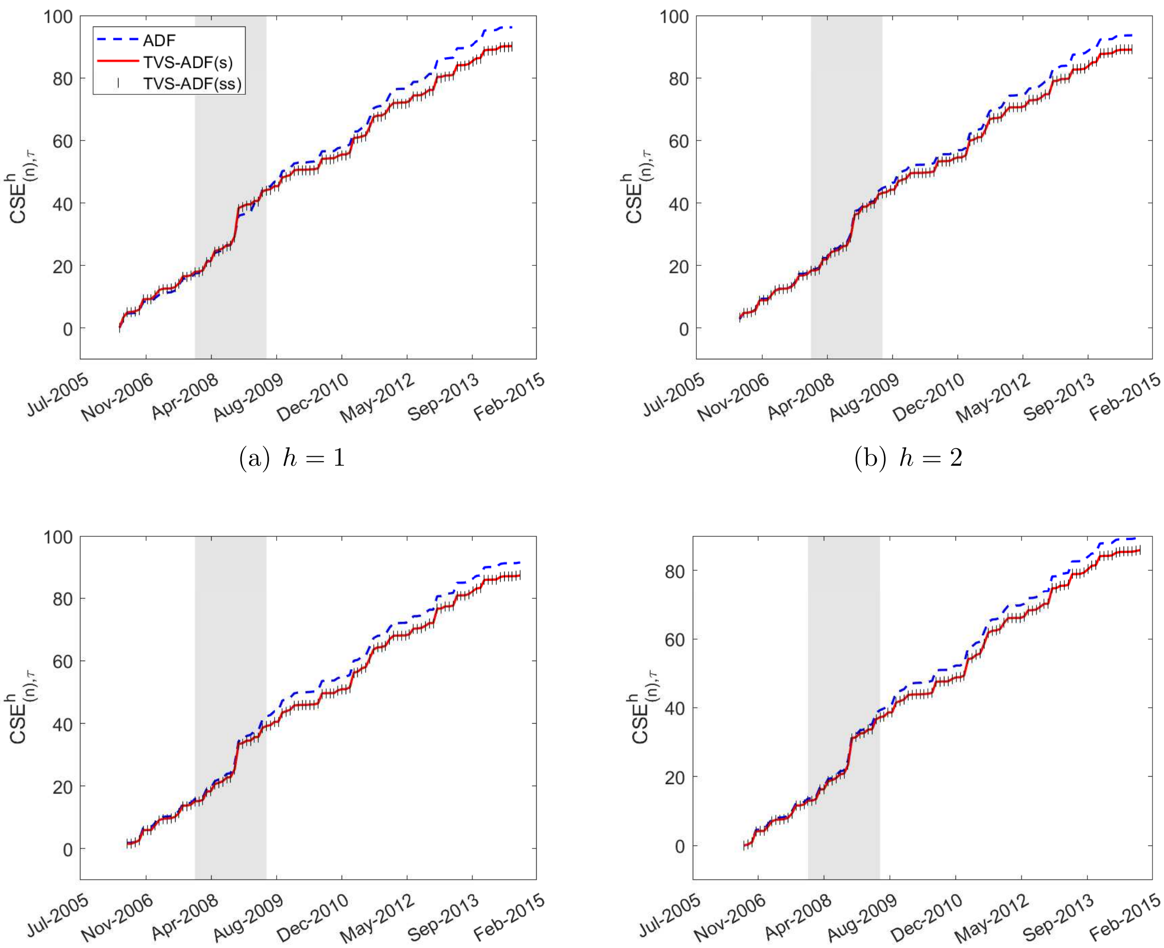
Table 4: $\text{MSE}_i^h$ of $h$-step-ahead point forecasts of three models for different sample sizes

| | | $T=100$ | | | $T=200$ | | |
|---|---|---|---|---|---|---|---|
| | | ADF | TVS-ADF(s) | TVS-ADF(ss) | ADF | TVS-ADF(s) | TVS-ADF(ss) |
| $h=3$ | **Group 1: Real Activity** | | | | | | |
| | CLF | 0.0000453 | 0.0000055 | 0.0000047 | 0.0001093 | 0.0000266 | 0.0000251 |
| | CE | 0.0000867 | 0.0000063 | 0.0000067 | 0.0013593 | 0.0007678 | 0.0002934 |
| | CUR | 0.3754399 | 0.0342800 | 0.0385781 | 6.1745659 | 1.2062096 | 1.2184667 |
| | RPI | 0.0003481 | 0.0000839 | 0.0000815 | 0.0016148 | 0.0010376 | 0.0008712 |
| | HSTNPO | 0.0201458 | 0.0151340 | 0.0311140 | 0.0357245 | 0.0168592 | 0.0288106 |
| | NPHP | 0.0322948 | 0.0123337 | 0.0246911 | 0.0353469 | 0.0095936 | 0.0173786 |
| | RPCE | 0.0002778 | 0.0000107 | 0.0000099 | 0.0010878 | 0.0003107 | 0.0003254 |
| | **Group 2: Money, Credit and Finance** | | | | | | |
| | TRDI | 0.1949322 | 0.0147544 | 0.0143345 | 0.0203652 | 0.0034294 | 0.0034656 |
| | M1MS | 0.0018272 | 0.0002616 | 0.0002645 | 0.0281965 | 0.0289940 | 0.0274075 |
| | M2MS | 0.0002312 | 0.0000217 | 0.0000221 | 0.0001184 | 0.0000671 | 0.0000459 |
| | CIL | 0.0001407 | 0.0000702 | 0.0000591 | 0.0004817 | 0.0003380 | 0.0002827 |
| | **Group 3: Exchange Rate** | | | | | | |
| | EXJPUS | 0.0076253 | 0.0005815 | 0.0005746 | 0.0012368 | 0.0004413 | 0.0004671 |
| | **Group 4: Price** | | | | | | |
| | PPI:CM | 0.0418529 | 0.0022302 | 0.0022453 | 0.0047940 | 0.0026880 | 0.0026381 |
| | PPI:IM | 0.0003606 | 0.0001385 | 0.0001256 | 0.0002095 | 0.0000990 | 0.0000946 |
| | PPI:FG | 0.0002498 | 0.0001044 | 0.0000965 | 0.0001544 | 0.0001004 | 0.0000945 |
| | CPI:AI | 0.0000300 | 0.0000248 | 0.0000112 | 0.0000185 | 0.0000122 | 0.0000076 |
| | PCE:CI | 0.0000310 | 0.0000090 | 0.0000053 | 0.0000087 | 0.0000112 | 0.0000039 |
| | **Group 5: Expectations** | | | | | | |
| | CSI | 18.5174800 | 18.2500645 | 18.2216503 | 23.4596572 | 18.1251140 | 18.0272474 |
| | NOCG | 0.0057085 | 0.0005706 | 0.0005451 | 0.0039113 | 0.0011162 | 0.0011223 |
| | TBI | 0.0007875 | 0.0000871 | 0.0000565 | 0.0001515 | 0.0000474 | 0.0000519 |
| | **Group 6: Monetary Policy (interest rate)** | | | | | | |
| | EFFR | 0.0278745 | 0.0272363 | 0.0274391 | 0.0433248 | 0.0353730 | 0.0369844 |
| | $\text{MSE}_{(n)}^h$ | 0.9156081 | 0.8741909 | 0.8743769 | 1.4196399 | 0.9253636 | 0.9221945 |
| $h=4$ | **Group 1: Real Activity** | | | | | | |
| | CLF | 0.0000460 | 0.0000054 | 0.0000047 | 0.0001487 | 0.0000253 | 0.0000251 |
| | CE | 0.0000846 | 0.0000076 | 0.0000067 | 0.0020646 | 0.0010337 | 0.0002787 |
| | CUR | 0.3646827 | 0.0376349 | 0.0394568 | 9.5628361 | 1.2190089 | 1.2144110 |
| | RPI | 0.0003609 | 0.0000805 | 0.0000812 | 0.0021144 | 0.0010297 | 0.0008724 |
| | HSTNPO | 0.0245225 | 0.0222007 | 0.0485147 | 0.0529737 | 0.0202204 | 0.0407916 |
| | NPHP | 0.0424145 | 0.0183835 | 0.0395700 | 0.0543019 | 0.0115179 | 0.0246383 |
| | RPCE | 0.0002742 | 0.0000104 | 0.0000101 | 0.0013917 | 0.0003227 | 0.0003078 |
| | **Group 2: Money, Credit and Finance** | | | | | | |
| | TRDI | 0.2019685 | 0.0159947 | 0.0150336 | 0.0196138 | 0.0035147 | 0.0034972 |
| | M1MS | 0.0017596 | 0.0002692 | 0.0002660 | 0.0277572 | 0.0286192 | 0.0268136 |
| | M2MS | 0.0002239 | 0.0000248 | 0.0000222 | 0.0001313 | 0.0000786 | 0.0000413 |
| | CIL | 0.0001092 | 0.0000764 | 0.0000568 | 0.0004999 | 0.0002845 | 0.0002803 |
| | **Group 3: Exchange Rate** | | | | | | |
| | EXJPUS | 0.0079293 | 0.0005797 | 0.0005770 | 0.0011942 | 0.0004779 | 0.0004706 |
| | **Group 4: Price** | | | | | | |
| | PPI:CM | 0.0410467 | 0.0022535 | 0.0022485 | 0.0050867 | 0.0026420 | 0.0026144 |
| | PPI:IM | 0.0004346 | 0.0001347 | 0.0001272 | 0.0002046 | 0.0000995 | 0.0000946 |
| | PPI:FG | 0.0002769 | 0.0000970 | 0.0000944 | 0.0001541 | 0.0001004 | 0.0000983 |
| | CPI:AI | 0.0000324 | 0.0000180 | 0.0000111 | 0.0000182 | 0.0000137 | 0.0000075 |
| | PCE:CI | 0.0000316 | 0.0000128 | 0.0000053 | 0.0000088 | 0.0000229 | 0.0000043 |
| | **Group 5: Expectations** | | | | | | |
| | CSI | 18.0989528 | 17.9136928 | 17.8914214 | 33.1687286 | 17.9011114 | 17.8782540 |
| | NOCG | 0.0054153 | 0.0005214 | 0.0005292 | 0.0043239 | 0.0010966 | 0.0010849 |
| | TBI | 0.0007362 | 0.0000604 | 0.0000450 | 0.0001730 | 0.0000492 | 0.0000537 |
| | **Group 6: Monetary Policy (interest rate)** | | | | | | |
| | EFFR | 0.0297801 | 0.0278271 | 0.0278923 | 0.0526073 | 0.0359239 | 0.0378348 |
| | $\text{MSE}_{(n)}^h$ | 0.8962420 | 0.8590422 | 0.8602845 | 2.0455396 | 0.9155806 | 0.9158321 |



