A neural network ensemble approach for GDP forecasting

Longo L., Riccaboni M., Rungi A.,

IMT School for Advanced Studies, Lucca

Article published in the *Journal of Economic Dynamics and Control*:
(https://www.sciencedirect.com/science/article/pii/S016518892100213X?dgcid=author)

Introduction

- We propose an ensemble learning methodology to forecast the future US GDP growth release. Our approach combines a Recurrent Neural Network (RNN) [2] [3] and a Dynamic Factor model [1] accounting for time-variation in the mean with a Generalized Autoregressive Score [5] (DFM-GAS).
- The primary motivation is to combine machine learning models with metrics-based approaches to deal with complexity in the forecasting exercise of the economic activity, primarily during a crisis.
- We find that combining neural networks with dynamic factor models pays off in estimating the future US GDP release.

Methods

The forecasting method combines RNN and DFM-GAS prediction with time-varying weights.

- DFM-GAS:

$$\hat{y}_{t+1}^{DFM-GAS} = \hat{\alpha}_t + \hat{\beta}'\hat{f}_t \qquad (1)$$

$\hat{f}_t$ is the vector of estimated factor from principal component analysis of a vector of macroeconomic indicators x_t, while α_t follows a GAS dynamics, depending on previous score s_{t-1}:

$$\hat{\alpha}_t = \hat{\gamma} s_{t-1} + \hat{\alpha}_{t-1} \qquad (2)$$

- RNN:

$$\hat{y}_{t+1} = \phi\left(h_t \cdot W_x + h_{t-1} \cdot W_y + b\right) \qquad (3)$$

h_t is a function of x_t and h_{t-1}, which is a function of x_{t-1} and h_{t-2} and so on.

- Ensemble: weights are based on the inverse of the mean squared error and defined as:

$$W^{*}_{M_1,t+1} = \frac{\frac{1}{MSE^{*}_{M_1,t+1}}}{\frac{1}{MSE^{*}_{M_1,t+1}} + \frac{1}{MSE^{*}_{M_2,t+1}}} \qquad (4)$$

The final forecast is:

$$\hat{y}_{t+1} = W_{DFM-GAS} * \hat{y}_{t+1}^{DFM-GAS} + W_{RNN} * \hat{y}_{t+1}^{RNN} \qquad (5)$$

Data

We use the data from [1] available on FRED database, for a total of 138 over 228 variables to generate version 1 of the model. The main analysis is on the timeline 1970Q1-2020Q1. The COVID crisis is added up to 2021Q1.

Model's inputs (version 1)

Groups of variables	Frequency	N. of variables
Labor markets	*M*	19
Manufacturing	*M*	7
Monetary	*M*	4
Reserves/federal surplus	*M*	3
Banking	*M*	5
Capacity/ industrial	*M*	22
Housing	*M*	5
Sales	*M*	11
CPI/PPI	*M*	21
Income and consumption	*M*	10
Interest rate and bonds	*M*	21
Trade	*M*	4
Other indicators	*M*	6

In the version 2, higher frequency indicators are added to version 1.

High-frequency indicators (version 2)

Groups of variables	Frequency	Missing values at the beginning	N. of variables
Initial claims	*W*	False	1
Assets/liabilities	*W*	False	8
Nasdaq stock market	*D*	True	10
Financial stress index	*W*	True	1
Uncertainty	*D,W,M,Q*	True	16
Exchange rates	*D*	False	3

Results

	RMSE			
	h = 1	h = 2	h = 3	h = 4
Random walk Model	1.007	1.007	1.006	1.005
Vector Autoregressive Model, VAR	0.926**	0.940	1.020	1.009
Dynamic Factor Model, DFM	1.065	1.084	1.065*	1.042
t.v. Dynamic Factor Model, DFM-GAS	0.922	0.975	0.966	0.908
XGBoost	0.897	0.917	0.950	0.937**
Random forest	0.928	0.940	0.935**	0.981
Long short-term Memory, LSTM	0.840*	**0.860***	**0.879**	0.872
Long short-term Memory, LSTM (dataset 2)	0.862*	0.869*	0.930	0.940
Ensemble LSTM/DFM-GAS	0.809**	0.870*	0.881	0.867
Ensemble LSTM/DFM-GAS (equal weights)	0.850*	0.900	0.890	**0.860**
Ensemble LSTM/VAR	0.867	0.863*	0.885	0.868
Ensemble LSTM/DFM-GAS (dataset 2)	**0.808****	0.888	0.931	0.932

Table 1: Forecast comparison in terms of root mean squared error (RMSE) normalized to AR(p). We report significance level of Diebold and Mariano test ($p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***).

