Discussion:

On the Interpretability of Deep Neural Networks in Macroeconomics

by George Kapetanios and Felix Kempf, King's College London

Andreas Joseph, Bank of England*

Modelling with Big Data and Machine Learning: Interpretability and Model Uncertainty

BoE, London, 5. Nov. 2019

*Disclaimer: The expressed views are my own and not necessarily those of the Bank of England (BoE) or its committees. All errors are mine.

Pros & Cons of machine learning (ML) relative to econometric approach

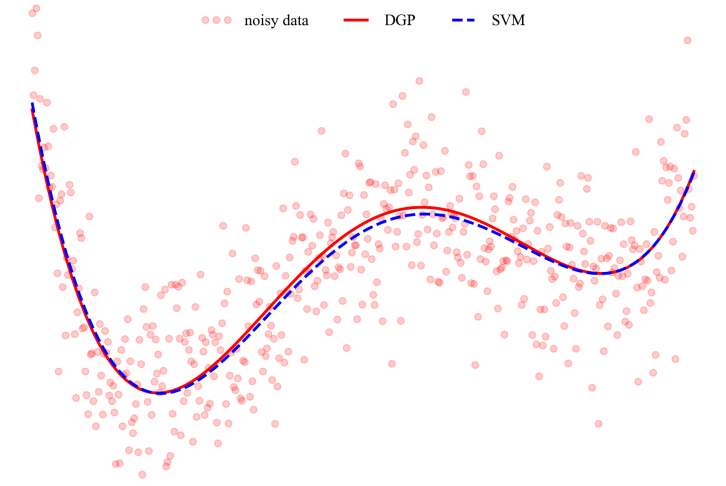
Advantages

- Often higher accuracy
- Lower risk of misspecification
- Return richer information set, e.g. non-linear functional form

Disadvantages

- Higher model complexity $\Rightarrow$ "black box critique"
- Less analytical guarantees, e.g. risk of overfitting
- Often larger data requirement

Example of toy DGP: Degree-4 polynomial with 500 noisy observations



This paper

Synopsis: Build accurate non-linear macroeconomic forecasting models (ANN) and tools to understand them.

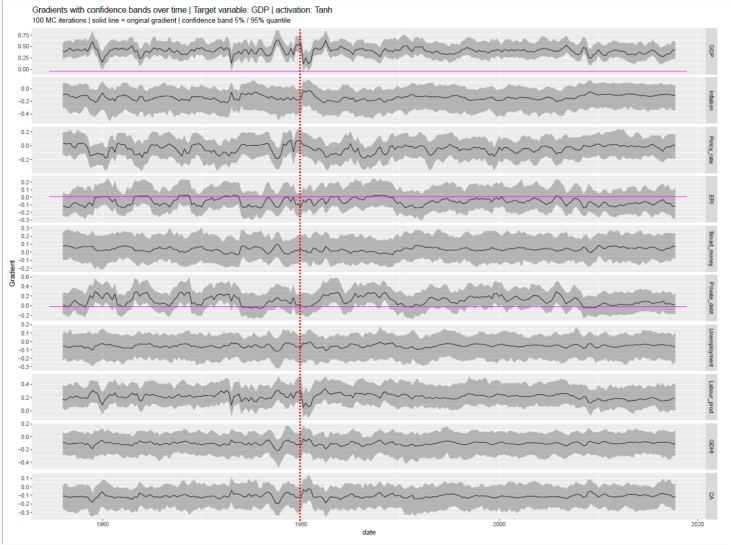
Why it matters: Good decisions, need good advise. E.g. macroeconomic projections should be accurate and transparent: Fundamental trade-off between structural and non-structural modelling $\Rightarrow$ ML may provide better trade-off!

What's being done: Out-of-bag[?] predictions and [later] expanding window forecasts of US/UK GDP, market returns, inflation and the policy rate (find non-linear[?] Taylor rule) and dynamic calculation of partial derivative and impulse response functions (novel!) to understand changing input-output relations:

$$PD: \quad I^k_{j,t} = \frac{\partial g^k(\mathbf{X}, \theta)}{\partial x_{j,t}} \Bigg|_{x_{i,t}, i \neq j}, \qquad IRF_{t+h?} = E[Y^k_{t++h?} | \nu_t, \Omega_{t-1}] - E[Y^k_{t++h?} | \Omega_{t-1}]$$

Is it worth it? Definitely!

