DSGE によるニューラルネットの正則化と経済予想

Combining a DSGE model with Neural Networks for Forecasting Economic Variables

塩野剛志 [1]

Takashi Shiono[1]

[1] クレディ・スイス経済調査部
[1]Credit Suisse Economics Research

Abstract: This paper examines the possibility of combining a DSGE model and neural networks to supplement each other, with regard to out-of-sample forecasts for economic variables. The aim is to build a model with theoretical interpretability and state-of-the-art performance. The novel neural-net structure of TDVAE (Temporal Difference Variational Auto-Encoder) proposed by Gregor et.al [2019] enables to realize this idea. TDVAE virtually replicates a Gaussian stochastic state-space model through combination of neural networks. Because a DSGE model provides theoretical restrictions on the state transition and observation matrices of a linear state-space model, I choose to transplant those DSGE-oriented matrices into the formulations of state transition and observation probabilities in TDVAE. This TDVAE-DSGE approach certainly achieved the superior performance in the task of out-of-sample forecasts on Japan's real GDP during 1Q/2011 and 4Q/2018.

1. Introduction

This paper examines the possibility of combining a DSGE model and neural networks to supplement each other, with regard to out-of-sample forecasts for economic variables. The aim is to build a model with theoretical interpretability and state-of-the-art performance.

1.1. Forecasts by DSGE models

The dynamic stochastic general equilibrium (DSGE) model is a central tool to analyze fluctuation of macroeconomic variables in the mainstream macroeconomics, and has increasingly been used for forecasting purpose. There is a plentiful amount of literature examining the forecast performances of DSGE models. Following the seminal work by Smets and Wouters [2007], Edge and Gurkaynak [2010] reported that the forecast accuracy of medium-scale DSGE model was not satisfactory but so are the other statistical model (Bayesian VAR) and professional forecasts.

The linearized DSGE model can virtually be regarded as a linear state-space model, with the theoretical specification of transition and observation matrixes. Indeed, Herbst and Schorfheide [2012] indicated that forecasts from a DSGE model could outperform those from an unrestricted VAR if the theoretical restriction on comovements among economic variables is well consistent with the actual data generation processes. Recently, Del Negro et.al [2013] showed that considering financial friction into a DSGE could contribute to the better forecast performance during the periods just after the financial crisis in 2008-09 in the US. The latest study of Hasumi et.al [2018] also found that an appropriate model could vary with underlying economic situations, through the introduction of time-varying pooling method of multiple models. They reported that the DSGE model with the financial friction outperformed the frictionless benchmark in terms of forecast performance during the 1980s financial bubble and following burst in Japan, while the frictionless model marked better scores for more recent periods. This may suggest that any single type of DSGE seems difficult to persistently outperform more flexible type of statistical models. Having said that, a DSGE model has own advantage as a descriptive tool, with the theoretical correspondence and high interpretability.

1.2. Forecasts by neural networks

One of the prominent progress in statistical modeling in these days is so-called artificial intelligence, more precisely, deep-neural-network type of models. Their superior performances to other competing statistical models were reported in many fields. Research applying neural networks to economic time-series are quite recent, and most of them deal with the forecasts of asset prices. The early research by Chen et al. [2003] applied a neural network to predict return of Taiwanese equity index, reporting the improved accuracy compared to a GMM approach and random walk. More recently, Fischer and Krauss [2017] compared the prediction performances of LSTM (Long Short Term Memory), Random Forests, and logistic regression, and reported the superiority of LSTM in respect of the daily S&P500 data from 1992 to 2015. Borovkova and Tsiamas [2018] employed an ensemble (dynamic pooling) method of multiple LSTM networks for the intraday stock price prediction, and reported the better performance against the logistic regression benchmarks. Those pieces of research show that neural-net type of models certainly tend to beat traditional statistical models in the context of forecasting financial data, and LSTM network particularly exhibits the superior performance. LSTM network can treat non-linear dynamics and long-memory feature of time series flexibly, which differentiate from traditional statistical models like simple Vector Auto-Regression (VAR).

Although neural network type of models tend to show state-of-the-art forecast performance, it has an disadvantage of so-called "black box" feature. For most neural networks, it is generally difficult to trace the transmission process of input information to outlay with their repeated non-linear transformations, which also hampers the models to hold theoretical correspondence.

