UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

November 2021

UNCTAD Research Paper No. 74
UNCTAD/SER.RP/2021/17

Performance of LSTM Neural Networks in Nowcasting during the COVID-19 Crisis

Daniel Hopp
Associate Statistician
Division on Globalisation and Development Strategies, UNCTAD
daniel.hopp@unctad.org

Abstract

The COVID-19 pandemic has demonstrated the increasing need of policymakers for timely estimates of macroeconomic variables. A prior UNCTAD research paper examined the suitability of long short-term memory artificial neural networks (LSTM) for performing economic nowcasting of this nature. Here, the LSTM's performance during the COVID-19 pandemic is compared and contrasted with that of the dynamic factor model (DFM), a commonly used methodology in the field. Three separate variables, global merchandise export values and volumes and global services exports, were nowcast with actual data vintages and performance evaluated for the second, third, and fourth quarters of 2020 and the first and second quarters of 2021. In terms of both mean absolute error and root mean square error, the LSTM obtained better performance in two-thirds of variable/quarter combinations, as well as displayed more gradual forecast evolutions with more consistent narratives and smaller revisions. Additionally, a methodology to introduce interpretability to LSTMs is introduced and made available in the accompanying *nowcast_lstm* Python library, which is now also available in R, MATLAB, and Julia.

Key words: Nowcasting, Economic forecast, Neural networks, Machine learning, Python, R, MATLAB, Julia, LSTM, COVID

Contents

Acknowledgements

I would like to thank Anu Peltola her valuable comments and feedback.

1. Introduction

The COVID-19 pandemic wrought havoc on the global economy in 2020. In contrast with other economic crises, such as the 2008 financial crisis, there were not primarily macroeconomic factors at play, but rather epidemiological ones. As the threat of contagion forced innumerous business closures, especially in the service and tourism sector (UN, 2020), economic contraction followed. In order to combat these events, unprecedented in modern times, many governments implemented extensive stimulus measures to help people through the crisis. In the months following initial widespread global closures in March 2020, the importance of timely information on the state of national economies and the global economy became essential in quickly assessing both the impact of existing policy measures, as well in guiding future ones. The months long publication delays typical of many macroeconomic series, especially globally aggregated ones, such as GDP or international trade, were rendered even more of a barrier for guiding policy during such a quickly developing crisis (Gerhard et al., 2021).

In this scenario, nowcasting, the estimation of the current or near-current state of a target variable using information that is available more quickly, could be an essential tool in gaining insight to the COVID-19 pandemic's effect on the global economy. The COVID-19 pandemic proved a stress-test for existing nowcasting models, most having never before been confronted with such an extreme and dynamic crisis. These circumstances make 2020 a particularly interesting case in which to examine the performance of different nowcasting methodologies. This paper seeks to do just that, assessing two methodologies, the dynamic factor model (DFM), currently a popular choice in economic nowcasting, and the long short-term memory neural network (LSTM), explored in-depth in Hopp (2021).

Additionally, the dynamic economic situation naturally leads to much larger revisions in model predictions over time than would be expected in normal economic circumstances. This increases the value of causal inference into what is driving the change in a model's predictions. To that end, this paper also explores a methodology to introduce such causal inference to the outputs of the LSTM. This functionality has been added to the *nowcast_lstm* Python library, which is discussed in the relevant section 4.1. Finally, in order to further increase accessibility to the use of LSTMs in economic nowcasting, wrappers for R, MATLAB, and Julia for the *nowcast_lstm* library have been introduced, enabling the use of library from these languages without the need for Python knowledge. More information is available from the following locations:

- R: https://github.com/dhopp1/nowcastLSTM
- MATLAB: https://github.com/dhopp1/nowcast_lstm_matlab
- Julia: https://github.com/dhopp1/NowcastLSTM.jl

The remainder of this paper is structured as follows: the next section will provide more background information on nowcasting, including during the COVID-19 pandemic, and the LSTM methodology; section three will examine the relative performance of DFMs and LSTMs in nowcasting three series during the pandemic: global merchandise trade exports expressed in both values and volumes and global services exports; section four will introduce and examine a methodology for introducing causal inference to LSTM predictions, as well as introduce the wrappers for the *nowcast_lstm* library; section five will conclude and examine areas of further research.