Electronic copy available at: https://ssrn.com/abstract=3996594

highlight the financial-crisis era or COVID-19-pandemic era using a shadow band in Figure 1 and 2, repsectively. For all forecast horizons $h$ with $T = 100$, the TVS-ADF(s) and TVS-ADF(ss) have much smaller increases in the predictive error relative to the ADF after August 2009. For $T = 200$, while the TVS-ADF(ss) and TVS-ADF(s) consistently beat the ADF after the initial months of the COVID-19 pandemic for all forecast horizons $h$. Moreover, the performance of the TVS-ADF(s) is similar to that of the TVS-ADF(ss).



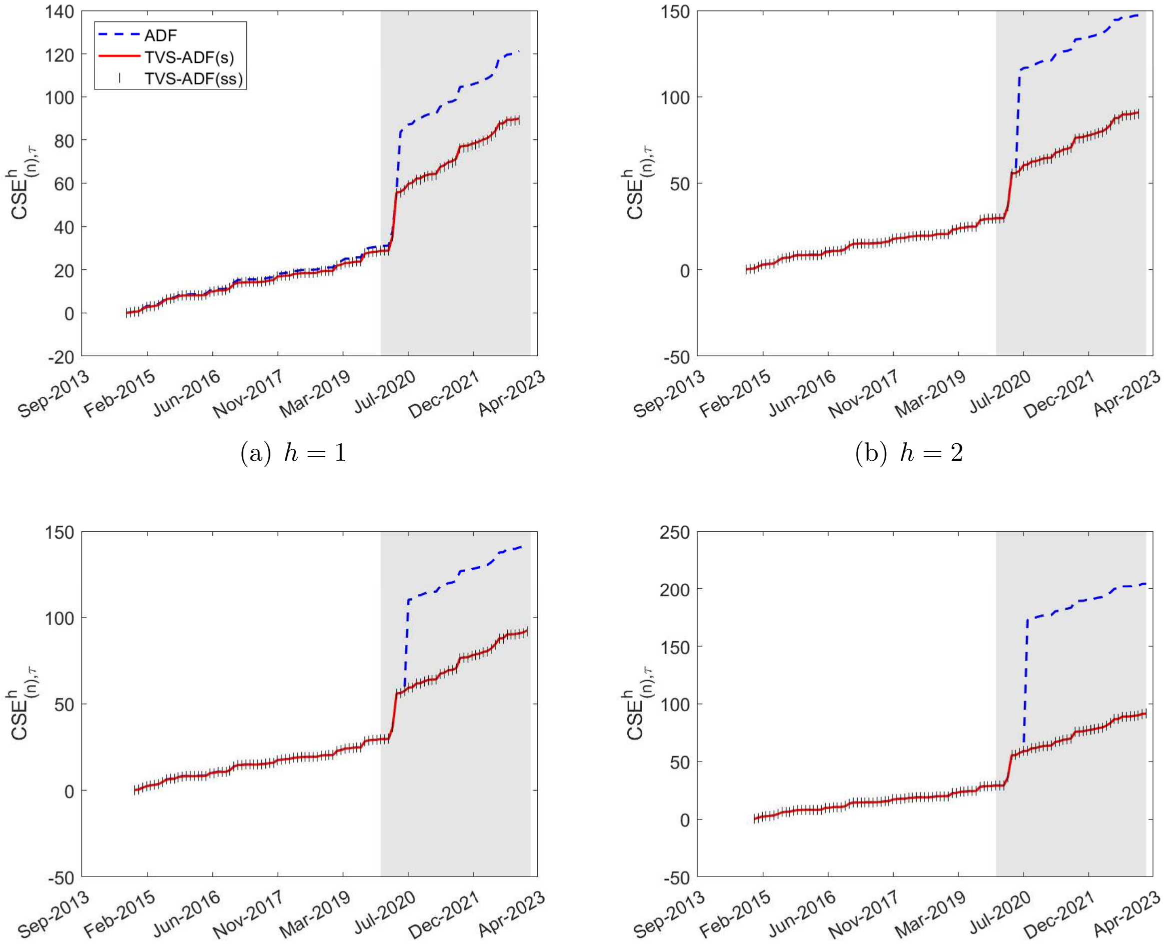

(a) $h = 1$ (b) $h = 2$ (c) $h = 3$ (d) $h = 4$

Figure 1: $\text{CSE}^{h}_{(n),\tau}$ of $h$-step-ahead point forecasts of three models for $T = 100$

In sum, compared to the ADF, the TVS-ADF(s) and TVS-ADF(ss) can better capture economic dynamic features and thus substantially improve the predictive accuracy for different forecast horizons no matter whether the sample size is small or large.



