Forecasting inflation in Latin American countries using a SARIMA–LSTM combination

Rodrigo Peirano · Werner Kristjanpoller · Marcel C. Minutolo

Accepted: 28 June 2021 / Published online: 10 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
Inflation forecasting has been and continues to be an important issue for the world’s economies. Governments, through their central banks, watch closely inflation indicators to make national decisions and policies. This study proposes to forecast the inflation rate in five Latin American emerging economies based on the commonly used seasonal autoregressive integrated moving average (SARIMA) approach combined with long short-term memory (LSTM). Additionally, we run forecasts based on fuzzy inference systems (FISs), artificial neural networks (ANNs), artificial neuro-FIS, and SARIMA ANN as benchmarks to compare the performance of the combines SARIMA–LSTM. The combined SARIMA–LSTM captures the linear aspects of the time series as well as the nonlinear aspects. The results indicate that the proposed model based on the combination of SARIMA and LSTM has higher accuracy in inflation forecasts over the SARIMA and LSTM separately.

Keywords Inflation forecasting · Econometrics models · Artificial intelligence models · Hybrid model · Time-series forecasting

1 Introduction
Inflation forecasting has been and remains an important area of research, due to inflation’s influence on national economies at both macro- and microeconomic levels. Due to its influence on national economies, decision makers pay particular attention to its forecasts to set national policies and agendas that drive economic development. Private firms as well as public policy makers use inflation forecasts as a key variable for decision making given expectations about the direction and rate, for instance whether to move forward or not on a particular project or a bank’s decision to grant or not a specific line of credit (Tinoco-Zermeño et al. 2014). Hence, improving the tools necessary for developing accurate forecasts is of particular concern.

There are essentially two approaches that are used to forecast inflation rate as noted by McKnight et al. (2020). The first approach is driven by theory such as the New Keynesian Phillips Curve (NKPC); the second uses time-series techniques such as autoregressive moving averages. This particular study is concerned with the second type of approach but with econometric modeling as explained in Sect. 3.1. However, we will briefly discuss NKPC approach in Sect. 2.

With respect to the second approach, many time-series econometric models assume that linearity exists between their variables. However, there is evidence that macroeconomic series have a nonlinear behavior due to the impact of shocks at different phases of an economy’s cycle (LeBaron 1994). Hence, forecasts based on the assumption of linearity do not provide sufficiently accurate responses when decision makers are modeling economic policy. To provide a model that captures both the linear nature and nonlinear nature of the time series is the concern of this work.

Given the nonlinear nature of the relationship between economic features and inflation, there has been an increased interest in the use of nonlinear models. For instance, models that use artificial neural networks are one area of interest
for their adaptability, nonlinearity, and free distribution (non-parametric) (Šestanović and Arnerić 2021; Xu et al. 2019); thus, they tend to offer more satisfactory and accurate results. Back-propagation ANN is one particular approach to forecast a macroeconomic series such as inflation (Maasoumi et al. 1994). When comparing a back-propagation ANN to traditional econometric methods, Moshiri and Cameron (2000) found that the ANN model outperformed traditional models with respect to long-term forecasting horizons. However, some studies have found that to the contrary artificial neural networks are not able to improve on an autoregressive model (Tkacz 2001). Hence, there are examples of differing results in the literature that need further investigation.

Another approach that has had promising results is the fuzzy inference system (FIS) model. FIS has provided robust forecasts even when the historical data are inaccurate, allowing for the inclusion of variables in the analysis that have not been accurately formulated. This technique has been applied for seasonality in time-series data (Chang 1997); temperature and humidity measurements from coffee crops (Bacani and de Barros 2017); and to forecast stock price (Vercher et al. 2007). Additionally, hybrid models that take advantages of both ANNs and fuzzy regressions’ strengths, which when combined overcome their limitations, have also been used (Khashei et al. 2008). The results from these machine learning techniques suggest that they can be an effective way of improving forecasting accuracy for inflation.

While there is little research on the use of ANN to forecast inflation in Latin America, there are promising studies of its use in other developed and developing countries. For instance, Işıççok et al. (2020) conducted a technical comparison of ARIMA and ANN to forecast inflation in Turkey and found that the results were similar. Conversely, Estiko and Wahyuddin (2019) found that ANN outperformed ARIMA when forecasting inflation in Indonesia. Finally, in a recent study forecasting inflation of the Euro, Šestanović and Arnerić (2021) found that the Jordan NN outperformed a feedforward NN.

Time-series data have a sequence characteristics that demand specific architectures to model it. In neural networks (NNs), long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) networks is a state-of-the-art architecture in sequence datasets that is widely used in different fields such as text, signals, and finance, among others (Shen et al. 2017; Qi et al. 2017; Young et al. 2017; Manning et al. 2017; Gamboa 2017; Wang et al. 2017). LSTM is a recurrent NN (RNN) (Connor et al. 1994; Graves et al. 2006), which means that it uses past values to adjust the architecture to improve results. There are other types of RNN (Lipton et al. 2015), but they do not perform as well as LSTM in terms of performance.

This paper seeks to contribute to the literature by advancing the study of inflation behavior and forecasting using a combination of two models: seasonal autoregressive integrated moving average (SARIMA) and LSTM. We combine the linear and nonlinear models together, which better capture key patterns in inflation forecasting, adding to the contribution made by the SARIMA and LSTM models. We propose a model that combines the good qualities of SARIMA as a linear model with those of the nonlinear LSTM. The rest of this paper is organized in the following manner: First, previous studies on the inflation of Latin American economies are presented. In “Empirical methods” section, we describe the econometric and machine learning models to be used, as well as the proposed model. In “Data” section, we specify the data from the countries used. In “Results”, we present and discuss the obtained results. Finally, we detail the conclusions of the performed study in “Conclusions” section.

