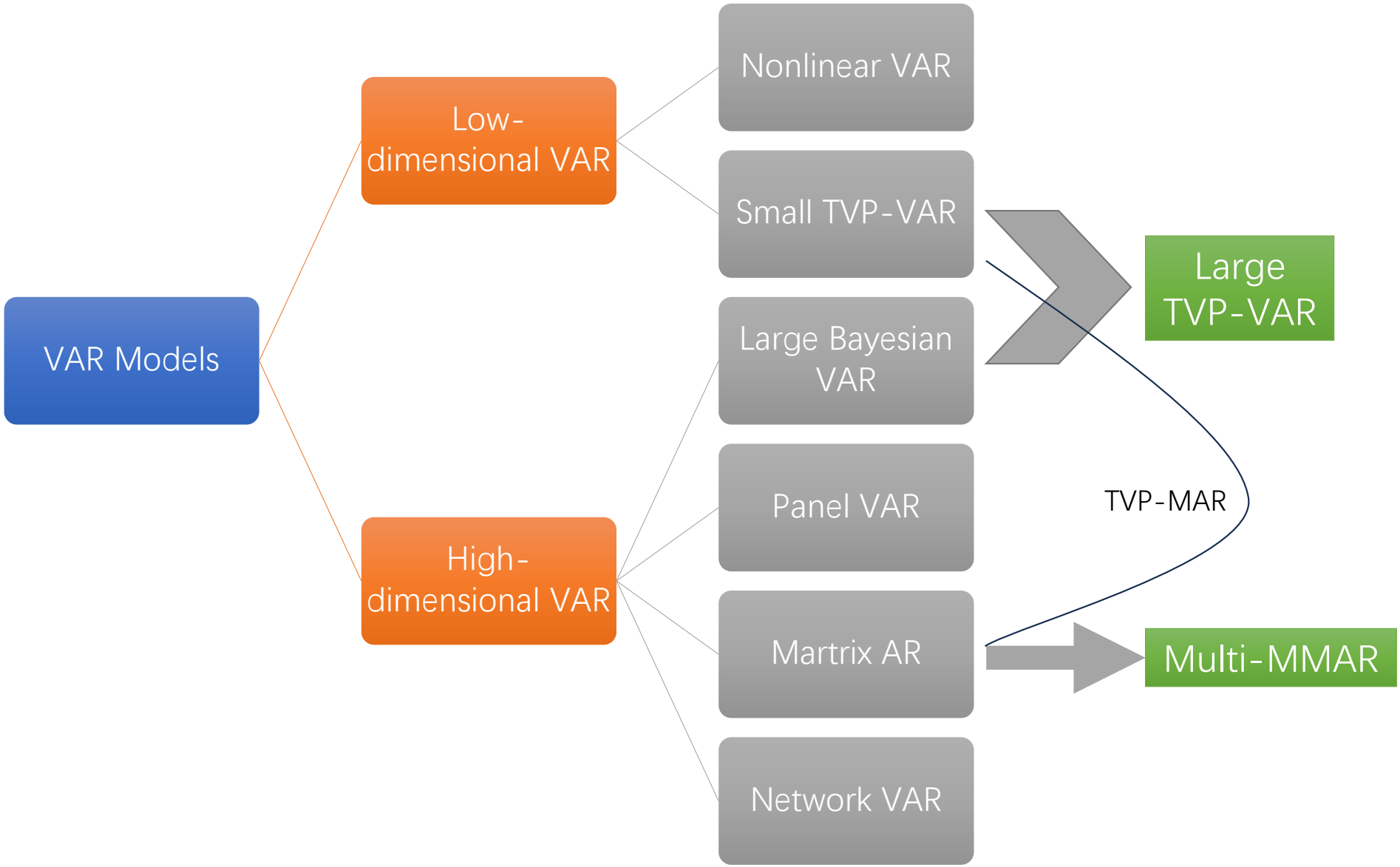
Chapter Nine

Big Data Macroeconometric Models

Tingguo Zheng (zhengtg@gmail.com)
School of Economics & WISE & Chow, Xiamen University

(For Chow & SOE & WISE Msc and PhD Program, 2026-2027, Autumn Semester)


VAR Models
Low-dimensional VAR
High-dimensional VAR
Nonlinear VAR
Small TVP-VAR
Large Bayesian VAR
Panel VAR
Martrix AR
Network VAR
Large TVP-VAR
TVP-MAR
Multi-MMAR

Outline

1. High-dimensional VAR Models

2. Panel VAR Model

3. Matrix VAR Model

4. Network VAR (NAR) Model

5. Large TVP-VAR Model

1. High-dimensional VAR Models

VAR models tend to have a lot of parameters, and large VAR models exacerbate this problem. For example, a VAR(4) with $n = 20$ dependent variables has 1,620 VAR coefficients (with intercepts), which is much larger than the number of observations in typical quarterly datasets.

- Without informative priors or regularization, it is not even possible to estimate the VAR coefficients.
- The Bayesian VAR (BVAR) approach is used for estimation since it helps to overcome the curse of dimensionality via priors. Litterman (1986a) suggests using a Minnesota prior.
- Lasso-VAR proposed by Hsu et al. (2008) and further explored by Song and Bickel (2011), Li and Chen (2014), and Davis et al. (2015).

2.1 Large Bayesian VAR model

Reference: Banbura, Giannone and Reichlin (2010), Koop and Korobilis (2010), and Chan (2019)

Assume that N is potentially large. A reduced-form VAR(p) model is

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad u_t \sim N(0, \Sigma), \tag{1}$$

where $c = (c_1, \ldots, c_N)$ and $A_1, \ldots, A_p$ are $N \times N$ AR matrices.

We can rewrite the VAR(p) as:

$$y_t = X_t \boldsymbol{\beta} + \varepsilon_t, \tag{2}$$

where $X_t = I_N \otimes (1, y_{t-1}', \ldots, y_{t-p}')$ and $\boldsymbol{\beta} = \mathrm{vec}((c, A_1, \ldots, A_p)')$ which is a $N(Np+1) \times 1$ vector.

Furthermore, stacking $\boldsymbol{y} = (y_1', \ldots, y_T')'$, we obtain its stacked version

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon}_t \sim N(0, I_T \otimes \Sigma), \tag{3}$$

where $\boldsymbol{X} = (X_1', \ldots, X_T')'$ is a $TN \times N(Np+1)$ matrix of regressors.

Likelihood function:

The likelihood function is given by

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\beta}, \Sigma) &= (2\pi)^{-\frac{TN}{2}} |(I_T \otimes \Sigma)|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(I_T \otimes \Sigma)^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right\} \\ &= (2\pi)^{-\frac{TN}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(I_T \otimes \Sigma^{-1})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right\} \end{aligned}$$

This implies that under a normal prior, the posterior for β is also normal.

Minnesota prior: The model parameters consist of two blocks: the VAR coefficients β and the error covariance matrix Σ.

For Σ: Instead of estimating Σ, the Minnesota prior replaces it with an estimate $\widehat{\Sigma}$ obtained as follows.

- We first estimate each of the N equations of the VAR separately, ignoring the error covariances across equations.
- Let s_i^2 denote the standard OLS estimate of the error variance for the i-th equation. Then, we set $\widehat{\Sigma} = \mathrm{diag}\{s_1^2, \ldots, s_N^2\}$.
  - The main advantage is that it simplifies the computations - often MCMC is not needed for posterior analysis or forecasts.
  - One main drawback is that it often produces inferior density forecasts.

For β: With Σ being replaced by an estimate, the only parameters are the VAR coefficients β. Now, consider the following normal prior:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_{\text{Minn}}, \boldsymbol{V}_{\text{Minn}}).$$

- $\boldsymbol{\beta}_{\text{Minn}}$**:** Banbura, Giannone and Reichlin (2010) impose a random walk prior mean $\boldsymbol{\beta}_{\text{Minn}} = \text{vec}(A^{*\prime})$, where

$$A^* = [0, I_N, 0, \ldots, 0],$$

  which corresponds to $A = [c, A_1, A_2, \ldots, A_p]$.

  - The prior mean of A_1 is identity matrix. *This prior has to be modified if there are variables that are not persistent.*
  - For the latter variables the prior mean of the corresponding diagonal element of A_1 is set to zero instead of 1.

- $\boldsymbol{V}_{\text{Minn}}$: This prior covariance matrix is set to be diagonal below.
  - The k-th diagonal element of $\boldsymbol{V}_{\text{Minn}}$ is given by

$$(\boldsymbol{V}_{\text{Minn}})_k = \begin{cases} \frac{c_1}{l^2}, & \text{for coefficients on own lag } l; \\ \frac{c_2 s_i^2}{l^2 s_j^2}, & \text{for coefficients on lag } l \text{ of variable } j \neq i; \\ 100 s_i^2, & \text{for the intercept,} \end{cases}$$

  - The exact values of the diagonal elements in turn depend on two key hyperparameters, c_1 and c_2, also called the degree of shrinkage.
  - Banbura et al. (2010) take one shrinkage parameter $\lambda = c_1 = c_2$.
  - There are by now many different variants of the Minnesota prior; see, e.g., Kadiyala and Karlsson (1997) and Karlsson (2013) for additional discussion.

Estimation: Estimation under the Minnesota prior is straightforward; that is one of the main appeals of the Minnesota prior.

- Given the stacked VAR representation in (3) and the normal prior $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_{\text{Minn}}, \boldsymbol{V}_{\text{Minn}})$, the posterior distribution of $\boldsymbol{\beta}$ is given by

$$(\boldsymbol{\beta} \mid \boldsymbol{y}) \sim N(\widehat{\boldsymbol{\beta}}, \boldsymbol{K}_{\beta}^{-1}),$$

where the precision matrix $\boldsymbol{K}_{\beta}$ and $\widehat{\boldsymbol{\beta}}$ can be expressed as

$$\begin{aligned}
\boldsymbol{K}_{\beta} &= \boldsymbol{V}_{\text{Minn}}^{-1} + \boldsymbol{X}'(I_T \otimes \widehat{\Sigma}^{-1})\boldsymbol{X}, \\
\widehat{\boldsymbol{\beta}} &= \boldsymbol{K}_{\beta}^{-1}\left(\boldsymbol{V}_{\text{Minn}}^{-1}\boldsymbol{\beta}_{\text{Minn}} + \boldsymbol{X}'(I_T \otimes \widehat{\Sigma}^{-1})\boldsymbol{y}\right),
\end{aligned}$$

and we have replaced Σ by the estimate $\widehat{\Sigma}$.

- The posterior mean of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}$, and we only need to compute this once instead of tens of thousands of times within a Gibbs sampler.

- When the number of variables N is large, however, computations might still be an issue. In this case,

  ▶ We first compute the Cholesky factor $\boldsymbol{C}_{K_\beta}$ of $\boldsymbol{K}_\beta$ such that

$$\boldsymbol{K}_\beta = \boldsymbol{C}_{K_\beta}\boldsymbol{C}'_{K_\beta}.$$

  ▶ Then, compute

$$\boldsymbol{C}'_{K_\beta} \backslash \left(\boldsymbol{C}_{K_\beta} \backslash \left(\boldsymbol{V}_{\text{Minn}}^{-1}\boldsymbol{\beta}_{\text{Minn}} + \boldsymbol{X}'(I_T \otimes \widehat{\Sigma}^{-1})\boldsymbol{y} \right) \right)$$

  by forward then backward substitution.[1]

[1] Since $\boldsymbol{V}_{\text{Minn}}$ is diagonal, its inverse is straightforward to compute.

2.2 LASSO-VAR

LASSO is short for “least absolute shrinkage and selection operator”, introduced in the seminal work of Tibshirani (1996).