Electronic copy available at: https://ssrn.com/abstract=3996594




(a) $h = 1$ (b) $h = 2$

(c) $h = 3$ (d) $h = 4$

Figure 2: $\text{CSE}^{h}_{(n),\tau}$ of $h$-step-ahead point forecasts of three models for $T = 200$

# 5 Conclusions

This study proposed a new factor model, TVS-ADF, by extending the ADF to capture the time-varying characteristics of economic data better. We detailedly discussed the identification scheme of the common factors in this model and provided the identification theory. Additionally, we also constructed an effective MCMC algorithm (seven-step Gibbs sampling) to estimate the TVS-ADF. Moreover, to avoid overparameterization, we provided shrinkage and sparsification methods for our model. Furthermore, we proposed several information criteria to determine the number of factors. Using an artificial data experiment, we showed that the TVS-ADF(s) and TVS-ADF(ss) always yield more precise forecasts



Electronic copy available at: https://ssrn.com/abstract=3996594

than the ADF for different cases. An empirical application to macroeconomic forecasting indicated that our model also captures the dynamic features of a real economic system better than the ADF. In our future research, we will consider a new decomposition for $\Gamma_t^\xi$ to model the dynamics of elements in $\Gamma_t^\xi$ in order to analyze financial time series better.

# References


Abbate, A., Eickmeier, S., Lemke, W., Marcellino, M., 2016. The changing international transmission of financial shocks: evidence from a classical time-varying favar. Journal of Money, Credit and Banking 48, 573–601.

Aharon, D.Y., Demir, E., 2022. Nfts and asset class spillovers: Lessons from the period around the covid-19 pandemic. Finance Research Letters 47, 102515.

Antolin-Diaz, J., Drechsel, T., Petrella, I., 2017. Tracking the slowdown in long-run gdp growth. Review of Economics and Statistics 99, 343–356.

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

Barigozzi, M., 2018. Dynamic factor models. Lecture notes. London School of Economics .

Barigozzi, M., Hallin, M., Soccorsi, S., von Sachs, R., 2021. Time-varying general dynamic factor models and the measurement of financial connectedness. Journal of Econometrics 222, 324–343.

Baumeister, C., Peersman, G., 2013. Time-varying effects of oil supply shocks on the us economy. American Economic Journal: Macroeconomics 5, 1–28.

Belviso, F., Milani, F., 2006. Structural factor-augmented vars (sfavars) and the effects of monetary policy. Topics in Macroeconomics 6.




Electronic copy available at: https://ssrn.com/abstract=3996594

Bhattacharya, A., Pati, D., Pillai, N.S., Dunson, D.B., 2015. Dirichlet–laplace priors for optimal shrinkage. Journal of the American Statistical Association 110, 1479–1490.

Bianchi, F., Mumtaz, H., Surico, P., 2009. Dynamics of the term structure of uk interest rates .

Bjørnland, H.C., Thorsrud, L.A., 2019. Commodity prices and fiscal policy design: Procyclical despite a rule. Journal of Applied Econometrics 34, 161–180.

Carter, C.K., Kohn, R., 1994. On gibbs sampling for state space models. Biometrika 81, 541–553.

Cataño, D.H., Rodríguez-Caballero, C.V., Chiann, C., Peña, D., 2021. Wavelet estimation for factor models with time-varying loadings. arXiv preprint arXiv:2110.04416 .

Chan, J.C., 2023. Large hybrid time-varying parameter vars. Journal of Business & Economic Statistics 41, 890–905.

Cogley, T., 2005. How fast can the new economy grow? a bayesian analysis of the evolution of trend growth. Journal of macroeconomics 27, 179–207.

Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post wwii us. Review of Economic dynamics 8, 262–302.

Del Negro, M., Otrok, C., 2008. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. FRB of New York Staff Report .

Eichler, M., Motta, G., Von Sachs, R., 2011. Fitting dynamic factor models to non-stationary time series. Journal of Econometrics 163, 51–70.

Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the black box: Structural factor models with large cross sections. Econometric Theory 25, 1319–1347.



Electronic copy available at: https://ssrn.com/abstract=3996594

Frühwirth-Schnatter, S., 1994. Data augmentation and dynamic linear models. Journal of time series analysis 15, 183–202.

Frühwirth-Schnatter, S., Wagner, H., 2010. Stochastic model specification search for gaussian and partial non-gaussian state space models. Journal of Econometrics 154, 85–100.

Galí, J., Gambetti, L., 2015. The effects of monetary policy on stock market bubbles: Some evidence. American Economic Journal: Macroeconomics 7, 233–57.

Gao, J., Peng, B., Yan, Y., 2023. Estimation, inference, and empirical analysis for time-varying var models. Journal of Business & Economic Statistics , 1–12.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis. CRC press.

Huber, F., Koop, G., Onorante, L., 2020. Inducing sparsity and shrinkage in time-varying parameter models. Journal of Business & Economic Statistics , 1–48.

Kadiyala, K.R., Karlsson, S., 1997. Numerical methods for estimation and inference in bayesian var-models. Journal of Applied Econometrics 12, 99–132.

Karakatsani, N.V., Bunn, D.W., 2008. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. International Journal of Forecasting 24, 764–785.

Koop, G., Korobilis, D., 2014. A new index of financial conditions. European Economic Review 71, 101–116.

Koop, G., Korobilis, D., Pettenuzzo, D., 2019. Bayesian compressed vector autoregressions. Journal of Econometrics 210, 135–154.

Koop, G., Leon-Gonzalez, R., Strachan, R.W., 2009. On the evolution of the monetary policy transmission mechanism. Journal of Economic Dynamics and Control 33, 997–1017.



Electronic copy available at: https://ssrn.com/abstract=3996594

Korobilis, D., 2013. Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. Oxford Bulletin of Economics and Statistics 75, 157–179.

Liu, P., Mumtaz, H., Theophilopoulou, A., 2011. International transmission of shocks: A time-varying factor-augmented var approach to the open economy .

Ma, S., Lan, W., Su, L., Tsai, C.L., 2020. Testing alphas in conditional time-varying factor models with high-dimensional assets. Journal of Business & Economic Statistics 38, 214–227.

Marcellino, M., Porqueddu, M., Venditti, F., 2016. Short-term gdp forecasting with a mixed-frequency dynamic factor model with stochastic volatility. Journal of Business & Economic Statistics 34, 118–127.

McCracken, M.W., Ng, S., 2016. Fred-md: A monthly database for macroeconomic research. Journal of Business & Economic Statistics 34, 574–589.

Mikkelsen, J.G., Hillebrand, E., Urga, G., 2019. Consistent estimation of time-varying loadings in high-dimensional factor models. Journal of Econometrics 208, 535–562.

Motta, G., Hafner, C.M., Von Sachs, R., 2011. Locally stationary factor models: Identification and nonparametric estimation. Econometric Theory 27, 1279–1319.

Mumtaz, H., Surico, P., 2012. Evolving international inflation dynamics: world and country-specific factors. Journal of the European Economic Association 10, 716–734.