2 Literature on inflation forecasting techniques for Latin American economies

McKnight et al. (2020) note in their work that there are broadly two approaches to the problem of forecasting inflation. The first approach, they claim, is “agnostic” to the theoretical underpinnings that drive inflation. The second approach is based on macroeconomic models with microeconomic foundations. The stream of research that this second line of work is based on focuses on the New Keynesian Phillips Curve (NKPC) (e.g., Galá and Gertler (1999); Atkeson and Ohanian (2001); Adolfson et al. (2007); Cogley and Sbordone (2008); Cai et al. (2019); and McKnight et al. (2020)). In particular, the NKPC is an aggregate inflation equation arising in dynamic stochastic general equilibrium (McKnight et al. 2020). McKnight et al. (2020) do note that while the NKPC approach is based on macroeconomic models with microeconomic foundation, the theoretical underpinnings that drive inflation, they have not performed as well as time-series approaches. In order to improve the NKPC models, McKnight et al. (2020) introduced variation in the trend inflation and a novel procedure using “time-varying” NKPC (TVT-NKPC). The first approach to forecasting that McKnight et al. (2020) mentioned uses various time-series analysis techniques. They claim that the TVT-NKPC captures the underlying theory that drives inflation and thereby enables theory-driven policy decisions. We argue that the times-series approach with machine learning techniques captures the features that drive inflation and therefore allows policy makers to make better informed policy decisions. To this end, the current work falls into the time-series analysis stream of research and extends the performance.

In the study of forecasting inflation for Latin America, various models have been used with varying results. In one study, inflation as well as other economic variables were forecasted for certain Latin American countries using a global
vector autoregressive (GVAR) model (Pesaran et al. 2009). The results of the GVAR were compared with those obtained by univariate autoregressive and random walk models. The findings indicate that the double-averaged GVAR forecasts have better results than the benchmark models, especially for the inflation variable. Another study used a similar approach but apply a Bayesian variant of the GVAR (B-GVAR) models to forecast inflation resulting in better performance in the forecast as compared to the univariate model (Cuaresma and Huber 2016). Broto (2011) also studies inflation for a range of Latin American countries using GARCH models. The author analyzes the benefits provided by considering inflation targets as a monetary policy framework finding favorable performance of inflation targets. There is one study in which the authors examined the predictive ability of SARIMA models applied to inflation in Chile, where the results indicate that the proposed model is superior over other univariate methods (Pincheira and García 2012). While the aforementioned study only considered the context of Chile, we seek to add additional Latin American countries to test the generalizability of the approach for developing countries.

In other studies focusing on only one country, the Phillips curve and a VAR model were used to forecast inflation in Argentina, comparing the two with a univariate model, where the results from the VAR model had better performance than the ARMA model (D’Amato et al. 2008). A similar situation occurs with the research performed by Pincheira and Gatty (2016), where they forecasted Chilean inflation using multivariate and univariate time-series models, augmented with different measures of international inflation, where the results were robust and efficient through a series of sensitivity analysis. In another Chilean application, state-space models were developed to forecast inflation, finding that the traditional models surpass the Phillips curve model with variable parameters for the first period (de Simona 2001). However, the traditional models’ superiority rapidly decreased when lengthening the projection period. Finally, with respect to the use of neural networks applied to the inflation of Latin American countries, Luis and Héctor (2013) implemented different NN models to forecast Mexican inflation, varying the number of hidden layers and the number of neurons in the hidden layers, thus constructing a model of 1 hidden layer and 49 neurons that predicted better than the Bank of Mexico in two of three occasions. In Colombia, Santana (2006) predicted Colombian inflation with various neural network models, showing that neural networks are able to forecast more accurately than econometric methodologies.

While there are some studies that apply various models to the Latin American context, there are none that look at how the combined linear and nonlinear models may improve forecast performance. As mentioned earlier, it is perhaps more important in this context that forecasting accuracy and precision is improved given the overall variability of inflation in the long run. Models that perform well in this context are likely to perform well were economies are more stable. In the next section, we develop our methodological approach.

### 3 Empirical methods

This paper analyzes and compares inflation forecast behavior through the proposal of new hybrid models, linear and nonlinear. The results are compared using the error metric mean square error (MSE), comparing the ARMA, artificial neural network, and the two proposed models. This section describes the methods used in this paper to forecast inflation.

#### 3.1 Econometric modeling

For a long time, Box–Jenkins with moving average linear models methodology was the prevailing technique for inflation forecasting (Liu et al. 1992). This methodology was able to optimally forecast in many studies that used this approach (e.g., Rojas et al. (2008) and Chan (1999)). Box and Jenkins (1976) popularized the use of ARMA (p,q) models through 4 steps (in order): (1) model identification; (2) parameter estimation; (3) model checking; and (4) forecast. This approach provides a theoretical framework and practical rules to determine appropriate values for p and q, specifying that any stationary time series $Y_t$ can be explained both by its autoregressive part (AR) as well as by its moving average part (MA); thus, combining both processes creates the autoregressive moving average process ARMA(p,q), where $p$ is the order of the autoregressive process and $q$ is the order of the moving average process. An ARMA(p,q) process follows the following structure:

$$Y_t = \theta + \sum_{i=1}^{p} \alpha_i * Y_{t-i} + \sum_{j=1}^{q} \beta_j * u_{t-j} + u_t$$  \hspace{1cm} (1)

where $\theta$ is a constant term; $Y_t$ is the time series; $\alpha_i$ represents the $i$th parameter that accompanies the lag $i$; $\beta_j$ is the $j$th parameter that accompanies the residual lag $j$; and, $u_t$ is the residual of time $t$.