(1) Equation-by-equation VAR estimation

- **LASSO (Tibshirani, 1996):** consider the least-squares estimation

$$\widehat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - \sum_{i} \beta_i x_{it} \right)^2$$

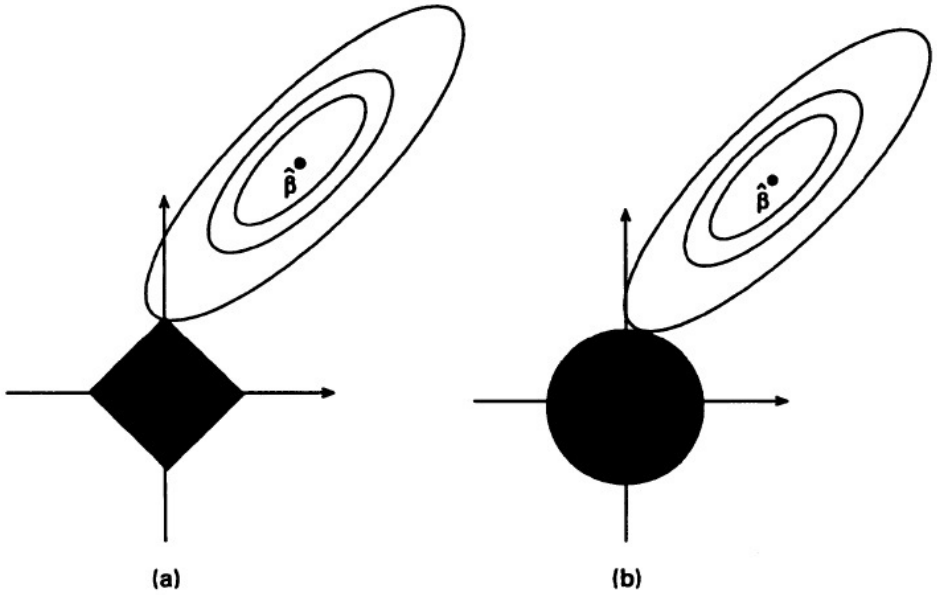
subject to the constraint:

$$\sum_{i=1}^{K} |\beta_i|^q \leq c.$$

Equivalently, consider the penalized estimation problem:

$$\widehat{\beta} = \arg\min_{\beta} \left(\sum_{t=1}^{T} \left(y_t - \sum_{i} \beta_i x_{it} \right)^2 + \lambda \sum_{i=1}^{K} |\beta_i|^q \right)$$

- LASSO (L1 regularization) if $q = 1$. Hence it shrinks and selects.
- Ridge regression (L2 regularization) if $q = 2$.



- **Adaptive Elastic Net (Zou and Zhang, 2009):** a simple extension of LASSO, *not only shrinks and selects, but also has the oracle property*.

- In Demirer, Diebold, Liu and Yilmaz (2018)'s implementation of the adaptive elastic net, they solve

$$\widehat{\beta}_{\text{AEnet}} = \arg\min_{\beta} \left(\sum_{t=1}^{T} \left(y_t - \sum_{i} \beta_i x_{it} \right)^2 + \lambda \sum_{i=1}^{K} w_i \left(\frac{1}{2}|\beta_i| + \frac{1}{2}\beta_i^2 \right) \right),$$

  where $w_i = 1/|\widehat{\beta}_{i,OLS}|$ and λ is selected equation-by-equation by 10-fold cross validation.

(2) Systematic VAR estimation

Recall the stacked VAR in Eq.(3),

$$\underbrace{\boldsymbol{y}}_{TN\times 1} = \underbrace{\boldsymbol{X}}_{TN\times N(Np+1)} \underbrace{\boldsymbol{\beta}}_{N(Np+1)\times 1} + \boldsymbol{\varepsilon}, \qquad \varepsilon_t \sim N(0, I_T \otimes \Sigma),$$

where $\boldsymbol{y} = (y_1', \ldots, y_T')'$ and $\boldsymbol{X} = (X_1', \ldots, X_T')'$.

- ℓ_1-penalized least squares (ℓ_1-LS) (Basu and Michailidis, 2015)

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{N}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 + \lambda_N||\boldsymbol{\beta}||_1$$

- ℓ_1-penalized log-likelihood (ℓ_1-LL) (Davis, Zhang and Zheng, 2012)

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{N}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(I_T \otimes \Sigma^{-1})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda_N||\boldsymbol{\beta}||_1$$

2. Large Panel VAR Model

Suppose N countries, M variables and T time periods. Thus $K = M \cdot N$.

(1) Panel VARs

General form: Let $y_{nt} = (y'_{1nt}, \ldots, y'_{Mnt})'$ be an $M \times 1$ vector of dependent variables where y_{int} is the ith variable of country n, $i = 1, \ldots, M$ and $n = 1, \ldots, N$. The general form model for y_{nt} is

$$y_{nt} = v_n + \boldsymbol{A}_n Y_{t-1} + u_{nt}, \tag{4}$$

where $Y_{n,t-1}$ denotes the vector lags of y_{nt}, $Y_{n,t-1} = (y'_{n,t-1}, \ldots, y'_{n,t-p})'$, and Y_{t-1} denotes the vector of all variables in the panel, $Y_{t-1} = (Y'_{1,t-1}, \ldots, Y'_{N,t-1})'$. The error term $u_t = (u'_{1t}, \ldots, u'_{Nt})'$ is assumed to be normal distributed, $u_t \sim N(0, \Sigma_u)$.

This panel structure has three characteristics.

- First, lags of all endogenous variables of all panel units are allowed to affect y_{it}.
- Second, the errors between countries may be correlated, i.e., u_{it} is in general correlated with other u_{jt}.
- Third, the model parameters may be specific to each i, allowing for cross-sectional heterogeneity.

In this general form, the model is not different from large-dimensional BVAR models. Therefore, it is difficult to estimate the panel VAR model in its general form without imposing additional prior structure.

Separate form: There is no dynamic interactions across the panel units. The panel VAR model can be restricted such that every panel unit corresponds to a separate VAR model

$$y_{nt} = v_n + A_n Y_{n,t-1} + u_{nt},$$

where the coefficient matrix A_n is much smaller that $\boldsymbol{A}_n$ in (4), with the interaction across units determined by the covariances $E(u_{nt}u'_{ms})$.

Homogeneous form: With the assumption of dynamic homogeneity, the panel VAR model has the same VAR coefficients such that

$$y_{nt} = v_n + AY_{n,t-1} + u_{nt}.$$

Suitable estimation methods for this panel VAR model are discussed in Canova and Ciccarelli (2013).

(2) Global VARs - Dees, Di Mauro, Pesaran, and Smith (2007)

Global VARs (GVARs) represent another class of models designed specifically to handle panels of time series data for many countries.

The model for the n-th country is set up as

$$y_{nt} = v_n + A_n Y_{n,t-1} + W(L) y_{nt}^* + u_{nt}, \tag{5}$$

where $W(L)$ is a matrix polynomial in the lag operator. y_{nt}^* is a weighted average of the variables from other countries corresponding to y_{nt},

$$y_{nt}^* = \sum_{\substack{j=1 \\ j \neq n}}^{N} \omega_{nj} y_{jt},$$

where ω_{nj} is the weight attached to country j for the n-th country. Dees et al. (2007) link the weights to *the share of country j in world trade*.

- **Advantages of GVAR analysis:**
  - One advantage of this approach is that several countries or regions can be modeled jointly as a global model (or GVAR model);
  - It is also possible to use standard tools such as impulse responses, provided shocks of interest can be identified.

- **Disadvantages of GVAR analysis:**
  - The difficulties in identifying structural shocks is one drawback.
  - Another drawback is that the transmission of the shocks is affected by the assumptions made when aggregating the variables from other countries.

3. Matrix AR (MAR) Model

Reference: Chen, Xiao and Yang (2021), Autoregressive models for matrix-valued time series, Journal of Econometrics, 222, p539-560.

Although it has been conventional to treat multiple observations as a vector, often *the inter-relationship among the time series exhibits some more structure*. When the time series are collected under the intersections of two classifications, they naturally form matrices.(当时间序列以两种分类交叉收集时，它们自然形成矩阵)

Figure 1 plots four economic indicators from five countries, resulting in a 4×5 matrix observed at each time point, with totally 20 time series.

In this example, the rows and columns correspond to different classifications (economic indicators and countries).

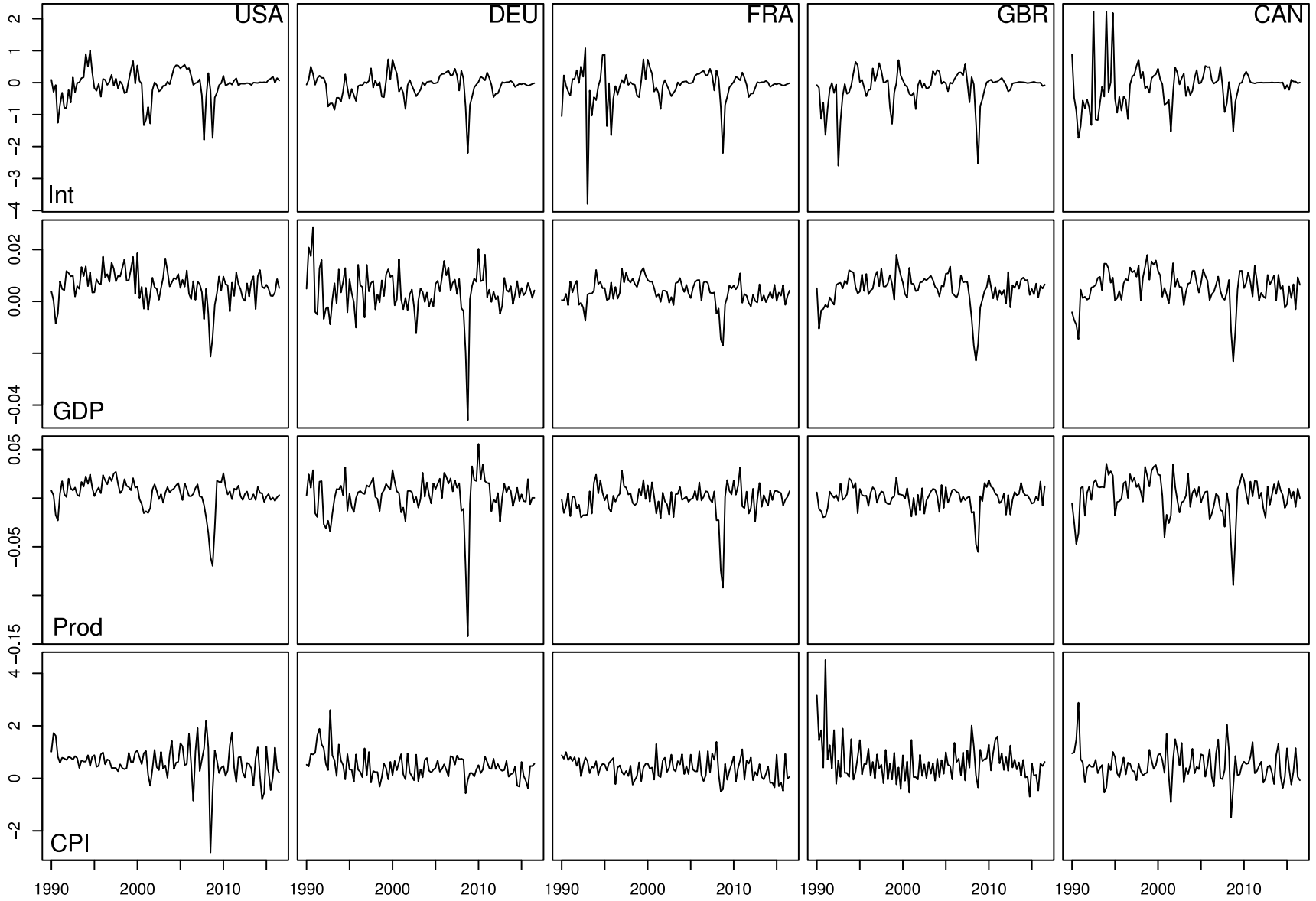


Figure 1 Time series of four economic indicators, sourced from Chen et al. (2021)

(1) Matrix autoregressive (MAR) models

Again suppose n countries, m variables and T time periods.

The traditional VAR(1) model is directly applicable for $\text{vec}(\boldsymbol{X}_t)$. That is,

$$\text{vec}(\boldsymbol{X}_t) = \Phi \text{vec}(\boldsymbol{X}_{t-1}) + \text{vec}(\boldsymbol{E}_t). \tag{6}$$

- The roles of rows and columns are mixed in the VAR model.
- Using the example shown in Fig. 1, the VAR model in (6) fails to recognize the strong connections within the columns (same country) and within the rows (same indicator).
- The (large) $mn \times mn$ coefficient matrix Φ does not have any assumed structure and is also very difficult to interpret.