Summing up the above literatures, the theoretical restriction by DSGE could make the linear state-space model interpretable, and contribute to the forecast accuracy as long as the restriction is consistent with the actual data generation process. Nonetheless, its absolute performance is said to not satisfactory. On the other hand, neural network models tend to outperform traditional statistical models in forecasts of financial time series, but its opacity or "black box" nature of information processing hinders the application to macroeconomic study.

1.3. State-space model constructed by neural networks

The novel structure of TDVAE (Temporal Difference Variational Auto-Encoder) proposed by Gregor et.al [2019] could help to address both issues of the opacity for neural networks and the poor accuracy for DSGE models, through combining the two approaches. The purpose of this paper involves this idea, namely, constructing theoretically interpretable and state-of-the-art performance model.

As mentioned more precisely later, TDVAE enables to replicate a Gaussian stochastic state-space model in a framework of end-to-end neural networks. Generally speaking, Variational Auto-Encoder (VAE) consists of encoder and decoder, where the former transforms data to lower-dimensional state variables (so-called representation) while the latter tries to reconstruct the data from representation, with assumption for the state variables to follow a normal distribution. TDVAE extends this concept to treat dynamic state variables by analogy with state space modeling. In this TDVAE model, a filtering probability constitutes the encoder which transforms observed data to contemporaneous states variables, while a transition probability describes dynamics of states variables. Then, a smoother probability validates the transition of state variables backward. These probabilities are parameterized by means and variances of Gaussian distribution, which are expressed as the outputs of deep neural networks trained by the data, inheriting the concept of Variational Auto-Encoder.

As seen above, TDVAE is virtually a semi-parametric non-linear state-space model, in which latent states, a low of motion, and an observation function can be specified in consistent with the theoretical restrictions derived from a DSGE model.

In the next section, I explain a basic structure of the proposed TDVAE-DSGE method. Then, results from forecasting back-tests are presented in Section 3, and the backgrounds of accuracy gains are discussed in Section4.

2. Proposed Methods

This section explains a basic structure of TDVAE, and the way to specify its components to be consistent with an orthodox DSGE model.

2.1. TDVAE structure

TDVAE assumes a general form of stochastic state-space model, where the joint likelihood of a state sequence $\mathbf{z} = (z_1, \dots, z_T)$ and an observation sequence $\mathbf{x} = (x_1, \dots, x_T)$ can be written as $p(\mathbf{x}, \mathbf{z}) = \prod_t p(z_t|z_{t-1}) p(x_t|z_t)$. Here, $p(z_t|z_{t-1})$ represents a transition model and $p(x_t|z_t)$ does a decoder or an observation function.

In order to optimize the data likelihood $p(\mathbf{x})$, a state-space model is commonly trained by computing a posterior $q(\mathbf{z}|\mathbf{x})$ over the states given the observations, which can be autoregressively decomposed as $q(\mathbf{z}|\mathbf{x}) = \prod_t p(z_t|z_{t-1}, \phi_t(\mathbf{x}))$, where ϕ_t is a function augmenting the observed data up to t $(x_1, \dots, x_t)$ for filtering posteriors, or the entire sequence $\mathbf{x}$ for smoothing posteriors.

For a given t, TDVAE evaluates the conditional likelihood $p(x_t|x_{<t})$ by inferring over only two latent states: z_{t-1} and z_t, and obtains the ELBO (Evidence Lower BOund) for a state-space model as follows:

$$\begin{aligned} \log p(x_t|x_{<t}) \geq \mathbb{E}_{(z_{t-1}, z_t) \sim q(z_{t-1}, z_t|x_{\leq t})}[&\log p(x_t|z_t) \\ &+ \log p(z_{t-1}|x_{<t}) + \log p(z_t|z_{t-1}) \\ &- \log q(z_t|z_{\leq t}) - \log q(z_{t-1}|x_{\leq t})]. \end{aligned} \tag{1}$$

Maximizing ELBO (the right-hand side) instead of the likelihood itself (the left hand side) is common approach to train encoder-decoder type of neural networks such as VAE (Variational Auto-Encoder). More formal derivation of ELBO for stochastic state-space model is explained in Buesing et al. [2018].