2. Background

2.1 Nowcasting in the context of the COVID-19 pandemic

Nowcasting is the forecasting of the current or near-current value of a variable, often using information that is published or made available more quickly than the variable of interest. Some commonly nowcasted series include GDP (Morgado et al., 2007; Giannone et al., 2009; Rossiter, 2010) and international trade (Cantú, 2018; Guichard and Rusticelli, 2011). These types of aggregated macroeconomic variables lend themselves well to the nowcasting paradigm, as they are often published later than some other economic indicators while still being of great interest to policymakers, investors, and firms. Some common methodologies to perform economic nowcasting include mixed data sampling (MIDAS) (Kuzin et al., 2009; Marcellino and Schumacher, 2010), dynamic factor models (DFM) (Guichard and Rusticelli, 2011; Corona et al., 2021), mixed-frequency vector autoregression (VAR) (Kuzin et al., 2009; Huber et al., 2020), and Bayesian vector autoregressions (Cimadomo et al., 2020). Hopp (2021) and Loermann and Maas (2019) examined neural networks' suitability to the application, more specifically long short-term memory (LSTM) networks in the case of the former. The LSTM methodology is explained further in section 2.2. For more information on nowcasting, including commentary on common data issues encountered in the field, see Hopp (2021) or Cimadomo et al. (2020).

Nowcasting became more relevant than ever in the wake of the economic fallout from the COVID-19 pandemic. Since March 2020, when many governments around the world began shutting down businesses and other forms of economic activity in response to the virus, transforming the crisis into a global one, the rate of change in the economy has been truly unprecedented (The World Bank, 2020). Furthermore, the epidemiological nature of the crisis and successive COVID-19 waves have meant that the economic recovery has not been one of monotonic recovery, as governments have often had to roll back and reinstate openings in response to the severity of local and national outbreaks. This has simultaneously increased the need for accurate, timely assessments of the economic situation to inform policy and mitigate economic impact on citizens, while making those assessments harder to acquire.

However, crisis often creates opportunity and breeds innovation, and the field of nowcasting has been no different. A wealth of papers relating to nowcasting during the COVID-19 pandemic have been published since March 2020. Many geographies are represented, including Canada (Chapman and Desai, 2021), Sub-Saharan Africa (Buell et al., 2021), the United States (Foroni et al., 2020), Mexico (Corona et al., 2021), and the Euro area (Huber et al., 2020), among others. Perhaps more interestingly, novel data sources have additionally been explored, for instance Google mobility data (Sampi and Jooste, 2020), retail payment system data (Chapman and Desai, 2021), Google search trends, and mobile payment data (Buell et al., 2021). Unfortunately, the longevity of the COVID-19 crisis to this point ensures that nowcasting its effects on the economy will remain fertile ground for new research in the coming months and years.

2.2 Long short-term memory neural networks

Having established the context in which the nowcasting exercise outlined in this paper takes place, this section will give a short background on the methodology employed. Artificial neural networks (ANNs) have risen in prominence in recent years due to their impressive performance in a variety of applications, including things like image classification and natural language processing. However, traditional feed-forward networks lack a temporal component, a frequent feature of many economic applications. The long short-term memory network architecture (LSTM) adds this component and renders them more suitable for application in the nowcasting context. For more information on how ANNs and LSTMs work, see: Hopp (2021), Singh and Prajneshu (2008), Sazli (2006), or Loermann and Maas (2019). For more detailed information on LSTMs' properties which make it suitable for nowcasting, see Hopp (2021), section 3.2.

3. Empirical analysis

3.1 Description of data and models

Hopp (2021) examined the LSTM's performance versus that of dynamic factor models (DFM) in nowcasting global merchandise and services trade. In that case, LSTMs were found to produce superior predictions. However, the test period was the fourth quarter of 2016 to the fourth quarter of 2019, a period when the target series' movements were much more muted than compared with 2020 and 2021. Furthermore, test performance was found using artificially simulated data vintages based on historical publication lags. The analysis performed in this paper seeks to build on those findings and further validate and stress test them with: A) a much more volatile and difficult to predict in context, and B) actual data vintages collected over the course of 2020 and 2021.

In this analysis, three target variables were again nowcast: global merchandise exports in both value (WTO, 2020) and volume (UNCTAD, 2021), and global services exports (UNCTAD, 2021). These are the same series examined in Hopp (2021). All target series were expressed in seasonally adjusted quarter over quarter growth rates. In total, 45 independent variables were used as inputs to estimate both a DFM and LSTM model for each target series: 17 for merchandise exports values, 17 for merchandise exports volumes, and 21 for services exports. Variables were sometimes used to estimate more than one target series. Input variables included things such as industrial production indices, manufacturing export order books, and retail trade indices, among others. See appendix 1 for a full list of input variables, including their geographies, frequencies, sources, and for which target series they were used. The same variables were used in estimating both the DFM and LSTM models to ensure maximum comparability. Input variables were a mix of monthly and quarterly frequencies expressed in period over period seasonally adjusted growth rates.

