Daniel Hopp

Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM)

*Daniel Hopp[1]
[1] UNCTAD; Geneva, Switzerland; daniel.hopp@un.org

Abstract:
Artificial neural networks (ANNs) have been the catalyst to numerous advances in a variety of fields and disciplines in recent years. Their impact on economics, however, has been comparatively muted. One type of ANN, the long short-term memory network (LSTM), is particularly wellsuited to deal with economic time-series. Here, the architecture's performance and characteristics are evaluated in comparison with the dynamic factor model (DFM), currently a popular choice in the field of economic nowcasting. LSTMs are found to produce superior results to DFMs in the nowcasting of three separate variables; global merchandise export values and volumes, and global services exports. Further advantages include their ability to handle large numbers of input features in a variety of time frequencies. A disadvantage is the inability to ascribe contributions of input features to model outputs, common to all ANNs. In order to facilitate continued applied research of the methodology by avoiding the need for any knowledge of deep-learning libraries, an accompanying Python library was developed using PyTorch: https://pypi.org/project/nowcast-lstm/.

Keywords:
nowcasting; economic forecast; machine learning; neural networks; python

1. Introduction:

A defining feature of the 21st century so far has been the explosion in both the volumes and varieties of data generated and stored (Domo, 2017). Simultaneously, rapid advancements in machine learning methods have been made, spurred on in part by the need for novel methods to analyze these new data quantities. Perhaps no methodology has gained greater prominence than the artificial neural network (ANN). ANNs are the engine behind tremendous leaps in fields as disparate as machine translation, image recognition, recommendation engines and even self-driving vehicles. Yet to date, their impact in the field of economic policy has been largely muted or exploratory in nature (Falat and Pancikova, 2015).

ANNs have been applied to economic nowcasting in the past (Loermann and Maas, 2019). However, due to the time-series nature of many economic nowcasting applications, the long short-term memory (LSTM) architecture is better suited to the problem than the traditional feedforward architecture explored in Loermann and Maas (2019). LSTMs are an extension of recurrent neural network (RNN) architecture, which introduces a temporal component to ANNs. LSTMs have been used to nowcast meteorological events (Shi et al., 2015) as well as GDP (Kurihara and Fukushima, 2019).

However, use of LSTMs in nowcasting economic variables remains in its infancy, perhaps partly due to high barriers to their implementation. Many common deep learning frameworks, including Keras and PyTorch, include provisions for LSTMs. However, the implementations are general and require knowledge of the frameworks to successfully implement. As such, Python and R libraries focused on using LSTMs for economic nowcasting have been published alongside Hopp (2021) (Hopp, 2020). Hopefully, more accessible libraries will help stimulate interest and expand the applications of these powerful models.

2. Methodology:

In order to assess the performance of LSTMs in economic nowcasting, three target variables were used: global merchandise exports in both value (WTO, 2020) and volume (UNCTAD, 2020), and global services trade (UNCTAD, 2020). To have a point of comparison, the LSTM's performance was compared with that of the DFM, as detailed in Cantu (2018). The target series are all quarterly. A large pool of 116 mixed-frequency monthly and quarterly economic series was used to estimate each of the target series. More information on these series is available in Hopp (2021). All series were converted to seasonally adjusted growth rates using the US Census Bureau's X13-ARIMA-SEATS methodology (USCB, 2017). For more technical details on ANNs and the LSTM architecture applied, see section 3 of Hopp (2021).

Hyper-parameter tuning of the LSTM and model performance was evaluated using a training set dating from the second quarter of 2005 to the third quarter of 2016. The test set dated from the fourth quarter of 2016 to the fourth quarter of 2019.

