

# Review of Large Time-varying Vector Autoregression Model

范小龙 数量经济学 WISE

ID: 27720201153473

[fanxiaolong98@gmail.com](mailto:fanxiaolong98@gmail.com)

2021 年 11 月 8 日

The paper mainly reviews the literature of the large time-varying vector autoregression(TVP-VAR). Section 1 introduces the basic of VAR, and section 2 introduces the TVP-VAR models, and section 3 introduces further introduced the large TVP-VAR models.

## 1 VAR

The vector autoregression model (Vector Autoregression) is a model widely used in economic analysis, which could capture complex dynamic interrelationships among macroeconomic variables. It is mainly used to analyze multiple macro-time series. [Sims \(1980\)](#) considered that traditional theories imposed many restrictions, and first proposed a VAR model without recognition restrictions. [Cooley and LeRoy \(1985\)](#) considered that structural recognition should be added to the structural shocks and impulse response of VAR, which caused the structure vector autoregressive (SVAR) was derived. The constant VAR model assumes that the coefficients do not change with time, and the macroeconomic characteristics of different eras are varying over time, which makes the model unable to effectively capture the nature of the economic system. Therefore, the research of the VAR model that allows the coefficient to change with time (Time-varying parameter) has begun to emerge.

## 2 Time-varying VAR

The time-varying vector autoregression (TVP-VAR) model is under the State space model, which was divided into a Gaussian linear state space model and a nonlinear state space model according to whether

the state equation is a linear relationship. Stochastic volatility widely used in financial measurement was firstly proposed by Black (1976), which has frequently appeared in the empirical analysis of macroeconomics in the past 10 years. Primiceri (2005) proposed a time-varying parameter vector autoregression model with stochastic volatility, which is now widely used in empirically macroeconomics research. The TVP-VAR model can capture the time-varying characteristics of the economic system. Primiceri (2005) assumes that the parameters of model are subject to a first-order random walk process, thus temporary and permanent changes in the parameters are allowed. The TVP-VAR model expression of Primiceri (2005) is in (1):

$$y_t = c_t + B_{1,t}y_{t-1} + \dots + B_{k,t}y_{t-k} + u_t \quad t = 1, \dots, T \quad (1)$$

Where  $y_t$  is the observable vector of  $n \times 1$ ,  $c_t$  is the time-varying intercept vector of  $n \times 1$ , and  $B_{i,t}$  is the time-varying coefficient of  $n \times n$ ,  $u_{i,t}$  is the unobservable shock of  $n \times 1$ , and its variance-covariance matrix is  $\Omega_t$ , heteroscedasticity is allowed, and we can get the reduced form of  $\Omega_t$ :

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t' \quad (2)$$

Where  $A_t$  is the lower triangular matrix, and  $\Sigma_t$  is the diagonal matrix, and  $\alpha_{ij,t}$  of  $A_t$  describes the impact of the  $i$ th variable on the  $j$ th variable. Koop and Korobilis (2010) divides TVP-VAR into homogeneous TVP-VAR ( $\Sigma_t = \Sigma$ ) and TVP-VAR with stochastic volatility according to whether  $\sigma_{i,t}$  of  $\Sigma_t$  varies over time.

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{n1,t} & \cdots & \alpha_{nn-1,t} & 1 \end{bmatrix} \quad \Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t} \end{bmatrix}$$

Stack the  $B_{i,t}$  of (1) into a vector  $B_t$  whose dimension is  $(kn + 1)n \times 1$ , (1) can be rewritten as (3), where  $\otimes$  represents the Kronecker product, and  $\text{Var}(\varepsilon_t) = I_n$ .

$$\begin{aligned} y_t &= X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t \\ X_t' &= I_n \otimes [1, y_{t-1}', \dots, y_{t-k}'] \end{aligned} \quad (3)$$

