BANK OF ALBANIA

Economic Review

2024 H2

You may use data of this publication, provided the source is acknowledged.
Published by: Bank of Albania,
Sheshi "Skënderbej", nr. 1, Tiranë, Shqipëri
Tel.: + 355 4 2419301/2/3; + 355 4 2419409/10/11
Fax: + 355 4 2419408
E-mail: public@bankofalbania.org
www.bankofalbania.org

www.bankofalbania.org

** Views expressed in this review are of the authors and do not necessarily reflect those of the Bank of Albania.*

C O N T E N T

SELECTING A NEURAL NETWORK MODEL WITH AN APPLICATION TO HOUSE PRICE DETERMINANTS IN ALBANIA

BLERINA VIKA, Department of Statistics and Applied Informatics,
Faculty of Economy, University of Tirana, Albania
ILIR VIKA, Research Department, Bank of Albania, Albania

ABSTRACT

The purpose of this article is twofold. Firstly, it discusses the key issues faced in designing, specifying and validating a neural network structure. Although the empirical literature continues to enlarge, the identification of the best neural network model still involves a lot of trial and error until an appropriate number of inputs, hidden layers and neurons is selected. After showing how those decisions are generally taken in practice, we then aim to investigate the driving forces behind the recent surge in Albania's house prices. The latter are analyzed in relation to a number of theory-based indicators, including GDP, bank credit, foreign purchases, rents, interest rates, construction costs, urban population, and house building permits.

A total of 112 candidate neural network structures is tested for their forecast accuracy of house prices from 2017Q1 to 2023Q2. The resulting optimal values for hyperparameters suggest that when building a neural network, practitioners may want to focus on model structures that consist of a couple of input lags, three hidden layers, three dozens of neurons, and around 67,600 parameters. Finally, our prediction-based analysis shows that past house price developments and per capita income are the most relevant indicators that help in explaining the recent run-up in house prices.

Keywords: neural network architecture; house price determinants; Albania.

1. INTRODUCTION

Much of the early research in neural networks (NN) applications have been sponsored by financial services institutions, as they realized the potentiality of using artificial NNs in rating financial investment risks, forecasting economic indicators, and analyzing the relevance of explanatory variables (Trippi and Turban, 1993). The empirical literature on neural networks continues to enlarge in all fields of economics, yet the methodology used for the construction and specification of NN models still involves a lot of trial and error in order to define and select the best structure that serves the matter in hand. Contrary to traditional econometric methods, neural networks can handle estimations that contain a much larger number of parameters than observations. Thus, the complexity of a neural network varies with the selection of the number of inputs (and their lags), hidden layers and neurons.

This article discusses the key issues faced in designing, specifying and validating a neural network structure to be used for analyzing or forecasting economic variables. After the methodological discussion, we demonstrate how the decisions are made in practice during the model selection procedure. As an illustration, we employ the long short-term memory (LSTM) network in an exercise that investigates the drivers of the recent surge in Albania's house prices. The latter are analyzed in relation to a number of theory-based indicators, such as GDP per capita, bank credit for real estate, foreign investment in housing market, home rental, interest rates, construction costs, urban population growth, and new house building permits. By varying the number of input lags (2, 4, 8, 16), hidden layers (1, 2, 3, 4), and nodes (16, 32, 64, 128, 256, 512, 1024), the total of candidate structures amount to 112 tested neural networks. The set of hyperparameters in each network is chosen such that it yields the lowest out-of-sample forecast error (RMSE) over a validation sample of 48 quarters. Finally, all NN structures are tested for their forecast ability of house prices from 2017Q1 to 2023Q2, and then ranked accordingly from best to worst performing models. Looking at their resulting optimal values for hyperparameters it is suggested that when building a neural network, practitioners may want to focus on model structures that consist of a couple of input lags, three hidden layers, three dozens of neurons, and around 67,600 parameters.

The article is organized as follows. Section 2 provides an overview of the framework for designing neural networks. It explains the main steps for constructing a neural network, starting from data collection and the need for preprocessing to model evaluation and selection issues. Section 3 draws on the steps outlined above and tries to provide some practical advice for beginners with regard to parameter selection tradeoffs, by building a neural network model to analyze house price determinants in Albania.