Here, TDVAE introduces an online belief state b_t as a summary statistics of the observed (past) data at any given t. This enables to rewrite the conditional data distribution at t as $p(x_{t+1}, \dots, x_T|x_1, \dots, x_t) \approx p(x_{t+1}, \dots, x_T|b_t)$, and the filtering posteriors at t as $q(z_t|x_1, \dots, x_t) \approx p_B(z_t|b_t)$, with $b_t = f(b_{t-1}, x_t)$. Now, the loss function of TDVAE derived from belief based ELBO is obtained as follows:

$$\begin{aligned} -\mathcal{L} = \mathbb{E}_{\substack{z_t \sim p_B(z_t|b_t) \\ z_{t-1} \sim q(z_{t-1}|b_t, b_{t-1})}} [&\log p(x_t|z_t) \\ &+ \log p_B(z_{t-1}|b_{t-1}) + \log p(z_t|z_{t-1}) \\ &- \log p_B(z_t|b_t) \\ &- \log q(z_{t-1}|z_t, b_t, b_{t-1})]. \end{aligned} \tag{2}$$

For actual training of TDVAE, functional forms need to be specified. The online belief state is expressed by a standard LSTM network: $b_t = LSTM(b_{t-1}, x_t)$. The mapping operator from x to a normal distribution with mean $\mu(x)$ and log-standard deviation $\log \sigma(x)$ is denoted as D, where $[\mu, \log \sigma] = \mathrm{W}_3 \tanh(W_1 x + B_1) \operatorname{sigmoid}(W_2 x + B_2) + B_3$, with W_1, W_2, W_3 as weight matrices and B_1, B_2, B_3 as biases. The size of hidden layer of the D maps is set to 100, and belief states have size 100. Then, I choose to use the Adam optimizer.

The set of equations describing the system and its loss function over the pair of two time periods t and $t-1$ are as follows:

$$\begin{aligned} &b_t = LSTM(b_{t-1}, x_t) \\ &z_t^B \sim p_B^t = D(b_t) \\ &p_B^{t-1} = D(b_{t-1}) \\ &z_{t-1}^S \sim q_S^{t-1|t} = D(b_t, z_t^B) \\ &z_t^T \sim p_T^{t|t-1} = D(z_{t-1}^S) \\ &p_D = D(z_t^B) \\ &L_{t-1} = KL\left(q_S^{t-1|t} | p_B^{t-1}\right) \\ &L_t = \log p_B^t(z_t^B) - \log p_T^{t|t-1}(z_t^T) \\ &L_x = -\log p_D(x_t) \\ &Loss = L_{t-1} + L_t + L_x. \end{aligned} \tag{3}$$

2.2. TDVAE-DSGE

As mentioned earlier, a standard New Keynesian DSGE model can be reduced to a linear state space representation, by solving its log-linearized equilibrium conditions via the method like Sims [2001] or Klein [2000]. In other words, a standard DSGE model can be summarized into a pair of the transition matrix on state variables and the observation (decoder) matrix to translate the states to observations. The values of those matrices are fully determined by the structural (deep) parameters of the DSGE model to be used.

The idea of this paper is to transplant those transition and decoder matrixes of a standard New Keynesian DSGE model into the aforementioned TDVAE framework. More precisely, means of a state transition probability

$p_T^{t|t-1}$ and a decoder probability p_D in TDVAE are to be calculated based on the given transition matrix F and the decoder matrix P of a reduced DSGE model. The system is described as follows:

$$\begin{aligned}
&b_t = LSTM(b_{t-1}, x_t) \\
&z_t^B \sim p_B^t = D(b_t) \\
&p_B^{t-1} = D(b_{t-1}) \\
&z_{t-1}^S \sim q_S^{t-1|t} = D(b_t, z_t^B) \\
\\
&\mu_{z^T} = F z_{t-1}^S \\
&\log \sigma_{z^T} = D(z_{t-1}^S) \\
&z_t^T \sim p_T^{t|t-1} = N(\mu_{z^T}, \sigma_{z^T}) \\
\\
&\mu_x = P z_t^B \\
&\log \sigma_x = D(z_t^B) \\
&p_D = N(\mu_x, \sigma_x).
\end{aligned} \tag{4}$$

This can also be trained by using the same ELBO loss function. Hereafter, this approach is named as TDVAE-DSGE.

2.3. The DSGE model to be used

The DSGE model to be used is a standard New-Keynesian model of Smets and Wouters [2007], which over the past decade has been a standard tool in mainstream macroeconomic research.

