Time-varying parameters with RNN

Zhendong Sun

May 2022

1 Introduction

Time-varying parameters (TVP) model has attracted much attention in different fields[1]. The TVP model is a type of linear regression model that allows parameters to change over time. The TVP model keeps the robustness and simplicity of the linear model while it can also capture some complicated function forms by adjusting the parameters at each time. Because of these properties, the TVP model could outperform the linear predictive model and some non-linear machine learning methods in economics and finance where the data is relatively noisy.

A conventional specification of the TVP model is to assume the parameters are unobserved variables that follow random walks. This specification generally suffers from two issues. First, the dimension of estimators is huge relative to the data because we have to estimate all the parameters in every period[2]. Second, the specification of model could not capture some sharp changes in the parameters because of the serial dependence of the random walk process[3].

[1]See Primiceri (2005),Koop and Korobilis (2010),Dangl and Halling (2012)

[2]Researchers have proposed different ways to simplify the model. For instance, Bitto and Frühwirth-Schnatter (2019) includes double-gamma prior distribution to shrink most parameters to zero. Chan et al. (2020) propose a factor-structure model that assumes the variations of parameters are driven by a smaller number of factors.

[3]Pettenuzzo and Timmermann (2017) investigate the performance of Markow switching and change point models that assume different functional forms of the evolution of parameters. Meanwhile, some researchers apply the methods that impose any presumption on evolution. For instance, Farmer et al. (2019) uses a kernel smooth method to estimate the parameters. Goulet Coulombe (2020) interpret the time-varying parameters as a special form of the tree-based model.

To tackle these problems, we propose a new TVP model that is built on the Recurrent Neural Network (RNN-TVP). The proposed method utilizes the ability of the Neural Network to capture the potentially complicated form of the time variations in parameters, while it keeps the advantage of linear models.

We investigate the performance of the proposed method with simulated data and economic data sets. In the simulation study, we find that the proposed RNN-TVP model can capture both smooth and sharp transitions in the parameters. In the practice of real economic data, we find that the RNN-TVP achieves better performance in both in-sample and out-of-sample applications relative to alternatives.

The rest of the paper is organized as follows. Section 2 discusses the proposed method in more detail. Section 3 reports the model performance with simulated and real economic data. Section 4 is the conclusion.

2 Methodology

We specify a generic TVP model as below:

$$y_t = x_t \beta_t + e_t \qquad e_t \sim N(0, \sigma) \tag{1}$$

$$\beta_t = f_\psi(\beta_{1:t-1}, x_{1:t}, y_{1:t}) \tag{2}$$

where the x_t is the vector of independent variables, y_t is the dependant variable, and β_t is the vector of parameters from function f with parameters ψ.

In the proposed RNN-TVP model, we aim to estimate the function f by RNN. For simplicity, we illustrate the RNN-TVP model by using a single layer of simple RNN to approximate the function. In practice, we could incorporate multiple layers of different types of Neural Network to find the best approximation for the function f.

We specify a simple RNN-TVP with a single layer of network for the parameters:

$$\beta_t = V h_t \tag{3}$$

$$h_t = tanh(a_t) \tag{4}$$

$$a_t = W^0 + W^1 h_{t-1} + W^2 x_t \tag{5}$$

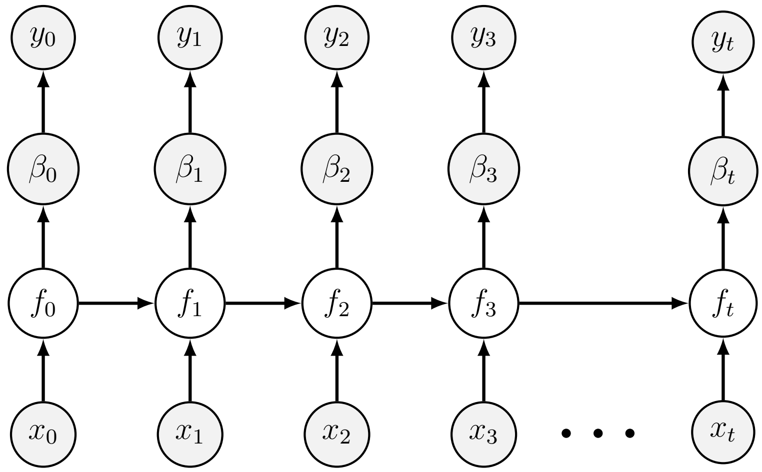

Figure 1: RNN-TVP diagram

where W and V are the parameters of the RNN, and *tanh* is the hyperbolic tangent function whose outputs range from -1 to 1.