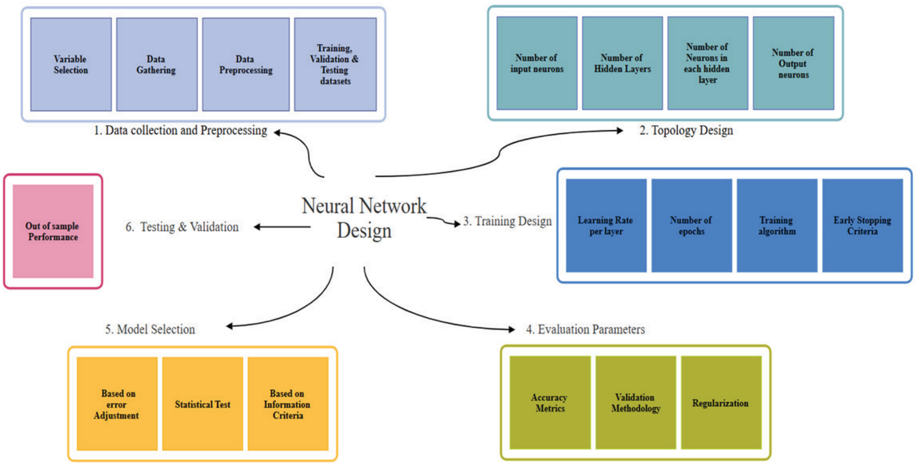
2. FRAMEWORK FOR DESIGNING NEURAL NETWORKS

The strategy for selecting the best neural network structure for time series forecasting should follow a set of general steps, regardless of the specific type of NN architecture (Kaastra and Boyd, 1996; Rahutomo and Gunawan, 2018; Vika and Tole, 2023). The general activities for selecting the best ANN model for time series prediction are presented in Figure 1.

2.1 DATA COLLECTION AND PREPROCESSING

The groundwork examination phase, as part of the data collection and preprocessing, involves preparation steps to comprehend the dataset and determine the key indicators for effective forecasting. In this phase, the data are preprocessed through several key activities with the purpose of cleaning and preparing the data to improve the accuracy of further analysis. The preprocessing techniques are evaluated based on their impact on the data and their impact on the prediction model. In this phase, the data, inputs and outputs are transformed to minimize noise, detect trends, and flatten distribution of variables to assist NN models in learning relevant patterns. Many techniques have been used in various research works and their selection has been made in accordance with the characteristics of the time series data (Tawakuli et al., 2024; Pisa et al., 2019). Rescaling is an important step when using neural network forecasting models. It can improve convergence speed, generalization, prevent vanishing gradients, and ensure that the learning process is not biased toward a particular time series. The time series data must be scaled between upper and lower bounds of transfer function. The most common data transformations include first differencing and taking natural logs. Others involve using ratios of inputs, or methods used in technical analysis. Removing outliers is considered good practice during the data cleaning process (Feng et al., 2014). Data splitting, in machine learning, is used to optimize model hyperparameters and evaluate the model's ability to generalize to unseen data. The time series, commonly, is divided into three different datasets: training, validation and testing set. The validation set can be used for fine-tuning the model performance such as for choosing hyper-parameters or regularization parameters in the model (Joseph and Vakayil, 2021).

Figure 1. Steps for constructing a neural network model for time series forecasting.



2.2 TOPOLOGY DESIGN

ANN is a group of interconnected units (or neurons), which resembles how information is processed in the nervous system. The three main features of an artificial neural network are: (i) the units; (ii) network architecture; and (iii) the training algorithm. The units, the basic processing elements, are organized in three different types of layers: input, hidden and output layer. The parameters that determine the neural network topology are: the type of neurons, how they are connected, number of neurons, number of hidden layers, number of neurons in each hidden layer and number of output neurons. The training algorithm is used to adjust the connection weights between neurons that are connected, trying to optimize the model's ability to learn from the training dataset.

Our research is based on the Long Short Memory (LSTM) architecture which is designed to work with sequence data. The LSTM model is a type of recurrent neural network, but the neuron design is able to surpass the vanishing gradient problem commonly encounter by classical RNNs. LSTMs architecture address the issue of "short-term memory" in RNNs by utilizing gates that allow the retention of relevant long-term memory and combining it with most recent input data. Throughout this process, LSTM are able to "memorize" or "forget" information by using a special memory cell state, which is regulated by three gates: the input gate, forget gate and output gate. These gates manage the flow of information into and out of the memory cell state.

