Deep Vector Autoregression for Macroeconomic Data

Patrick Altmeyer[1] Marc Agusti[2] Ignacio Vidal-Quadras Costa[3]

31 January, 2022

[1]Delft University of Technology, p.altmeyer@tudelft.nl

[2]European Central Bank, marc.agusti _i _torres@ecb.europa.eu

[3]European Central Bank, ignacio.vidalquadrascosta@barcelonagse.eu

Motivation

Can we leverage the power of deep learning in VAR models?

- We propose **Deep VAR**: a novel approach towards VAR that leverages the power of deep learning in order to model non-linear relationships.
- Worked under the following premise: **maximize performance** of an existing and trusted framework under **minimal intervention**.
- We maintain the additive structure of the VAR, but relax the assumption of linearity by modelling each equation of the VAR system as a recurrent neural network.
- By staying methodologically as close as possible to the original benchmark, we hope that our approach is more likely to find acceptance in the economics domain.

Key contributions

- Simple methodology close in spirit to conventional benchmark.
- Significant improvement in model fit and forecasting accuracy.
- Open source R package `deepvars` to facilitate reproducibility.

Work-in-progress:

- Master's thesis was selected for publication by Universitat Pompeu Fabra.
- Feedback rounds with Eddie Gerba (Bank of England, LSE) and Chiara Osbat (ECB).
- Presented an updated version of the paper at NeurIPS 2021 MLECON workshop in December.

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.

Methodology

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

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

- Each single variable is model is modelled as a recurrent neural network:

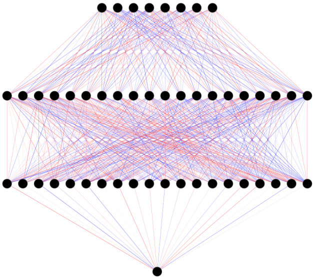

Figure 1: Neural Network Architecture.

Data

- To evaluate our proposed methodology empirically we use the FRED-MD data base to collect a sample of monthly US data on:
  - output (IP)
  - unemployment (UR)
  - inflation (CPI)
  - interest rates (FFR)
- Our sample spans the period from January 1959 through March 2021.

Model fit

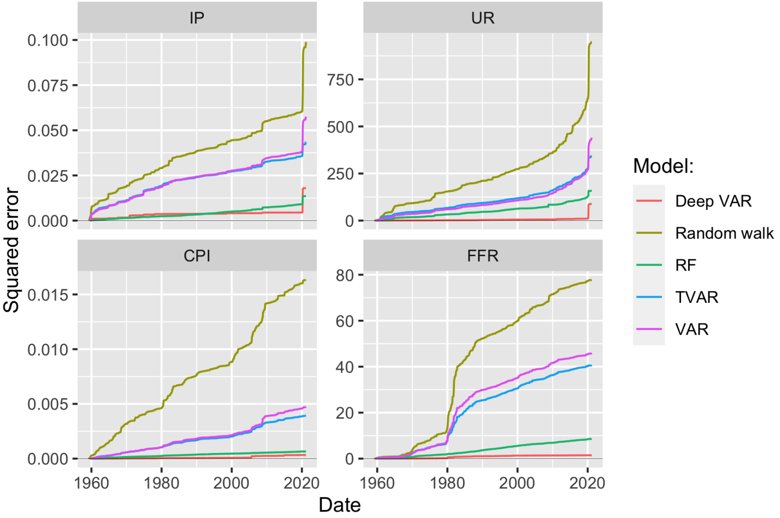

Figure 2: Comparison of cumulative loss over the entire sample period for Deep VAR and benchmarks.

Forecasting

Question: recursive forecasts like in conventional VAR or training on n outputs?

- Initially we opted for the former approach and provided anecdotal evidence that Deep VAR outperforms
- Have since tested this more rigorously using rolling window:
  - Deep VAR still outperforms VAR especially at short horizon
  - Currently investigating if training on n outputs provides additional edge.

Concluding remarks

- Simple framework that relies on the premise of minimal intervention in the conventional and trusted framework.
- Deep learning appears to do a good job at capturing non-linear dependencies.

But...

- Added complexity is (often) coupled with lack of interpretability:
  - No analytical expressions for impulse response functions and variance decompositions
  - Verstyuk (2020) manages to recover IRFs numerically; should be readily applicable to our Deep VAR framework.
- Uncertainty estimation can be done through Bayesian methods: deep ensemble, MC dropout, Variational Inference:
  - All of the above entail an added layer (layers really!) of computational complexity.
  - Laplace Redux for effortless Bayesian Deep Learning (Daxberger et al. 2021) holds promise, but not yet implemented.

Your questions and comments

References I

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

Daxberger, Erik, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. 2021. "Laplace Redux-Effortless Bayesian Deep Learning." *Advances in Neural Information Processing Systems* 34.

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."

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

References II

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.

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

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)

$$
\begin{aligned}
\mathbf{f}_t &= \sigma\left(\mathbf{b}_f + \mathbf{W}_f \mathbf{h}_{t-1} + \mathbf{U}_f \mathbf{h}_{-1}\right) \\
\mathbf{i}_t &= \sigma\left(\mathbf{b}_i + \mathbf{W}_i \mathbf{h}_{t-1} + \mathbf{U}_i \mathbf{h}_{-\mathbf{1}}\right) \\
\mathbf{o}_t &= \sigma\left(\mathbf{b}_o + \mathbf{W}_o \mathbf{h}_{t-1} + \mathbf{U}_o \mathbf{h}_{-1}\right) \\
\mathbf{C}_t &= \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tanh\left(\mathbf{b}_C + \mathbf{W}_C \mathbf{h}_{t-1} + \mathbf{U}_C \mathbf{h}_{-1}\right) \\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \\
\hat{y}_{it} &= c + \mathbf{v}^T \mathbf{h}_t
\end{aligned}
\tag{2}
$$

Rolling window forecasts