In some cases, we prefer to decompose the parameters in the TVP model into a fixed term and a time-varying term. By doing so, the model could perform better if the true parameters are close to fixed (e.g.,Bitto and Frühwirth-Schnatter (2019)). Specifically, the model can be rewritten as:

$$y_t = g_t + x_t\beta_t + e_t \qquad e_t \sim N(0, \sigma) \tag{6}$$

$$g_t = x_t\beta \tag{7}$$

$$\beta_t = f_\psi(\beta_{1:t-1}, x_{1:t}, y_{1:t}) \tag{8}$$

Figure (1) shows the architecture of the RNN-TVP model. In a toy model, we only include a single layer of simple RNN for function f, but it could contain multiple layers of more complicated structure such as LSTM.

3 Simulation Study

In the simulation study, we generate data sets with two different processes of time-varying coefficients. The first one assumes coefficients evolve smoothly with sine and cosine functions:

$$
\begin{aligned}
l_i &= \frac{4\pi}{200} * i \qquad\qquad i = 1, 2, \cdots 200 &(9)\\
\beta_{1,i} &= sin(l_i) &(10)\\
\beta_{2,i} &= cos(l_i) &(11)\\
\beta_{3,i} &= 0 &(12)\\
y_i &= x_i * \beta_i + \epsilon_i \qquad\qquad \epsilon_i \sim N(0, 0.1) &(13)
\end{aligned}
$$

where each element in x_i are generated from N(0,10) independently. In the second experiment, we assume coefficients have a single change point:

$$
\begin{aligned}
\beta_{1,i} &= 0.5 + 0.5 * \mathbb{1}(i < 100) &(14)\\
\beta_{2,i} &= \mathbb{1}(i > 100) &(15)\\
\beta_{3,i} &= 0 &(16)\\
y_i &= x_i * \beta_i + \epsilon_i \qquad\qquad \epsilon_i \sim N(0, 0.1) &(17)
\end{aligned}
$$

where $\mathbb{1}$ is an indicator function.

We compare the proposed RNN-TVP with the conventional TVP that assumes a random walk process of parameters (RW-TVP). The RW-TVP model is estimated by a conventional MCMC method with Kalman filter (e.g., Primiceri (2005), Koop and Korobilis (2010)). For the RNN-TVP, we incorporate two layers of RNN that consist of a simple RNN layer and an LSTM.

Table (1) reports the Mean Squared Error(MSE) of in-sample fit of coefficients β and output y. As shown in the table, the RNN-TVP performs better than the RW-TVP for fitting coefficients and output. The advantage of RNN-TVP is larger in the experiment with the jump process of coefficients, which implies the Neural Network could better capture the large variations compared with the random walk assumption.

Table 1: Mean squared error of in-sample fit

Case 1: smooth process of coefficients				
		coefficients		output
RNN-TVP	0.0001	0.0001	0.0000	0.0002
RW-TVP	0.0011	0.0018	0.0000	0.0032
Case 2: jump process of coefficients				
		coefficients		output
RNN-TVP	0.0013	0.0031	0.0000	0.0021
RW-TVP	0.0032	0.0129	0.0000	0.0700

Figure (2) shows how the RNN-TVP and RW-TVP fit the coefficients over time in the experiment of the smooth process of coefficients. Generally, we can see the fitted coefficients by RNN-TVP are closer to the true value compared with the RW-TVP, while the difference is trivial when the truth is fixed at zero. Based on the Cumulative Squared Error(CSE) differential between RNN-TVP and RW-TVP, we observe the CSE differentials keep increasing for the first two coefficients, which implies the out-performance of RNN-TVP for these two coefficients is persistent.

Figure (3) shows how the RNN-TVP and RW-TVP fit the coefficients over time in the experiment of jump process of coefficients. We can see the difference between RNN-TVP and RW-TVP is not large when the true coefficients is fixed, but the RNN-TVP captures the variations much better around the break points. Based on the CSE differential in the lower row of the figure, Both RNN-TVP and RW-TVP yields similar level of performance when the true coefficients is fixed, while RNN-TVP performs significantly between around the points of break.