MAR model: To overcome the drawback of the direct VAR modeling that requires vectorization, and to take advantage of the original matrix structure, the MAR(1) model is given by

$$\boldsymbol{X}_t = A\boldsymbol{X}_{t-1}B' + \boldsymbol{E}_t, \tag{7}$$

where $A = (a_{ij})$ and $B = (b_{ij})$ are $m \times m$ and $n \times n$ autoregressive coefficient matrices, and $E_t = (e_{ij,t})$ is a $m \times n$ matrix white noise.

Clearly the model can be extended to a MAR(p) model

$$\boldsymbol{X}_t = A_1\boldsymbol{X}_{t-1}B_1' + \ldots + A_p\boldsymbol{X}_{t-p}B_p' + \boldsymbol{E}_t, \tag{8}$$

The MAR(1) model in (7) can be represented in the form of a VAR model

$$\mathrm{vec}(\boldsymbol{X}_t) = (B \otimes A)\mathrm{vec}(\boldsymbol{X}_{t-1}) + \mathrm{vec}(\boldsymbol{E}_t). \tag{9}$$

Model comparison

$$\text{vec}(\boldsymbol{X}_t) = \Phi \text{vec}(\boldsymbol{X}_{t-1}) + \text{vec}(\boldsymbol{E}_t)$$

$$\text{vec}(\boldsymbol{X}_t) = (B \otimes A)\text{vec}(\boldsymbol{X}_{t-1}) + \text{vec}(\boldsymbol{E}_t)$$

- The MAR(1) model can be viewed as a special case of the classical VAR(1) model, with its AR coefficient matrix given by a Kronecker product.

- The MAR(1) requires $m^2 + n^2$ coefficients as the entries of A and B, while an unrestricted VAR(1) needs $m^2 n^2$ coefficients for Φ. The latter can be much larger when both m and n are large.

Parameter identification

$$\boldsymbol{X}_t = A\boldsymbol{X}_{t-1}B' + \boldsymbol{E}_t,$$

There is an obvious identifiability issue with the MAR(1) model in (7) regarding the two coefficient matrices A and B. To avoid ambiguity, A is normalized so that its Frobenius norm is one, i.e., $||A||_F = 1$.

(2) Model interpretations

$$\boldsymbol{X}_t = A\boldsymbol{X}_{t-1}B' + \boldsymbol{E}_t,$$

The MAR(1) model is not a straightforward model.

- The left matrix A reflects row-wise interactions, and the right matrix B' introduces column-wise dependence, and therefore the conditional mean combines the row-wise and column-wise interactions.

- We can interpret the model from a row-wise and column-wise VAR model point of view.

Example 1. We consider two special cases:

(i) If $B = I$, then the model becomes

$$\boldsymbol{X}_t = A\boldsymbol{X}_{t-1} + \boldsymbol{E}_t.$$

In this case, each column of $\boldsymbol{X}_t$ follows

$$\boldsymbol{X}_{t,.j} = A\boldsymbol{X}_{t-1,.j} + \boldsymbol{E}_{t,.j}.$$

That is, each column of $\boldsymbol{X}_t$ follows the same VAR(1) model of dimension m.

For the first two columns in the example of Fig. 1, the models are

$$\begin{pmatrix} \text{USA} \\ \text{Int} \\ \text{GDP} \\ \text{Prod} \\ \text{CPI} \end{pmatrix}_t = A \begin{pmatrix} \text{USA} \\ \text{Int} \\ \text{GDP} \\ \text{Prod} \\ \text{CPI} \end{pmatrix}_{t-1} + \begin{pmatrix} \text{USA} \\ \text{e\_Int} \\ \text{e\_GDP} \\ \text{e\_Prod} \\ \text{e\_CPI} \end{pmatrix}_t$$

and

$$\begin{pmatrix} \text{DEU} \\ \text{Int} \\ \text{GDP} \\ \text{Prod} \\ \text{CPI} \end{pmatrix}_t = A \begin{pmatrix} \text{DEU} \\ \text{Int} \\ \text{GDP} \\ \text{Prod} \\ \text{CPI} \end{pmatrix}_{t-1} + \begin{pmatrix} \text{DEU} \\ \text{e\_Int} \\ \text{e\_GDP} \\ \text{e\_Prod} \\ \text{e\_CPI} \end{pmatrix}_t$$

For each country, its economic indicators follow a VAR(1) model of dimension m; and different countries would follow the same VAR(1) model.

(ii) If $A = I$, then

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} B' + \boldsymbol{E}_t.$$

In this case, each row of $\boldsymbol{X}_t$ (same indicator from different countries) would follow a VAR(1) model,

$$\boldsymbol{X}_{t,j.} = \boldsymbol{X}_{t-1,j.} B' + \boldsymbol{E}_{t,j.}$$
$$\boldsymbol{X}'_{t,j.} = B \boldsymbol{X}'_{t-1,j.} + \boldsymbol{E}'_{t,j.}.$$

And the coefficient matrices corresponding to different rows would be the same.

(3) Estimation methods

- **Projection estimator:** This estimator is implemented by projecting $\widehat{\Phi}$ onto the space of Kronecker products under the Frobenius norm:

$$(\widehat{A}_1, \widehat{B}_1) = \arg\min_{A,B} ||\widehat{\Phi} - B \otimes A||_F^2. \tag{10}$$

  - The Frobenius norm of a matrix only depends on the elements in the matrix, but not the arrangement. Let $\widetilde{\Phi} = \mathcal{G}(\widehat{\Phi})$ is the re-arranged $\widehat{\Phi}$, then

$$\begin{aligned}
\min_{A,B} ||\widehat{\Phi} - B \otimes A||_F^2 &= \min_{A,B} ||\mathcal{G}(\widehat{\Phi}) - \mathcal{G}(B \otimes A)||_F^2 \\
&= \min_{A,B} ||\mathcal{G}(\widehat{\Phi}) - \text{vec}(A)\text{vec}(B)'||_F^2 \\
&= \min_{A,B} ||\widetilde{\Phi} - \text{vec}(A)\text{vec}(B)'||_F^2
\end{aligned}$$

- The solution of (10) can be obtained through a singular value decomposition (SVD)

$$\text{vec}(\widehat{A})\text{vec}(\widehat{B})' = d_1 \boldsymbol{u}_1 \boldsymbol{v}_1'$$

  where d_1 is the largest singular value of $\widetilde{\Phi}$, and $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the left and right singular vectors, respectively.
- By converting the vectors into matrices, we obtain the estimators of A and B, denoted by $\widehat{A}$ and $\widehat{B}$.

- **Iterated least squares:** If the entries of $\boldsymbol{E}_t$ are assumed to be i.i.d. normal with mean zero and a constant variance, the maximum likelihood estimator is the solution of the least squares problem

$$(\widehat{A}_2, \widehat{B}_2) = \arg\min_{A,B} \sum_t ||\boldsymbol{X}_t - A\boldsymbol{X}_{t-1}B'||_F^2.$$

To solve this, two matrices $\widehat{A}$ and $\widehat{B}$ are iteratively updated by updating one of them while holding the other one fixed, starting with some initial A and B.

- The iteration of updating B given A is

$$\widehat{B} \leftarrow \left(\sum_t \boldsymbol{X}_t' A \boldsymbol{X}_{t-1} \right) \left(\sum_t \boldsymbol{X}_t' A' A \boldsymbol{X}_{t-1} \right)^{-1}$$

- Similarly the iteration of updating A given B is

$$\widehat{A} \leftarrow \left(\sum_t \boldsymbol{X}_t B \boldsymbol{X}_{t-1}' \right) \left(\sum_t \boldsymbol{X}_{t-1} B' B \boldsymbol{X}_{t-1}' \right)^{-1}$$

(4) An application

Consider the example shown in Fig. 1, ranges from 1991 to 2016. The data consists of quarterly observations of 4 economic indicators from 5 countries:

- Four economic indicators:
  - 3-month interbank Interest Rate (first order differenced series)
  - GDP growth (first order differenced log of GDP series)
  - total manufacturing Production growth (first order differenced log of Production series)
  - Consumer Price Index core inflation (first order differenced log of core inflation series)
- Five countries: Canada, France, Germany, United Kingdom and United States

Table 3
Estimated left coefficient matrix $\boldsymbol{A}$ of MAR(1) using LS method. Standard errors are shown in the parentheses. The right panel indicates the positively significant, negatively significant and insignificant parameters at 5% level using symbols $(+, -, 0)$, respectively.

	Int	GDP	Prod	CPI	Int	GDP	Prod	CPI
Int	0.177	0.215	0.132	−0.171	+	+	0	−
	(0.061)	(0.082)	(0.088)	(0.063)				
GDP	−0.19	0.341	0.346	−0.08	−	+	+	0
	(0.05)	(0.086)	(0.081)	(0.062)				
Prod	−0.223	0.318	0.424	−0.095	−	+	+	0
	(0.054)	(0.092)	(0.087)	(0.068)				
CPI	−0.028	0.048	−0.045	0.502	0	0	0	+
	(0.05)	(0.07)	(0.078)	(0.052)				

Table 4
Estimated right coefficient matrix $\boldsymbol{B}$ of MAR(1) using LS method. Standard errors are shown in parentheses. The right panel indicates the positively significant, negatively significant and insignificant parameters at 5% level using symbols $(+, -, 0)$, respectively.

	USA	DEU	FRA	GBR	CAN	USA	DEU	FRA	GBR	CAN
USA	0.878	−0.044	0.15	0.359	−0.043	+	0	0	+	0
	(0.134)	(0.202)	(0.138)	(0.132)	(0.156)					
DEU	0.722	0.072	0.801	0.308	−0.212	+	0	+	+	−
	(0.076)	(0.124)	(0.083)	(0.078)	(0.092)					
FRA	0.44	0.064	0.438	0.208	0.024	+	0	+	0	0
	(0.12)	(0.197)	(0.136)	(0.125)	(0.148)					
GBR	0.545	0.032	0.272	0.406	−0.018	+	0	+	+	0
	(0.089)	(0.153)	(0.101)	(0.101)	(0.118)					
CAN	0.553	0.023	−0.002	0.531	0.324	+	0	0	+	+
	(0.079)	(0.13)	(0.087)	(0.085)	(0.1)					

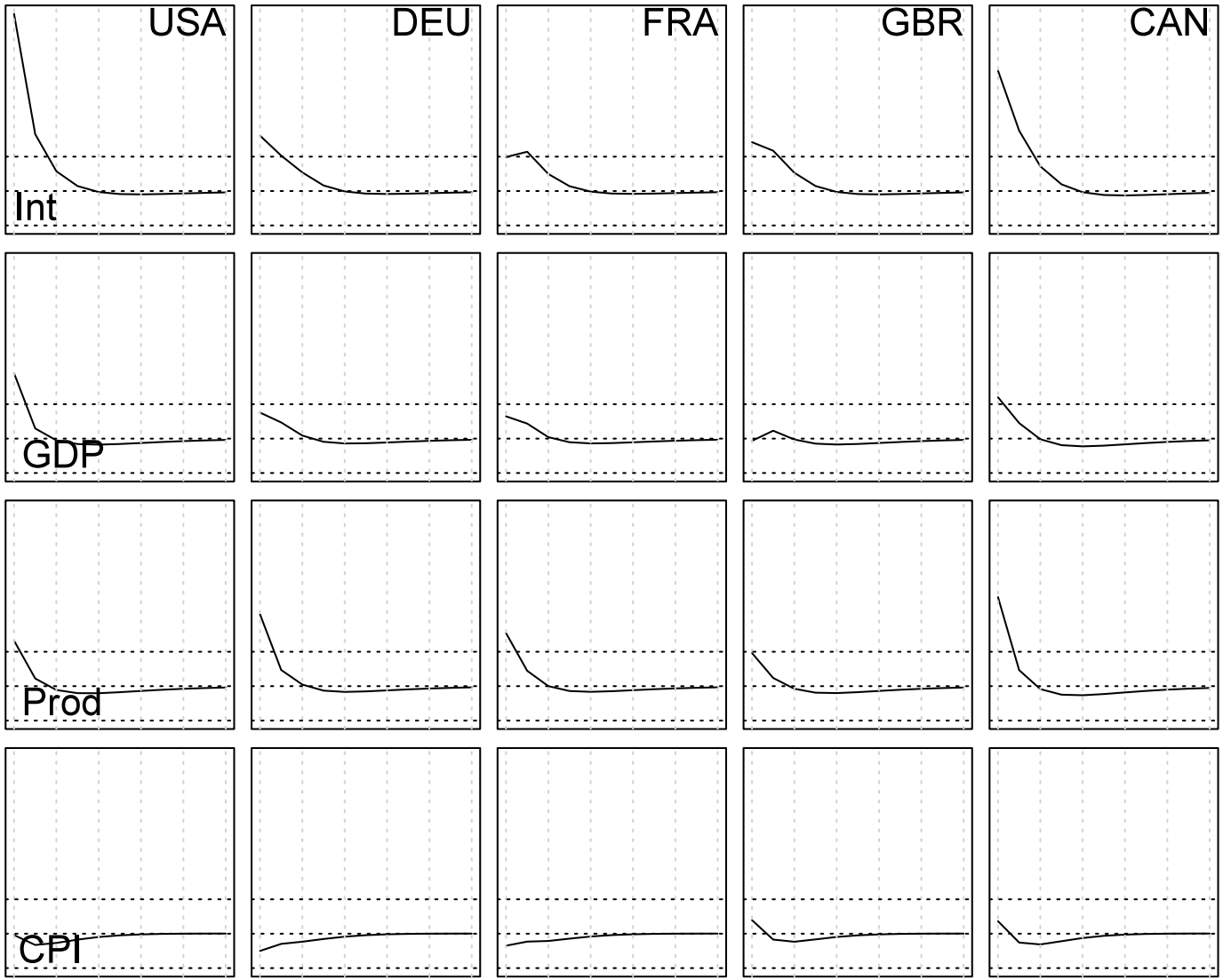


Figure 2 The impulse response functions with orthogonal innovations with one standard deviation shock on US interest rate