The data are split in the first 70% - in-sample that serves to train and validate the models and corresponds to 1970Q1-2005Q1 - and the last 30% - out-of-sample used to test the estimated/validated models and corresponds to 2005Q2-2020Q1. There is a **trade-off** between model complexity and forecast gains: for a short-term forecast the combination of the models in the ensemble improves the performance of the single models, and the Diebold-Mariano test shows that these improvements are significant when we consider LSTM and DFM-GAS.

A fluctuation test [4] is computed over the out-of-sample window (2005Q2-2020Q1) to evaluate significant differences between compelling forecasts.

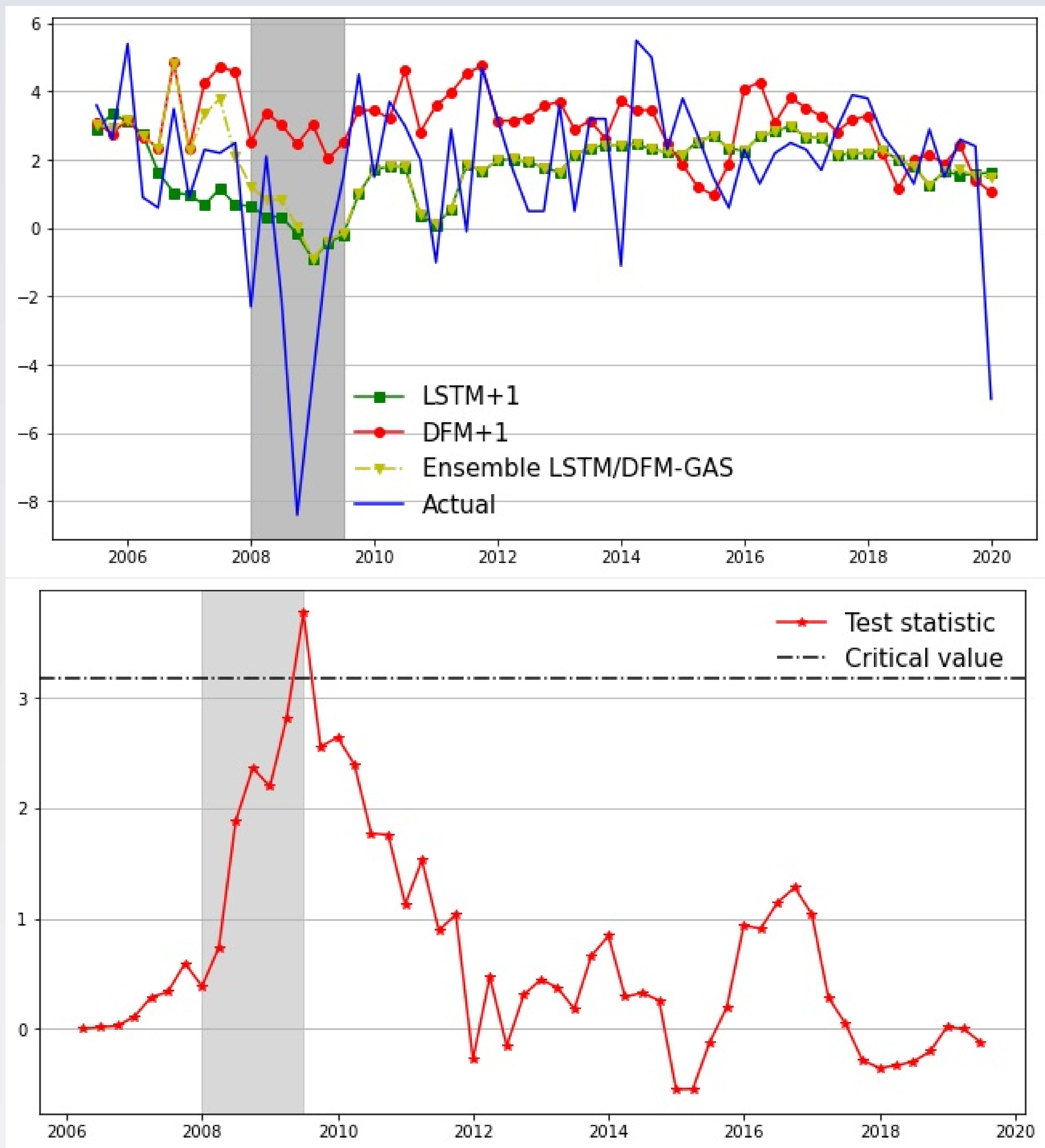

Figure 1: US GDP growth rate one quarter ahead forecast (top panel); Fluctuation test statistics for h= 1, comparing DFM-GAS and ensemble LSTM/DFM-GAS. Significance level of 0.05

The fluctuation test shows that the ensemble model significantly outperforms the DFM-GAS during the 2008-09 Great Recession. In figure 1 we show that most of the forecast gains come from the RNN component of the ensemble during the crisis.