LSTMs are capable of modeling complex systems even without precise knowledge of underlying rules, thus being a suitable tool to address obstacles in analyzing and forecasting economic times series, such as house prices in Albania, which stem from the absence of a consensus model, disputes over variables to be included, and data measurement challenges.

2.3 TRAINING DESIGN

The training algorithm adjusts the weights that connect neurons, in pursuit of optimizing the model's ability to learn from the training dataset. LSTM architectures that are used here have a more complex structure as compared to traditional recursive neural networks (RNNs). They are computationally intensive and require careful design and tuning of hyperparameters for optimal performance. The aim of designing an ANN is to determine the set of weights between the nodes/neurons that minimize the error function, enhancing the model's ability to perform accurate forecasts. Ideally, this set of weights should allow the model to generalize well, as long as it is not overfitting or underfitting. The training design process includes fine-tuning different settings to improve the model's training efficiency and accuracy. Important hyperparameters include the learning rate, epoch size, early stopping criteria, training algorithm and the number of units (aka nodes or neurons). Defining an effective learning rate is crucial for the model to achieve a better generalization. It is usually recommended to start with a high learning rate and then gradually reduce it once the model's performance stabilizes on the training set (Li, Wei and Ma, 2019).

A popular technique to prevent overfitting in deep learning models is the so-called early stopping, which is employed to stop the training process when the model's performance stops improving on the validation sample data. Increasing training time, by using a large epoch size does not necessarily guarantee an improvement in validation accuracy, and early stopping can effectively prevent overfitting (Song et al., 2020).

The size of the dataset is important to achieve a good generalization of the model. A small dataset produces a significant disparity between the performance evaluations during validation and testing phases. To decrease the gap, it is needed to have more data series for the training or validation set. For an accurate performance evaluation, a good balance between the size of the training and validation set is necessary (Xu and Goodacre, 2018).

2.4 EVALUATION OF PARAMETERS

To evaluate a deep learning model, it is important to measure its predictive ability, generalization capability and its general quality. Frequently used metrics for measuring model forecast capabilities include the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Coefficient of Determination ($R2$).

For time series forecasting tasks, various strategies have been proposed, and no consensus exists on the methods for performance estimation. Out-of-sample and cross-validation are frequently used. The out-of-sample approach, in contrast to cross-validation, is never assessed on past data. In their research, Cerquira et al. (2020) conclude that cross-validation method demonstrates competitive estimation ability in stationary time series. However, when non-stationarity is present, this method consistently produces worse estimations compared to the out-of-sample approach.

Regularization is another technique used to improve the generalization ability model. Constraints are applied on the complexity and smoothness of the model, in order to achieve a good generalization (Tian and Zhang, 2022). Some of the regularization techniques used in deep learning models for time series forecasting include Dropout, L2&L1 Regularization and Early Stopping.

2.5 MODEL SELECTION

To identify the optimal configuration of a deep learning model for a given time series, it is important to use an accurate training algorithm that effectively minimizes errors as additional parameters are introduced. However, it is crucial to compare the performance of different architectures and select the one that achieves optimal levels among the errors and number or parameters.

The process of finding the best model is a trial-and-error process. Changing the number of parameters in the structure of a neural network generates different models, among which it is necessary to select the best model. Selection on error correction is based on selecting the one that gives the lowest measure of error, as evaluated by MSE, MAE, RMSE and MAPE. Another approach that is widely used for model selection is based on the information criteria, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC). These criteria are deeply based on foundational thoughts in statistics and information theory. In this regard, AIC can represent a group of information criteria that are efficient in nonparametric scenarios, whereas BIC can be more suitable for a parametric scenario (Zhang, Yang, and Ding, 2023).

2.6 TESTING AND VALIDATION

Defining the best model for out-of-sample forecasting, which follows the general activities of the design framework, requires a trial-and-error process. During this process, one ought to determine the structure, training algorithm, weights and also the data transformation that impact the performance of the model. In real-world forecasting, the deep learning model must be periodically re-trained to maintain and also improve the performance over time (Lopes et al., 2023). A well-known advantage of ANNs is their capacity to adapt to changing economic trends through periodic retraining. The performance during the retraining can be compromised if the independent variable becomes less significant.