Nakajima, J., et al., 2011. Time-varying parameter var model with stochastic volatility: An overview of methodology and empirical applications .

Pelger, M., Xiong, R., 2021. State-varying factor models of large dimensions. Journal of Business & Economic Statistics , 1–19.



Electronic copy available at: https://ssrn.com/abstract=3996594

Polson, N.G., Scott, J.G., 2010. Shrink globally, act locally: Sparse bayesian regularization and prediction. Bayesian statistics 9, 105.

Primiceri, G.E., 2005. Time varying structural vector autoregressions and monetary policy. The Review of Economic Studies 72, 821–852.

Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for var analysis.

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

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

Thorsrud, L.A., 2020. Words are the new numbers: A newsy coincident index of the business cycle. Journal of Business & Economic Statistics 38, 393–409.



Electronic copy available at: https://ssrn.com/abstract=3996594

# Appendix A: Proof of Theorem 1

*Proof.* Let $r_{it} = y_{it} - \beta'_{it} x_{it} = \lambda'_{it} F_t + \xi_{it}$. Since $\frac{1}{n}\sum_{i=1}^{n} r_{it} = \frac{1}{n}\sum_{i=1}^{n} \lambda'_{it} F_t + \frac{1}{n}\sum_{i=1}^{n} \xi_{it}$, then it follows that

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} r_{it}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\text{Cov}\left(\lambda'_{it}F_t, \lambda'_{jt}F_t\right)\right] + \frac{1}{T^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\text{Cov}\left(\xi_{it}, \xi_{jt}\right)\right]. \quad \text{(A.1)}$$

First, let us consider the first term of the right-hand side of (A.1). Note that

$$\begin{aligned}\text{Cov}\left(\lambda'_{it}F_t, \lambda'_{jt}F_t\right) &= \text{E}(\lambda'_{it}F_t\lambda'_{jt}F_t) = \text{E}\left(\sum_{s=1}^{r}\sum_{l=1}^{r}\lambda_{is,t}\lambda_{jl,t}F_{st}F_{lt}\right) \\ &= \sum_{s=1}^{r}\sum_{l=1}^{r}\text{E}\left(\lambda_{is,t}\lambda_{jl,t}\right)\text{E}\left(F_{st}F_{lt}\right), \end{aligned} \quad \text{(A.2)}$$

where the second equation holds because $E(F_t) = 0$ when the absolute value of the maximum eigenvalue of $C$ is smaller than 1, and $\lambda_{is,t}$ and $F_{st}$ mean the $s$-th elements of $\lambda_{it}$ and $F_t$, respectively (similarly for $\lambda_{jl,t}$ and $F_{lt}$). According to Assumption 1 and $\lambda_t = \lambda_{t-1} + \tilde{d}_t$, then we have

$$\begin{aligned}\text{E}\left(\lambda_{is,t}\lambda_{jl,t}\right) &= \text{E}\left[(\lambda_{is,t-1} + \tilde{d}_{is,t})(\lambda_{jl,t-1} + \tilde{d}_{jl,t})\right] = \text{E}\left[(\lambda_{is,0} + \sum_{k=0}^{t-1}\tilde{d}_{is,t-k})(\lambda_{jl,0} + \sum_{k=0}^{t-1}\tilde{d}_{jl,t-k})\right] \\ &= \begin{cases}\text{E}\left(\lambda_{is,0}^2\right) + t\tilde{D}_{is}, & \text{if } i = j, s = l \\ \text{E}\left(\lambda_{is,0}\right)\text{E}\left(\lambda_{il,0}\right), & \text{if } i = j, s \neq l \\ \text{E}\left(\lambda_{is,0}\right)\text{E}\left(\lambda_{jl,0}\right), & \text{if } i \neq j\end{cases},\end{aligned} \quad \text{(A.3)}$$

where $\tilde{D}_{is}$ denotes the $s$-th diagonal element of $\tilde{D}_i$.

Let $C_s$ and $\eta_{st}$ denote the $s$-th row of $C$ and $\eta_t$, respectively. Then, for $\text{E}(F_{st}F_{lt})$, we



Electronic copy available at: https://ssrn.com/abstract=3996594

have

$$
\begin{aligned}
&\mathrm{E}(F_{st}F_{lt}) = \mathrm{E}\left[(C_sF_{t-1}+\eta_{st})(C_lF_{t-1}+\eta_{lt})\right] \\
=&\mathrm{E}\left\{\left[C_sC^{t-1}F_0+\sum_{j=1}^{t-1}\left(C_sC^{j-1}\eta_{t-j}\right)+\eta_{st}\right]\left[C_lC^{t-1}F_0+\sum_{j=1}^{t-1}\left(C_lC^{j-1}\eta_{t-j}\right)+\eta_{lt}\right]\right\} \\
=&C_sC^{t-1}\mathrm{E}(F_0F_0')C^{t-1\prime}C_l'+\sum_{i=1}^{t-1}\left[C_sC^{t-i-1}\mathrm{E}(\eta_i\eta_i')C^{t-i-1\prime}C_l'\right]+\mathrm{E}(\eta_{st}\eta_{lt}) \\
=&C_sC^{t-1}\mathrm{E}(F_0F_0')C^{t-1\prime}C_l'+\sum_{i=1}^{t-1}\left[C_sC^{t-i-1}C^{t-i-1\prime}C_l'\right]+\mathrm{E}(\eta_{st}\eta_{lt}) \qquad \text{(A.4)}
\end{aligned}
$$

Note that $\mathrm{E}(F_0F_0') = P_0M_0P_0'$ and $C = PMP^{-1}$ by Assumptions 2(b) and 3(b). Let $P_s$ denote the $s$-th row of $P$, then $C_s = P_sMP^{-1}$. Now, for (A.4), it follows that

$$
\begin{aligned}
\mathrm{E}(F_{st}F_{l\tau}) =&P_sP^{-1}P_0M^{2t}M_0P_0'P^{-1\prime}P_l'+\sum_{i=1}^{t-1}\left[P_sP^{-1}M^{2t-2i}P^{-1\prime}P_l'\right]+\mathrm{E}(\eta_{st}\eta_{lt}) \\
=&P_sP^{-1}\left[P_0M^{2t}M_0P_0'+\sum_{i=1}^{t-1}M^{2t-2i}\right]P^{-1\prime}P_l^{-1\prime}+\mathrm{E}(\eta_{st}\eta_{lt}) \\
\equiv&\mathbf{H} \qquad \text{(A.5)}
\end{aligned}
$$

Combining (A.3) and (A.5), we can rewrite (A.2) as

$$
\begin{aligned}
\mathrm{Cov}\left(\lambda_{it}'F_t, \lambda_{jt}'F_t\right) &= \sum_{s=1}^{r}\sum_{l=1}^{r}\mathrm{E}\left(\lambda_{is,t}\lambda_{jl,t}\right)\mathrm{E}\left(F_{st}F_{lt}\right) \\
&= \begin{cases} \sum_{s=1}^{r}\sum_{l=1}^{r}\left[\mathrm{E}\left(\lambda_{is,0}^2\right)+t\tilde{D}_{is}\right]\mathbf{H}, & \text{if } i=j, s=l \\ \sum_{s=1}^{r}\sum_{l=1}^{r}\mathrm{E}\left(\lambda_{is,0}\right)\mathrm{E}\left(\lambda_{jl,0}\right)\mathbf{H}, & \text{otherwise} \end{cases}, \qquad \text{(A.6)}
\end{aligned}
$$

According to (A.6) and Assumptions 1, 2(b), and 3(b), it follows that

$$
\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathrm{Cov}\left(\lambda_{it}'F_t, \lambda_{jt}'F_t\right)\right] = O(1) \qquad \text{(A.7)}
$$



Electronic copy available at: https://ssrn.com/abstract=3996594

as $n \to \infty$.