Results under COVID-19

The results for the sample with COVID crisis show that the machine learning component allows to better catch the economic rebound in the US starting in the third quarter of 2020.

Model interpretation

To make the results of the RNN interpretable, we use integrated gradients (IGs). An interpretable forecasting routine is essential in understanding the role of the features over the out-of-sample window.

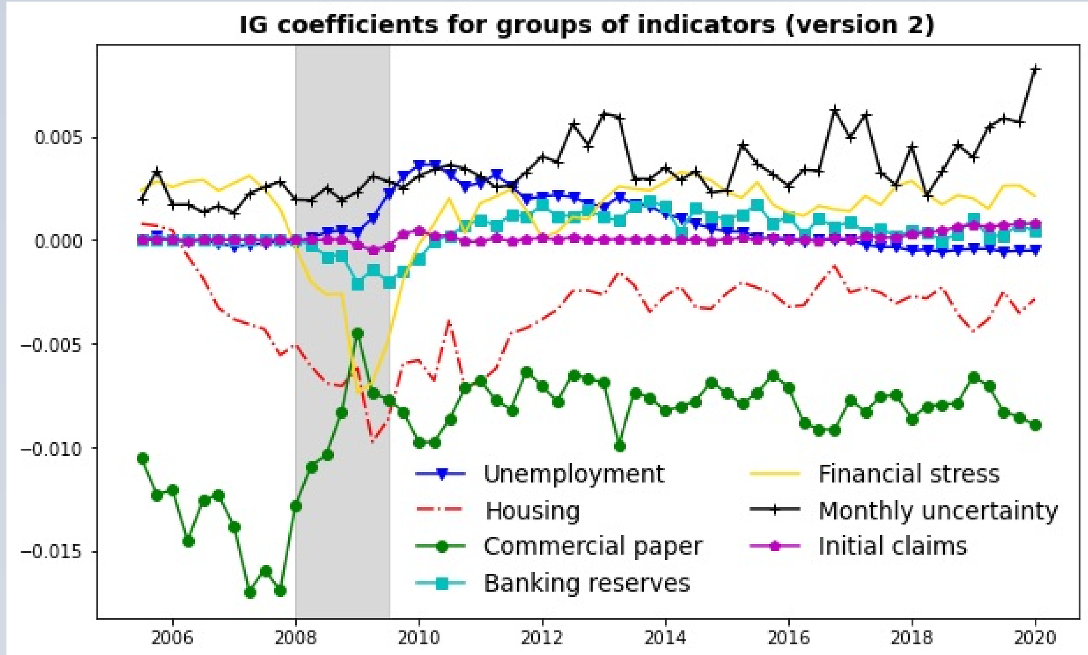

Figure 2

The IG coefficients should be interpreted as features or group of features contribution to the prediction of the US GDP at a specific point in time. A high coefficient means that the feature has a significant impact on the predicted variable.

Conclusions

- An approach encompassing a metrics-based and a machine learning component allows us to better catch the fluctuations of business cycles, especially in the case of structural breaks, when in-sample information has a more limited predictive power.
- During economic recessions and fast recoveries, the data generating process is rapid enough that the analyst cannot be aware of neither the intensity nor the impact of the (positive or negative) shock on the business cycle, be it permanent or temporary. In this context, a highly non-linear model like LSTM combined with a DFM-GAS helps predicting sudden mean-shifting in macroeconomic indicators.
- The ensemble is especially effective in the short-term because it is able to catch shifts in the mean of the GDP growth rate that actually last for no more than one period, as was the case of the 2008-09 crisis.
- Integrated gradients are important to evaluate the role of different features in forecasts, and how they vary along the entire timeline.

References

[1] Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: the real-time informational content of macroeconomic data. J Monet Econ 55 (4), 665–676.

[2] Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., et al., 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[3] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput 9 (8), 1735–1780.

[4] Giacomini, R., Rossi, B., 2010. Forecast comparisons in unstable environments. Journal of Applied Econometrics 25 (4), 595–620.

[5] Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28 (5), 777–795.

Contact

Luigi Longo (corresponding author). Email: *luigi.longo@imtlucca.it*