The DFM and LSTM models were trained on data dating from the second quarter of 2005 to the fourth quarter of 2019, representing the maximum extent of information a forecaster or policymaker would have had in the run up to the COVID-19 pandemic. Actual data vintages collected over the period from March 2020 to October 2021 were then used to assess model performance in nowcasting the target series from the second quarter of 2020 to the second quarter of 2021, an exceptionally volatile and difficult period to nowcast due to the unprecedented impacts on the global economy of the COVID-19 pandemic. Actual data vintages were collected on a monthly basis from March to July 2020, then on a weekly basis from August 2020 to October 2021.

The LSTM model used was the same examined in Hopp (2021), using the averaged output of 10 networks. For the logic of using the average of multiple networks' outputs, see Hopp (2021) sections 4.1 and 5, or Stock and Watson (2004). Hyperparameters were found by using the period from the second quarter of 2005 to the third quarter of 2016 as a training period, and the fourth quarter of 2016 to the fourth quarter of 2019 as a test period. Ragged edges were filled using the mean of each series, see Hopp (2021) section 3.2 for more information.

The DFM model used was that described in Cantú (2018), where a state-space representation is used to model the DFM under the assumption that the target and independent variables share a common underlying factor, as well as containing their own idiosyncratic component. Subsequently, the Kalman filter is applied and maximum likelihood estimation used to obtain parameter estimates. For more information on this specific DFM methodology, see Bańbura and Rünstler (2011) and Bok et al. (2018).

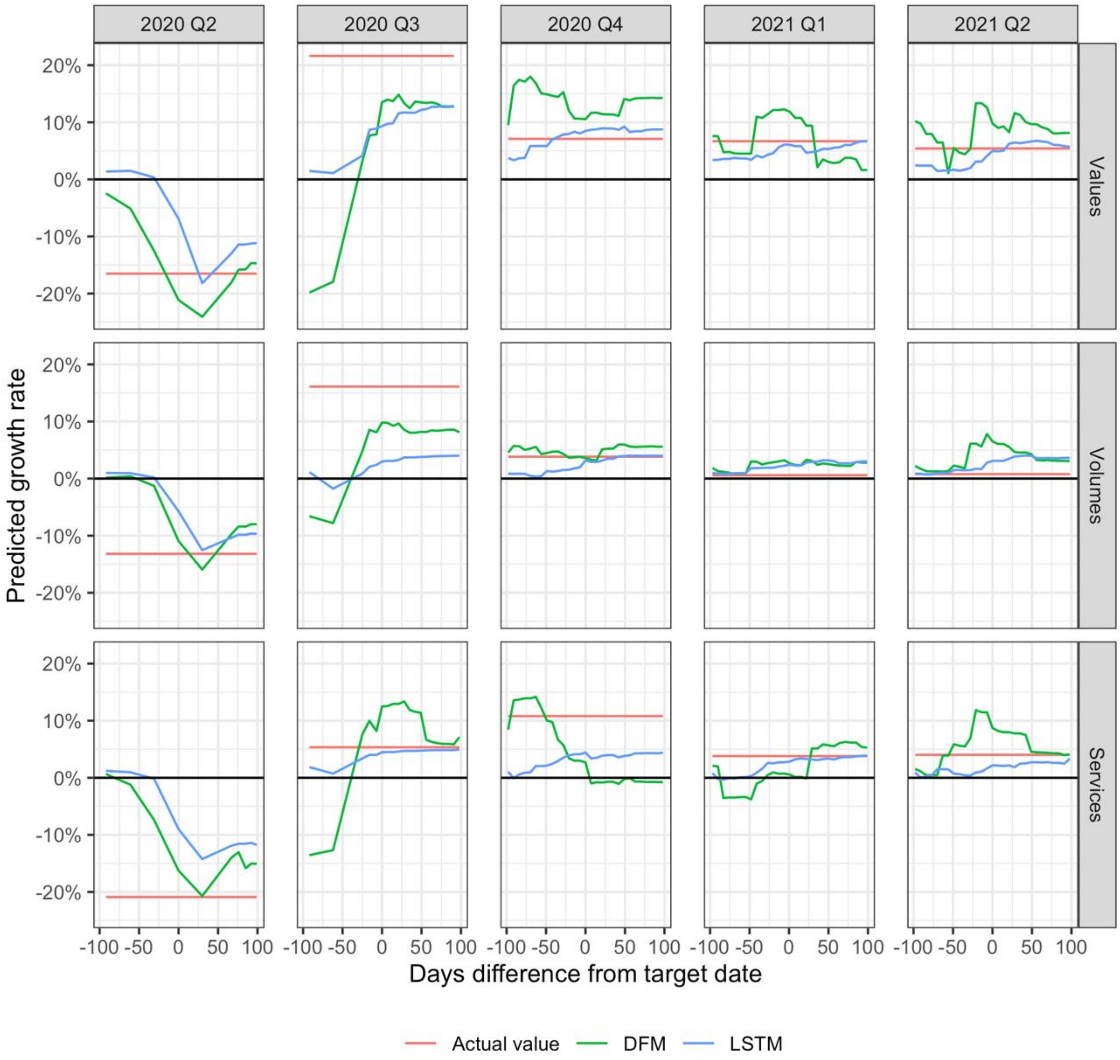
Once DFM and LSTM models were trained for each target series with data up until the fourth quarter of 2019, predictions could be obtained on actual monthly and weekly data vintages to see how the models' forecasts would have developed over time as the pandemic unfolded and its economic repercussions began to appear in the data. In this way, we can see what narratives and guidance the nowcasts would have provided to policy makers and analysts as well as assess their errors over time and final performance.