Based on the simulation experiments of two different processes, we conclude that the RNN-TVP could generally outperform the RW-TVP in terms of in-sample fit. The advantage of RNN-TVP gets more clear when the process of coefficients is more variant such as the existence of jump points in the second experiment.

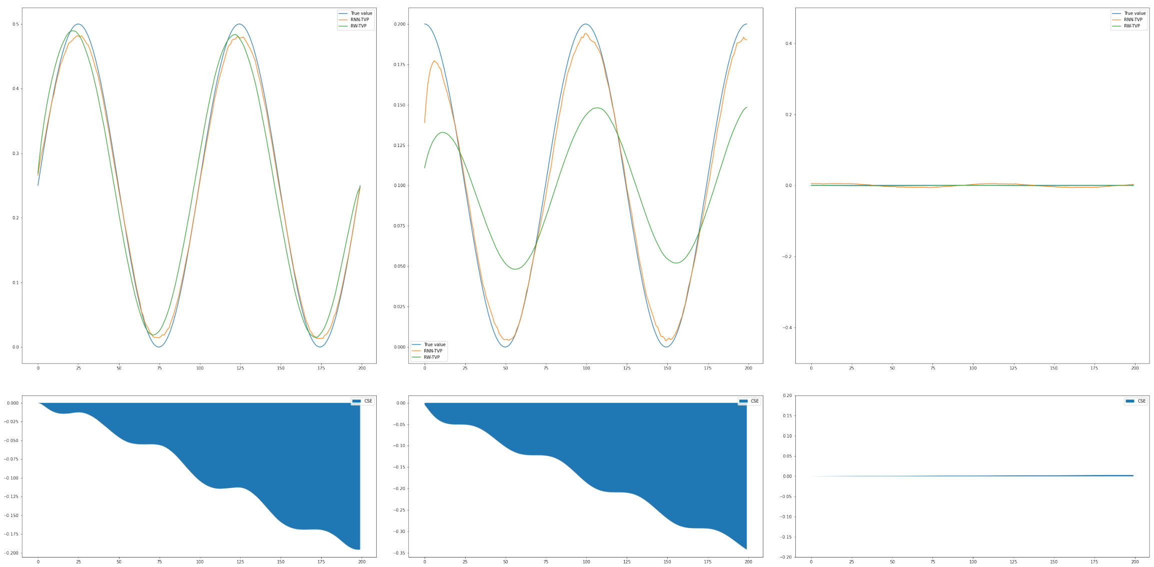

Figure 2: Smooth process: in-sample fit of coefficient. The figure shows how the RNN-TVP and RW-TVP fit the coefficients over time in the experiment of smooth process of coefficients. The upper row shows the true coefficients and fitted value by RNN-TVP and RW-TVP. The lower row shows the Cumulative Squared Error differentials between RNN-TVP and RW-TVP. Specifically, if the CSE area is below zero, it implies the RNN-TVP outperforms the RW-TVP up to the particular time.