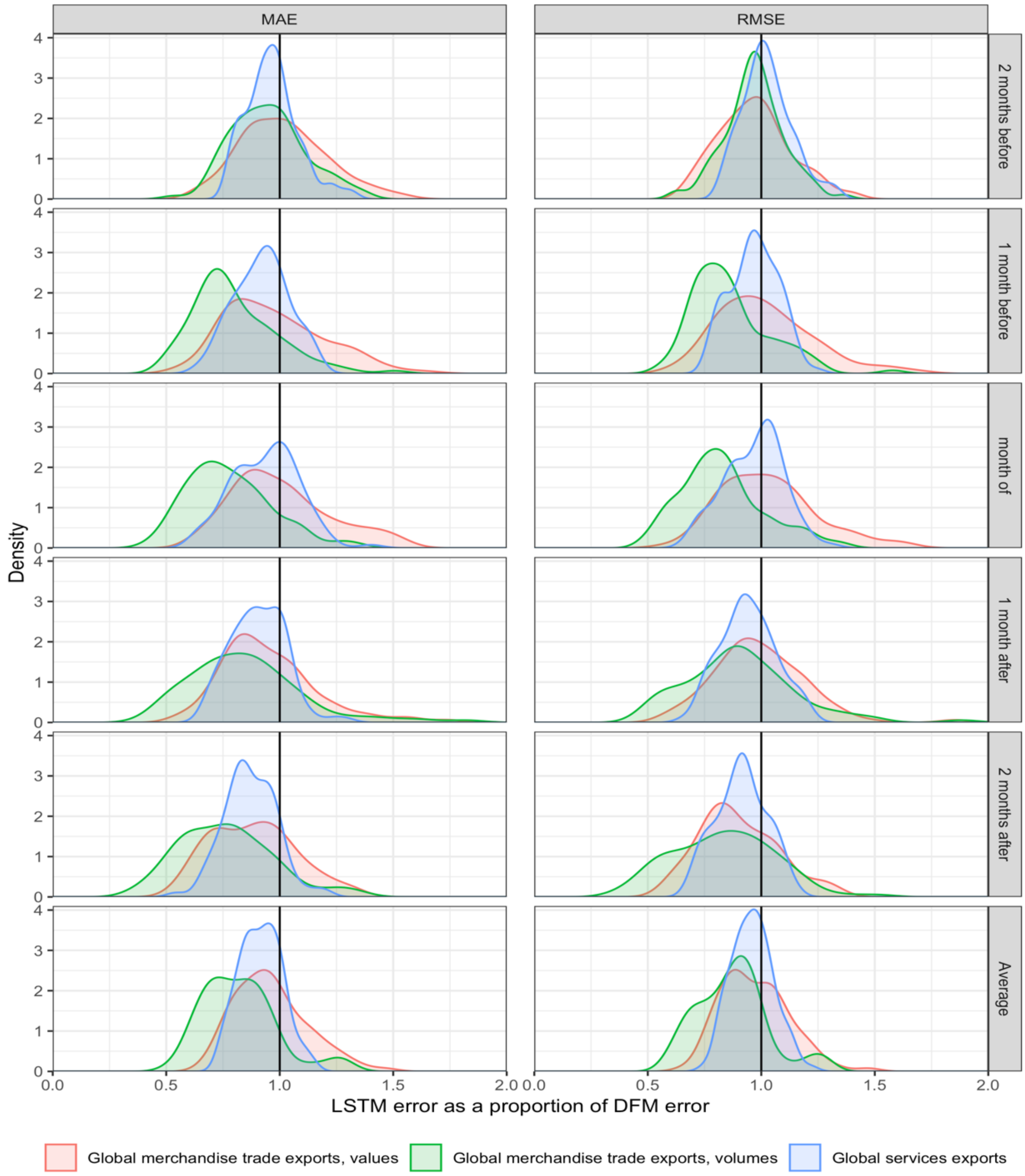
A pool of independent variables was used to ensure the robustness of results, as either model could perform better on a single set of features due to chance. As such, the models' performance was evaluated by taking random samples of between five and 20 features, then fitting both an LSTM and DFM model on this same sample. Both methods' performance was then evaluated on the test set via mean absolute error (MAE) and root-mean-square error (RMSE) on five different data vintages, repeating the process 100 times for each of the three target variables. In this manner, a distribution of relative performance over a wide breadth of independent variables could be obtained. The number of features was restricted to a maximum of 20 due to the high computational time of estimating DFMs with more than this number.

Data vintages in this case refer to the artificial withholding of data to simulate what the availability of data would have looked like at different points in the past. This is important in evaluating model performance in the nowcasting context, as in real life series have varying publication schedules which nowcasting models must be robust to. The five vintages simulated were: two months before the target period, e.g. if the target was the second quarter of 2019, the data as it would have appeared in April 2019; one month before; the month of; a month afterwards; and two months afterwards. The model continues to be evaluated even after the target period has theoretically "passed" as data continue to be published for a given month well after it has passed, depending on the series' individual publication schedule. For example, two months after the second quarter of 2019 simulates being in August 2019, when much more data on the second quarter is available. The variables' publication lags were obtained based on empirical observations from the period from April to November 2020.