$\alpha_t = (\alpha_{2,1}, \alpha_{3,1} \cdots, \alpha_{k,k-1})'$  is a stacked vector of the lower-triangular elements, and  $\sigma_t$  is the element

on the diagonal in  $\Sigma_t$ . The time-varying coefficients of the model are assumed to satisfy the following equation (4), where  $B_t$  and  $\alpha_t$  are assumed to satisfy the random walk process, and  $\sigma_t$  is assumed to satisfy the geometric random walk process, while  $\sigma_t$  is an unobservable variable.

$$\begin{aligned} B_t &= B_{t-1} + \nu_t \\ \alpha_t &= \alpha_{t-1} + \zeta_t \\ \log \sigma_t &= \log \sigma_{t-1} + \eta_t \end{aligned} \tag{4}$$

The shocks in the model are assumed to follow the joint normal distribution (5), and the dimension of  $\alpha_t$  is  $\frac{n(n-1)}{2}$ , and the dimension of  $\sigma_t$  is  $n \times 1$ , and the model needs to estimate the parameters  $B_t$ ,  $\alpha_t$ ,  $\sigma_t$  and their variance-covariance matrix. We can find that the number of parameters that need to be estimated will increase exponentially as the number of variables ( $n$ ) increases.

$$V = \text{Var} \left( \begin{bmatrix} \varepsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t \end{bmatrix} \right) = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W_{n \times n} \end{bmatrix} \tag{5}$$

Nakajima et al. (2011) introduced the univariate TVP model and the multivariate TVP model in detail. The univariate TVP model proposed a new unobservable shock equation, in which the random volatility assumes that AR(1) process, and  $\sigma_t^2$  is the element on the diagonal of  $\Sigma_t$  in (6). The multivariate TVP model uses SVAR as the benchmark model, and other settings are the same as the Primiceri (2005) model. The settings are basically the same.

$$\sigma_t^2 = \gamma \exp(h_t), \quad h_{t+1} = \phi h_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad t = 0, \dots, n-1 \tag{6}$$

Nakajima et al. (2011) used Markov chain Monte Carlo Method(MCMC) method and Kalman Filter to estimate the parameters of the model, sampling  $\beta; \alpha; h; \omega$  respectively, where  $\beta; \alpha$  is sampled by Kalman Filter; and sampling  $h$  by a multi-step moving sampling;  $\omega$  including  $\Sigma_\beta, \Sigma_\alpha, \Sigma_h$ , which Nakajima et al. (2011) uses MCMC method to estimate. However, the calculation cost of the MCMC method to estimate the parameters will increase exponentially as the dimensionality increases. Koop and Korobilis (2013) used forgetting factors to replace the MCMC methods, which could handle the computation burden.

The data generation process of economic variables is related to the impact of time-varying coefficients

and stochastic volatility. The estimated values of the parameters can only be obtained through approximate solutions since the analytic expression of the likelihood function of random volatility is difficult to obtain. The TVP-VAR model can use MCMC method for parameter estimation for low-dimensional variables under the Bayesian framework. [Koop and Korobilis \(2010\)](#) considers that a large number of parameters in the time-varying coefficients VAR model may make the estimation over-fitting. [Bańbura et al. \(2010\)](#) shows that VAR with Bayesian shrinkage is an appropriate tool for large dynamic models. Soft and hard constraints can effectively solve these problems in the Bayesian method. The soft limit is to shrink the model parameters (Shrinkage) to a specific value, generally set to 0, thereby reducing the computational complexity of the model parameters; the hard limit relates to the exact model theoretical restriction.

### 3 Large Time-varying VAR

Large-dimensions time-varying coefficient VAR model can accommodate more economic variables into economic system analysis, but it also increases the complexity of parameter estimation. Big data <sup>1</sup> may change the foundation of empirical macroeconomics. More information in big data will improve the predictive ability of the model and help us better understand the macroeconomics. [Koop \(2017\)](#) believes that in the era of big data, current research faces three major challenges: a large number of explanatory variables; a large number of dependent variables; a large number of model parameters that change over time. [Koop \(2016\)](#) believes that the model inference obtained by processing fat data using traditional methods (Bayesian method without prior information, MLE, etc.) is inaccurate. [Bańbura et al. \(2010\)](#) consider Bayesian VAR is a valid alternative to factor models or panel VARs for the analysis of large dynamic systems. When the number of explanatory variables is more than the number of observations, the least squares regression (OLS) is invalid. When we use the new method, we are facing a new problem: model uncertainty, there are  $2^K$  models that contain these explanatory variables for K explanatory variables. When we select only the best model and ignore other models, it will inevitably lead to inaccurate model parameters.