4. Network VAR Model

With the rapid advance of social network sites such as Facebook, Twitter, Weixin, Sina Weibo and many others, network data are becoming increasingly available.

In order to incorporate the network information among individuals, Zhu et al. (2017) developed a network vector autoregression (NAR) model.

Assume that the response Y_{it} is in influenced by:
(a) its lagged value $Y_{i(t-1)}$,
(b) its socially connected neighbors (社交关系邻居) $n_i^{-1} \sum_j a_{ij} Y_{j(t-1)}$,
(c) a set of node-specific (节点特异) covariates $Z_i = (Z_{i1}, \ldots, Z_{ip})' \in \mathbb{R}$,
(d) and an independent noise ε_{it}.

As a result, the NAR model in Zhu et al. (2017) is specified as

$$Y_{it} = \beta_0 + Z_i'\gamma + \beta_1 n_i^{-1} \sum_j a_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \varepsilon_{it}, \tag{11}$$

- $n_i = \sum_{j \neq i} a_{ij}$ is the total number of nodes that i follows, and it is called out-degree.
  - ▶ This model implicitly assumes that the focal node i is unlikely to be affected by another node j, unless i follows j;
- The term $Y_{i(t-1)}$ is the standard autoregressive impact, and β_2 is referred to as the momentum effect.

$$Y_{it} = \beta_0 + Z_i'\gamma + \beta_1 n_i^{-1} \sum_j a_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \varepsilon_{it},$$

- the term $\beta_0 + Z_i'\gamma$ characterizes the nodal impact of the i-th node, where $\beta_0 \in \mathbb{R}$ is the intercept and $\gamma = (\gamma_1, \ldots, \gamma_p)$ is the associated coefficient (i.e., nodal effect).
  - For convenience, we write $\beta_{0i} = \beta_0 + Z_i\gamma$;
  - The nodal impact (i.e., β_{0i}) is invariant over time t.
- The quantity $n_i^{-1} \sum_j a_{ij} Y_{j(t-1)}$ is the average impact from i-th neighbors.
  - Its associated parameter β_1 is referred to as the network effect.

$$Y_{it} = \beta_0 + Z_i'\gamma + \beta_1 n_i^{-1} \sum_j a_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \varepsilon_{it},$$

- To reduce the dimensionality, it is assumed that the error term

$$\varepsilon_{it} \sim N(0, \sigma^2),$$

so a diagonal covariance structure is assumed in this work.

5. Large TVP-VAR Model

(1) Small TVP-VAR model

We begin with a basic structural VAR model defined as

$$Ay_t = F_1 y_{t-1} + \ldots + F_p y_{t-p} + u_t, \quad u_t \sim N(0, \Sigma\Sigma), \tag{12}$$

where $\Sigma = \text{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$ and A is a lower-triangular matrix with the diagonal elements being ones. Rewrite it as the reduced-form model

$$y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + A^{-1}\Sigma\varepsilon_t, \quad \varepsilon_t \sim N(0, I)$$

Furthermore, defining $\boldsymbol{X}_t = I_N \otimes (y'_{t-1}, \ldots, y'_{t-p})$, the model can be rewritten as

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + A^{-1}\Sigma\varepsilon_t.$$

Consider the time varying parameter (TVP) VAR model specified by

$$y_t = \boldsymbol{X}_t \boldsymbol{\beta}_t + A_t^{-1} \Sigma_t \varepsilon_t. \tag{13}$$

where $\boldsymbol{X}_t = I_N \otimes (y'_{t-1}, \ldots, y'_{t-p})$, $\boldsymbol{\beta}_t = \mathrm{vec}((B_{1t}, B_{2t}, \ldots, B_{pt})')$,

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_{N1,t} & \cdots & a_{NN-1,t} & 1 \end{bmatrix}, \quad \Sigma_t = \begin{bmatrix} \sigma_{1t} & 0 & \cdots & 0 \\ 0 & \sigma_{2t} & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{Nt} \end{bmatrix}.$$

Following Primiceri (2005), let

$$\boldsymbol{\alpha}_t = (a_{21,t}, a_{31,t}, a_{32,t}, a_{41,t}, \ldots, a_{N,N-1,t})',$$

$$\boldsymbol{h}_t = (h_{1t}, \ldots, h_{Nt})', \quad \text{with } h_{jt} = \log \sigma_{jt}^2 \text{ for } j = 1, \ldots, N.$$

It is assumed that the parameters in (13) follow a random walk process

$$\begin{aligned}\boldsymbol{\beta}_{t+1} &= \boldsymbol{\beta}_t + \boldsymbol{u}_{\beta t},\\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{\alpha}_t + \boldsymbol{u}_{\alpha t},\\ \boldsymbol{h}_{t+1} &= \boldsymbol{h}_t + \boldsymbol{u}_{ht},\end{aligned} \qquad \begin{pmatrix} \varepsilon_t \\ \boldsymbol{u}_{\beta t} \\ \boldsymbol{u}_{\alpha t} \\ \boldsymbol{u}_{ht} \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & O & O & O \\ O & \Sigma_\beta & O & O \\ O & O & \Sigma_\alpha & O \\ O & O & O & \Sigma_h \end{pmatrix}\right)$$

where $\Sigma_h = \mathrm{diag}\{\sigma_{h1}^2, \ldots, \sigma_{hN}^2\}$ is a diagonal matrix.

Estimation method

Given the data $\boldsymbol{y}$, we draw samples from the posterior distribution, $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{h}, \Sigma_\beta, \Sigma_\alpha, \Sigma_h | \boldsymbol{y})$, by the following MCMC algorithm:

1. Initialize $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{h}$ and $\Sigma_\beta, \Sigma_\alpha, \Sigma_h$.
2. Sample $\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{h}, \Sigma_\beta, \boldsymbol{y}$ by FFBS (forward filtering and backward sampling).
3. Sample $\Sigma_\beta|\boldsymbol{\beta}$ from Inverse Wishart posterior.
4. Sample $\boldsymbol{\alpha}|\boldsymbol{\beta}, \boldsymbol{h}, \Sigma_\alpha, \boldsymbol{y}$ by FFBS.
5. Sample $\Sigma_\alpha|\boldsymbol{\alpha}$ from Inverse Wishart posterior.
6. Sample $\boldsymbol{h}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_h, \boldsymbol{y}$ by single-move or multi-move sampling.
7. Sample $\sigma_{hi}^2|\boldsymbol{h}$ from Inverse Gamma posterior, $i = 1, \ldots, N$.
8. Go to 2.

Applications

- The original intention is to capture the transmission path of monetary policies, see for example Cogley and Sargent (2005), Primiceri (2005), Koop et al. (2009) and Nakajima (2011).
- In a study of network spillovers among financial institutions, Geraci and Gnabo (2018) demonstrate the utility of TVP-VAR models for a system of four sectors.
- Similarly, Ciccarelli and Rebucci (2007) use a TVP-VAR to study contagion and interdependence among exchange rates.
- In psychology, the TVP-VAR is used to model emotion dynamics and has been proposed for the study of networks in psychopathology (精神病理学网络) (see Bringmann et al. (2018) and references therein).

(2) Large TVP-VAR model

Consider the reduced form large TVP-VAR model given by:

$$y_t = \boldsymbol{X}_t \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t) \tag{14}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + u_t, \quad u_t \sim N(0, Q_t). \tag{15}$$

The MCMC method works well with small TVP-VARs, but can be computationally very demanding in larger VARs.

Given the Minnesota prior $\boldsymbol{\beta}_0 \sim N(b_0, P_0)$ and initial values Σ_0, Koop and Korobilis (2013) estimate it using forgetting factors and EWMA volatility

$$\widehat{\Sigma}_t = \kappa \widehat{\Sigma}_{t-1} + (1-\kappa)\widetilde{\varepsilon}_t \widetilde{\varepsilon}_t',$$

where $\widetilde{\varepsilon}_t = y_t - \boldsymbol{X}_t \boldsymbol{\beta}_{t|t-1}$ is produced by the Kalman filter.

Kalman filter with forgetting factor

- **Prediction step**
  - Set $\boldsymbol{\beta}_{t|t-1} = \boldsymbol{\beta}_{t-1|t-1}$
  - Set $P_{t|t-1} = \lambda^{-1} P_{t-1|t-1}$ (forgetting factor $Q_t = (\lambda^{-1} - 1) P_{t-1|t-1}$)

  where for $t = 1$ we use the fact that $\boldsymbol{\beta}_{0|0} = b_0$ and $P_{0|0} = P_0$.

- **Update step**
  - Estimate $\widehat{\varepsilon}_t = y_t - \boldsymbol{X}_t \boldsymbol{\beta}_{t|t-1}$ (measurement residual)
  - Estimate $\widehat{\Sigma}_t = \kappa \widehat{\Sigma}_{t-1} + (1-\kappa) \widehat{\varepsilon}_t \widehat{\varepsilon}_t'$ by EWMA, where for $t = 1$ it simply holds that $\widehat{\Sigma}_1 = \kappa \Sigma_0$.
  - Estimate $\boldsymbol{\beta}_{t|t} = \boldsymbol{\beta}_{t|t-1} + P_{t|t-1} \boldsymbol{X}_t' (\widehat{\Sigma}_t + \boldsymbol{X}_t P_{t|t-1} \boldsymbol{X}_t')^{-1} \widetilde{\varepsilon}_t$.
  - Estimate $P_{t|t} = P_{t|t-1} - P_{t|t-1} \boldsymbol{X}_t' (\widehat{\Sigma}_t + \boldsymbol{X}_t P_{t|t-1} \boldsymbol{X}_t')^{-1} \boldsymbol{X}_t P_{t|t-1}$.

(3) Fast estimation of a large TVP-VAR with score-driven volatilities

- Zheng, Ye and Hong (Journal of Economic Dynamics & Control, 2023)

- A large TVP-VAR model is reparameterized as follows:

$$\boldsymbol{A}_t\boldsymbol{y}_t = \boldsymbol{c}_t + \boldsymbol{B}_{1,t}\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{B}_{p,t}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_t), \quad (16)$$

where $\boldsymbol{\Sigma}_t = \text{diag}\{h_{1,t}, \ldots, h_{N,t}\}$ is a diagonal matrix, and $\boldsymbol{A}_t$ is a time-varying $N \times N$ lower triangular matrix with the form:

$$\boldsymbol{A}_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_{N1,t} & \cdots & a_{NN-1,t} & 1 \end{bmatrix}.$$

- The i-th equation in the TVP-SVAR system (16) is then given by:

$$y_{i,t} = c_{i,t} + \widetilde{\boldsymbol{x}}_{1,t}'\boldsymbol{\beta}_{i,t}^{(1)} + \cdots + \widetilde{\boldsymbol{x}}_{p,t}'\boldsymbol{\beta}_{i,t}^{(p)} + \widetilde{\boldsymbol{w}}_{i,t}'\boldsymbol{\alpha}_{i,t} + \varepsilon_{i,t}$$

where $\boldsymbol{\beta}_{i,t}^{(k)} = (b_{i1,t}^{(k)}, \ldots, b_{iN,t}^{(k)})'$, $\widetilde{\boldsymbol{x}}_{k,t} = (y_{1,t-k}, \ldots, y_{N,t-k})'$, $\boldsymbol{\alpha}_{i,t} = (a_{i1,t}, \ldots, a_{ii-1,t})'$ and $\widetilde{\boldsymbol{w}}_{i,t} = (-y_{1,t}, \ldots, -y_{i-1,t})'$, for $i = 2, \ldots, N$.