Now, let us consider the second term of the right-hand side of (A.1). Note that

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathrm{Cov}\left(\xi_{it},\xi_{jt}\right)\right] = \frac{1}{n^2} l_n' \Gamma_t^{\xi} l_n \leq \frac{1}{n^2} n \gamma_t^{\xi,(n)} = o_p(1), \tag{A.8}$$

where $l_n$ is an $n$-dimensional vector of ones, and the last inequality holds by Assumption 5.

Combining (A.7) and (A.8), we can conclude that the common component $\Lambda_t F_t$ and idiosyncratic component $\xi_t$ can be seperated in probability as $n \to \infty$. □

# Appendix B: MCMC algorithm (7-step Gibbs sampling)

## B.1 *Step 1: drawing $B_t$ and $\Lambda_t$*

$B_t$ and $\Lambda_t$ are drawn together, conditional on the remaining parameters. To simplify the drawing, shrinkage, and sparsification, we first undertake some transformations and introduce some additional notations. Note that in this step, we apply shrinkage to $B_t$ and $\Lambda_t$, while their sparsification will be done in step 6.

According to (1), for $i = 1, \cdots, n$, we have $y_{it} = \beta_{it}' x_{it} + \lambda_{it}' F_t + \xi_{it}, \xi_{it} \sim N(0, \Gamma_{t,ii}^{\xi})$. Then, by conflating $\beta_{it}$ and $\lambda_{it}$ and combining (4) and (5), we have

$$\begin{aligned} y_{it} &= b_{it}' p_{it} + \xi_{it}, \qquad \xi_{it} \sim N(0, \Gamma_{t,ii}^{\xi}), \\ b_{it} &= b_{it-1} + v_{it}, \qquad v_{it} \sim N(0, D_i), \end{aligned} \tag{B.1}$$

where $b_{it}' = (b_{i1t}, b_{i2t}, \cdots, b_{ik_it}) = (\beta_{it}', \lambda_{it}')$, $k_i = m_i + r$, $p_{it} = (x_{it}', F_t')'$, and $D_i = \mathrm{diag}(\bar{D}_i, \tilde{D}_i)$.

We introduce an $k_i \times 1$ vector $b_i = (b_{i1}, b_{i2}, \cdots, b_{ik_i})'$. Then, following Frühwirth-



Electronic copy available at: https://ssrn.com/abstract=3996594

Schnatter and Wagner (2010), we can transfer (B.1) to:

$$y_{it} = \tilde{b}'_{it} D_i^{1/2} p_{it} + b'_i p_{it} + \xi_{it}, \qquad \xi_{it} \sim N(0, \Gamma^{\xi}_{t,ii})$$
$$\tilde{b}_{it} = \tilde{b}_{it-1} + \tilde{v}_{it}, \qquad \tilde{v}_{it} \sim N(0, I_{k_i}), \tag{B.2}$$

where $I_{k_i}$ is an $k_i \times k_i$ identity matrix, $\tilde{b}_{i0} = 0$ and

$$\tilde{b}_{it} = (\tilde{b}_{i1t}, \cdots, \tilde{b}_{ik_it})', \qquad \tilde{b}_{ijt} = \frac{b_{ijt} - b_{ij}}{\sqrt{D_{ij}}}, \qquad j = 1, \cdots, k_i, \tag{B.3}$$

where $D_{ij}$ is the $j$-th diagonal element of $D_i$. Then, (B.2) also can be written as

$$y_{it} = \alpha'_i z_{it} + \xi_{it}, \tag{B.4}$$

with $\alpha_i = (\sqrt{D_{i1}}, \cdots, \sqrt{D_{ik_i}}, b_{i1}, \cdots, b_{ik_i})'$, $z_{it} = ((\tilde{b}_{it} \odot p_{it})', p'_{it})'$, and $\odot$ denotes element-wise multiplication.

**Prior** As we want to shrink parameters $\beta_{it}$, $\lambda_{it}$, and $D_i$, which have been collected and transformed to $\alpha_i$, toward zero, we use a special prior, namely the Dirichlet–Laplace prior proposed by Bhattacharya et al. (2015). Specifically, $\alpha_{ij}$, $j = 1, \cdots, 2k_i$ denotes the $j$-th element of $\alpha_i$ and follows a Gaussian distribution:

$$\alpha_{ij} \mid \omega_{ij}, \epsilon_{ij}, J_i \sim N(0, \omega_{ij}\epsilon^2_{ij}\zeta^2_i), \tag{B.5}$$

$$\omega_{ij} \sim e(1/2) \quad \epsilon_{ij} \sim D(a, \cdots, a) \quad \zeta_i \sim G(2k_i a, 1/2), \tag{B.6}$$

where $e(\cdot)$ denotes the exponential distribution, $a$ is specified as $(2k_i)^{-(1+\phi)}$ with $\phi$ being a positive number close to zero, $D(\cdot)$ is the Dirichlet distribution, and $G(\cdot)$ refers to the Gamma distribution. This prior is adopted for the especially popular method of global–



Electronic copy available at: https://ssrn.com/abstract=3996594

local shrinkage (e.g., Polson and Scott, 2010), which is both global (i.e., common to all parameters) and local (i.e., specific to each parameter). The shrinkages of the $2k_i$ parameters (i.e., global shrinkage) are controlled by $\zeta_i$, while $\omega_{ij}$ and $\epsilon_{ij}$ handle the shrinkage of the $j$-th parameter (i.e., local shrinkage). Regarding equation (B.6), $\zeta_i$ and $\omega_{ij}$ both take very small positive values with a high probability, given the properties of the exponential and Gamma distributions; so does $\epsilon_{ij}$ because the marginal distributions of $D(a, \cdots, a)$ are beta distributions with $a < 1/2$. Therefore, in this type of setup, the value of $\alpha_{ij}$ will be close to zero with a high probability. Note that $a$ plays an important role in determining the shrinkage behavior of the Dirichlet–Laplace prior. Following Huber et al. (2020), we draw $a$ from its posterior distribution, which is obtained based on the prior of a uniform distribution bounded between $(2k_i)^{-1}$ and $1/2$.

**Drawing process** Now, we show how to simulate the full history of $B_t$ and $\Lambda_t$, which have been transformed into $\tilde{b}_{it}$ in (B.2), using the Dirichlet–Laplace prior. We need some additional notations, as follows. Let $b_i^T = (b_{i1}, \cdots, b_{iT})$, $b^T = (b_1^T, \cdots, b_n^T)$, $\tilde{b}_i^T = (\tilde{b}_{i1}, \cdots, \tilde{b}_{iT})$, and $\tilde{b}^T = (\tilde{b}_1^T, \cdots, \tilde{b}_n^T)$, similarly for $F^T$, $A^T$, $H^T$, $Y^T$, $X^T$. Furthermore, let $\omega_i = (\omega_{i1}, \cdots, \omega_{i2k_i})$, $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{i2k_i})$, and $M_1 = (\omega_i, \epsilon_i, \zeta_i, a)$. According to Carter and Kohn (1994) and Frühwirth-Schnatter (1994), the conditional probability density function of $\tilde{b}_{it}$ can be factorized as

$$f(\tilde{b}_i^T|\Theta_1^T) = f(\tilde{b}_{iT}|\Theta_1^T) \prod_{t=1}^{T-1} f(\tilde{b}_{it}|\tilde{b}_{it+1}, \Theta_1^T), \tag{B.7}$$

where $\Theta_1^T = (F^T, A^T, H^T, \alpha_i, M_1, Y^T, X^T)$ and $f(\cdot|\cdot)$ stands for the conditional probability density function. Obviously, all conditional density functions in the equation are normal distributions. To conduct the drawing process, we first need to obtain the mean and variance of each conditional distribution.