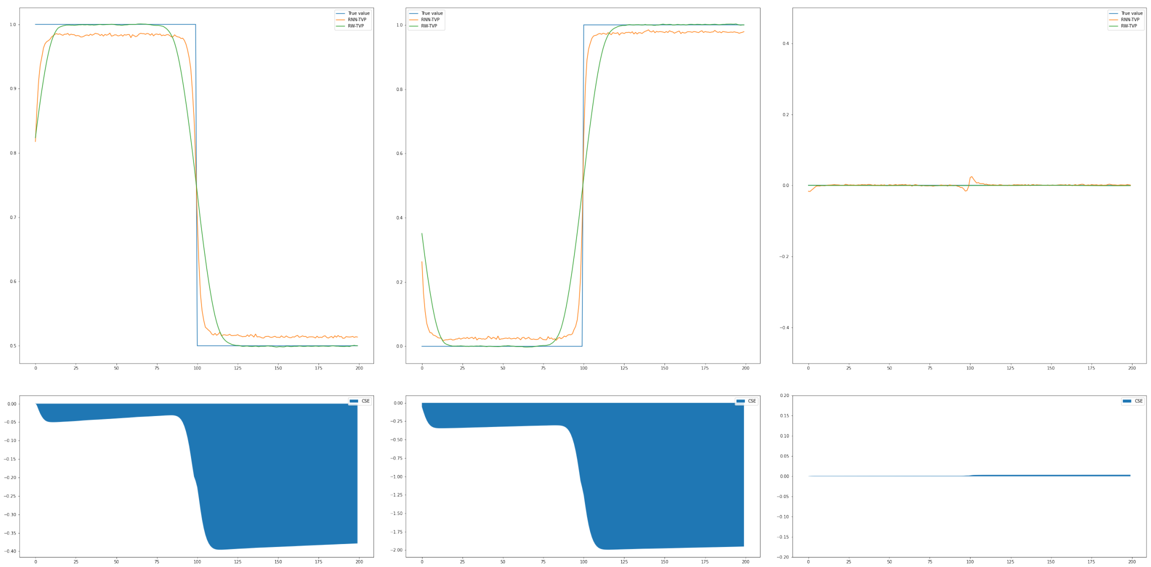

Figure 3: Jump process: in-sample fit of coefficient. The figure shows how the RNN-TVP and RW-TVP fit the coefficients over time in the experiment of smooth process of coefficients. The upper row shows the true coefficients and fitted value by RNN-TVP and RW-TVP. The lower row shows the Cumulative Squared Error differentials between RNN-TVP and RW-TVP. Specifically, if the CSE area is below zero, it implies the RNN-TVP outperforms the RW-TVP up to the particular time.

4 Empirical Result

In this section, we applied the RNN-TVP into a real world problem that is to predict GDP in US. In this experiment, we will focus on the out-of-sample result in stead of in-sample fit in order to avoid the potential overfitting issue.

We include the GDP growth rate, inflation rate and interest rate as the inputs for the model. Specifically, our target is the GDP growth rate, and we use the two lags of three variables as the predictors. The data series range from 1954Q3 to 2020Q1, and the evaluation period starts at 1989Q3.

Figure (4) shows the out-of-sample performance of the RNN-TVP and RW-TVP. Based on the lower row of the figure, we observe that the RNN-TVP yields a generally better performance during the whole evaluation period. Moreover, the advantage of the RNN-TVP mainly comes from the middle of the period, which is the financial crisis of 2008. This implies that there exist some sharp changes behind the data and the RNN-TVP could capture these changes more timely than RW-TVP, which leads to better out-of-sample performance. Meanwhile, the RNN-TVP does not show a significant advantage during other times, probably because the economy evolves more stably most time.

5 Conclusion

This study introduces a time-varying parameters model based on RNN(RNN-TVP). The RNN-TVP inherits the robustness and interpretability of the linear model as well as the flexibility of the neural network. We compare the RNN-TVP model with the time-varying parameters model that assumes a random walk process of parameters(RW-TVP). In the simulation experiment, we find the RNN-TVP capture the variation in parameters better especially when there exists sharp changes in the process. In the practice of forecasting US GDP, we find the RNN-TVP could achieve better out-of-sample performance.