3. A LSTM ANALYSIS ON THE RECENT SURGE IN HOUSE PRICES

The recent rise in housing costs has been one of the most contentious economic issues around the world, including Albania. While many people rely on houses as the main source of wealth and borrowing, many others are getting disillusioned by watching a decrease in their ability to purchase or rent a house. For that reason, understanding the forces that drive the housing market is important to maintain financial stability and ensure a healthy society.

3.1 DETERMINANTS OF HOUSE PRICES IN ALBANIA

During the 2008-22 period, house prices in Albania rose much quicker than in Western Balkans and most of the European Union countries. Although they fluctuated around a "steady" level for several years in the aftermath of the global financial crisis, a strong upward trend emerged in the years prior to, and post the Covid-19 pandemic. Expressed in euro currency, Albanian house prices increased at an average annual rate of 18.1 percent from 2017 to 2022 (or 9.4 percent if converted in local currency and adjusted for inflation), which was in line particularly with the income growth and real estate purchases by non-residents. On the other hand, the steep decline in annual urban population growth in combination with rising building permits for new dwellings (in square kilometer) should have eased demand pressures on house prices (please see Vika and Vika (2023) for a broader view on housing market developments and some long-run stylized facts in Albania).

Different studies have inquired into property market developments in search of establishing a long-run relationship between house prices and their fundamentals in Albania. They generally point out that house prices entered a period of "correction" after 2012 (Yzeiraj, 2016), and a strong link might exist with financial leverage and the exchange rate (Suljoti, 2017), mortgage loans, long-term lending rates and construction costs (Marku, Lleshaj and Lleshaj, 2020), and remittances (Lleshaj and Korbi, 2019). Nevertheless, the estimation of a theoretically-relevant framework with many variables seems to be hard to identify (Koprencka, Bundo and Qarri, 2014), and the reaction of house prices to certain variables might change over time (Halili, 2022).

The empirical literature uses a number of demand and supply fundamental variables in modeling house prices. Typical factors on the demand side include income, mortgage loans and interest rates, financial wealth, demographic and labor market indicators, expected return, and other institutional factors. On the supply side, house prices depend on profitability, which can be a function of house price expectations and costs of construction. Assuming that demand and supply are always equal in the housing market, we model Albanian house prices as the following function:

$$RHP_t = f(RHP_{t-p}, YPC_{t-p}, CRE_{t-p}, FIRE_{t-p}, REN_{t-p}, RI2Y_{t-p}, RCC_{t-p}, POPCM_{t-p}) \qquad (1)$$

where

RHP = house prices expressed in lek currency, adjusted for inflation and seasonality;
YPC = real GDP per capita;
CRE = bank lending for real estates, calculated as annual change in stock, in % of GDP;
FIRE = real estate sales to non-residents, calculated as 4-quarter cumulative sales, in % of GDP;
REN = house rental, adjusted for inflation;
RI2Y = interest rate on two-year government bonds, adjusted for inflation;
RCC = construction costs, adjusted for inflation and seasonality;
POPCM = urban population changes (calculated as cumulative changes in the past three years) as a ratio of new house building permits in the last three years.

The above model postulates that the real house price index is explained by the past performance, t-p, of its own and the theory-based indicators. If house prices are persistent, then current values should be correlated with the past behavior. The real (per capita) income depicts the household purchasing power and their borrowing capacity, therefore it should be positively related with RHP. A similar link is expected with the volume of mortgage debt, as it discloses whether credit to households is rationed or not. Foreign investment in domestic real estate would further stimulate demand and put upward pressures on house prices. House rental enters the model following the user cost theory postulation that house prices and rental share a cointegrating relationship, which is linked to dividend pricing hypothesis. On the other hand, RHP should be negatively associated with higher real interest rates, as they reduce households' capacity to borrow, or make them rotate their portfolio in favor of alternative investments, such as government securities. Construction costs, such as wages of construction workers and material costs, could also have a negative impact on the supply of housing which is described to be a positive function of profitability in the construction business. Finally, house prices tend to rise if population density increases and urban immigration is not met in due time with appropriate expanded areas of newly-built houses.