Electronic copy available at: https://ssrn.com/abstract=3996594

For $f(\tilde{b}_{iT}|\Theta_1^T)$, we use the Kalman filter for (B.2) as follows:

$$\begin{aligned}
\tilde{b}_{it|t} =& \tilde{b}_{it|t-1} + \Gamma^{\tilde{b}_i}_{t|t-1} D_i^{1/2} p_{it} [p'_{it} D_i^{1/2} \Gamma^{\tilde{b}_i}_{t|t-1} D_i^{1/2} p_{it} + \Gamma^{\xi}_{t|t-1,ii}]^{-1} [y_{it} - \tilde{b}'_{it|t-1} D_i^{1/2} p_{it} - b'_i p_{it}],\\
\Gamma^{\tilde{b}_i}_{t|t} =& \Gamma^{\tilde{b}_i}_{t|t-1} - \Gamma^{\tilde{b}_i}_{t|t-1} D_i^{1/2} p_{it} [p'_{it} D_i^{1/2} \Gamma^{\tilde{b}_i}_{t|t-1} D_i^{1/2} p_{it} + \Gamma^{\xi}_{t|t-1,ii}]^{-1} p'_{it} D_i^{1/2} \Gamma^{\tilde{b}_i}_{t|t-1},\\
\tilde{b}_{it+1|t} =& \tilde{b}_{it|t},\\
\Gamma^{\tilde{b}_i}_{t+1|t} =& \Gamma^{\tilde{b}_i}_{t|t} + I_{k_i},
\end{aligned} \tag{B.8}$$

where $\tilde{b}_{i1|0} = 0$ and $\Gamma^{\tilde{b}_i}_{1|0}$ is a positive number close to zero. The final iteration of the Kalman filter provides the mean and variance of $f(\tilde{b}_{iT}|\Theta_1^T)$.

Following Carter and Kohn (1994), to obtain the mean and variance of $f(\tilde{b}_{it}|\tilde{b}_{it+1}, \Theta_1^T)$, we first conduct some equation transformations. We consider $\tilde{b}_{it+1} = \tilde{b}_{it} + \tilde{v}_{it+1}$ as additional observations on $\tilde{b}_{it}$. We then pre-multiply $\tilde{b}_{it+1} = \tilde{b}_{it} + \tilde{v}_{it+1}$ by $L^{-1}$, where $L$ is from the Cholesky decomposition of $I_{k_i} = L'\Delta_i L$, and we have $L^{-1}\tilde{b}_{it+1} = L^{-1}\tilde{b}_{it} + L^{-1}\tilde{v}_{it+1}$. We define $\bar{b}_{it+1} = L^{-1}\tilde{b}_{it+1}$ and $\bar{v}_{it+1} = L^{-1}\tilde{v}_{it+1}$. Then, for the $j$-th row of $\bar{b}_{it+1}$, $\bar{b}_{it+1,j} = L_j^{-1}\tilde{b}_{it} + \bar{v}_{it+1,j}$, $\bar{v}_{it+1,j} \sim N(0, \Delta_{i,jj})$. For $j = 1, \cdots, k_i$, let $\tilde{b}_{it|t,j} = E[\tilde{b}_{it}|\Theta_1^t, \tilde{b}_{it+1,1}, \cdots, \tilde{b}_{it+1,j-1}]$ and $\Gamma^{\tilde{b}_i}_{t|t,j} = Var[\tilde{b}_{it}|\Theta_1^t, \tilde{b}_{it+1,1}, \cdots, \tilde{b}_{it+1,j-1}]$, where $\tilde{b}_{it+1,j}$ denotes the $j$-th row of $\tilde{b}_{it+1}$. It is straightforward to obtain the following observation update equations using the Kalman filter for $\tilde{b}_{it|t,j-1}$:

$$\begin{aligned}
\tilde{b}_{it|t,j} &= \tilde{b}_{it|t,j-1} + \Gamma^{\tilde{b}_i}_{t|t,j-1} L_j^{-1'} [L_j^{-1} \Gamma^{\tilde{b}_i}_{t|t,j-1} L_j^{-1'} + \Delta_{i,jj}]^{-1} [\bar{b}_{it+1,j} - L_j^{-1} \tilde{b}_{it|t,j-1}],\\
\Gamma^{\tilde{b}_i}_{t|t,j} &= \Gamma^{\tilde{b}_i}_{t|t,j-1} - \Gamma^{\tilde{b}_i}_{t|t,j-1} L_j^{-1'} [L_j^{-1} \Gamma^{\tilde{b}_i}_{t|t,j-1} L_j^{-1'} + \Delta_{i,jj}]^{-1} L_j^{-1} \Gamma^{\tilde{b}_i}_{t|t,j-1},
\end{aligned}$$

where $\tilde{b}_{it|t,0} = \tilde{b}_{it|t}$ and $\Gamma^{\tilde{b}_i}_{t|t,0} = \Gamma^{\tilde{b}_i}_{t|t}$, which are the outcomes of the Kalman filter in (B.8). To run the updated equations above, we need to obtain $\bar{b}_{it,j}$ for $t = 1, \cdots, T$. To do this, we first draw $\tilde{b}_{iT}$ from $f(\tilde{b}_{iT}|\Theta_1^T)$; then, $\bar{b}_{iT,j}$ can be obtained by $\bar{b}_{iT} = L^{-1}\tilde{b}_{iT}$. Based on $\tilde{b}_{iT}$, $\tilde{b}_{iT-1}$ can be drawn from $f(\tilde{b}_{iT-1}|\tilde{b}_{iT}, \Theta_1^T)$, and $\bar{b}_{iT-1,j}$ can be obtained as $\bar{b}_{iT-1} = L^{-1}\tilde{b}_{iT-1}$.



Electronic copy available at: https://ssrn.com/abstract=3996594

The process is similar for $\bar{b}_{iT-2,j}, \cdots, \bar{b}_{i1,j}$. Now, we can run the above update equations $k_i$ times. The final iteration gives the expectation and variance of $\tilde{b}_{it}|\tilde{b}_{it+1}, \Theta_1^T$. As (B.7) is a product of Gaussian densities, we can easily draw $\tilde{b}_{it}$ from it and then transform $\tilde{b}_{it}$ back to obtain $b_{it}$ based on (B.3).

## B.2 *Step 2: drawing* $F_t$

$F_t$ is drawn from its conditional distribution:

$$f(F^T|\Theta_2^T) = f(F_T|\Theta_2^T) \prod_{t=1}^{T-1} f(F_t|F_{t+1}, \Theta_2^T), \tag{B.9}$$

where $\Theta_2^T = (b^T, A^T, H^T, C, Y^T, X^T)$. Combining (1) and (2), we can simulate the full history of $F_t$ by following a similar drawing process as the one in step 1.