This model includes seven structural (i.e. theoretically identified) shock processes, namely (1) a productivity shock, (2) a shock to household (investor) preferences (the time discount rate), (3) a shock to exogenous demand (government spending and foreign demand), (4) a shock to capital investment efficiency, (5) a price mark-up shock, (6) a wage mark-up shock, and (7) a discretionary monetary policy shock.

Given the materialized values of seven shocks and the deep parameters specifying behaviors of agents, the ten variables of interest, namely capital, labor inputs, production, consumption, investment, the real capital yield, the real capital price, the inflation rate, the real wage, and the nominal interest rate, are determined through the structural equations which represents optimizing behavior of the six types of economic agents, namely households (end investors), employment agencies, intermediate good producers, final good producers, capital producers, and the public sector (central bank and government).

By solving the structural equations into reduced forms, the DSGE is expressed in state-space representation, and its deep parameters, endogenous variables, and shock processes are simultaneously estimated by a Particle Filter algorithm so as to obtain the best replication of actual fluctuations in macroeconomic data. This paper denotes this standard estimation procedure as PF-DSGE for the later use. Obtaining filtered time series of these structural shock processes will enable us to identify the exogenous causes of macroeconomic fluctuations.

One major distinction of TDVAE-DSGE from PF-DSGE is an existence of belief state, which enables to compress a large number of observable data in efficient way through the LSTM network and conditions the filtering probability of state variables (i.e. shock processes). There are some pieces of research tackling data-rich estimation of a DSGE model through Dynamic Factor Model (Iiboshi [2015] for example), which reported, however, no significant information gain by expanding dataset more than core observable variables. The factor loading matrix of DFM, which is linear and only considers contemporaneous correlations, seems too restrictive to extract potentially valuable information from the big dataset.

2.4. TDVAE-DSGE with correction term

In the framework of TDVAE-DSGE described above, a law of motion (or transition model) of state variables is fully specified by the values of corresponding DSGE matrix. As discussed in the first section, this might be too restrictive to capture potentially non-stationary nature of the economic data. In other words, actual developments of economy (or true data generation process) could often deviate from the assumptions of a plain-vanilla linearized New Keynesian DSGE model. In order to address this problem, I choose to insert a neural-net-based correction term into the transition model of TDVAE-DSGE. Thus, the transition probability with a correction term is described as follows:

$$\begin{aligned}
&[CT, \log \sigma_{z^T}] = D(z_{t-1}^S), \\
&\mu_{z^T} = F z_{t-1}^S + CT, \\
&z_t^T \sim p_T^{t|t-1} = N(\mu_{z^T}, \sigma_{z^T}).
\end{aligned}$$

(5)

With its non-linear nature of a neural network, the correction term CT could help to capture flexibly the residual dynamics of state variables, which accounts for the movement of observed data not identified by the DSGE model. Hereafter, this modified structure is denoted as TDVAE-DSGE-CT.

3. Empirical Results

This section compares performances of the models introduced above and some benchmarks with regard to forecasting back-test. Precisely, following three versions of TDVAE-DSGEs and three benchmarks are tested on the equal-footing basis.

(1) TDVAE-DSGE-7V (7V): this version of TDVAE-DSGE uses only the seven core observation variables (i.e. production, consumption, investment, labor input, wage, price, interest rate) as inputs to construct belief states.

(2) TDVAE-DSGE-DR (DR): this is the data-rich version of TDVAE-DSGE where belief states compress a large number (more than 500) of economic and financial data in addition to the seven core variables. Decoder probability (i.e. observation likelihood) remains to be calculated with the seven core variables.

(3) TDVAE-DSGE-CT (CT): this version has a correction term in its transition model as explained in the section 2.4, where belief states are constructed from the big data as with the DR version.

(4) VAR: a simple reduced-form VAR (2) model, augmenting the core seven variables.

(5) PF-DSGE (PF): the estimated DSGE model by particle filter explained above. Forecasts are constructed by randomly sampled structural shocks for quarters ahead.

(6) LSTM: a standard LSTM network taking lagged values of the seven core variables to explain the current values of them, intended to capture a potentially non-linear low of motion.