The interaction of house prices with other socio-economic factors is pretty complex. The absence of a consensus model, disputes over which variables to include, and data measurement challenges have instigated many authors to resort to machine learning techniques in the hope of improving their forecast analysis in many advanced and developing countries (see for instance Alfaro-Navarro et al., 2020; Banerjee and Dutta, 2017; Behr et al., 2023; Ceh et al., 2018; Chatzidis, 2019; Fan, Cui and Zhong, 2018; Hacıevliyagil, Drachal and Eksi, 2022; Ho, Tang, and Wong, 2021; Hong, 2020; Mora-Garcia, Cespedes-Lopez, and Perez-Sanchez, 2022; Park and Bae, 2015; Vika and Vika, 2023; and Wang and Li, 2019). Indeed, the literature on artificial neural network offers a number of techniques that evolve quite fast. Common tools that are used in different fields of natural and social sciences include the feed-forward neural network (FFNN), recurrent neural networks (RNN), the long short-term memory (LSTM), and generalized regression neural networks (GRNN). In our illustrative exercise below on the practical issues that are faced throughout the estimation of a neural network we have employed the LSTM technique, which is encountered in many recent studies that focus on the prediction of economic and financial time series.

3.2 ESTIMATION OF THE LSTM NEURAL NETWORK

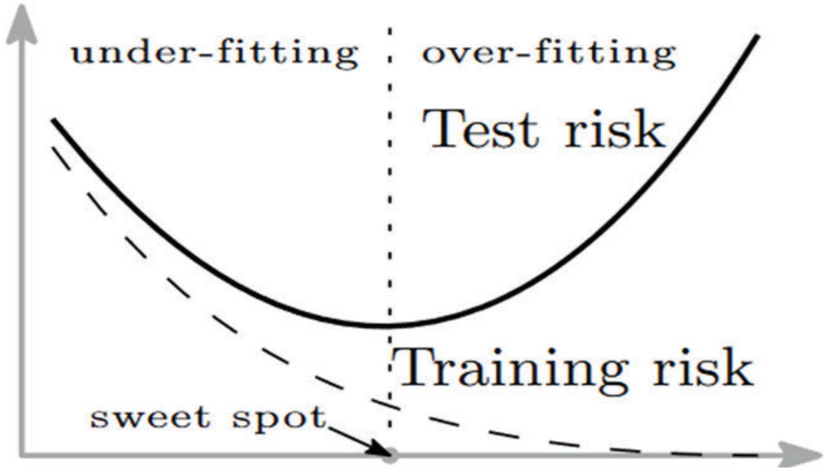
A recurrent neural network focuses on the past data and their correlation with current values. The LSTM method enhances the long-term memory of an RNN by storing only relevant information and shrinking forecast errors over extended periods. As in other neural network techniques, parameters are estimated by minimizing a loss function between the observed and fitted values in the in-sample period. Neural network parameters are often optimized by applying the (stochastic) gradient descent method, where parameters are updated iteratively at a learning rate that determines the size of the step towards optimization.

We train and update the parameters of our network by using the Adam optimization algorithm, which is an extension of the gradient-descent procedure and has become popular due its ability to combine the advantages of other standard algorithms in the literature. As an adaptive learning algorithm Adam reduces the need to adjust the learning rate hyperparameter, therefore we rely on its default value of 0.001. Nevertheless, many practitioners do not update the network biases and weights up to the point where errors in the training set declines and equals the learning rate hyperparameter. Instead, they give more attention to the issue of model overfitting. As the deep learning neural networks try to generalize well, learning too many details in the training sample may result in a poor forecast performance on unseen data or the validation set.

In order to tackle overfitting we have focused on three methods, namely the i) weight regularization, ii) dropout regularization (a.k.a. the ensemble technique), and iii) early stopping regularization. The first method reduces overfitting by penalizing the large values of weights in the network, since a small change in the input could cause large changes in the output. A common practice in this regard is to transform input variables such that they have the same scale.

Therefore, all input data in our exercise enter the model in a standardized form. In the second method, one modifies the network by either dropping some neurons during training in each iteration, or assigning a probability to all neurons so that they are temporarily ignored from computations. We have chosen the latter technique, by setting the dropout rate to 0.2. The final regularization method is useful when deciding on the number of epochs that can be run before the network estimation starts to overfit. An epoch, in our exercise, is defined as the number of iterations of training data in one cycle for training the model. Since the learning algorithms may take thousands of epochs to minimize the model error to the desired extent, the number of epochs is considered an important hyperparameter in order to make an earlier stopping if the update of parameters does not provide further improvements in forecast errors or the given model is shown over-learned. The network performance on the validation (or test) sample often improves up to a certain point. However, the validation error might start to go up afterwards (defined as the sweet spot in Figure 2) even though the training error could continue to decline. We initially tried up to six hundred epochs. After screening out the model loss values in the training and validation samples we adjusted the number of epochs whenever there were clear signs of under-, or over-fitting.

