Deep Vector Autoregression for Macroeconomic Data

Masters Thesis - BGSE

Marc Agustí (marc.agusti@barcelonagse.eu) Patrick Altmeyer (patrick.altmeyer@barcelonagse.eu) Ignacio Vidal-Quadras Costa (ignacio.vidalquadrascosta@barcelonagse.eu)

23 June, 2021

At a glance

We effectively extend the conventional linear VAR model to the broader class of Deep VARs. By fitting each equation of the VAR system with a deep neural network the Deep VAR is able to capture non-linear dependencies across time and variables. We find that this leads to significant performance improvements in terms of in-sample fit, out-of-sample fit and point forecasting accuracy.

Motivation

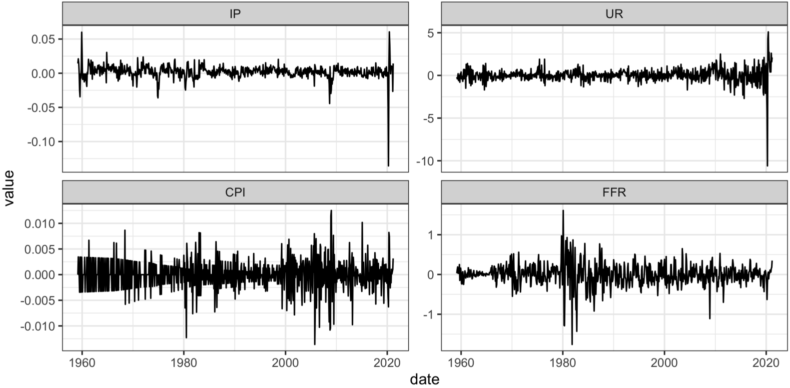
- Bulk of research on the forecasting of time series to assess the effect of current policy decisions on future economic variables. Accurately modelling and understanding the transmission mechanism is of utmost importance to central bankers.
- The conventional VAR is still one of the most common tools used by practitioners:
  - Advantages include its simplicity, closed-form solutions for standard errors and straight-forward interpretability.
  - **Shortfall**: imposes linearity.
- While alternatives and extensions of the VAR exist (Bayesian VAR, non-linear VAR, large econometric models, ...) little work has so far been done on integrating deep learning.
- Deep learning has been shown to be remarkably useful for modelling highly non-linear functions of essentially arbitrary form (Goodfellow, Bengio, and Courville 2016).

Previous literature

- Non-linear dependencies are likely to form part of the data generating process of variables commonly used to model the monetary transmission mechanism (Brock et al. 1991).
- A range of machine learning models has previously been used in the context of time series forecasting Kihoro, Otieno, and Wafula (2004). Deep learning has been shown to be particularly successful at capturing non-linearities G. P. Zhang (2003).
- Joseph et al. (2021) review both machine learning and deep learning methods for forecasting inflation and find that neural networks in particular are useful for forecasting especially at a longer horizon.

Data

- We use the relatively novel FRED-MD data base which is updated monthly and publicly available (McCracken and Ng 2016).
- The sample spans from March, 1959 to March, 2021 providing us with a relatively rich data set of macroeconomic time series with $T = 745$ observations.



Deep VAR methodology

Form conventional VAR ...

- Recall the conventional VAR(p) with p lags and a constant deterministic term:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t \qquad (1)$$

- Can be estimated through ordinary least-squared and considered as a **seemingly unrelated regression** (SUR) model composed of individual regressions with common regressors (Greene 2012):

$$y_{it} = c_i + \sum_{m=1}^{p}\sum_{j=1}^{K} a_{jm}y_{jt-m} + u_{it} \quad , \quad \forall i = 1, ..., K \qquad (2)$$

. . . to Deep VAR

- Relax the assumption of linearity and instead model the process as system of potentially highly non-linear equations:

$$y_{it} = f_i\left(\mathbf{y}_{t-1:t-p};\theta\right) + v_{it} \quad , \quad \forall i = 1, ..., K \tag{3}$$

- Can restate this more compactly as:

$$\mathbf{y}_t = \mathbf{f}(\mathbf{y}_{t-1:t-p};\theta) + \mathbf{v}_t \tag{4}$$

Long Short-Term Memory

- The most common choice of neural networks architectures for modelling persistent time series is the LSTM:

  "The LSTM [has] the ability to remove or add information to the cell state, carefully regulated by structures called gates. Gates are a way to optionally let information through." — Olah (2015)

$$
\begin{aligned}
\mathbf{f}_t &= \sigma\left(\mathbf{b}_f + \mathbf{W}_f \mathbf{h}_{t-1} + \mathbf{U}_f \mathbf{h}_{-1}\right) \\
\mathbf{i}_t &= \sigma\left(\mathbf{b}_i + \mathbf{W}_i \mathbf{h}_{t-1} + \mathbf{U}_i \mathbf{h}_{-1}\right) \\
\mathbf{o}_t &= \sigma\left(\mathbf{b}_o + \mathbf{W}_o \mathbf{h}_{t-1} + \mathbf{U}_o \mathbf{h}_{-1}\right) \\
\mathbf{C}_t &= \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tanh\left(\mathbf{b}_C + \mathbf{W}_C \mathbf{h}_{t-1} + \mathbf{U}_C \mathbf{h}_{-1}\right) \\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \\
\widehat{\mathbf{y}}_t &= \mathbf{c} + \mathbf{V}\mathbf{h}_t
\end{aligned}
\tag{5}
$$