In a back-test procedure, forecasts for real GDP by each model have been assessed, and all generated with no future information except for the use of revised data. Back-test periods are from 1Q2011 to 4Q2018, in which all models are re-estimated by every quarter, with the data that had been available at the release date of GDP statistics. The corresponding release dates to the values of each economic indicators are corrected from the sources, but for the periods of unavailable pasts, complemented by the largest days of delay in the recent known release dates.

For every quarters of back testing, forecasts up-to eight quarters ahead were generated after the re-estimation. All models take real GDP in the form of log-detrended level, so do their forecasts. In order to assess the performance of each model, Root Mean Squared Error (RMSE) is calculated for two, four and eight quarters-ahead forecasts. When evaluating forecasts of a level variable with some persistency, it is known that simple autoregressive model like AR(1) or even side-slide forecasts could record a high RMSE accuracy, which is so called echo-effects. The actual forecasts by PF-DSGE and VAR indeed exhibited the patterns of side-slide, in which their RMSEs are thought of too low with the existence of such echo-effects. Therefore, I choose to additionally test the performances in terms of quarter by quarter (qoq) changes. In this case, RMSEs up-to two, four or eight quarters ahead are calculated first at each period, and then averaged over all back-test periods. All those results are summarized in Table 1. Overall score of each model is evaluated by the ranking over four patterns of those RMSEs.

Table 1. Summary of the back-test results

RMSE 1Q2011 to 4Q2018

	PF	VAR	LSTM	TDVAE-7V	TDVAE-DR	TDVAE-CT
q+2	1.40	1.49	1.48	2.09	1.58	1.03
q+4	1.70	2.31	1.88	2.26	1.57	1.05
q+8	0.82	3.05	2.29	2.11	2.09	0.96
qoq_2q	0.75	0.71	0.63	0.55	0.49	0.48
qoq_4q	0.71	0.80	0.68	0.60	0.59	0.62
qoq_8q	0.68	0.74	0.65	0.63	0.64	0.64
Total Ranking	4	6	5	3	2	1

4. Discussions

4.1. Role of correction term

Assessing with the overall ranking, the best score was achieved by TDVAE-DSGE-CT (CT) as anticipated. This outperformance of CT against TDVAE-DSGE-DR (DR) suggests that the correction term constructed through neural networks played an important role to capture a potentially non-linear dynamics (or underlying data generation process) of the state variables, corresponding to the structural shocks of macroeconomy under the DSGE restriction.

Comparing the actual filtered values of the seven state variables under CT and DR, some variables show acute differences in particular periods despite the considerable resemblances globally for most variables. Concretely, the productivity process under CT shows a peak and trough during the late 1980s and early 1990s (the Japan's financial bubble and burst), and also a much deeper correction in 2008-2009 (the Lehman shock), which are not appeared in the process under DR. The difference between CT and DR is also visible in the wage-markup process in most periods. In addition, the process of monetary policy shock has been broadly non-similar.

These may suggest the modeling of financial sector, labor market, and monetary policy in the plain-vanilla DSGE of Smets and Wouters [2007] are not adequate to explain the underlying data generation process of Japan's macroeconomy during the corresponding periods. Indeed, the importance of richer description of financial sector was also emphasized in Hasumi et.al [2018].