Figure 2. Curves for training set (dashed line) and validation set (solid line).



3.3 LSTM MODEL SELECTION AND FORECAST PROCEDURE

We now turn to the issue of establishing a neural network's architecture. The level of its complexity varies with the number of neurons and hidden layers that are needed to estimate the best-fitting model. Using a simple model with just a few lags, layers or neurons may result in a poor curve fitting, while using too many of them might cause an overfitting. Until lately, determining the optimal number of neurons and layers remains one of the major difficulties as there is no conventional analytical approach to select to proper NN size and architecture. For example, Blum (1992) offers a rule of thumb to select "somewhere between the input layer size and output layer size", while Boger and Guterman (1997) advocate a broader approach of choosing "as many hidden nodes as dimensions needed to capture 70-90% of the variance [in] the input data." Indeed, neural network estimations allow for a much larger number of parameters than observations. For that reason, we try various combinations from a plentiful number of input lags (p), hidden layers (h) and hidden nodes (n) in order to choose the set that is more appropriate in explaining our data. The neural network candidate structures estimated in the selection procedure are specified as a combination of p = {2, 4, 8, 16}, h = {1, 2, 3, 4}, and n = {16, 32, 64, 128, 256, 512, 1024}. Table 1 displays the list of all LSTM specifications, where the number of trials for input lags pi, hidden layers hi and hidden nodes ni equals 28, 28, and 16, respectively. By keeping the number of nodes unchanged in each hidden layer of a particular model, all possible combinations of different layers and neurons amounted to 112 neural network structures.

Table 1: Summary List of All LSTM Network Specifications

Model No.	NN Structure	Model No.	NN Structure	Model No.	NN Structure	Model No.	NN Structure
1	h1;p2;n16	29	h2;p2;n16	57	h3;p2;n16	85	h4;p2;n16
2	h1;p2;n32	30	h2;p2;n32	58	h3;p2;n32	86	h4;p2;n32
3	h1;p2;n64	31	h2;p2;n64	59	h3;p2;n64	87	h4;p2;n64
4	h1;p2;n128	32	h2;p2;n128	60	h3;p2;n128	88	h4;p2;n128
5	h1;p2;n256	33	h2;p2;n256	61	h3;p2;n256	89	h4;p2;n256
6	h1;p2;n512	34	h2;p2;n512	62	h3;p2;n512	90	h4;p2;n512
7	h1;p2;n1024	35	h2;p2;n1024	63	h3;p2;n1024	91	h4;p2;n1024
8	h1;p4;n16	36	h2;p4;n16	64	h3;p4;n16	92	h4;p4;n16
9	h1;p4;n32	37	h2;p4;n32	65	h3;p4;n32	93	h4;p4;n32
10	h1;p4;n64	38	h2;p4;n64	66	h3;p4;n64	94	h4;p4;n64
11	h1;p4;n128	39	h2;p4;n128	67	h3;p4;n128	95	h4;p4;n128
12	h1;p4;n256	40	h2;p4;n256	68	h3;p4;n256	96	h4;p4;n256
13	h1;p4;n512	41	h2;p4;n512	69	h3;p4;n512	97	h4;p4;n512
14	h1;p4;n1024	42	h2;p4;n1024	70	h3;p4;n1024	98	h4;p4;n1024
15	h1;p8;n16	43	h2;p8;n16	71	h3;p8;n16	99	h4;p8;n16
16	h1;p8;n32	44	h2;p8;n32	72	h3;p8;n32	100	h4;p8;n32
17	h1;p8;n64	45	h2;p8;n64	73	h3;p8;n64	101	h4;p8;n64
18	h1;p8;n128	46	h2;p8;n128	74	h3;p8;n128	102	h4;p8;n128
19	h1;p8;n256	47	h2;p8;n256	75	h3;p8;n256	103	h4;p8;n256
20	h1;p8;n512	48	h2;p8;n512	76	h3;p8;n512	104	h4;p8;n512
21	h1;p8;n1024	49	h2;p8;n1024	77	h3;p8;n1024	105	h4;p8;n1024
22	h1;p16;n16	50	h2;p16;n16	78	h3;p16;n16	106	h4;p16;n16
23	h1;p16;n32	51	h2;p16;n32	79	h3;p16;n32	107	h4;p16;n32
24	h1;p16;n64	52	h2;p16;n64	80	h3;p16;n64	108	h4;p16;n64
25	h1;p16;n128	53	h2;p16;n128	81	h3;p16;n128	109	h4;p16;n128
26	h1;p16;n256	54	h2;p16;n256	82	h3;p16;n256	110	h4;p16;n256
27	h1;p16;n512	55	h2;p16;n512	83	h3;p16;n512	111	h4;p16;n512
28	h1;p16;n1024	56	h2;p16;n1024	84	h3;p16;n1024	112	h4;p16;n1024
Legend: h = hidden layers; p = lags; n = nodes in h.							