Model selection

- With respect to the lag order p we initially use the Akaike Information Criterion for both the VAR and Deep VAR.
- We check for stability of the process by looking at the Kp eigenvalues of the companion matrix of the VAR.
- For the neural networks underlying the baseline Deep VAR we use a depth of two layers and a constant width of 100 neurons.
- The neural networks are regularized through dropout.

Findings

In-sample loss

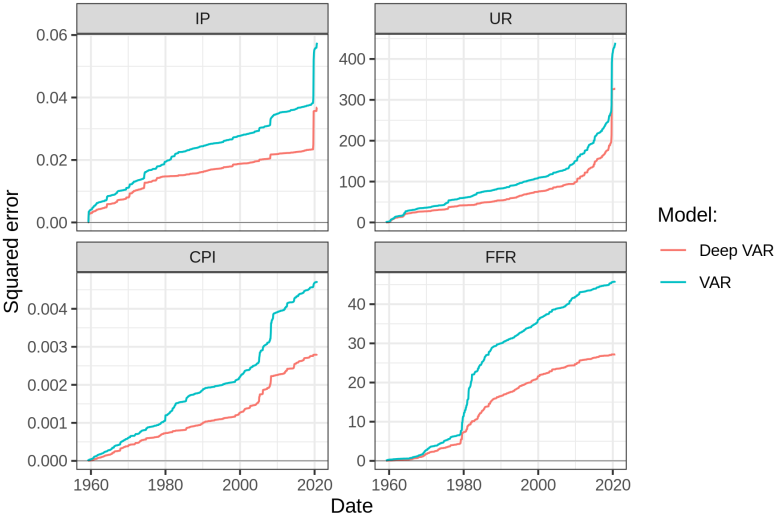


Figure 2: Comparison of cumulative loss over the entire sample period for conventional VAR and proposed Deep VAR.

Out-of-sample loss

Table 1: Root mean squared error (RMSE) for the two models across subsamples and variables.

Sample	Variable	DVAR	VAR	Ratio (DVAR / VAR)
test	IP	0.00608	0.01484	0.40957
test	UR	0.97022	1.65170	0.58741
test	CPI	0.00273	0.00342	0.79869
test	FFR	0.19964	0.23974	0.83271
train	IP	0.00528	0.00727	0.72610
train	UR	0.32325	0.43322	0.74615
train	CPI	0.00149	0.00232	0.64019
train	FFR	0.16115	0.25780	0.62509

Out-of-sample forecasting errors

Table 2: Comparison of n-step ahead pseudo out-of-sample forecasts.

Variable	VAR FRMSE	Deep-VAR FRMSE	VAR correlations	Deep-VAR correlations
IP	0.01870	0.01602	-0.30409	-0.65279
UR	0.85984	0.82785	-0.10093	0.27425
CPI	0.00946	0.00708	-0.33567	0.07823
FFR	0.52321	0.39161	-0.55935	0.01161

Varying hyperparameters

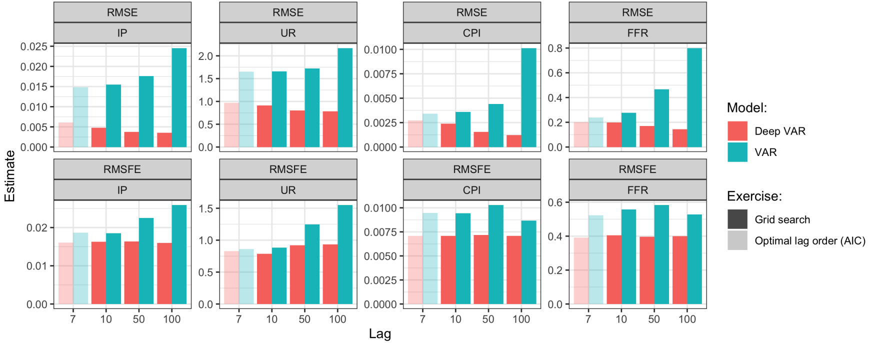


Figure 3: Pseudo out-of-sample RMSE and RMSFE for both models across the different lag choices. For the sake of completeness we also include the performance measures we obtained when we initially ran both models using the optimal lag order as determined by the AIC.

Discussion

Caveats and extensions

- Policy-makers rarely base their decisions solely on point estimates, so ultimately our proposed Deep VAR framework needs to be able to quantify uncertainty around estimates.
  - **Monte Carlo dropout**: distribution over point estimates through the introduction of stochasticity
- Support for the estimation of impulse response functions and variance decompositions is another missing cornerstone in the current version of our proposed framework.
  - Verstyuk (2020) manages to recover IRFs numerically; should be readily applicable to our Deep VAR framework.
- To allow for a ‘fairer’ comparison we need to let Bayesian VAR, non-linear VAR and large econometric models enter into the horse race.