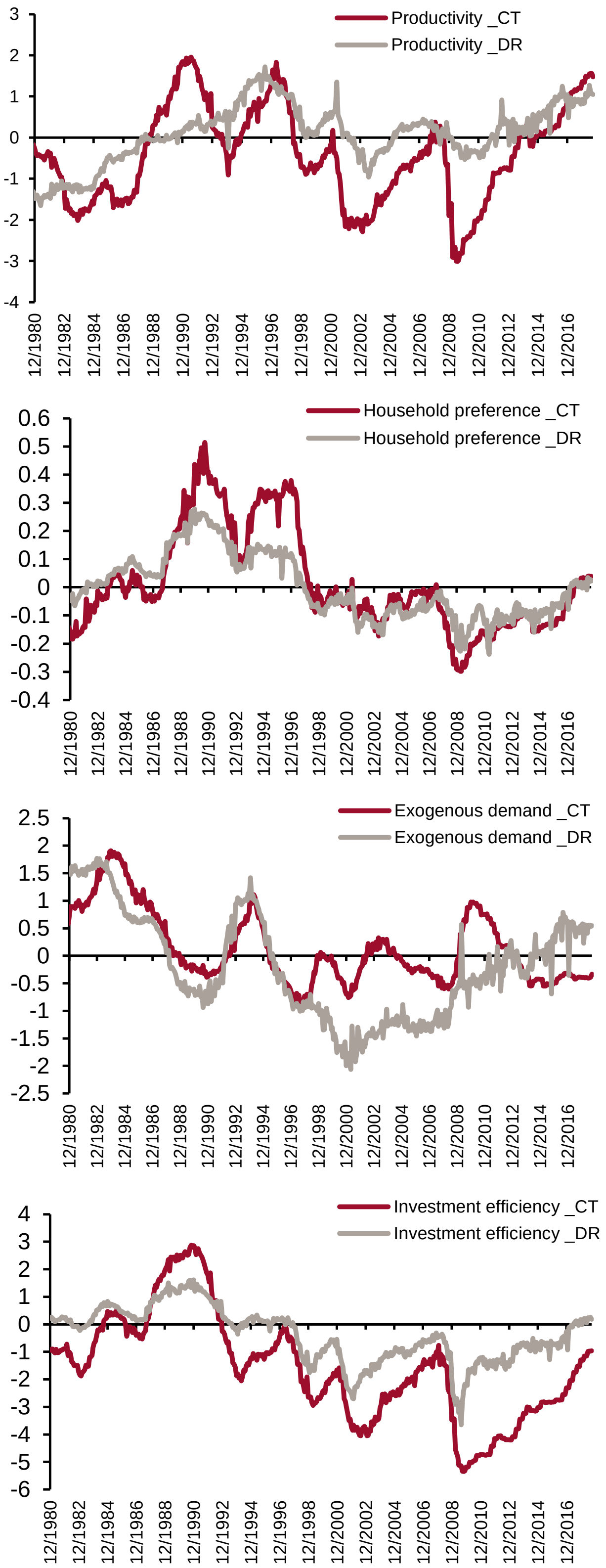
4.2. Data-rich beliefs

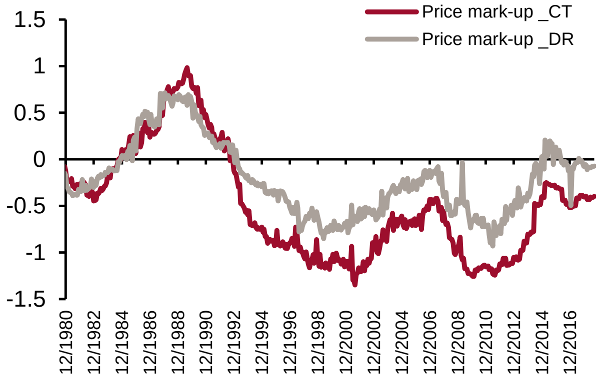
The better performance of TDVAE-DSGE-DR (DR) to TDVAE-DSGE-7V (7V) implies that the use of richer dataset to form belief states leads to more accurate inference of the state variables, and their dynamics. The 7V did not so significantly outperform the standard particle filter estimation of DSGE (PF) in the overall accuracy ranking, which suggest that the advantage of TDVAE methodology tends only materialized with a richer dataset. These results imply that non-linear flexible structure of LSTM could efficiently extract valuable information from large number of economic indicators, which is not contained in the seven core observable variables.