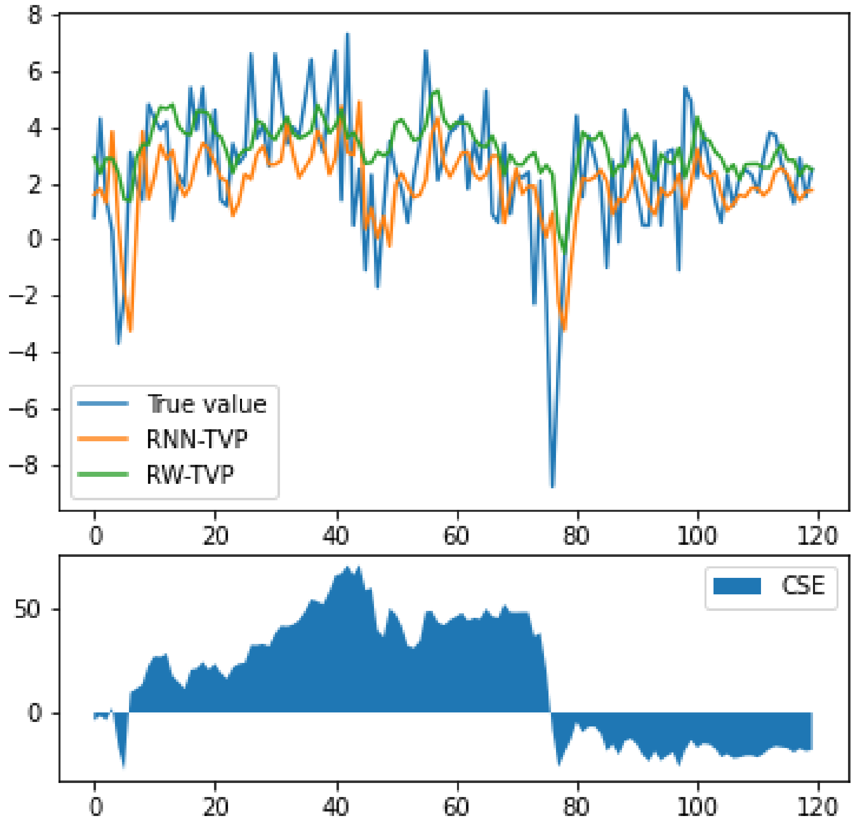

Figure 4: Out-of-sample prediction. The figure shows how the RNN-TVP and RW-TVP predict US GDP in the out-of-sample scenario. The upper row shows the true GDP and predicted value by the RNN-TVP and RW-TVP. The lower row shows the Cumulative Squared Error differentials between RNN-TVP and RW-TVP. Specifically, if the CSE area is below zero, it implies the RNN-TVP outperforms the RW-TVP up to the particular time.

References

Amir-Ahmadi, P., C. Matthes, and M.-C. Wang (2018). Choosing prior hyperparameters: with applications to time-varying parameter models. *Journal of Business & Economic Statistics*, 1–13.

Amir-Ahmadi, P. and Z. Sun (2021). Forecasting with time-varying parameter vars: An efficient sequential monte carlo approach. *none*.

Arias, J. E., J. F. Rubio-Ramírez, M. Shin, et al. (2021). *Macroeconomic Forecasting and Variable Ordering in Multivariate Stochastic Volatility Models*. Research Department, Federal Reserve Bank of Philadelphia.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Nowcasting and the real-time data flow. In *Handbook of economic forecasting*, Volume 2, pp. 195–237. Elsevier.

Banbura, M., D. Giannone, and L. Reichlin (2010). Nowcasting.

Bates, J. M. and C. W. Granger (1969). The combination of forecasts. *Journal of the Operational Research Society 20*(4), 451–468.

Beskos, A., D. Crisan, and A. Jasra (2014). On the stability of sequential monte carlo methods in high dimensions. *The Annals of Applied Probability 24*(4), 1396–1445.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. Van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics 177*(2), 213–232.

Bitto, A. and S. Frühwirth-Schnatter (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association 112*(518), 859–877.

Bognanni, M. and J. Zito (2020). Sequential bayesian inference for vector autoregressions with stochastic volatility. *Journal of Economic Dynamics and Control*, 103851.

Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics 10*, 615–643.

Bryzgalova, S., J. Huang, and C. Julliard (2019). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Available at SSRN 3481736*.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy 107*(2), 205–251.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies 21*(4), 1509–1531.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Realtime nowcasting with a bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(4), 837–862.

Carriero, A., T. E. Clark, and M. Marcellino (2019). Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics 212*(1), 137–154.

Chan, J. C., E. Eisenstat, and R. W. Strachan (2020). Reducing the state space dimension in a large tvp-var. *Journal of Econometrics 218*(1), 105–118.

Chan, J. C., G. Koop, and X. Yu (2021). Large order-invariant bayesian vars with stochastic volatility. *arXiv preprint arXiv:2111.07225*.

Cimadomo, J., D. Giannone, M. Lenza, F. Monti, and A. Sokol (2021). Nowcasting with large bayesian vector autoregressions. *Journal of Econometrics*.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics 138*(1), 291–311.

Cochrane, J. H. (2005). Financial markets and the real economy.

Cogley, T. and T. J. Sargent (2005). Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic dynamics 8*(2), 262–302.

Conflitti, C., C. De Mol, and D. Giannone (2015). Optimal combination of survey forecasts. *International Journal of Forecasting 31*(4), 1096–1103.

Creal, D. (2012). A survey of sequential monte carlo methods for economics and finance. *Econometric reviews 31*(3), 245–296.

D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic forecasting and structural change. *Journal of applied econometrics 28*(1), 82–101.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157–181.

Del Negro, M. and G. E. Primiceri (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The review of economic studies 82*(4), 1342–1345.

Diebold, F. X. and M. Shin (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting 35*(4), 1679–1691.

Dong, X., Y. Li, D. Rapach, and G. Zhou (2021). Anomalies and the expected market return. *Baruch College Zicklin School of Business Research Paper* (2020-02), 02.