Concluding remarks

- To the best of our knowledge we are the first to propose a framework for Deep VARs that barely deviates from the conventional framework.
- We show that relaxing the linearity constraint leads to a massive boost in modelling performance.
- To allow other researchers straight-forward access the methodology has been easily avialable through an R package:

```
devtools::install_github(
  "pat-alt/deepvars",
  build_vignettes=TRUE
)
```

- We continue to be in close contact with Eddie Gerba, research manager at the Bank of England, and plan to submit to an upcoming conference on ML for Monetary Policy.

Your questions and comments

References I

Brock, William A, David Arthur Hsieh, Blake Dean LeBaron, William E Brock, and others. 1991. *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. MIT press.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

Greene, William H. 2012. "Econometric Analysis 7th Ed (International)." Pearson.

Hamzaçebi, Coşkun. 2008. "Improving Artificial Neural Networks' Performance in Seasonal Time Series Forecasting." *Information Sciences* 178 (23): 4550–59.

Joseph, Andreas, Eleni Kalamara, George Kapetanios, and Galina Potjagailo. 2021. "Forecasting Uk Inflation Bottom Up."

References II

Kihoro, J, RO Otieno, and C Wafula. 2004. "Seasonal Time Series Forecasting: A Comparative Study of ARIMA and ANN Models."

McCracken, Michael W, and Serena Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business & Economic Statistics* 34 (4): 574–89.

Olah, Chris. 2015. "Understanding LSTM Networks." https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Verstyuk, Sergiy. 2020. "Modeling Multivariate Time Series in Economics: From Auto-Regressions to Recurrent Neural Networks." *Available at SSRN 3589337*.

Zhang, G Peter. 2003. "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model." *Neurocomputing* 50: 159–75.

References III

Zhang, Guoqiang, B Eddy Patuwo, and Michael Y Hu. 1998. "Forecasting with Artificial Neural Networks:: The State of the Art." *International Journal of Forecasting* 14 (1): 35–62.

Hiddens

Cumulative loss for random walk

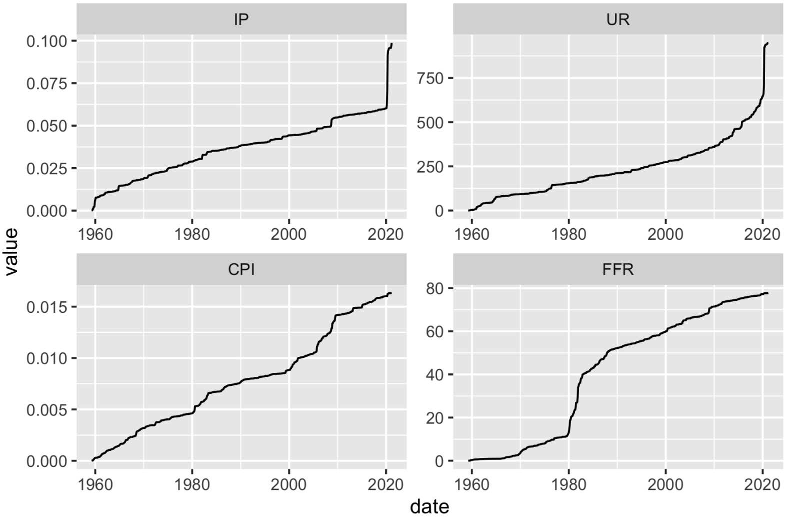
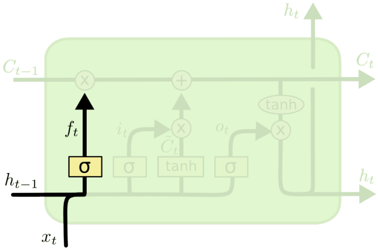


Figure 4: Cumulative mean squared error for naive (random walk) prediction.

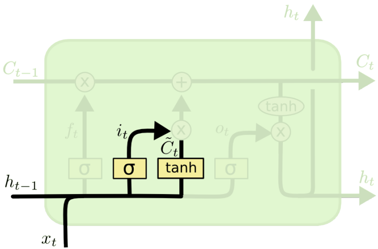
LSTM - Forget gate



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

Figure 5: Forget gate. Source: Olah (2015)

LSTM - Input gate

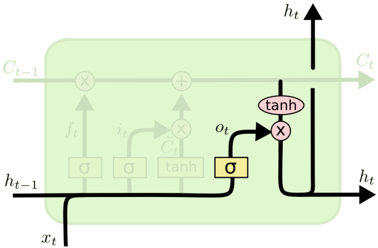


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Figure 6: Input gate. Source: Olah (2015)

LSTM - Output gate



$$o_t = \sigma\left(W_o \; [\, h_{t-1}, x_t] \; + \; b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 7: Output gate. Source: Olah (2015)