3. Result:

Figure 1. shows the distribution of the LSTM's error as a proportion of the DFM's for each target variable. The columns indicate two different performance metrics, mean absolute error (MAE) and root mean square error (RMSE), while the rows indicate different data vintages. A value less than one for an individual model indicates better performance on the test set for the LSTM, while a value greater than one indicates worse performance. Consequently, a distribution centered around one, i.e. the vertical line, indicates comparable performance between the two models, while one to the left of the vertical line indicates better performance on average for the LSTM model.

Figure 1. LSTM error as a proportion of DFM error



The results clearly favor the LSTM model, obtaining better average performance for both performance metrics across all data vintages and target variables, with the sole exception of RMSE for the two months before services exports vintage. Tables 1, 2 and 3 display the average performance metrics for the two models over the sample of 100 different feature combinations, as well as the results using a simple autoregressive model as a benchmark. A one-tailed t-test was performed on the LSTM and DFM errors to ascertain the significance of these differences in performance, with the alternative hypothesis that the LSTM errors were lower. Results are displayed in the LSTM columns.

Table 1. Average performance metrics, global merchandise trade exports, values

Vintage	ARMA MAE	LSTM MAE	DFM MAE	ARMA RMSE	LSTM RMSE	DFM RMSE
2 months before	0.0177	0.0149	0.0150	0.0233	0.0176**	0.0185
1 month before	0.0177	0.0112*	0.0117	0.0233	0.014	0.0141
month of	0.0177	0.0115	0.0118	0.0233	0.0142	0.0142
1 month after	0.0168	0.0108***	0.0117	0.0217	0.0138	0.0142
2 months after	0.0168	0.0094***	0.0109	0.0217	0.0119***	0.0135
Average	0.0173	0.0115**	0.0122	0.0227	0.0143*	0.0149

*Note: *p<.05 **p<.01 ***p<.001*

Table 2. Average performance metrics, global merchandise trade exports, volumes

Vintage	ARMA MAE	LSTM MAE	DFM MAE	ARMA RMSE	LSTM RMSE	DFM RMSE
2 months before	0.0085	0.006**	0.0064	0.0097	0.0075**	0.0078
1 month before	0.0085	0.0051***	0.0066	0.0097	0.0066***	0.0079
month of	0.0085	0.0049***	0.0065	0.0097	0.0063***	0.0079
1 month after	0.0084	0.0045***	0.0057	0.0108	0.0059***	0.0069
2 months after	0.0084	0.0042***	0.0056	0.0108	0.0054***	0.0067
Average	0.0085	0.0049***	0.0062	0.0101	0.0063***	0.0074

*Note: *p<.05 **p<.01 ***p<.001*

Table 3. Average performance metrics, global services exports

Vintage	ARMA MAE	LSTM MAE	DFM MAE	ARMA RMSE	LSTM RMSE	DFM RMSE
2 months before	0.0119	0.0123***	0.0129	0.0151	0.0154	0.0152
1 month before	0.0119	0.0103***	0.0113	0.0151	0.0135**	0.0140
month of	0.0119	0.0103***	0.0111	0.0151	0.0135**	0.0141
1 month after	0.0119	0.0103***	0.0115	0.0151	0.0137***	0.0146
2 months after	0.0119	0.0101***	0.0117	0.0151	0.0134***	0.0147
Average	0.0119	0.0107***	0.0117	0.0151	0.0139***	0.0145

*Note: *p<.05 **p<.01 ***p<.001*

4. Discussion and Conclusion:

The fact that the LSTM performed better than the DFM on average for all three target variables across almost all vintages and both performance metrics is strong evidence for their relevance in the economic nowcasting space. Of course, it does not indicate that LSTMs are superior to DFMs in every instance. The results rather provide some evidence that they can be a competitive alternative to DFMs and have the potential to become a more commonly used methodology in nowcasting.