- The i-th structural equation is further rewritten as:

$$y_{i,t} = \boldsymbol{z}_{i,t}'\boldsymbol{\beta}_{i,t} + \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim \mathcal{N}(0, h_{i,t}), \tag{17}$$

where $\boldsymbol{z}_{i,t}' = (1, \widetilde{\boldsymbol{x}}_{1,t}', \ldots, \widetilde{\boldsymbol{x}}_{p,t}', \widetilde{\boldsymbol{w}}_{i,t}')$ and $\boldsymbol{\beta}_{i,t} = (c_{i,t}, \boldsymbol{\beta}_{i,t}^{(1)\prime}, \ldots, \boldsymbol{\beta}_{i,t}^{(p)\prime}, \boldsymbol{\alpha}_{i,t}')'$, which is of dimension $k_i = Np + i$, $i = 1, \ldots, N$.

- For $\boldsymbol{\beta}_{i,t}$, it is specified as the following random walk process:

$$\boldsymbol{\beta}_{i,t+1} = \boldsymbol{\beta}_{i,t} + \boldsymbol{v}_{i,t}. \tag{18}$$

We assume that the stacked vector $\boldsymbol{v}_t = (\boldsymbol{v}_{1t}', \ldots, \boldsymbol{v}_{Nt}')' \sim \mathcal{N}(0, \boldsymbol{Q}_t)$.

State space representation:

Stacking all the equations, we have the following forward-form state space representation:

$$\boldsymbol{y}_t = \boldsymbol{Z}_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_t), \tag{19}$$

$$\boldsymbol{\beta}_{t+1} = \boldsymbol{\beta}_t + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim \mathcal{N}(0, \boldsymbol{Q}_t), \tag{20}$$

where $\boldsymbol{y}_t$ and $\boldsymbol{\beta}_t$ are the stacked vectors of $y_{i,t}$ and $\boldsymbol{\beta}_{i,t}$, $i = 1, \ldots, N$, respectively, and $\boldsymbol{Z}_t$ is the stacked matrix with the form:

$$\boldsymbol{Z}_t = \begin{bmatrix} \boldsymbol{z}'_{1,t} & 0 & \cdots & 0 \\ 0 & \boldsymbol{z}'_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{z}'_{N,t} \end{bmatrix}_{N \times \left(N^2 p + \frac{N(N+1)}{2}\right)}.$$

Score-driven log-volatilities, $\log \boldsymbol{h}_t$

- Following the generalized autoregressive or GAS model proposed by Creal et al. (2013), the vector of score-driven log-volatilities, $\boldsymbol{\mu}_t = (\log h_{1,t|t-1}, \ldots, \log h_{N,t|t-1})'$, is specified as:

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{\mu}_{t|t-1} + A\boldsymbol{s}_t, \qquad \boldsymbol{s}_t = \mathcal{I}_t^{-1}\nabla_t,$$

where $\nabla_t = \partial \ell_t / \partial \boldsymbol{\mu}_{t|t-1}$, and $\mathcal{I}_t = \mathrm{E}_{t-1}[\nabla_t \nabla_t']$.

  - We let $A = \alpha I_N$, where α is a common positive parameter representing an average speed of innovation.
  - The log likelihood function: $\ell_t = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{F}_t| - \frac{1}{2}\widetilde{\boldsymbol{y}}_t'\boldsymbol{F}_t^{-1}\widetilde{\boldsymbol{y}}_t$, where $\widetilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{Z}_t\boldsymbol{\beta}_{t|t-1}$ and $\boldsymbol{F}_t = \boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t' + \boldsymbol{\Sigma}_t$ are produced by the Kalman filter.

By simple computation, we have

$$\boldsymbol{s}_t = \boldsymbol{I}_t^{-1}\boldsymbol{\nabla}_t = \begin{bmatrix} (\widetilde{y}_{1t}^2 - \sigma_{1t}^2)h_{1t}^{-1} \\ (\widetilde{y}_{2t}^2 - \sigma_{2t}^2)h_{2t}^{-1} \\ \vdots \\ (\widetilde{y}_{Nt}^2 - \sigma_{Nt}^2)h_{Nt}^{-1} \end{bmatrix} := (\widetilde{\boldsymbol{y}}_t \circ \widetilde{\boldsymbol{y}}_t - \boldsymbol{\sigma}_t \circ \boldsymbol{\sigma}_t) \circ \boldsymbol{h}_t^{-1},$$

where the operator ($\circ$) denotes the Hadamard's product.

Finally, we obtain the following vector form GAS(1,1) updating equation for the vector of time-varying log-volatilities:

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{\mu}_{t|t-1} + \alpha(\widetilde{\boldsymbol{y}}_t \circ \widetilde{\boldsymbol{y}}_t - \boldsymbol{\sigma}_t \circ \boldsymbol{\sigma}_t) \circ \boldsymbol{h}_t^{-1}.$$

Algorithm 1 (Score-driven Kalman filter (SDKF-SYS)). *Given the initial values* $\boldsymbol{\beta}_{1|0} = \boldsymbol{b}_1$, $\boldsymbol{P}_1$, *and* $\boldsymbol{h}_1$ *(thus* $\boldsymbol{\Sigma}_1$*). Then for* $t = 2, \ldots, T$, *we have the following filtering recursion:*

- *At each time* t, *run the Kalman filter and the score-driven updating:*

$$\widetilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{Z}_t \boldsymbol{\beta}_{t|t-1}, \qquad \boldsymbol{F}_t = \boldsymbol{Z}_t \boldsymbol{P}_t \boldsymbol{Z}_t' + \boldsymbol{\Sigma}_t,$$

$$\boldsymbol{K}_t = \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1}, \qquad \boldsymbol{L}_t = I - \boldsymbol{K}_t \boldsymbol{Z}_t,$$

$$\boldsymbol{\beta}_{t+1|t} = \boldsymbol{\beta}_{t|t} = \boldsymbol{\beta}_{t|t-1} + \boldsymbol{K}_t \widetilde{\boldsymbol{y}}_t, \qquad \boldsymbol{P}_{t+1} = \boldsymbol{P}_{t|t}/\lambda = \boldsymbol{P}_t \boldsymbol{L}_t'/\lambda,$$

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_t + \alpha(\widetilde{\boldsymbol{y}}_t \circ \widetilde{\boldsymbol{y}}_t - \boldsymbol{\sigma}_t \circ \boldsymbol{\sigma}_t) \circ \boldsymbol{h}_t^{-1}, \quad \boldsymbol{h}_{t+1} = \exp(\boldsymbol{\mu}_{t+1|t}).$$

- *Store* $\widetilde{\boldsymbol{y}}_t$, $\boldsymbol{\beta}_{t+1|t}$, $\boldsymbol{P}_{t+1}$, *and* $\boldsymbol{\mu}_{t|t-1}$ *for smoothing estimation.*

Algorithm 2 (Score-driven Kalman filter (SDKF-EBE)). *Given the initial values* $\boldsymbol{\beta}_{1|0} = \boldsymbol{b}_1$, $\boldsymbol{P}_1$, *and* $\boldsymbol{h}_1$ *(thus* $\boldsymbol{\Sigma}_1$*). Then for* $t = 2, \ldots, T$, *we have :*

- *At each time point* t, *for* $i = 1, \ldots, N$, *run the Kalman filter and score-driven updating in equation-by-equation:*

$$\widetilde{y}_{i,t} = y_{i,t} - \boldsymbol{z}_{i,t}'\boldsymbol{\beta}_{i,t|t-1}, \qquad \sigma_{i,t}^2 = \boldsymbol{z}_{i,t}'\boldsymbol{P}_{i,t}\boldsymbol{z}_{i,t} + h_{i,t},$$

$$\boldsymbol{k}_{i,t} = \boldsymbol{P}_{i,t}\boldsymbol{z}_{i,t}\sigma_{i,t}^{-2}, \qquad \boldsymbol{L}_{i,t} = I - \boldsymbol{k}_{i,t}\boldsymbol{z}_{i,t}',$$

$$\boldsymbol{\beta}_{i,t+1|t} = \boldsymbol{\beta}_{i,t|t} = \boldsymbol{\beta}_{i,t|t-1} + \boldsymbol{k}_{i,t}\widetilde{y}_{i,t}, \qquad \boldsymbol{P}_{i,t+1} = \boldsymbol{P}_{i,t|t} = \boldsymbol{P}_{i,t}\boldsymbol{L}_{i,t}'/\lambda,$$

$$\mu_{i,t+1|t} = \mu_{i,t|t} = \mu_{i,t|t-1} + \alpha(\widetilde{y}_{i,t}^2 - \sigma_{i,t}^2)/h_{i,t}, \qquad h_{i,t+1} = \exp(\mu_{i,t+1|t}).$$

- *Store the smoothing quantities, such as* $\widetilde{\boldsymbol{y}}_t = (\widetilde{y}_{1,t}, \ldots, \widetilde{y}_{N,t})'$, $\boldsymbol{\beta}_{t+1|t} = (\boldsymbol{\beta}_{1,t+1|t}', \ldots, \boldsymbol{\beta}_{N,t+1|t}')'$, $\boldsymbol{P}_{t+1} = \operatorname{diag}\{\boldsymbol{P}_{1,t+1}, \ldots, \boldsymbol{P}_{N,t+1}\}$, *and* $\boldsymbol{h}_t = (h_{1,t}, \ldots, h_{N,t})'$.

Algorithm 3 (Score-driven Kalman smoother (SDKS)). *For* $t = T-1, \ldots, 1$, *the SDKS can be written as a backward recursion:*

$$\widehat{\boldsymbol{\beta}}_t = \lambda \widehat{\boldsymbol{\beta}}_{t+1} + (1-\lambda)\boldsymbol{\beta}_{t+1|t}, \qquad \widehat{\boldsymbol{P}}_t = \lambda^2 \widehat{\boldsymbol{P}}_{t+1} + (\lambda - \lambda^2)\boldsymbol{P}_{t+1},$$

$$\boldsymbol{r}_{t-1} = \boldsymbol{\nabla}_t^{\mu} + (1-\alpha)\boldsymbol{r}_t, \qquad \widehat{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_{t|t-1} + \alpha \boldsymbol{S}_t^{\boldsymbol{\mu}} \boldsymbol{r}_{t-1},$$

where $\boldsymbol{\nabla}_t = ((\widetilde{y}_{1,t}^2 - \sigma_{1,t}^2)h_{1,t}/2\sigma_{1,t}^4, \ldots, (\widetilde{y}_{N,t}^2 - \sigma_{N,t}^2)h_{N,t}/2\sigma_{N,t}^4)$, $\boldsymbol{S}_t^{\mu} = \operatorname{diag}\{2\sigma_{1,t}^4 h_{1,t}^{-2}, \ldots, 2\sigma_{N,t}^4 h_{N,t}^{-2}\}$, *and the smoothed estimators of* $\boldsymbol{h}_t$ *and* $\boldsymbol{\Sigma}_t$ *are obtained by letting* $\widehat{\boldsymbol{h}}_t = \exp(\widehat{\boldsymbol{\mu}}_t)$ *and* $\widehat{\boldsymbol{\Sigma}}_t = \operatorname{diag}\{\widehat{\boldsymbol{h}}_t\}$.