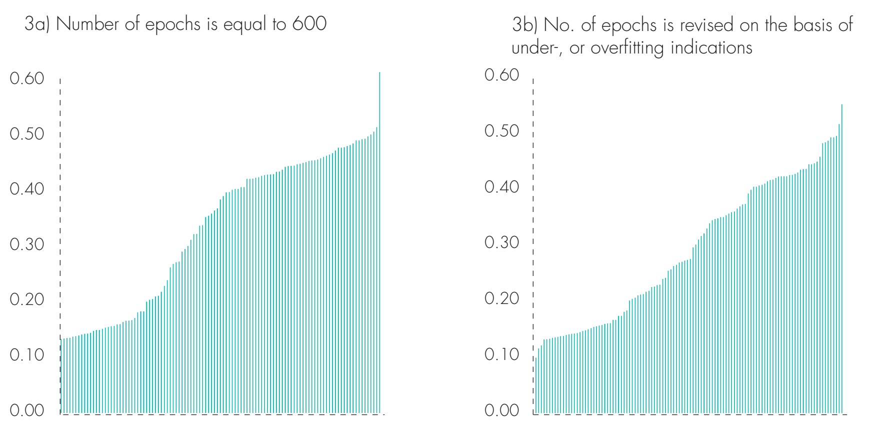
We selected the best model(s) by following the error adjustment strategy, where one ought to choose the combination that yields the lowest out-of-sample forecast error over the validation or test sample. Because our interest is also to examine the relevance of all our preferred variables in forecasting house prices, we proceed with the estimation of hyperparameters as follows: a) split the data into three consecutive samples, namely the training set, validation set, and test set; b) estimate each LSTM specification over a training sample of 60 quarters (1998Q1-2012Q4) and make predictions over the validation sample of 16 quarters (2013Q1-2016Q4); and c) repeat this process for

600 epochs (or adjust its number by stopping at the point where the mean square error (MSE) curve is decreasing for the training set but is starting to increase for the validation set, as discussed above in the early stopping regularization technique). After estimating and validating the hyperparameters of each NN structure, we finally compare their forecasting power that is evaluated in the (out-of-sample) test period consisting of the last 26 quarters in our dataset (2017Q1-2023Q2). The forecast evaluation at this stage will reveal the optimal number of parameters that is necessary for the selection of best model(s).

3.4 WHICH LSTM MODEL TO SELECT?

The quality of the house price predictions has been measured by the root mean square error (RMSE), which is computed on forecasts derived from the test period 2017Q1 to 2023Q2. Following the banking sector stress-testing period conducted at the Bank of Albania, we evaluate the models' ability to forecast house prices over a horizon of two years. Figure 3a-3b shows all LSTM structures ranking from those with the lowest RMSE to those with the highest errors. It appears that model selection is quite an important step to go through for forecasting purposes, as the worst performing models deliver forecast errors that are more than three times higher than the best performers. We needed to revise the number of epochs for about half of estimated models (57 out of 112) in order to satisfy the overfitting concerns. a visual inspection at the RMSE levels in Figures (a) and (b) reveals that adjusting the number of epochs improves the general forecasting power, albeit not significantly.