The seasonal autoregressive integrated moving average (SARIMA) model extends the autoregressive lags or moving average in $S$ periods, related to the seasonality cycle. For example, in the case of the inflation rate modeling is used a yearly seasonality, ($S=12$), Equ. 2.

$$Y_t = \theta + \sum_{i=1}^{p} \alpha_i * Y_{t-i} + \sum_{j=1}^{q} \beta_j * u_{t-j} + \gamma_S * Y_{t-S} + u_t$$  \hspace{1cm} (2)

where $\gamma_S$ is the coefficient of the seasonality $S$. 

© Springer
3.2 Artificial neural network

Research with artificial neural networks can be traced back to McCulloch and Pitts (1943), who in their study discuss how a neuron functions and developed the first artificial neuron model. Years later, Hebb (1949) demonstrated that artificial neural network learning occurs when certain changes in neurons are activated. Rosenblatt (1958) introduced the perceptron model, which is capable of learning a series of patterns to then recognize other similar ones. This work is based on what is explained by Rumelhart et al. (1985), where they apply the backpropagation neural network (BPN) methodology, in which a neural network is able to learn from the different relationships that exist between input patterns and goal values.

A neural network is defined as a collection of parallel processors connected together in the form of a directed graph (Freeman and Skapura 1991). The network structure is made up of three components: input layer, hidden layers, and output layer. Within a layer, there is a finite amount of artificial neurons that are connected depending on the flow of information.

The most widely used form of ANN is the multilayer perceptron (Rosenblatt 1958), in which a set of inputs is propagated through a series of layers containing the processing units. Consider $I$ input units, $J$ number of neurons per layer, $L$ hidden layers and activation function $\theta$, so in its standard form, the mathematical interpretation of the first layer is:

$$ S_1^j = \sum_{i=1}^{I} w_{ij}^1 x_i $$

where $w_{ij}^1$ is the weight of the unit $j$ of the first layer related to the $i$ unit on the input layer; $x_i$ are the input vectors; $S_1^j$ is the input for the activation function $\theta$ of each neuron $j$; $b_j^1$ is the bias of each $j$ neuron in the first layer; and, $Z_1^j$ is the output of each neuron $j$ of the first hidden layer. Then, Eqs. 3 and 4 for the rest of the $L$ layers are given by:

$$ S_L^j = \sum_{k=1}^{J} w_{jk}^L Z_{k}^{L-1} $$

$$ Z_L^j = \theta(S_L^j + b_j^L) $$

where $w_{jk}^l$ is the weight of the unit $j$ of the hidden layer $l$ related to the output of the unit $k$ of the hidden layer $l - 1$; $Z_{k}^{L-1}$ is the output of each neuron $j$ of the hidden layer $l - 1$; $S_L^j$ is the input for the activation function $\theta$ of each neuron $j$ of the hidden layer $l$; $l, b_j^l$ is the bias of each $j$ neuron in the hidden layer $l$. Finally, the output layer is calculated as:

$$ S_{L+1}^j = \sum_{k=1}^{J} w_{jk}^{L+1} Z_{k}^{L} $$

$$ Z_{L+1}^j = \psi(S_{L+1}^j) $$

where $Z_{L+1}^j$ is the final output and $\psi$ is a linear function.

This is the classical forward pass for a MLP. It is then adjusted and trained by backpropagation (Rumelhart et al. 1988).

3.3 Fuzzy inference system

The fuzzy inference system is a system that allows decision making based on fuzzy, inaccurate, or incomplete data, connected through input and output variables. The fuzzy method can perform robust forecasts even when historical data are not accurate (Chen 1996). The FIS consists of fuzzy rules, an inference system unit, and operations of fuzzification and defuzzification. There are three different types of inference systems, the two most used and important are developed in Mamdani (1976) and Takagi and Sugeno (1983). The latter is used in this study, since it uses linear functions in the fuzzy system rules. The Sugeno method is computationally efficient and works well with the optimization and adaptive techniques, particularly for nonlinear dynamic systems.

In this study, we use a FIS with a subtractive clustering algorithm. This clustering method was proposed by Chiu (1994) and determines the number of rules and membership functions as well as the initial location of the cluster centers. Cluster estimates, which are obtained from this method, are used to initialize clusters based on iterative optimization, using linear least squares estimation to determine the consecutive equations of each rule; this function provides a FIS structure containing a set of fuzzy rules.

3.4 Adaptive neuro-fuzzy inference system

The fuzzy logic approach to analyze problems proposes a different interpretation of the task to solve. The fuzzy logic describes the task through possibility functions that are very easy to understand. These functions are called membership functions and characterize the potential output in possibilities in a straightforward way. To make the selection and optimization of this function, a NN is implemented in the adaptive neuro-fuzzy inference system (Jang 1993). Here the neural network has a predefined architecture with 5 layers, where 4 are fixed and they do not dependent of the input data, regardless of the problem. Just the first layer has to change according to the problem, as it depends directly on the inputs and their characteristics. The other layers relate to rule generation (second layer), normalization (third layer), linear relation (fourth layer) and output calculation (fifth layer). This is the basic notion of the ANFIS framework, as some studies try to reduce...
the architecture (from 5 to 4 layers), but overall results are similar.