Predictions were made for each quarter on data vintages dating 100 days either forwards or backwards in time, to assess performance both early on, when little data for the period was available, and later on, when data on most independent series had been published.

3.2 Results

Figure 1 shows the development of the two models' predictions over time for the period from the second quarter of 2020 to the second quarter of 2021. The X axis shows the days difference from the target period. E.g., 0 days difference for 2020 Q2 refers to 1 June 2020, to 1 September for 2020 Q3, etc. The Y axis displays the quarter over quarter growth rate. The red line displays the actual observed growth rate, while the blue and green lines represent the predictions of the LSTM and DFM models, respectively. Each point making up the blue and green lines represents what the two models predicted the growth rate of the target series would be given the data available at that point in time. Generally, the predictions should move closer to the actuals line as time goes on and more data is released.

Figure 1. Nowcast evolution over time



Note: For brevity, "Values" refers to global merchandise exports in values, "Volumes" refers to global merchandise exports in volumes, and "Services" refers to global services exports.

2020 Q2

The first column of figure 1 details predictions for 2020 Q2. This was the first quarter where the full effects of the pandemic were reflected in economic data globally. While China was already experiencing lock downs in the first quarter of 2020, most other places did not until COVID-19 was declared a pandemic by the WHO on 11 March 2020 (WHO, 2020). The first quarter of 2020 was not assessed in this modelling exercise as UNCTAD did not begin the systematic gathering of actual data vintages until after this period had elapsed.

Global merchandise exports expressed in values dropped 16.5 per cent quarter over quarter in the second quarter of 2020. Between 2005 and 2021, this was the second largest decline recorded, second only to the fourth quarter of 2008, during the height of the financial crisis. While the DFM already began to pick up on contraction in March and April, it began severely revising its predictions downwards in May and June (day difference of 0 on the X axis), actually severely overshooting the eventual number by nearly 8 percentage points in July, as negative figures from April and May had been published, but not as many more positive ones from June had been. As data continued to be released through July, August, and September, it revised its predictions upwards before settling quite close to the actual figure in the beginning of September. The LSTM took longer to reflect the downturn, only heavily revising its predictions downwards in June. It displayed a similar shape to the DFM, with steep downward revision followed by upward correction. However, its post-July trough to peak delta was only 7 percentage points, compared with nearly 10 percentage points for the DFM.

Global merchandise exports expressed in volumes dropped by 13.2 per cent in the quarter, representing the largest decline recorded between 2005 and 2021, even greater than declines observed during the financial crisis. In this series, the DFM was slower to pick up on the decline compared with values, only revising predictions strongly downwards in June. Again, it overshot the mark and revised itself upwards after hitting its nadir in July. This time, however, it overshot the mark on the way up as well, and it finished predicting a decline that was only about 60 per cent as large as the actual observed decline. The LSTMs' predictions followed a similar pattern, declining sharply in June and July, then revising upwards afterwards. Its revisions, however, were significantly smaller than the DFMs', with a post-July trough to peak delta of only 3 percentage points, compared with the DFM's of 8 percentage points. Its final predictions also ended closer to the actual value.

Both models had a hard time picking up on the degree of decline for global services exports. Perhaps understandable, considering the series experienced its greatest decline in the second quarter of 2020 in the period from 2005 to 2021, almost doubling the next largest downturn experienced during the global financial crisis. Both models' predictions displayed similar shapes to the merchandise export series, with big revisions downwards followed by corrections. Again, the LSTM displayed smaller corrections, with a post-July trough to peak delta of 3 percentage points compared with the DFM's of 8 percentage points, though the DFM's final prediction was closer to the observed value.

2020 Q3

The third quarter of 2020, represented in the second column of figure 1, experienced strong recovery after astounding contractions in the second quarter. Though recovery had already begun in May and June of 2020, as summer in the northern hemisphere brought about partial economic reopening combined with adaptation to the circumstances of the pandemic, it was visible in earnest in the third quarter.

Global merchandise exports expressed in values ended up growing an impressive 21.6