## B.3 *Step 3: drawing* $A_t$

In this step, we implement shrinkage on $A_t$, while the sparsification for $A_t$ is illustrated in step 6. As preparation, we perform some equation transformations. From (3), we have $A_t \xi_t = H_t^{1/2} e_t$. This means:

$$\begin{aligned} \xi_{1t} &= \sqrt{h_{1t}} e_{1t} \\ a_{21,t}\xi_{1t} + \xi_{2t} &= \sqrt{h_{2t}} e_{2t} \\ a_{31,t}\xi_{1t} + a_{32,t}\xi_{2t} + \xi_{3t} &= \sqrt{h_{3t}} e_{3t} \\ &\vdots \\ a_{n1,t}\xi_{1t} + a_{n2,t}\xi_{2t} + \cdots + a_{nn-1,t}\xi_{n-1t} + \xi_{nt} &= \sqrt{h_{nt}} e_{nt}. \end{aligned} \tag{B.10}$$



Electronic copy available at: https://ssrn.com/abstract=3996594

Then, for $i = 2, \cdots, n$, we have

$$\frac{\xi_{it}}{\sqrt{h_{it}}} = -\frac{1}{\sqrt{h_{it}}}\xi^{i-1t\prime} a_{it} + e_{it}, \qquad e_{it} \sim N(0,1), \tag{B.11}$$

where $\xi^{i-1t\prime} = (\xi_{1t}/\sqrt{h_{it}}, \cdots, \xi_{i-1t}/\sqrt{h_{it}})$ and $a_{it} = (a_{i1,t}, \cdots, a_{ii-1,t})'$. Moreover, (6) can be rewritten as

$$a_{it} = \rho_i a_{it-1} + u_{it}, \qquad u_{it} \sim N(0, S_i), \tag{B.12}$$

where $\rho_i$ is a $r \times r$ diagonal matrix and $S_i$ is an $(i-1) \times (i-1)$ diagonal matrix.

Now, we introduce an $(i-1) \times 1$ vector $a_i = (a_{i1}, \cdots, a_{ii-1})'$ and let $S_{ij}$ denote the $j$-th diagonal element of $S_i$. Then, we can rewrite (B.11) and (B.12) as

$$\frac{\xi_{it}}{\sqrt{h_{it}}} = -\frac{1}{\sqrt{h_{it}}}\tilde{a}'_{it}\sqrt{S_i}\xi^{i-1t} - \frac{1}{\sqrt{h_{it}}}a'_{it}\xi^{i-1t} + e_{it}, \qquad e_{it} \sim N(0,1), \tag{B.13}$$

$$\tilde{a}_{it} = \rho_i \tilde{a}_{it-1} + S_i^{-1/2}(\rho_i - I_i)a_i + \tilde{u}_{it}, \qquad \tilde{u}_{it} \sim N(0, I_i), \tag{B.14}$$

where $I_i$ is an $(i-1) \times (i-1)$ identity matrix:

$$\tilde{a}_{it} = (\tilde{a}_{i1,t}, \cdots, \tilde{a}_{ii-1,t})', \qquad \tilde{a}_{ij,t} = \frac{a_{ij,t} - a_{ij}}{\sqrt{S_{ij}}}, \qquad j = 1, \cdots, i-1, \tag{B.15}$$

and $\tilde{a}_{i0} = 0$. We then collect the parameters together and define the following new notations: $\psi_i = (\sqrt{S_{i1}}, \cdots, \sqrt{S_{ii-1}}, a'_i)'$ and $c_{it} = ((-1/\sqrt{h_{it}}\tilde{a}_{it} \odot \xi^{i-1t})', -1/\sqrt{h_{it}}\xi^{i-1t\prime})'$. The following transformation of (B.13) is used subsequently: $\xi_{it}/\sqrt{h_{it}} = \psi'_i c_{it} + e_{it}$.

**Prior** For shrinking $A_t$, we use the Dirichlet–Laplace prior for $\psi_i$.

**Drawing process** We simulate the full history of $\tilde{a}_{it}$ using the Dirichlet–Laplace prior. Specifically, we define $\tilde{a}_i^T = (\tilde{a}_{i1}, \cdots, \tilde{a}_{iT})$, and the conditional probability density function



Electronic copy available at: https://ssrn.com/abstract=3996594

of $\tilde{a}_{it}$ can be expressed as follows:

$$f(\tilde{a}_i^T|\Theta_3^T) = f(\tilde{a}_{iT}|\Theta_3^T)\prod_{t=1}^{T-1} f(\tilde{a}_{it}|\tilde{a}_{it+1}, \Theta_3^T), \tag{B.16}$$

where $\Theta_3^T = (b^T, F^T, H^T, \psi_i, M_2, \rho_i, Y^T, X^T)$ and $M_2$ denotes the hyperparameter in the prior of $\psi_i$. Based on (B.16), (B.13), and (B.14), we can obtain $\tilde{a}_i^T$ using a drawing process similar to that used in step 1. Finally, we transform $\tilde{a}_{it}$ back to get $a_{it}$ based on (B.15).

## B.4 *Step 4: drawing* $H_t$

We implement shrinkage in this step and sparsification in step 6 on $H_t$. As a preparation, we make the following equation transformations. We define $m_t = A_t\xi_t$; then, from $A_t\xi_t = H_t^{1/2}e_t$, for $i = 1, \cdots, n$, we have the $i$-th element of $m_t$, $m_{it} = \sqrt{h_{it}}e_{it}$. Consequently,

$$\begin{aligned} log(m_{it}^2) &= log(h_{it}) + log(e_{it}^2), \qquad e_{it}^2 \sim \chi^2(1), \\ &\approx -1.27 + log(h_{it}) + \phi_{it}, \qquad \phi_{it} \sim N(0, \frac{\pi^2}{2}). \end{aligned} \tag{B.17}$$

Next, we introduce element $h_i$ and let $\sigma_i^2$ denote the $i$-th diagonal element of $\Sigma$. Combining (7) and (B.17), we have

$$log(m_{it}^2) = -1.27 + log(\tilde{h}_{it})\sigma_i + h_i + \phi_{it}, \qquad \phi_{it} \sim N(0, \frac{\pi^2}{2}), \tag{B.18}$$

$$log(\tilde{h}_{it}) = \phi_i log(\tilde{h}_{it-1}) + (\phi_i - 1)h_i/\sqrt{\sigma_i^2} + \tilde{\gamma}_{it}, \qquad \tilde{\gamma}_{it} \sim N(0, 1), \tag{B.19}$$

where

$$log(\tilde{h}_{it}) = \frac{log(h_{it}) - h_i}{\sqrt{\sigma_i^2}}, \tag{B.20}$$

and $log(\tilde{h}_{i0}) = 0$. Further, we can rewrite (B.18) as $log(m_{it}^2) + 1.27 = \tau_i' l_{it} + \phi_{it}$, where $\tau_i' = (\sigma_i, h_i)$ and $l_{it} = (log(\tilde{h}_{it}), 1)'$.