The first layer of the ANFIS framework establishes the membership function related to every input, where the function type (triangular, Gaussian, and others) are preselected by the user, together with the number of membership functions that represent each input. The second layer, based on the membership function defined previously, rules are created to represent the fuzzy relationship between the inputs, in the form of $x=A_1$ and $y=B_1$ then $z_1=C_1$ and their predictive power is assigned by weights. The third layer normalizes the previous layer, to make the training and learning of the neural network smoother. The fourth layer takes the output of the last layer and makes a linear (Sugeno) or constant (Mamdani) prediction. Finally, this is passed to the fifth layer to compute the final defuzzificated value.

### 3.5 Long-short term memory (LSTM) network

We use LSTM network in this study due to the inflation sequential dependence of the data generation process. LSTM, as stated before, is widely used in many fields for its ability to better model data that have sequence characteristics. Since the seminal paper of Hochreiter and Schmidhuber (1997), there have been many different variations of LSTM. We make use of the original LSTM algorithm, developed in Greff et al. (2017). Using the description presented in that study, Fig. 1 shows a graphical representation of the general algorithm. The mathematical definition of the network is described in Eq. 9.

\[
\begin{align*}
    f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\
    i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\
    \tilde{C}_t &= \text{tanh}(W_C[h_{t-1}, x_t] + b_C) \\
    C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\
    o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\
    h_t &= o_t \cdot \text{tanh}(C_t)
\end{align*}
\]  

(9)

where $f_t, i_t, o_t$ are the forget, input and exit gates, respectively; $C_t$ is the state of the unit; and $h_t$ is the output of the unit. It is important to note that the mathematical formulation is for one unit, and the LSTM network work similar to the classical NN, this is, has layers and several units per layer.

### 3.6 Proposed combination SARIMA–LSTM

As was presented previously, the hybrid models gain the benefits of the econometric model as well as the advantages of machine learning. One of the requirements to improve results, greater accuracy of forecasts, is to have the enough data to train the hybrid model. First, part of the data is used to train the econometric model, and the forecasting results of the training set of the econometric model are used as additional input to the machine learning. The LSTM is known for its ability to model and predict time series, but fitting the LSTM requires a long time series. Given these issues, we propose a model combining the forecasts of the SARIMA and LSTM to predict inflation. Inflationary data is reported monthly and, for Latin American countries, the data available is very limited, about 730 observations. For this reason, we propose to obtain the final forecasts by developing a single average between the forecasts of SARIMA and LSTM. This proposed model has the advantage of using the available data to train both models taking advantage of the two best features of each, one linear and the other nonlinear, to build a hybrid model. Figure 2 describes the experiment proposed for the 5 economies analyzed, showing the different models (benchmarks) and the proposed model.

All the models described in Fig. 2 are fitted in rolling windows, except the LSTM. The rolling windows consist of a window ($L$ months or window length) to fit the model and then with this model fitted forecasts the inflation one month ahead. After this, the window is advanced one month keeping the length and fit again and forecast the next period. For example, if the window has a length of 180 months (months 1–180), the inflation fore-
casting will be for the month 181. Then, the window advances one month (months 2–181), the model is fitted, and the inflation forecast will be the inflation of the month 182.

To determine the accuracy of the forecasts, we use the mean-square-error (MSE) loss function. In order to understand which model provides a better forecast over another:

$$MSE = \frac{1}{T} \sum_{t=1}^{T} (\pi_t - \hat{\pi}_t)^2$$ (10)

where $\pi_t$ corresponds to the real value of the series at time $t$ and $\hat{\pi}_t$ corresponds to the value forecasted by the proposed model at time $t$.

To determine which model is better statistically, we use the model confidence set (MCS), which is an improvement of the superior predictive ability (SPA) model (Hansen 2005), showing the robustness of the results. The MCS test does not require a comparison model to be chosen; its goal is to determine whether the set $M^*$ where the model $i$ is preferred in terms of loss $L$ or is preferred by an alternative $j$. The null hypothesis establishes $H_0 : \mathbb{E}_{d_{i,j}} = 0$ where $d_{i,j} = L_i - L_j$. The model $i$ is incorporated into the set $M^*$ if $\mathbb{E}_{d_{i,j}} \leq 0$ for all $j$ in the set $M^0$ that contains all the models to be analyzed. The entire experiment is described in Fig. 2, incorporating the measure of the accuracy and the test to analyze the forecasting accuracy superiority.

4 Data

The data used in this study correspond to the inflation of 5 Latin American countries: Brazil, Mexico, Chile, Colombia, and Peru, in monthly terms. The choice of these countries is based on the challenge that being able to forecast emerging economies is more difficult as compared to developed economies due to the volatile nature of the data. The inflation data corresponds to the months from January 1958 to June 2019, with a total of 738 observations, except for Mexico where the inflation data start from February 1969 (605 observations). The inflation data were divided into periods, the first period until June 2009 to train the models. The rest of the months from July 2009 to June 2019, 10 years or 120 months, is the out of sample used to forecast the inflation of the five Latin American economies. All the models are applied across rolling windows (except LSTM); then, to forecast the next month’s inflation, the windows move one month ahead and calculate all again. All the configurations of each models are fitted 120 times to have a one inflation forecasting ahead in each time.

The use of the rolling windows approach is consistent with the literature for forecasting inflation. For instance, Faust and Wright (2013) used a single-quarter approach rather that cumulative inflation rates. McKnight et al. (2020) note that if one uses cumulative inflation rates over the time horizons distortions are likely to occur since the magnitude of the cumulative effects, positive or negative, may cause greater errors in the forecast accuracy (Stock and Watson 2007).
The five inflation time series analyzed are illustrated in Fig. 3.

Through an analysis of Table 1, it can be seen that in all countries, inflation series are asymmetrically positive, with a high concentration of values close to the mean, i.e., they present a leptokurtic distribution. These values tend to gather more on the left side than on the right of the mean. All the countries, with the exception of Colombia, had hyperinflation periods as illustrated in Fig. 3.