One of the pros relative to the DFM independent of predictive accuracy is LSTMs' ability to handle many more features than the DFM before coming up against computational bottlenecks. This could be beneficial by lessening the need for variable selection in the early stages of an analysis, easing the obtainment of initial results. Additionally, a model is able to be reliably trained on any given set of features and values, which is not the case for the DFM, the training of which may fail if input matrices are non-invertible.

Another advantage compared with the DFM is the ability to easily use mixed-frequency variables with no corresponding change in the underlying modeling and formulas. Annual, quarterly, monthly, and even theoretically daily data can be combined in a single model just by changing the structure or frequency of the input data, explained in more detail in section 3.2 of Hopp (2021).

Disadvantages of the LSTM approach include the stochastic nature of ANNs and the lack of interpretability of their coefficients. Due to the initial randomization of weights in ANNs, two different models trained on the same data will produce two different predictions. This problem can be mitigated by taking the average predictions of several networks. The “black box” nature of ANNs is well-known, however there is opportunity for further research into applying existing ANN interpretability methods, such as activation maximization or sensitivity analysis (Montavon et al., 2018), to the nowcasting LSTM framework.

The events of 2020 have shown the value of timely, accurate estimates of macroeconomic series to help inform policy decisions. The findings here provide evidence for stronger consideration of LSTMs for this purpose. LSTMs were shown to produce superior predictions compared with DFMs on three different target series: global merchandise trade exports expressed in both values and volumes and global services exports, and over five different data vintages.

In addition to better empirical performance for the three target series, LSTMs provide advantages over DFMs by being able to handle large numbers of features without computational bottlenecks, not relying on the invertibility of any matrices, thus being able to be fit on any dataset, and the ability to use any mixture of frequencies in features or target. Disadvantages relative to DFMs include LSTMs’ stochastic nature, the lack of interpretability in their coefficients, and opacity regarding feature contribution to predictions.

The *nowcast_lstm* library can facilitate the use of LSTMs in economic nowcasting by lowering the barrier to experimentation. LSTMs’ ability to reliably generate predictions on a large number of input features makes it easier to quickly verify whether or not a given series has the potential to be nowcast, a characteristic that could help expand the variety and quantity of economic variables monitored via nowcasting.

There remains much scope for future research and development on this topic. Further testing should be performed to verify LSTMs’ performance on a wider variety of series and frequency mixtures. More hyper-parameter tuning could be performed to see if tweaking other aspects of model architecture could result in even better results. There is also much scope for exploring different methods of filling missing values beyond ARMA or mean-filling. Finally, methods for interpreting LSTMs and ascertaining feature contribution to predictions would increase the method’s viability as a policy-informing instrument. The library could then be extended in the future to incorporate any improvements to performance or functionality deriving from future research, continuing to facilitate the adoption and development of the methodology in the nowcasting domain.

References:

Cantu F (2018). Estimation of a coincident indicator for international trade and global economic activity. UNCTAD Research Paper No. 27.

Domo (2017). Data Never Sleeps 5.0. Available at https://www.domo.com/learn/data-never-sleeps-5 (accessed 1 September 2021).

Falat L and Pancikova L (2015). Quantitative Modelling in Economics with Advanced Artificial Neural Networks. *Procedia Economics and Finance*. 34194–201.

Hopp D (2020). nowcast_lstm. Available at https://github.com/dhopp1/nowcast_lstm.

Hopp D (2021). Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM). UNCTAD Research Paper No. 62. UNCTAD.

Kurihara Y and Fukushima A (2019). AR Model or Machine Learning for Forecasting GDP and Consumer Price for G7 Countries. *Applied Economics and Finance*. 61.

Loermann J and Maas B (2019). Nowcasting US GDP with artificial neural networks. MPRA Paper No. 95459. University Library of Munich, Germany.

Montavon G, Samek W and Müller K-R (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*. 731–15.

Shi X et al. (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*.

UNCTAD (2020). UNCTADStat. Available at https://unctadstat.unctad.org/EN/Index.html (accessed 11 January 2021).

USCB (2017). The X-13ARIMA-SEATS Seasonal Adjustment Program. Available at https://www.census.gov/srd/www/x13as/.

WTO (2020). Statistics on merchandise trade. Available at https://www.wto.org/english/res_e/statis_e/merch_trade_stat_e.htm (accessed 1 November 2020).