Some results: Dynamic PD of ANN for modelling UK inflation



Main comment: Do not reinvent the wheel but build on existing ML literature

There are a large number of explanatory approaches in ML, many very similar to yours:

- Partial dependence (general)
- Gain, Saabas, permutation (tree models)
- LIME (General) [Ribeiro et al. (2016)]
- Integrated gradient (ANN) [Sundararajan et al. (2017)]
- DeepLift (ANN) [Shrikumar et al. (2017)]
- Shapley values & regressions (general) [Strumbelj and Kononenko (2010); Joseph (2019)]

Shapley values have the best properties

$$\Phi^S\big(\hat{f}(x_i)\big) = \phi_0 + \sum_{k=1}^{m} \phi_{ik}^S \overset{\textit{lin. model}}{=} x_i\hat{\beta}$$

Desirable properties: Efficiency/accuracy, missingness, symmetry, strong monotonicity/consistency and linearity.

Proposition: PD can be defined using Shapley values

$$I_{j,t}^{\phi_k} \equiv \lim_{\Delta x_j \to 0} \frac{\phi_k(x_j + \Delta x_j \,|\, x_i, i \neq j, t) - \phi_k(x_j \,|\, x_i, i \neq j, t)}{\Delta x_j}$$

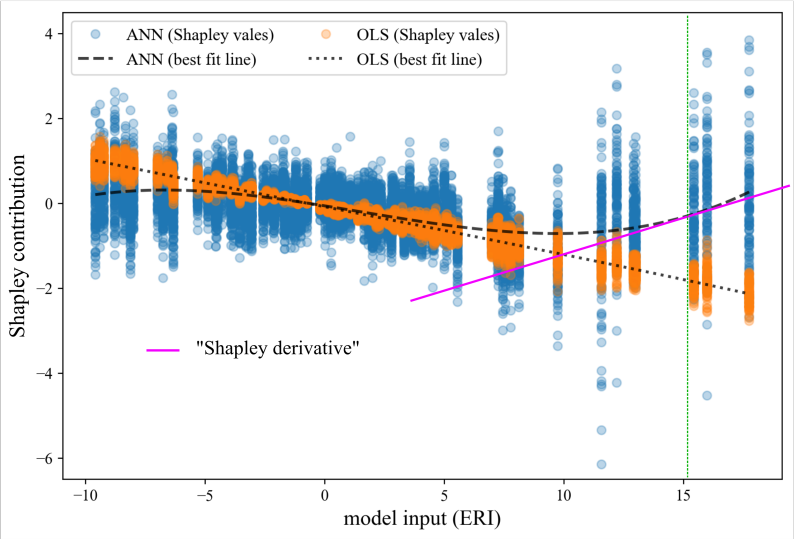
Surely this will inherit some nice properties.

Shapley regression inflation forcasting using ANN (US) [Joseph (2019)]

variable	rank	Γ_k^S	β^S	t_{β^S}
Inflation$_{-4q}$	1	$+0.22$	-0.17	-1.13
Private debt	2	$+0.15$	0.01	0.06
ERI	3	-0.11^{***}	0.31	2.79
Policy rate	4	$+0.10^{**}$	0.37	1.65
Broad money	5	-0.10	-0.25	-2.43
GDP	6	-0.09	0.13	0.76
Labour prod	7	$+0.07$	-0.32	-2.07
CA	8	$+0.07$	0.16	0.66
GDHI	9	-0.05	0.04	0.33
Unemployment	10	-0.04	0.12	0.73

Significance levels: * (10%), ** (5%), *** (1%). Source: Author's calculation.

ANN functional form and derivative using Shapley values



Additional comments

- Modelling setup not clear. VAR-like MIMO structure would leverage most on ANN advantages
- Quantifying information content of changes in PD very interesting:
  E.g. PD change $\iff$ change in structural parameters?
- Discuss high-dim. Cholesky decomposition in more detail
- Cross-validation, training & test setup not clear, especially sample splitting strategy. Watch out for information leakage (data standardisation)
- Suggest using bagged model, instead of single "best" model & calculate PD within ensemble instead of separate bootstrap
- Clearer presentation of results, e.g. NO 3-D plots

Overall, very interesting approach with lots of potential applications. Thanks!

References i

Joseph, A. (2019). Shapley regressions: A universal framework for statistical inference on machine learning models. *Bank of England Staff Working Paper Series*, (784).

Ribeiro, M., Singh, S., and Guestrin, C. (2016). "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD*, pages 1135–1144.

Shrikumar, A., Greenside, P., and Anshul, K. (2017). Learning important features through propagating activation differences. *CoRR*, abs/1704.02685.

Strumbelj, E. and Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18.

Sundararajan, M., Taly, A., and Yan, Q. (2017). Axiomatic attribution for deep networks. *CoRR*, abs/1703.01365.