As demonstrated in Table 1, Colombia had the lowest average inflation over the study period with only 1.09%. Brazil on the other hand had the greatest average inflation over the study period with 4.97%. While not the highest average, Peru experienced an average inflation rate of 3.59%. Peru’s high average was caused by a period of hyperinflation between 1987 and 1990 as illustrated in Fig. 3e. With the exception of Colombia, the other countries also present cases of hyper-inflation, where two digits were surpassed. In the case of Chile, there was a peak in October 1973 with a value of 87%, as a result of a collapse amid a strong economic crisis that the country experienced. Mexico presents a maximum for two consecutive months, November and December, with monthly inflation of 15%. Of particular interest, it can be seen in Fig. 3 that the 3 countries did not experience major effects on inflation during the subprime crisis of 2008 which globally affected markets.

5 Results

In this section, we present the results obtained by the proposed models and compare them with the results provided by SARIMA, the machine learning models, and the hybrid models, as described in Fig. 2. The first forecasts that were...
performed were those of the SARIMA models, in order to use it as a reference point to compare the others, as the classical econometric model used to forecasting the inflation rate. These values are presented in Table 2, with Colombia having the lowest MSE and Brazil the highest. It is interesting to observe that the best SARIMA configuration is not reached with the longest window length. This can be explained, given that the inflation rate behavior is changing frequently and that it is not necessary to have long memory to forecast. Also, the results show the superiority of the SARIMA models over the ARMA models, indicating the importance of the seasonal effect in the inflation rate. Low values of AR and MA reached the best performance to predict the inflation rate in the countries studied. Following Atkeson and Ohanian (2001), the random-walk (RW) inflation forecasting and performance was computed to use it as benchmark. As can observed in Table 2, SARIMA models had better accuracy to forecast the inflation than the random walk model.

The results obtained by the FIS are worse than the SARIMA forecasting performance. The FIS model has the capacity to fit with short time series, but in this case the MSE obtained in the prediction out of sample is higher than SARIMA for all the economies studied (see Table 3).

The ANFIS model is similar to the FIS performance. All the configurations tested obtained worse accuracy than the SARIMA model (Table 4).

The different configurations tested to the artificial neural networks did not beat the accuracy of the SARIMA forecast-
Table 5 Artificial neural networks forecasting results

| Country | MSE         | Configuration |
|---------|-------------|---------------|
| Brazil  | 2.1971E-05  | LA:1; N:2; L:360; AR:1 |
| Mexico  | 3.1832E-05  | LA:1; N:5; L:420; AR:6 |
| Chile   | 9.9037E-06  | LA:1; N:5; L:420; AR:6 |
| Colombia| 1.7414E-05  | LA:1; N:5; L:420; AR:12 |
| Peru    | 7.2020E-05  | LA:1; N:2; L:360; AR:1 |

The best MSE values obtained by the forecasts of different artificial neural networks’ configurations. LA is the number of layers; N is the number of neurons per layer; L is the window length; and AR is the number of autoregressive terms. The grid of hyperparameters was: LA: 1-6; N:2,5,10,15,20; L:360,420,480; AR: 1, 3, 6, 12 months

Table 6 Seasonal autoregressive moving average–artificial neural networks forecasting results

| Country | MSE         | Configuration |
|---------|-------------|---------------|
| Brazil  | 1.2405E-05  | LA:5; N:2; L:360; AR:3 |
| Mexico  | 1.4027E-05  | LA:4; N:5; L:360; AR:3 |
| Chile   | 8.7292E-05  | LA:4; N:5; L:360; AR:6 |
| Colombia| 1.0713E-05  | LA:3; N:2; L:360; AR:1 |
| Peru    | 7.2730E-05  | LA:3; N:2; L:420; AR:6 |

The best MSE values obtained by the forecasts of different artificial neural networks’ configurations including the best SARIMA forecasting as input. LA is the number of layers; N is the number of neurons per layer; L is the window length; and AR is the number of autoregressive terms. The grid of hyperparameters was: LA: 1-4; N:2, 4, 6, 10, 12, 15, 18, 20, 40, 60, 100, 150, 200, 250; L:360,420,480; M: 0.2, 0.3, 0.4

Table 7 Long short-term memory forecasting results

| Country | MSE         | Configuration |
|---------|-------------|---------------|
| Brazil  | 9.9368E-06  | LA:2; N:250; L:360; M:0.4 |
| Mexico  | 9.1370E-06  | LA:1; N:18; L:360; M:0.4 |
| Chile   | 9.0082E-06  | LA:4; N:20; L:360; M:0.4 |
| Colombia| 4.59318E-06 | LA:2; N:150; L:420; M:0.4 |
| Peru    | 7.4845E-06  | LA:2; N:2; L:360; M:0.2 |

The best MSE values obtained by the forecasts of different LSTM configurations. LA is the number of layers; N is the number of units per layer; L is the window length; and M is the probability that the dropout layer randomly sets input elements to zero. The grid of hyperparameters was: LA: 1-4; N:2, 4, 6, 10, 12, 15, 18, 20, 40, 60, 100, 150, 200, 250; L:360,420,480; M: 0.2, 0.3, 0.4

Table 8 Seasonal autoregressive moving average–long short-term memory forecasting results

| Country | MSE         | Configuration |
|---------|-------------|---------------|
| Brazil  | 9.4940E-06  | LA:3; N:200; L:480; M:0.3 |
| Mexico  | 7.6323E-06  | LA:2; N:250; L:360; M:0.4 |
| in Chile| 8.9394E-06  | LA:3; N:200; L:360; M:0.4 |
| Colombia| 4.1976E-06  | LA:1; N:250; L:360; M:0.3 |
| Peru    | 7.4169E-06  | LA:1; N:6; L:420; M:0.2 |