The Bayesian estimation method uses prior information with data information to estimate model parameters. Dense parameterization will lead to unsound inferences and inaccurate out-of-sample predictions. [Koop \(2017\)](#) believes that the use of Bayesian methods can solve the above problems (over

---

<sup>1</sup>[Varian \(2014\)](#) thinks that big data includes Tall Data and Fat Data, and Tall Data means more numbers of observations than explanatory variables, and Fat Data means more explanatory variables.

parameterization and calculation problems). Bayesian methods can use prior information to provide a method for shrinkage of parameters. The setting of the prior (Prior) is crucial to the estimation of the parameters, and many prior parameter shrinkages are data-driven; [Giannone et al. \(2015\)](#) introduced the choice of prior information for the VAR model. [Giannone et al. \(2015\)](#), [Koop and Korobilis \(2010\)](#) and [Chan \(2020\)](#) introduced the details of the Minnesota prior, natural conjugate priors and Stochastic Search Variable Selection (SSVS).

The Minnesota prior appeared a recent boom in popularity because of its simplicity and success in many applications. [Koop and Korobilis \(2010\)](#) considers the big advantage of the Minnesota prior was that it leads to simple posterior inference involving only the Gaussian distribution. [Karlsson \(2013\)](#) considers the Minnesota prior captures widely held beliefs about the long run properties of the data. Natural conjugate priors lead to analytical results, which can greatly reduce the computational burden. The natural conjugate prior does not allow us to use prior information of this more lagged form. SSVS often used with Fat Data specifies a hierarchical prior which is a mixture of two Normal distributions, while LASSO is just a different Normal hierarchical prior for the regression coefficients in [Koop \(2016\)](#).

Bayesian approach allow for the uncertainty of models, which means entire forecasting model to vary over time. Bayes Model Average (BMA) and Dynamic Model Average (DMA) are effective methods for solving uncertainty of models. The posterior probability of the model is weighted with the model, which means that the model will change over time. [Koop and Korobilis \(2012\)](#) rewrite TVP-VAR under DMA setting to (7), which set  $z_t$  of different model(k) as explanatory variables, we can get  $M = 2^K$  models, Where  $k = 1, 2, \dots, M$ .

$$\begin{aligned} y_t &= z_t^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)} \\ \theta_{t+1}^{(k)} &= \theta_t^{(k)} + \eta_t^{(k)} \end{aligned} \tag{7}$$

The standard Bayesian approach to model selection is based on the marginal likelihood. [Andersson and Karlsson \(2008\)](#) suggest replacing the marginal likelihood with the predictive likelihood for the variables of interest, which implies to calculate the posterior probabilities or model weights. We regard the models as a random variable, which means obtaining the expectation of the estimated parameters for the model. Where  $p(M_r | y)$  in (8) is the posterior probability of the model,  $p(\phi | y, M_k)$  in (8) is the estimated parameter under each model The posterior probability,  $p(\phi | y)$  in (8) is the posterior probability of the estimated parameters after the model is averaged. The model k will receive more weight at time t if it has forecast well in the recent past.

$$p(\phi | y) = \sum_{r=1}^M p(\phi | y, M_k) p(M_k | y) \quad (8)$$

Raftery et al. (2010) improved Kalman filtering proceeds by using forgetting factor( $\lambda$ ) replaced by (9), which could reduce computational burden from  $K^T$  times to  $K$  times, and we always consider  $\lambda \in (0.95, 0.99)$  in our empirical work. Koop and Korobilis (2013) used forgetting factor to select the optimal value of the shrinkage parameter at different points in time and the dimension of the TVP-VAR models.