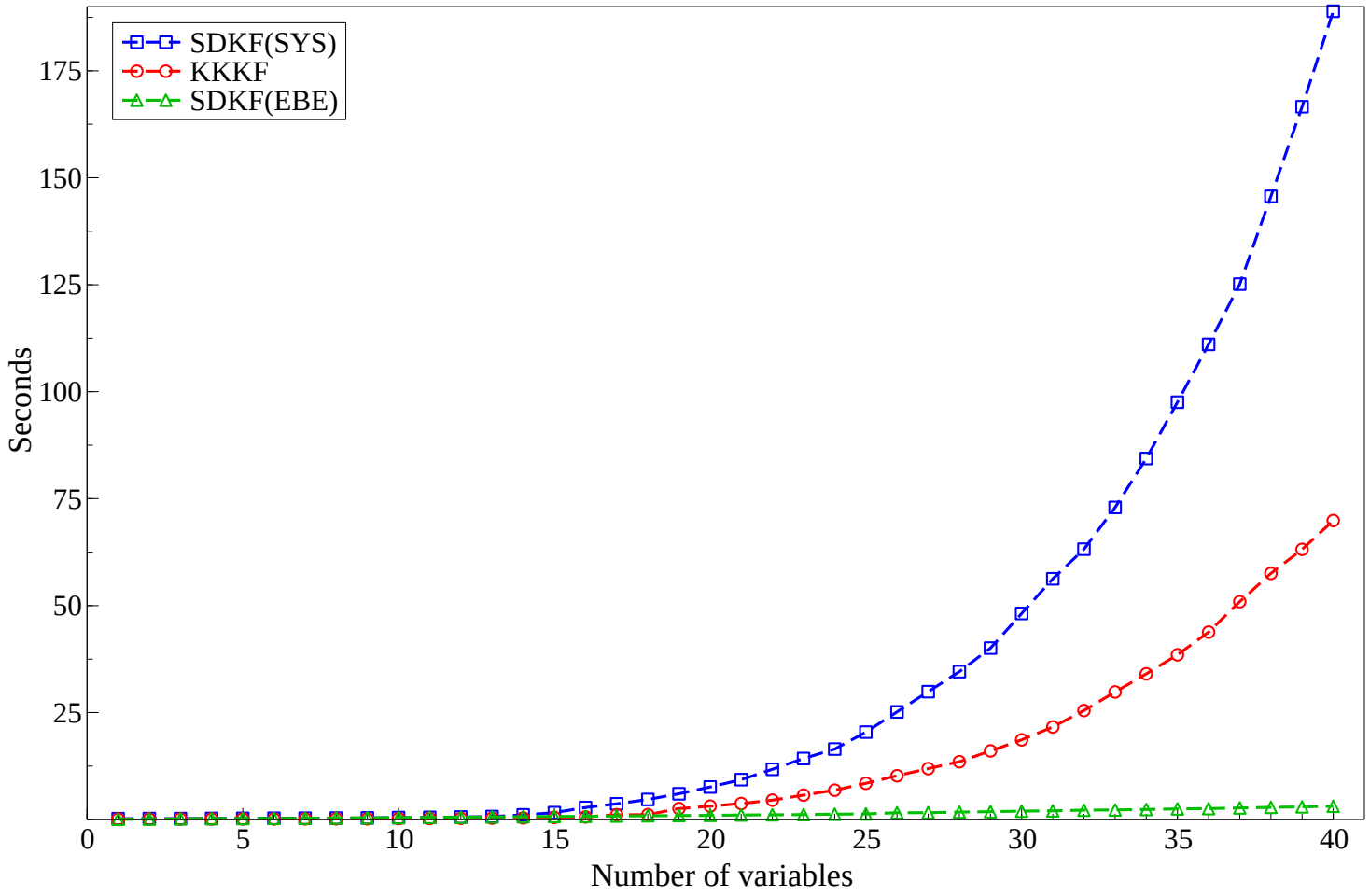


Figure 3 Estimation time for DGP1 under different model size N. In all cases, we set $p = 1$ and $T = 1000$.

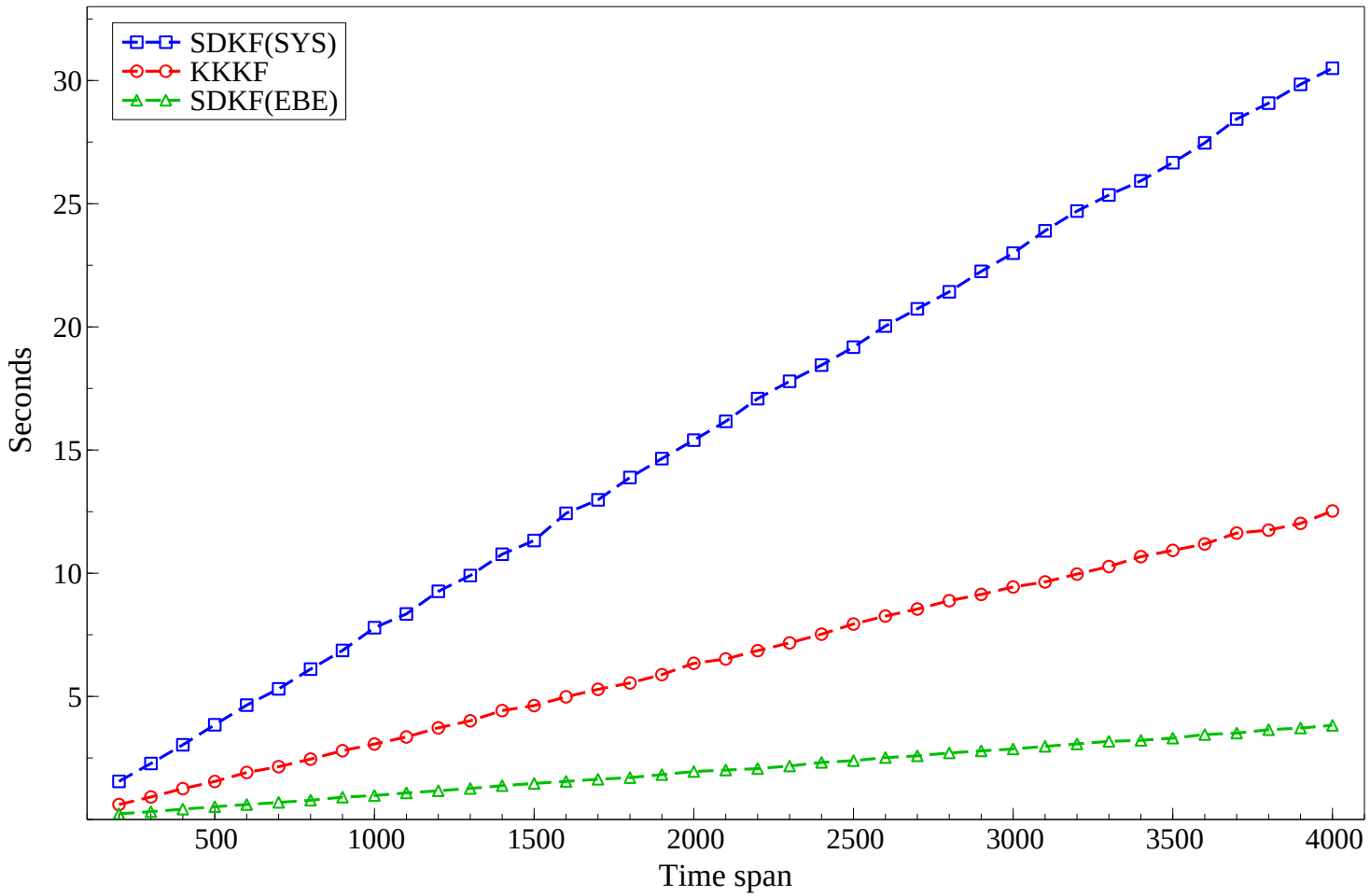


Figure 4 Estimation time for DGP1 under different time span T. In all cases, we set $p = 1$ and $N = 20$.

(4) Other estimation procedures

- Rolling window Lasso-VAR estimation used by Demirer, Diebold, Liu, and Yilmaz (2018)

- Nonparametric estimation proposed by Kapetanios, Marcellino, and Venditti (2019)

- A dimension reduction method by Chan, Eisenstat, and Strachan (2020)

  ▶ They propose a specification that uses the singularity of the covariance matrix of the state equation to develop a factor-like structure to estimate a TVP-SVAR for many variables.

(5) Connectedness

The connectedness index (market risk) can be directly calculated using the estimated results from the reduced form TVP-VAR model.

Given $\widetilde{\boldsymbol{c}}_t$, $\widetilde{\boldsymbol{B}}_{k,t}$ for $k = 1, \ldots, p$ and $\widetilde{\boldsymbol{\Sigma}}_t$, we first transform it to the TVP-VMA form at each time t and obtain the MA coefficient matrices

$$\boldsymbol{\Psi}_{h,t} = \widetilde{\boldsymbol{B}}_{1,t}\boldsymbol{\Psi}_{h-1,t} + \ldots + \widetilde{\boldsymbol{B}}_{p,t}\boldsymbol{\Psi}_{h-p,t}$$

with for $h = 1, \ldots, H$ with $\boldsymbol{\Psi}_{0,t} = I$ and $\boldsymbol{\Psi}_{i,t} = 0$ for $i < 0$.

Following Diebold and Yilmaz (2012), variable j's contribution to variable i's H-step-ahead generalized forecast error variance is

$$\theta_{ij,t}(H) = \frac{\widetilde{\sigma}_{jj,t}^{-1} \sum_{h=0}^{H-1} (e_i' \boldsymbol{\Psi}_{t,h} \widetilde{\boldsymbol{\Sigma}}_t e_j)^2}{\sum_{h=0}^{H-1} (e_i' \boldsymbol{\Psi}_{t,h} \widetilde{\boldsymbol{\Sigma}}_t \boldsymbol{\Psi}_{t,h}' e_i)}.$$

- The sum of the elements of each row of the generalized variance decomposition (GVD) matrix is not necessarily equal to 1: $\sum_{j=1}^{N} \theta_{ij,t}(H) \neq 1$.

- Normalize each entry of the GVD matrix by:

$$d_{ij,t}^{H} = \frac{\theta_{ij,t}(H)}{\sum_{j=1}^{N} \theta_{ij,t}(H)}.$$

Dynamic connectedness: The total connectedness or spillover index at time t is then defined as:

$$C_t^H = \frac{1}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} d_{ij,t}^{H} * 100. \tag{21}$$

Figure 5 Connectedness Table

	x_1	x_2	...	x_N	From Others
x_1	d_{11}^H	d_{12}^H	$\cdots$	d_{1N}^H	$\Sigma_{j=1}^N d_{1j}^H, j \neq 1$
x_2	d_{21}^H	d_{22}^H	$\cdots$	d_{2N}^H	$\Sigma_{j=1}^N d_{2j}^H, j \neq 2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_N	d_{N1}^H	d_{N2}^H	$\cdots$	d_{NN}^H	$\Sigma_{j=1}^N d_{Nj}^H, j \neq N$
To Others	$\Sigma_{i=1}^N d_{i1}^H$ $i \neq 1$	$\Sigma_{i=1}^N d_{i2}^H$ $i \neq 2$	$\cdots$	$\Sigma_{i=1}^N d_{iN}^H$ $i \neq N$	$\frac{1}{N}\Sigma_{i,j=1}^N d_{ij}^H$ $i \neq j$

Application 1 - Exchange rate connectedness

Antonakakis, Chatziantoniou, and Gabauer (2020), *"Refined* Measures of Dynamic Connectedness Based on Time-Varying Parameter Vector Autoregressions", *Journal of Risk and Financial Management*, 13.

They estimated time-varying or dynamic connectedness (see Part 1) based on TVP-VAR model and compare it with the results from rolling window methods with different window sizes.

Only four data series are considered, including: *EUR/USD, GBP/USD, Swiss franc (CHF)/USD, and JPY/USD*, range from February 1975 until January 2019. Daily returns are utilized.