The best MSE values obtained by the forecasts of different LSTM configurations including the best SARIMA forecasting as input. LA is the number of layers; N is the number of units per layer; L is the window length; and M is the probability that the dropout layer randomly sets input elements to zero. The grid of hyperparameters was: LA: 1-4; N:2, 4, 6, 10, 12, 15, 18, 20, 40, 60, 100, 150, 200, 250; L:360,420,480; M: 0.2, 0.3, 0.4

It is interesting to observe that the best ANN configuration is reached with one layer and a low number of neurons for all the countries analyzed. This finding indicates that the ANN does not have the ability to model the inflation behavior of the countries analyzed.

The hybrid models’ SARIMA–ANN obtained better accuracy performance than the ANN forecasts, indicating the synergy of the hybrid models. However, this improvement in all countries is not enough to beat the precision in the forecasts of the SARIMA models (Table 6). All the previous results demonstrate the superiority of econometric model, in particular the SARIMA model to predict the inflation rate.

The forecasts predicted by the LSTM are better for the inflation rate series studied than the other machine learning models (Table 7). The MSE performance of the LSTM was slightly higher than the MSE obtained by the SARIMA. Even in the case of Chile, the LSTM forecasts had better precision than the SARIMA model. The number of layers is low for the best configurations, because LSTM with more layers needs more data to fit.

The hybrid SARIMA–LSTM forecasts had better accuracy than the LSTM predictions. These results indicate that the use of the SARIMA forecasts as input for the LSTM improved the forecast precision. The hybrid SARIMA–LSTM obtained lower MSE than the SARIMA and LSTM for the inflation rate of Brazil and Chile (Table 8).

The next step was to apply the proposed model, calculating the single average between the SARIMA forecasts and LSTM forecasts. The results of the proposed model obtained better accuracy over all the benchmark models (SARIMA, FIS, ANFIS, ANN, SARIMA–ANN, LSTM, SARIMA–LSTM) for all the economies analyzed. Taking the best previous forecast model, the proposed model reduced the MSE by 0.91% for Brazil, 5.51% in for Mexican, 3.33% for the Chilean inflation rate, 4.15% in the Colombian case, and slightly higher than the MSE obtained by the SARIMA. Even in the case of Chile, the LSTM forecasts had better precision than the SARIMA model. The number of layers is low for the best configurations, because LSTM with more layers needs more data to fit.

The hybrid SARIMA–LSTM forecasts had better accuracy than the LSTM predictions. These results indicate that the use of the SARIMA forecasts as input for the LSTM improved the forecast precision. The hybrid SARIMA–LSTM obtained lower MSE than the SARIMA and LSTM for the inflation rate of Brazil and Chile (Table 8).

The next step was to apply the proposed model, calculating the single average between the SARIMA forecasts and LSTM forecasts. The results of the proposed model obtained better accuracy over all the benchmark models (SARIMA, FIS, ANFIS, ANN, SARIMA–ANN, LSTM, SARIMA–LSTM) for all the economies analyzed. Taking the best previous forecast model, the proposed model reduced the MSE by 0.91% for Brazil, 5.51% in for Mexican, 3.33% for the Chilean inflation rate, 4.15% in the Colombian case, and 5.61% for the Peruvian inflation rate.

Finally, the model confidence set (MCS) test was conducted for each country to test the robustness of the results. For this evaluation, the best forecasts of each of the seven models and the proposed hybrid model were taken of each country. This test was applied to all the evaluated models. These results can be seen in Table 10. For the inflation rate prediction of Mexico, Colombia, and Peru, the superiority of
Table 9  Proposed model forecasting results

| Country | Proposed | SARIMA–LSTM | Improv(%) | SARIMA | Improv(%) | LSTM | Improv(%) |
|---------|----------|-------------|-----------|-------|-----------|------|-----------|
| Brazil  | 9.4074 E-06 | 9.4940 E-06 | -0.91%    | 9.6158 E-06 | -2.17%    | 9.9368 E-06 | -5.33%    |
| Mexico  | 7.1810 E-06 | 7.6323 E-06 | -5.91%    | 7.6000 E-06 | -5.51%    | 9.1370 E-06 | -21.41%   |
| Chile   | 8.6414 E-06 | 8.9394 E-06 | -3.33%    | 9.0504 E-06 | -4.52%    | 9.0082 E-06 | -4.07%    |
| Colombia| 3.7895 E-06 | 4.1976 E-06 | -9.72%    | 3.9533 E-06 | -4.15%    | 4.5932 E-06 | -17.50%   |
| Peru    | 6.4554 E-06 | 7.4169 E-06 | -12.96%   | 6.8393 E-06 | -5.61%    | 7.4845 E-06 | -13.75%   |

The improvement of the proposed model MSE over the best forecast models

Table 10  Model confidence test results

| Model | Brazil | Mexico | Chile | Colombia | Peru |
|-------|--------|--------|-------|----------|------|
| Proposed | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| SARIMA | 0.4790 | 0.2819 | 0.6092 | 0.2639 | 0.2289 |
| SARIMA–LSTM | 0.6344 | 0.2819 | 0.6243 | 0.0589 | 0.0413 |
| LSTM | 0.3143 | 0.0961 | 0.6092 | 0.0105 | 0.0258 |
| SARIMA–ANN | 0.0671 | 0.0001 | 0.8455 | 0.0000 | 0.0763 |
| ANN | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| FIS | 0.0000 | 0.0000 | 0.0006 | 0.0000 | 0.0050 |
| ANFIS | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 |

The table shows the best model for each methodology and the p-value of MCS Test.