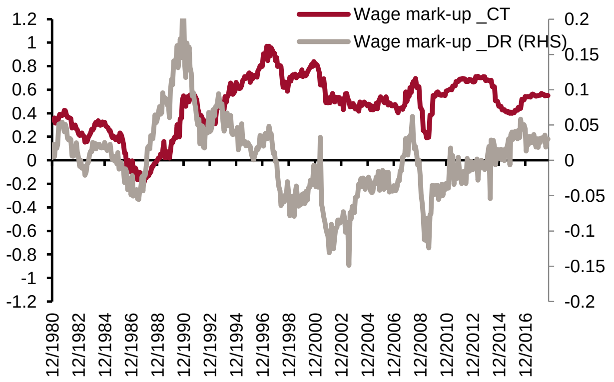
4.3. Theoretical restriction

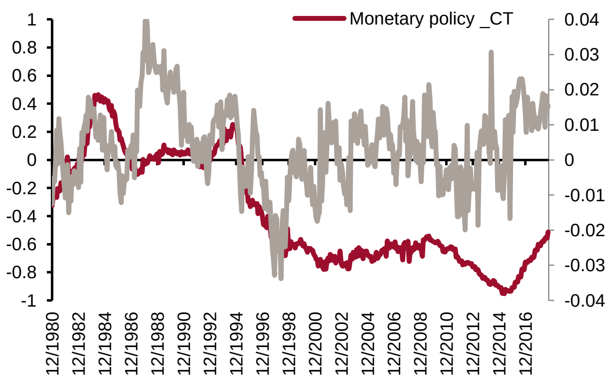
Both the higher accuracy of TDVAE-DSGEs against LSTM and the same of PF-DSGE against VAR suggest that theoretical restrictions by a DSGE model certainly enhanced the out-of-sample forecasting performances of Japan's GDP, regardless of the type of models, namely, neural networks or linear autoregression.

Figure 1. Filtered values of the states by TDVAE-DSGE-CT vs. DR









5. Conclusion

This paper examined the possibility of combining a DSGE model and neural networks to supplement each other, with regard to out-of-sample forecasts for economic variables. The aim is to build a model with theoretical interpretability and state-of-the-art performance. The novel neural-net structure of TDVAE (Temporal Difference Variational Auto-Encoder) proposed by Gregor et.al [2019] enables to realize this idea. TDVAE virtually replicates a Gaussian stochastic state-space model by combination of neural networks. Because a DSGE model provides theoretical restrictions on the state transition and observation matrices of a linear state-space model, I choose to transplant those DSGE-oriented matrices into the formulations of state transition and observation probabilities in TDVAE. This TDVAE-DSGE approach certainly achieved the superior performance in the task of out-of-sample forecasts on Japan's real GDP during 1Q/2011 and 4Q/2018.

References

[1] Borovkova, S. and Tsiamas, I. (2019). An ensemble of LSTM neural networks for high - frequency stock market classification. Journal of Forecasting, 2019;1-20. https://doi.org/10.1002/for.2585

[2] Buesing, L., Weber, T., Racaniere, S., Ali Eslami, S., Rezende, D., Reichert, D., Viola, F., Besse, F., Gregor, K., Hassabis, D., and Wierstra, D. (2018). Learning and querying fast generative models for reinforcement learning. arXiv preprint arXiv:1802.03006.

[3] Chen A., Leung M.T., and Daouk H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. Computers & Operations Research, 30: 901-923.

[4] Del Negro, M., Eusepi, S., Giannoni, M., Sbordone, A., Tambalotti, A., Cocci, M., Hasegawa, R., Linder, and M. (2013). The FRBNY DSGE model, Staff Reports 647. Federal Reserve Bank of New York.

[5] Rochelle, M., and Refet, G. (2010). How useful are estimated DSGE model forecasts for central bankers?. CEPR Discussion Papers, No 8158.

[6] Fischer, T., and Krauss, C. (2017).Deep learning with long short-term memory networks for financial market predictions. FAU Discussion Papers in Economics 11/2017.

[7] Gelfer, D.G. (2016). Financial crises and labor market dynamics: evidence from a data-rich DSGE model. Mimeo.

[8] Gregor, K., Papamakarios, G., Besse, F., Buesing, L., and Weber, T. (2019). Temporal Difference Variational Auto-Encoder. arXiv preprint arXiv:1806.03107.

[9] Hasumi et.al [2018]

[10] Hasumi, R., Iiboshi, H., Matsumae, T., and Nakamura, D. (2018). Does a financial accelerator improve forecasts during financial crises?: Evidence from Japan with Prediction Pool Methods. MPRA Paper, No. 85523.

[11] Herbst, E., and Schorfheide, F. (2012). Evaluating DSGE model forecast of comovements. Journal of Econometrics, 171(2), 152-166.

［１２］ Iiboshi, H., Matsumae, T., Namba, R., and Nishiyama, S. (2015). Estimating a DSGE model for Japan in a data-rich environment. Journal of the Japanese and International Economies, vol. 36(C), pages 25-55.
［１３］ Klein, P., (2000). Using the generalized schur form to solve a multivariate linear rational expectations model. Journal of Economic Dynamics and Control, 24, 1405–1423.
［１４］ Sims, C.A., (2001). Solving linear rational expectations models. Journal of Computational Economics, 20 (1–2), 1–20.
［１５］ Smets,F., and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. American Economic Review, 97 (3): 586-606. DOI: 10.1257/aer.97.3.586.