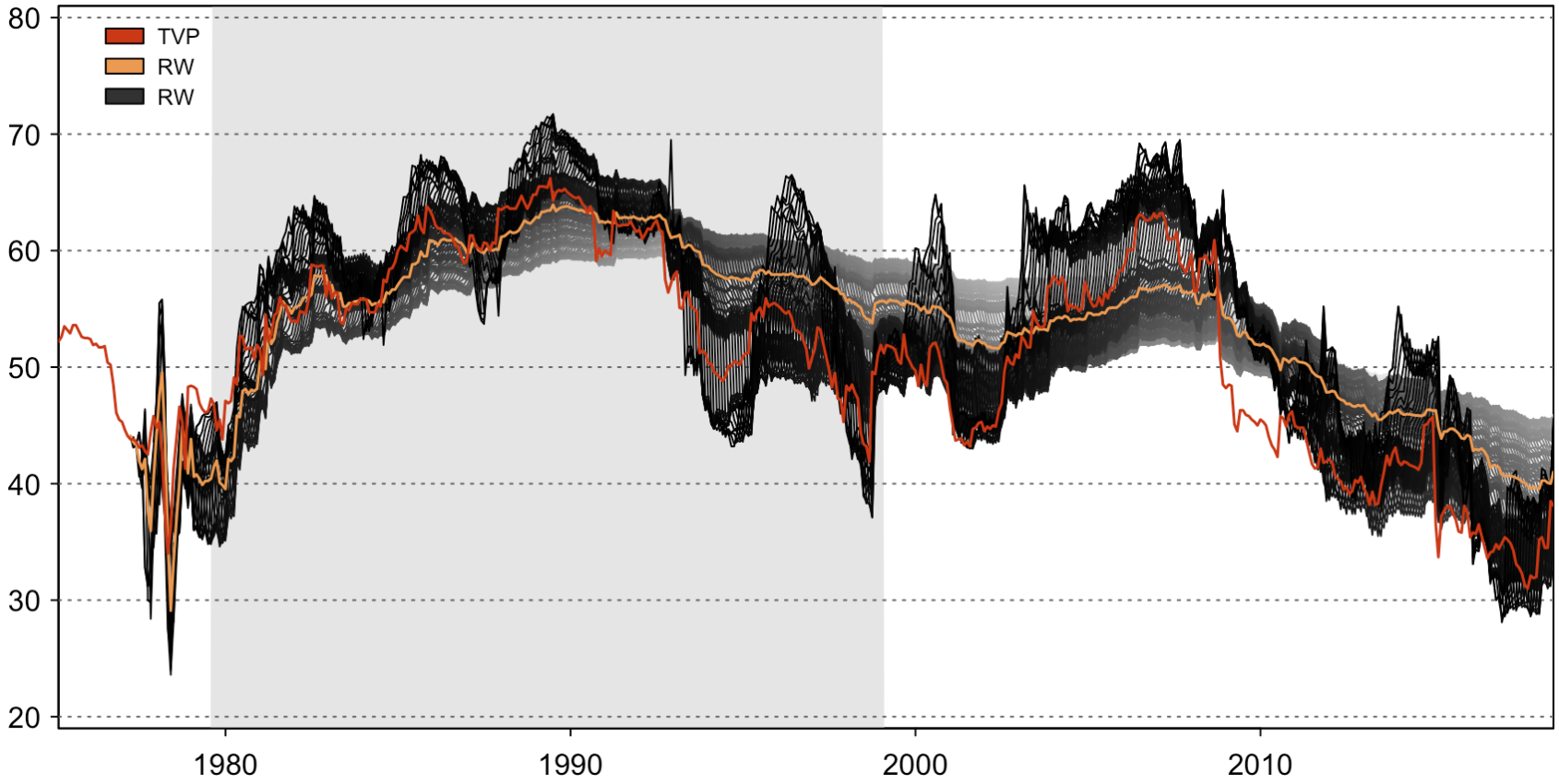


Figure 6 Dynamic total connectedness. the darker the series, the smaller the window size, and vice versa (25, 26, . . . , 274, 275)

Application 2 - Stock volatility connectedness - small data

Korobilis and Yilmaz, 2018, "Measuring Dynamic Connectedness with Large Bayesian VAR Models." Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3099725`.

They study stock return volatilities of 35 major financial institutions; 17 of these are American financial institutions while the remaining 18 are European.

The sample covers 2 January 2004 - 22 July 2016 with 3235 daily observations.

For a given financial institution on a given day, they use a daily range-based volatility estimate.

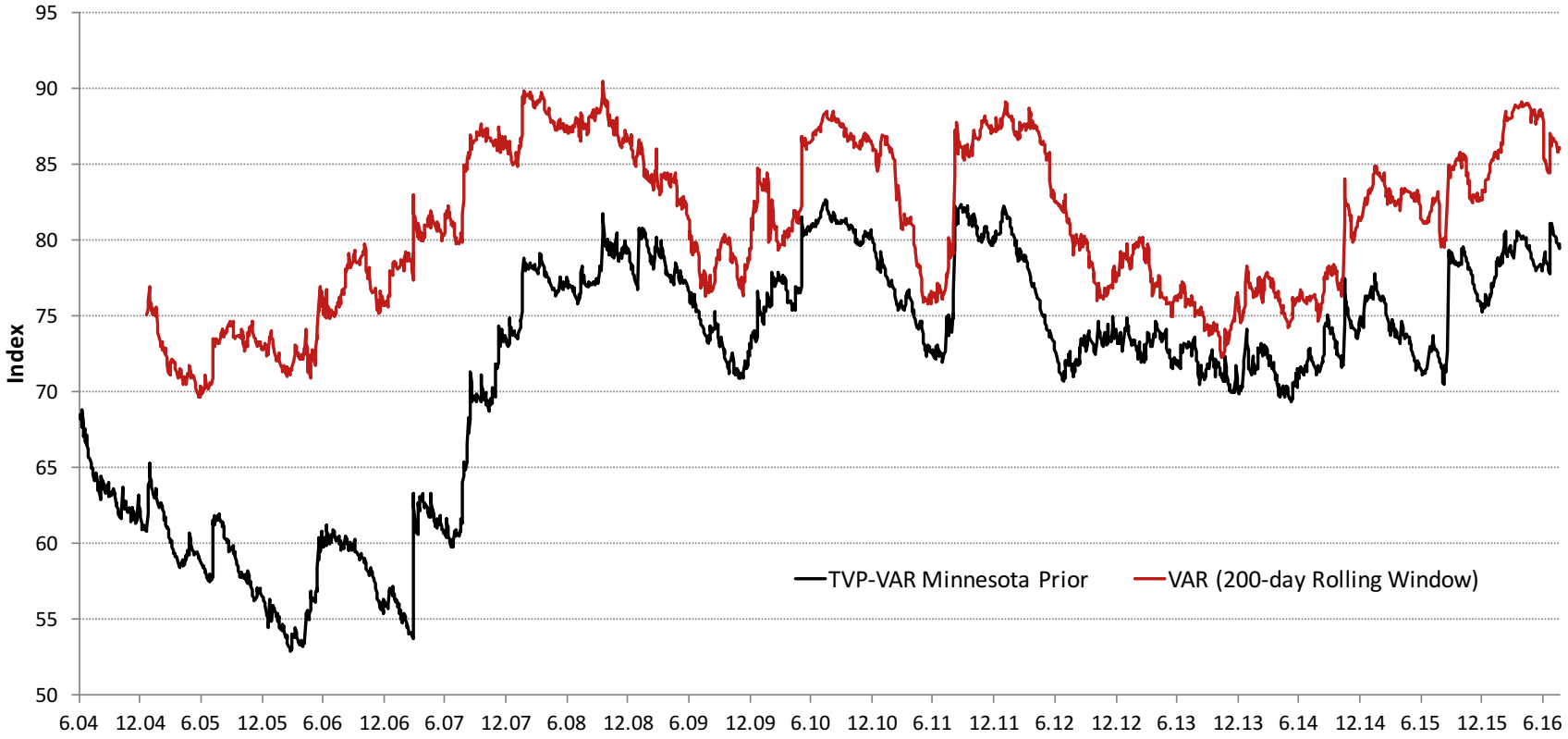


Figure 7 Total Connectedness: TVP-VAR vs VAR with 200-day Rolling Window

Application 3 - Stock volatility connectedness - large data

- We collect a large-scale data series including 62 daily stock indices from 50 different countries around the world, sourced from the website of Investing.com and EastMoney Choice Database.

- The data can be divided into 9 different regions according to their geographical locations.

- The daily log return series are used: $r_{i,t} = \log(P_{i,t}/P_{i,t-1}) * 100$, where $P_{i,t}$ is the closing price of index i at day t.

- The final series is sampled from August 1, 2001 to February 21, 2021.

- Running time comparison

1) SDKF: the proposed filtering estimation
2) SDKFS: SDKF + smoothing
3) KKKF: Koop and Korobillis (2013)'s Kalman filter + EWMA volatility
4) RW-150: Demirer et al. (2018)'s rolling window Lasso-VAR estimation

Runtime comparison of real data exercise ($N = 62; T = 5103$).

	SDKS(EBE)	KKKF	RW-150
Time spent	$\approx$ 30 seconds	$\approx$ 48 minutes	> 17 hours

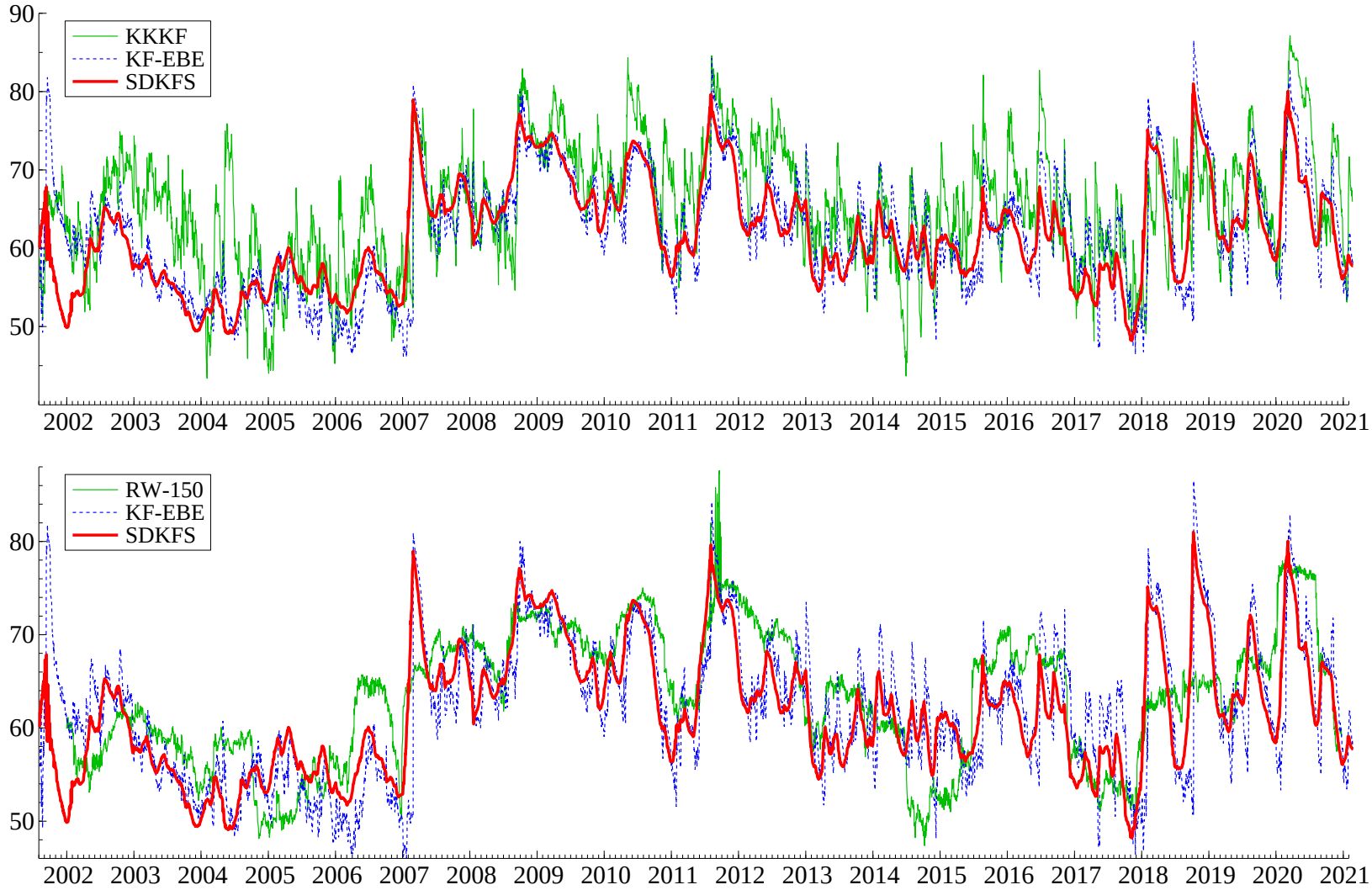


Figure 8 Time-varying total return connectedness with small sample size (N=10), ranges from 2001/8/2 – 2021/2/2.