6 Discussion

As stated in the literature review, there are two general approaches to forecasting inflation: those based on macro- and microeconomic models such as the NKPC approach and those that focus on time series (McKnight et al. 2020). The approach that we focused on extending is the approach that analyzes the time series. Within the time-series modeling approach, we suggested several modeling techniques such as SARIMA, ANN; SARIMA–ANN; FIS; and ANFIS.

Pincheira and García (2012) found that a SARIMA model was better able to forecast inflation in Chile which served as part of the motivation of the current work. We applied the SARIMA model across all five countries in our study to (1) try to replicate the earlier study and (2) use it as a benchmark for the other models. We did find similar results to the earlier study in that it outperformed a traditional ARMA model, but the performance was mixed across the countries as noted in the results.

Consistent with the results that Işıççök et al. (2020) but contrary to Estiko and Wahyuddin (2019) when comparing ANN and ARIMA models, we found that ANN did not outperform ARIMA for most of the countries in our study. While we did not include any test for what may drive the similarity dissimilarity, we think that there may be some structural characteristics that are similar between the countries in our study and the Turkey but not Indonesia.

While Luis and Héctor (2013) and Santana (2006) both found that ANN outperformed the Bank of Mexico and other econometric techniques, respectively, our results suggest something different. We do note that the SARIMA–ANN model outperformed both the ANN and the SARIMA models alone. This suggests augmentation of the ANN improves the overall performance and aligns more closely with the results of these two aforementioned studies.

7 Conclusions

This paper explores the implementation of a single combination of SARIMA and LSTM to forecast inflation rate in emerging economies. The model was applied in five Latin American countries: Brazil, Mexico, Chile, Colombia, and Peru. The time frame for the study was from January 1958 to June 2019. Despite important role that inflation plays for governments and central banks, this macroeconomic variable
is a topic that lacks new research to improve forecasts, especially in developing countries. Thus, this study analyzes the performance of the average of SARIMA and LSTM forecasting as final prediction. The results indicate that there is a significant improvement as measured by a reduction of the accuracy measure used. The results suggest that the proposed approach demonstrates improvement and better forecasting accuracy for Mexico, Colombia, and Peru as compared with all other models. For Brazil and Chile, however, the MSE decreased with the proposed model, indicating a higher precision but that the improvement is not statistically significant as compared with SARIMA, SARIMA–LSTM and LSTM.

The proposed model reached better accuracy as a combined model as compared to only averaging the SARIMA and LSTM forecasts. The proposition was to keep the benefits of SARIMA as a linear model and capture the nonlinear aspects with LSTM. The LSTM transforms the series, and as such, the linear nature of the series captured by SARIMA is lost. Hence, when the output from the SARIMA is used as input for the LSTM model (SARIMA–LSTM), the linear aspect of the data is lost that would otherwise have been captured if the data were not preprocessed. For this reason, the single average of the two forecasts (SARIMA and LSTM) maintains the linear and nonlinear nature mixing both, and no one is lost. However, while the proposed model has demonstrated improved performance in some instance, it did not demonstrate improvement in all cases as suggested. Therefore, there needs to be continued work in this domain to tune the models and understand better why the hybrid model performs well in some developing economies and not in others. Since inflation forecasts are so important to setting national direction and monetary policies, continued work is encouraged. In fact, it may be the case that there are economic features that may be included in the models, which result in improved performance and ought to be considered.

As future work, the results obtained in this study are promising, but the scope is constrained to only one month ahead. One of the challenges is to test and propose models to forecast to inflation to longer periods. Additionally, it may be worth considering a combination of the NKPC with the machine learning techniques proposed here.