Farmer, L., L. Schmidt, and A. Timmermann (2019). Pockets of predictability. *Available at SSRN 3152386*.

Ferreira, I. H. and M. C. Medeiros (2021). Modeling and forecasting intraday market returns: a machine learning approach. *arXiv preprint arXiv:2112.15108*.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics 193*(2), 294–314.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE proceedings F (radar and signal processing)*, Volume 140, pp. 107–113. IET.

Goulet Coulombe, P. (2020). The macroeconomy as a random forest. *Available at SSRN 3633110*.

Granger, C. W. and R. Ramanathan (1984). Improved methods of combining forecasts. *Journal of forecasting 3*(2), 197–204.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Gupta, N. and R. Hauser (2007). Kalman filtering with equality and inequality state constraints. *arXiv preprint arXiv:0709.2791*.

Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics 14*(4), 429–434.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of financial economics 99*(3), 560–580.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian estimation of DSGE models*. Princeton University Press.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies 28*(3), 791–837.

Huang, D., J. Li, and L. Wang (2021). Are disagreements agreeable? evidence from information aggregation. *Journal of Financial Economics 141*(1), 83–101.

Huber, F., G. Koop, and L. Onorante (2020). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, 1–15.

Huber, F., G. Koop, L. Onorante, M. Pfarrhofer, and J. Schreiner (2020). Nowcasting in a pandemic using non-parametric mixed frequency vars. *Journal of Econometrics*.

Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika 76*(2), 297–307.

Koop, G. and D. Korobilis (2010). *Bayesian multivariate time series methods for empirical macroeconomics*. Now Publishers Inc.

Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. *Journal of Econometrics 177*(2), 185–198.

Koop, G. and D. Korobilis (2020). Bayesian dynamic variable selection in high dimensions. *Available at SSRN 3246472*.

Koop, G., D. Korobilis, and D. Pettenuzzo (2019). Bayesian compressed vector autoregressions. *Journal of Econometrics 210*(1), 135–154.

Ludvigson, S. C. and S. Ng (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2020). Machine learning advances for time series forecasting. *arXiv preprint arXiv:2012.12802*.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *The Journal of Finance 71*(1), 5–32.

Medeiros, M. C., G. F. Vasconcelos, Á. Veiga, and E. Zilberman (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics 39*(1), 98–119.

Minka, T. (2000). Automatic choice of dimensionality for pca. *Advances in neural information processing systems 13*.

Pettenuzzo, D. and A. Timmermann (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics 35*(2), 183–201.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association 94*(446), 590–599.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Rapach, D. and G. Zhou (2013). Forecasting stock returns. In *Handbook of economic forecasting*, Volume 2, pp. 328–383. Elsevier.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies 23*(2), 821–862.

Rapach, D. E. and G. Zhou (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–33.

Ročková, V. (2018). Particle em for variable selection. *Journal of the American Statistical Association 113*(524), 1684–1697.

Rockova, V. and K. McAlinn (2021). Dynamic variable selection with spike-and-slab process priors. *Bayesian Analysis 16*(1), 233–269.

Safikhani, A. and A. Shojaie (2020). Joint structural break detection and parameter estimation in high-dimensional nonstationary var models. *Journal of the American Statistical Association*, 1–14.

Särkkä, S. (2013). *Bayesian filtering and smoothing*, Volume 3. Cambridge University Press.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency var. *Journal of Business & Economic Statistics 33*(3), 366–380.

Schorfheide, F. and D. Song (2021). Real-time forecasting with a (standard) mixed-frequency var during a pandemic. Technical report, National Bureau of Economic Research.

Simon, D. (2010). Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications 4*(8), 1303–1318.

Su, X. and Y. Chen (2021). Variational approximation for importance sampling. *Computational Statistics*, 1–30.

Surace, S. C., A. Kutschireiter, and J.-P. Pfister (2019). How to avoid the curse of dimensionality: Scalability of particle filters with and without importance weights. *Siam Review 61*(1), 79–91.

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting 1*, 135–196.

Timmermann, A. (2018). Forecasting methods in finance. *Annual Review of Financial Economics 10*, 449–479.

Wang, J., Y. Ouyang, Y. Ji, and F. Liang (2021). An ensemble em algorithm for bayesian variable selection. *Bayesian Analysis 1*(1), 1–22.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.