$$Q_t = (1 - \lambda^{-1}) \Sigma_{t-1|t-1} \quad (9)$$

Koop and Korobilis (2012) consider stochastic volatility or ARCH specification for  $H_t^{(k)}$ , which greatly added to the computational burden, thus they used Exponentially Weighted Moving Average (EWMA) estimator, which was commonly used to model time-varying volatilities in finance. Koop and Korobilis (2013) draw on ideas from the DMA which achieve reductions in the computational burden through the use forgetting factors for high-dimensions variables, and the parameters of the shrinkage priors commonly-used with large VAR models.

How should we compare the performance of different models of VAR? Chan and Eisenstat (2018) improved the importance sampling method with calculating the marginal likelihood and deviance information criterion (DIC) of the time-varying coefficient vector as Bayesian model criterion, and comparing the performance between TVP-VAR and the constant coefficient model based on these two standards. Chan and Eisenstat (2018) found that the constant VAR model with stochastic volatility performs better than the TVP-VAR model with stochastic volatility.

What is the future development direction of TVP-VAR? I think the answer is that both the computational speed, accuracy of approximation and the specification of the TVP-VAR models. Koop and Korobilis (2013) developed methods for estimation and forecasting in large TVP-VAR models, which allowed for model switching. Chan and Eisenstat (2018) did not rule out the possibility that the VAR model of some coefficients may be constant, while some other coefficients may be time-varying coefficients performs better. Koop and Korobilis (2019) proposed methods for estimating and forecasting in Bayesian panel vector autoregressions (PVAR) of large dimensions with time-varying parameters and stochastic volatility. Koop and Korobilis (2020) proposed a variational Bayesian method for computationally efficient posterior and predictive inference in time-varying parameters models, and specified a new dynamic variable/model selection strategy for TVP models with the presence of Fat Data.

## References

- Michael K Andersson and Sune Karlsson. *Bayesian forecast combination for VAR models*. Emerald Group Publishing Limited, 2008.
- Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin. Large bayesian vector auto regressions. *Journal of applied Econometrics*, 25(1):71–92, 2010.
- Fischer Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1-2):167–179, 1976.
- Joshua CC Chan. Large bayesian vector autoregressions. In *Macroeconomic Forecasting in the Era of Big Data*, pages 95–125. Springer, 2020.
- Joshua CC Chan and Eric Eisenstat. Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of Applied Econometrics*, 33(4):509–532, 2018.
- Thomas F Cooley and Stephen F LeRoy. Atheoretical macroeconometrics: a critique. *Journal of Monetary Economics*, 16(3):283–308, 1985.
- Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451, 2015.
- Sune Karlsson. Forecasting with bayesian vector autoregression. *Handbook of economic forecasting*, 2:791–897, 2013.
- Gary Koop. Bayesian methods for fat data. *Manuscript available at: garykoop/research*, 2016.
- Gary Koop. Bayesian methods for empirical macroeconomics with big data. *Review of Economic Analysis*, 9(1):33–56, 2017.
- Gary Koop and Dimitris Korobilis. *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc, 2010.
- Gary Koop and Dimitris Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012.
- Gary Koop and Dimitris Korobilis. Large time-varying parameter vars. *Journal of Econometrics*, 177(2):185–198, 2013.

- Gary Koop and Dimitris Korobilis. Forecasting with high-dimensional panel vars. *Oxford Bulletin of Economics and Statistics*, 81(5):937–959, 2019.
- Gary Koop and Dimitris Korobilis. Bayesian dynamic variable selection in high dimensions. *Available at SSRN 3246472*, 2020.
- Jouchi Nakajima et al. Time-varying parameter var model with stochastic volatility: An overview of methodology and empirical applications. 2011.
- Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.
- Adrian E Raftery, Miroslav Kárný, and Pavel Ettler. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1):52–66, 2010.
- Christopher A Sims. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48, 1980.
- Hal R Varian. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28, 2014.