Acknowledgements The authors would like to thank the Dirección General de Investigación, Innovación y Postgrado (DGIP), UTFSM, Valparaíso, Chile, for financial support provided through the "Programa de Incentivos a la Iniciación Científica (PIIC)" initiative.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Adolfson M, Lindé J, Villani M (2007) Forecasting performance of an open economy dsge model. Econ Rev 26(2–4):289–328
Atkeson A, Ohanian LE et al (2001) Are Phillips curves useful for forecasting inflation? Federal Reser Bank Minneapolis Quar Rev 25(1):2–11
Bacanli F, de Barros LC (2017) Application of prediction models using fuzzy sets: a Bayesian inspired approach. Fuzzy Sets Syst 319:104–116
Box GE, Jenkins GM (1976) Time series analysis: forecasting and control, revised. Holden-Day
Broto C (2011) Inflation targeting in Latin America: empirical analysis using GARCH models. Econ Modell 28(3):1424–1434
Cai M, Del Negro M, Giannoni MP, Gupta A, Li F, Moszkowski E (2019) Dsge forecasts of the lost recovery. Int J Forecast 35(4):1770–1789
Chan W-S (1999) A comparison of some of pattern identification methods for order determination of mixed ARMA models. Stat Prob Lett 42(1):69–79
Chang P-T (1997) Fuzzy seasonality forecasting. Fuzzy Sets Syst 90(1):1–10
Chen S-M (1996) Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst 81(3):311–319
Chiu SL (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2(3):267–278
Cogley T, Sbordone AM (2008) Trend inflation, indexation, and inflation persistence in the new keynesian Phillips curve. Am Econ Rev 98(5):2101–26
Connor JT, Martin RD, Atlas LE (1994) Recurrent neural networks and robust time series prediction. IEEE Trans Neural Netw 5(2):240–254
Jesus Cuaresma, Cresp FM, Huber F (2016) Forecasting with global vector autoregressive models: a Bayesian approach. J Appl Econ 31(7):1371–1391
D’Amato L, Garegnani L, Blanco E (2008) Forecasting inflation in Argentina: Individual models or forecast pooling? Technical report, Working Paper, Central Bank of Argentina (BCRA)
de Simona FN (2001) Proyección de la Inflación en Chile. Economía Chilena 4(3):59–85
Estiko FI, Wahyuddin S (2019) Analysis of Indonesia’s inflation using arima and artificial neural network. Econ Dev Anal J 8(2):151–162
Faust J, Wright JH (2013) Forecasting inflation, volume 2 of Handbook of Economic Forecasting, pages 2–56. Elsevier
Freeman JA, Skapura DM (1991) Algorithms, applications, and programming techniques. Addison-Wesley Publishing Company, USA
Gali J, Gertler M (1999) Inflation dynamics: a structural econometric approach. J Monetary Econ 44:195–222
Gambou JCB (2017) Deep Learning for Time-Series Analysis. arXiv preprint arXiv:1701.01887
Graves A, Fernández S, Gomez F, Schmidhuber J (2006) Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In Proceedings of the 23rd international conference on Machine learning, pages 369–376. ACM
Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space odyssey. IEEE Trans Neural Netw Learn Syst 28(10):2222–2232
Hansen PR (2005) A test for superior predictive ability. J Bus Econ Stat 23(4):365–380
Heeb D (1949) The organization of behavior. New York
Hoehreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
Işığçok E, Öz R, Tarkan S (2020) Forecasting and technical comparison of inflation in turkey with box-jenkins (arima) models and the
artificial neural network. Int J Energy Opt Eng (IJEOE) 9(4):84–103
Jang J-S (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685
Khashei M, Hejazi SR, Bijari M (2008) A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. Fuzzy Syst Syst 159(7):769–786
LeBaron B (1994) Chaos and nonlinear forecastability in economics and finance. Philos Trans Royal Soc London A Math Phys Eng Sci 348(1688):397–404
Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019
Liu L-M, Hudak GB, Box GE, Muller ME, Tiao GC (1992) Forecasting and time series analysis using the SCA statistical system, vol 1. Scientific Computing Associates DeKalb, IL
Luis J, Héctor J, et al. (2013) Forecasting Mexican inflation using neural networks. In Electronics, Communications and Computing (CONIELECOMP), 2013 International Conference on, pages 32–35. IEEE
Maasoumi E, Khotanzed A, Abaye A (1994) Artificial neural networks for some macroeconomic series: a first report. Econ Rev 13(1):105–122
Mamdani EH (1976) Advances in the linguistic synthesis of fuzzy controllers. Int J Man-Mach Stud 8(6):669–678
Manning C, Socher R, Fang GG, Mundra R (2017) CS224n: Natural Language Processing with Deep Learning1
McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biol 5(4):115–133
McKnight S, Mihailov A, Rumler F (2020) Inflation forecasting using the New Keynesian Phillips Curve with a time-varying trend. Econ Modell 87:383–393
Pesaran MH, Schuermann T, Smith LV (2009) Forecasting economic and financial variables with global VARs. Int J Forecast 25(4):642–675
Pincheira P, García Á (2012) En busca de un buen marco de referencia predictivo para la inflación en Chile. El trimestre económico 79(313):85–123
Pincheira P, Gatty A (2016) Forecasting Chilean inflation with international factors. Emp Econ 51(3):981–1010
Qi CR, Su H, Mo K, Guibas LJ (2017) Pointnet: Deep learning on point sets for 3D classification and segmentation. Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 1(2):4
Rojas I, Valenzuela O, Rojas F, Guillén A, Herrera LJ, Pomares H, Marquez L, Pasadas M (2008) Soft-computing techniques and ARMA model for time series prediction. Neurocomputing 71(4):519–537
Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386
Rumelhart DE, Hinton GE, Williams RJ (1985) Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science
Rumelhart DE, Hinton GE, Williams RJ et al (1988) Learning representations by back-propagating errors. Cognit Model 5(3):1
Santana JC (2006) Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. Revista Colombiana de Estadística 29(1):77–92
Šestanović T, Arnerić J (2021) Neural network structure identification in inflation forecasting. J Forecast 40(1):62–79
Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis. Ann Rev Biomed Eng 19:221–248
Stock JH, Watson MW (2007) Why has US inflation become harder to forecast? J Money Credit Bank 39:3–34
Takagi T, Sugeno M (1983) Derivation of fuzzy control rules from human operator’s control actions. IFAC Proc Vol 16(13):55–60
Tinoco-Zermeño MÁ, Venegas-Martínez F, Torres-Preciado VH (2014) Growth, bank credit, and inflation in Mexico: evidence from an ARDL-bounds testing approach. Latin Am Econ Rev 23(1):8
Tkacz G (2001) Neural network forecasting of Canadian GDP growth. Int J Forecast 17(1):57–69
Vercher E, Bermúdez JD, Segura JV (2007) Fuzzy portfolio optimization under downside risk measures. Fuzzy Sets Syst 158(7):769–782
Wang B, Yin P, Bertozzi AL, Brantingham PJ, Osher SJ, Xin J (2017) Deep Learning for Real-Time Crime Forecasting and its Ternarization. arXiv preprint arXiv:1711.08833
Xu Q, Zhuo X, Jiang C, Liu Y (2019) An artificial neural network for mixed frequency data. Exp Syst Appl 118:127–139
Young T, Hazarika D, Poria S, Cambria E (2017) Recent trends in deep learning based natural language processing. arXiv preprint arXiv:1708.02709

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.