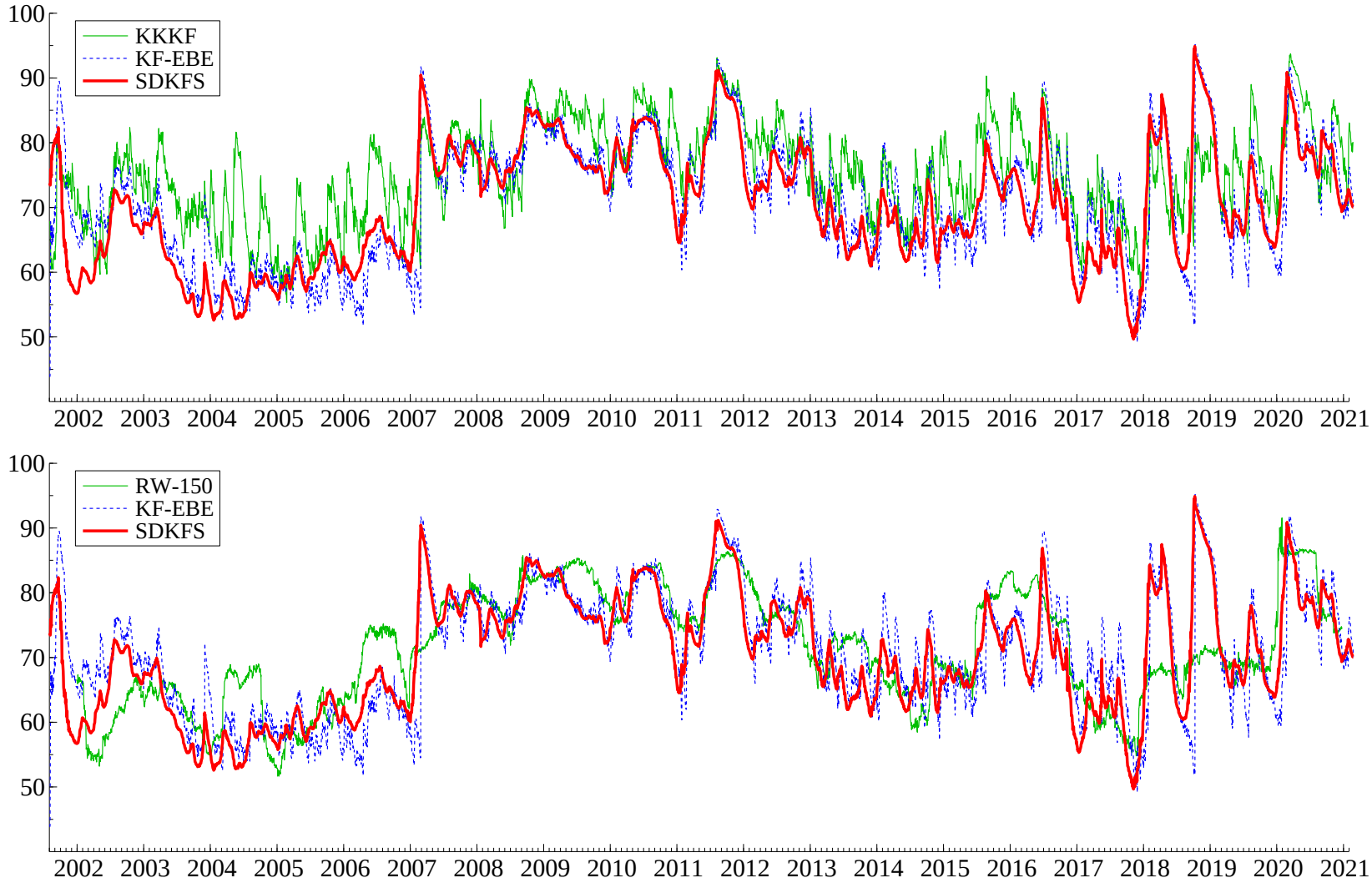


Figure 9 Time-varying total return connectedness with medium sample size (N=24), ranges from 2001/8/2 – 2021/2/2.

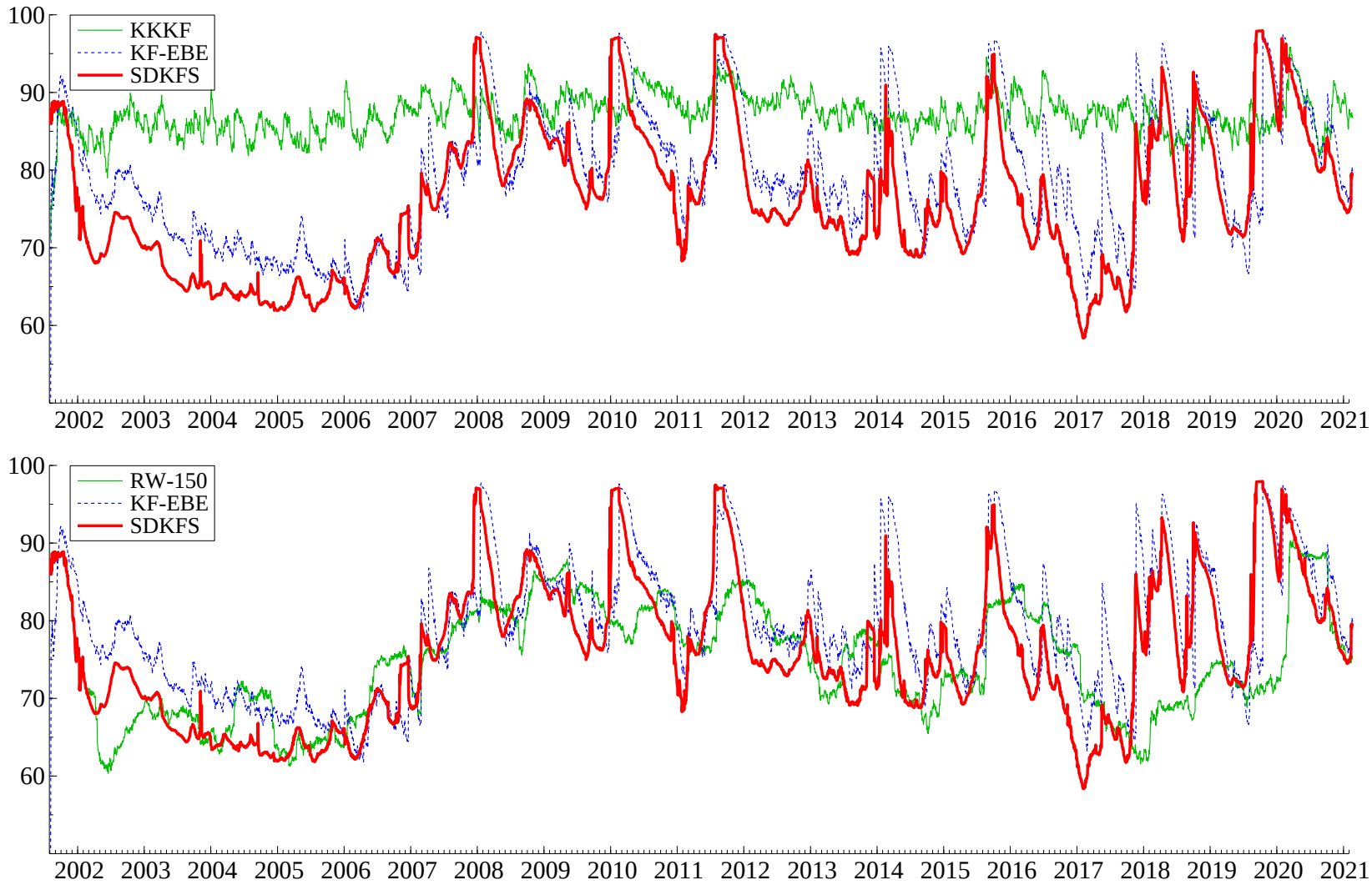
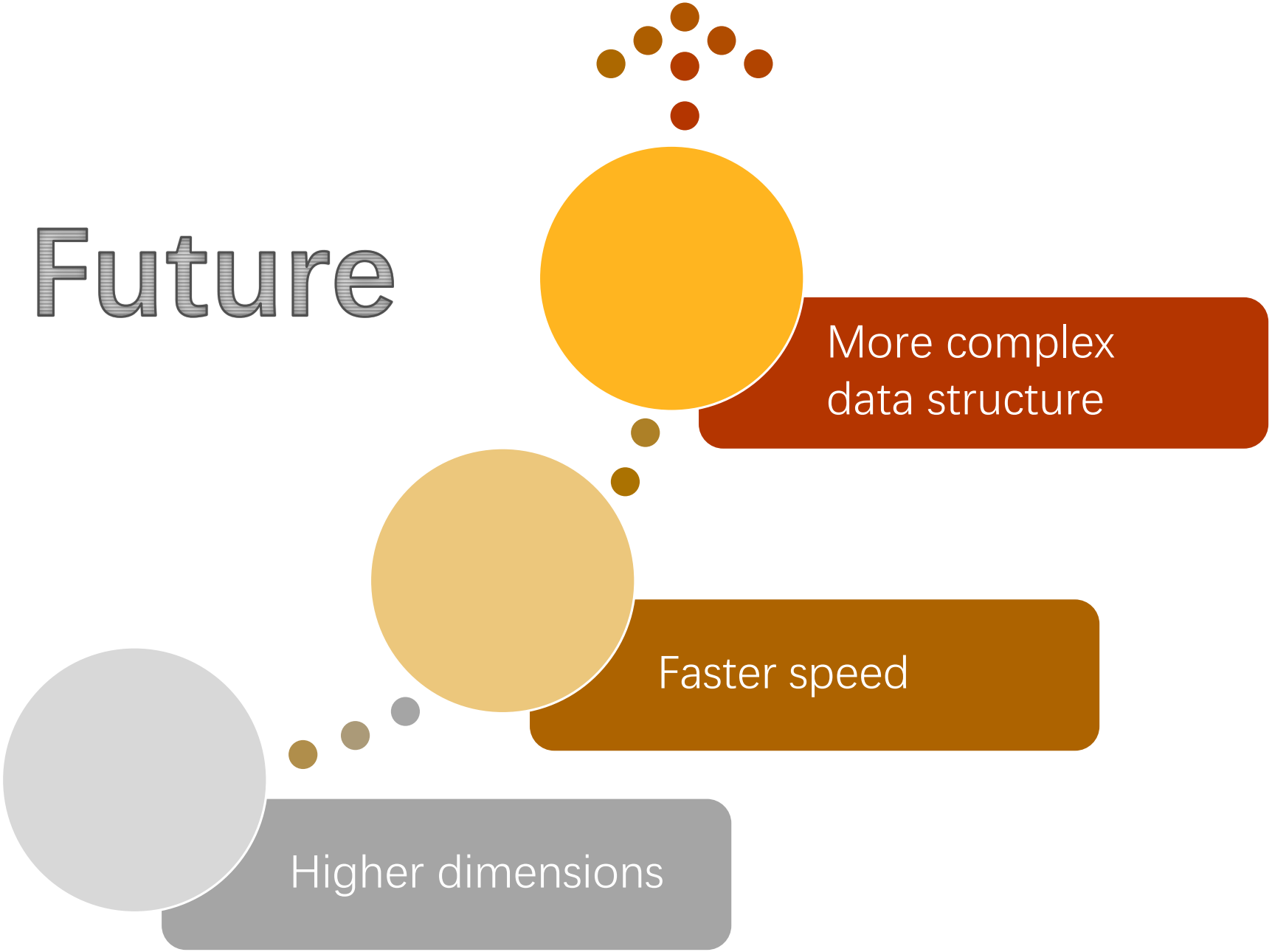


Figure 10 Time-varying total return connectedness with large sample size (N=62), ranges from 2001/8/2 – 2021/2/2.

Future


More complex
data structure
Faster speed
Higher dimensions