Electronic copy available at: https://ssrn.com/abstract=3996594

**Prior** To shrink $H_t$, we use the Dirichlet–Laplace prior for $\tau_i$.

**Drawing process** We simulate the full history of $log(\tilde{h}_{it})$ based on the conditional probability density function of $log(\tilde{h}_{it})$:

$$f(log(\tilde{h}_i)^T|\Theta_4^T) = f(log(\tilde{h}_{it})|\Theta_4^T)\prod_{t=1}^{T-1} f(log(\tilde{h}_{it})|log(\tilde{h}_{it+1}), \Theta_4^T), \tag{B.21}$$

where $log(\tilde{h}_i)^T = (log(\tilde{h}_{i1}), \cdots, log(\tilde{h}_{iT}))$, $\Theta_4^T = (b^T, F^T, A^T, \tau_i, M_3, \phi_i, Y^T, X^T)$ and $M_3$ refers to the hyperparameter in the prior of $\tau_i$. Applying (B.18), (B.19), and (B.21), and following a similar drawing process to that in step 1, $log(\tilde{h}_{it})$ can be obtained. Finally, we transform $log(\tilde{h}_{it})$ back to obtain $log(h_{it})$ based on (B.20).

## B.5 *Step 5: drawing* $C$*,* $\rho$*, and* $\phi$

Let $C_s$ and $\eta_{st}$ denote the $s$-th row of $C$ and $\eta_t$, respectively. From (2), for $s = 1, \cdots, r$, $F_{st} = C_s F_{t-1} + \eta_{st}$, $\eta_{st} \sim N(0, 1)$. Note that (2) is a VAR. There are several types of prior distributions available that can be used for VAR models, as summarized by Kadiyala and Karlsson (1997) and Gelman et al. (2013). Among these, we select the standard noninformative prior for $C_s$. Then, we draw $C_s$ from its conditional posterior distribution, $C_s \mid F^T \sim N((F_{t-1}^{T'} F_{t-1}^T)^{-1} F_{t-1}^{T'} F_{st}^T, (F_{t-1}^{T'} F_{t-1}^T)^{-1})$, where $F_{t-1}^T = (F_1, \cdots, F_{T-1})'$ and $F_{st}^T = (F_{s2}, \cdots, F_{sT})'$. Similarly for $\rho$ and $\phi$.

## B.6 *Step 6: drawing* $\alpha_i$*,* $\psi_i$*, and* $\tau_i$

In this step, we first show how to draw $\alpha_i$, $\psi_i$, and $\tau_i$ from their posteriors, and then illustrate how to sparsify them.



Electronic copy available at: https://ssrn.com/abstract=3996594

**Posterior** $\alpha_i$, $\psi_i$, and $\tau_i$ are drawn from their posteriors, which can be obtained straightforwardly. Specifically,

$$\begin{aligned}
\alpha_i|\Theta_5^T &\sim N((\Omega_i^{-1} + z_i'\Gamma_i^{-1}z_i)^{-1}z_i'\Gamma_i^{-1}y_i,\ (\Omega_i^{-1} + z_i'\Gamma_i^{-1}z_i)^{-1}), \qquad \text{(B.22)}\\
\psi_i|\Theta_6^T &\sim N((\bar{\Omega}_i^{-1} + c_i'c_i)^{-1}c_i'\bar{y}_i,\ (\bar{\Omega}_i^{-1} + c_i'c_i)^{-1}),\\
\tau_i|\Theta_7^T &\sim N((\tilde{\Omega}_i^{-1} + l_i'\check{\Omega}_i^{-1}l_i)^{-1}l_i'\check{\Omega}_i^{-1}\tilde{y}_i,\ (\tilde{\Omega}_i^{-1} + l_i'\check{\Omega}_i^{-1}l_i)^{-1}),
\end{aligned}$$

where $\Theta_5^T = (b^T, F^T, H^T, A^T, M_1, Y^T, X^T)$, $\Omega_i$ refers to the variance of the prior of $\alpha_i$, $\Gamma_i = \text{diag}(\Gamma_{1,ii}^{\xi}, \cdots, \Gamma_{T,ii}^{\xi})$, $z_i = (z_{i1}, \cdots, z_{iT})'$, $\Theta_6^T = (b^T, F^T, H^T, A^T, M_2, Y^T, X^T)$, $\bar{\Omega}_i$ refers to the variance of the prior of $\phi_i$, $c_i = (c_{i1}, \cdots, c_{iT})'$, $\bar{y}_i = (\frac{\xi_{i1}}{\sqrt{h_{i1}}}, \cdots, \frac{\xi_{iT}}{\sqrt{h_{iT}}})'$, $\Theta_7^T = (b^T, F^T, H^T, A^T, M_3, Y^T, X^T)$, $\tilde{\Omega}_i$ refers to the variance of the prior of $\tau_i$, $l_i = (l_{i1}, \cdots, l_{iT})'$, $\check{\Omega}_i = \pi^2/2$, and $\tilde{y}_i = (log(m_{i1}^2) + 1.27, \cdots, log(m_{iT}^2) + 1.27)'$.

**Sparsification** In this study, we conduct sparsification following Huber et al. (2020) in which the sparsification is done on the resulting estimates of the shrinkage (i.e., the posterior results of the shrinkage). Specifically, given a draw $\alpha_i^{(nT)} = (\alpha_{i1}^{(nT)}, \cdots, \alpha_{i2k_i}^{(nT)})$ from (B.22), the sparsified $\alpha_i$ is obtained as

$$\bar{\alpha}_{ij} = sign(\alpha_{ij}^{(nT)}) \parallel z_{i,j} \parallel^{-2} (\mid \alpha_{ij}^{(nT)} \mid\parallel z_{i,j} \parallel^2 - \kappa_{ij})_+, \quad j = 1, \cdots, 2k_i, \qquad \text{(B.23)}$$

where $\kappa_{ij} = \mid \alpha_{ij}^{(nT)} \mid^{-2}$, $sign(x)$ returns the sign of $x$, $z_{i,j}$ denotes the $j$-th column of $z_i$, and $(x)_+ = \max(x, 0)$. Note that equation (B.23) is a soft-thresholding approach, in which the value of $\bar{\alpha}_{ij}$ below a certain value is set to zero. Sparsification can be conducted similarly for $\psi_i$ and $\tau_i$.



Electronic copy available at: https://ssrn.com/abstract=3996594

## B.7 *Step 7: drawing* $M_1$, $M_2$, $M_3$

In this step, we draw hyperparameters $M_i$ for $i = 1, 2, 3$. For $M_1$, which includes $\omega_i$, $\epsilon_i$, $\zeta_i$, and $a$ in the prior of $\alpha_i$, we have

$$\omega_{ij} \mid \alpha_{ij}, \epsilon_{ij}, \zeta_i \sim IG(\zeta_i \frac{\epsilon_{ij}}{\mid \alpha_{ij} \mid}, 1),$$

$$\zeta_i \mid \alpha_i, \epsilon_i \sim GIG(2k_i(a-1), 1, 2\sum_{j=1}^{2k_i} \frac{\mid \alpha_{ij} \mid}{\epsilon_{ij}}),$$

$$\epsilon_{ij} = \frac{T_{ij}}{\sum_{j=1}^{2k_i} T_{ij}}, \quad T_{ij} \mid \alpha_{ij} \sim GIG(a-1, 1, 2 \mid \alpha_{ij} \mid),$$

where $IG$ denotes the inverse Gaussian distribution and $GIG$ refers to the generalized inverted Gaussian distribution (see Bhattacharya et al., 2015 and Huber et al., 2020). Following Huber et al. 2020, we obtain the conditional posterior of $a$ using a Metropolis–Hastings algorithm with a Gaussian proposal distribution truncated between $(2k_i)^{-1}$ and $1/2$. In the simulation and empirical application, the variance of the proposal distribution is tuned during the first 20% of the burn-in stage of the MCMC sampler, such that the acceptance rate is between 20% and 40%.

This is similar for $M_2$ and $M_3$.



Electronic copy available at: https://ssrn.com/abstract=3996594