Figure 3. Ranking of LSTM structures from lowest to highest RMSE



But what do these models tell us about the optimal numbers of input lags, hidden layers and hidden nodes? Are there clear propositions on what to select?

Figure 4 shows the occurrence of time lags (p) in the top (and bottom) 5, 10, 15 and 20 models. Obviously, the lag length of 2 quarters shows a crushing dominance in all groups of top performing networks. Particularly in the best

5 models category, there appears no other lags except for lag number 2. Moreover, the short-term lags up to 4 quarters dominate about 95 percent of all top 20 models (out of 112), indicating that information contained in the last year of data is much more relevant for the current period predictions. This result is quickly confirmed by taking a look at the worst 20 performers in part b, which are dominated by longer lag lengths, such as 8 and 16 quarters.

Figure 4. Distribution of lags (p) in best and worse LSTM performers (in %)

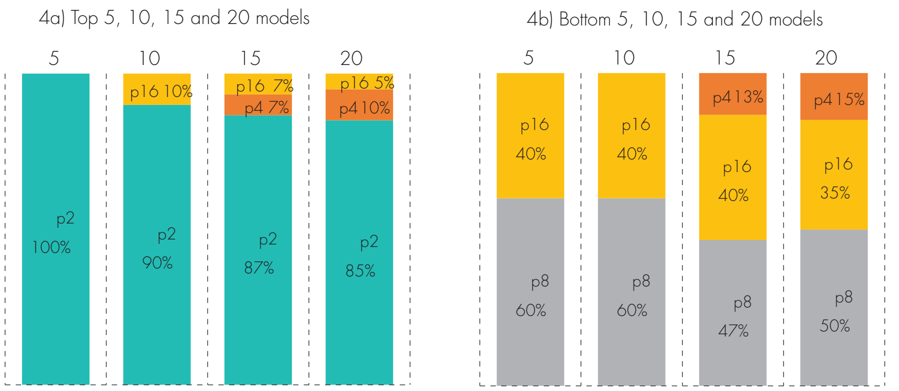
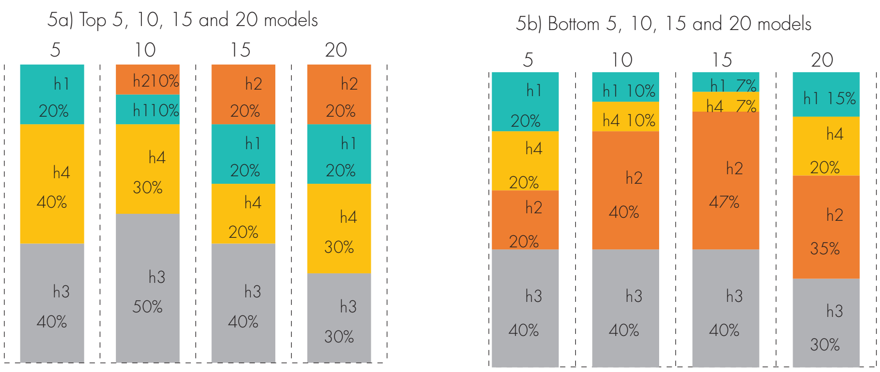


Figure 5. Occurrence of hidden layers (h) in best and worse LSTM performers (in %)



Selecting the number of hidden layers, however, is not so clear cut. A glimpse at the occurrence of hidden layers (h) from 1 to 4 in Figure 5 gives us the impression that using a couple of them could underperform in predicting house prices, as it is suggested by their unsatisfactory share of only 20 percent in the group of best ten models. Thus, a "deeper learning" neural network (defined as more than one hidden layer) might be needed to achieve the full potential of their forecasting power. Interestingly, three hidden layers exhibit the highest occurrence of 40 percent in both, best and worst 15 performers; whereas the occurrence of a single h declines from 20 percent in the best 15 to just 7 percent in the worst 15.

Similarly, testing different numbers of hidden nodes (n = 16, 32, 64, 128, 256, 512, 1024) suggests that a rather two-digit number might be sufficient to obtain the best fitting results (Figure 6). This is particularly so in the top 5 performing models, where the occurrence of 32 and 64 nodes reaches 40 percent for each of them. Enhancing the top list with up to 20 models (out of 112) diversifies the range of nodes by raising the importance of three-digit numbers, such as 256 and 512.

Figure 6. Occurrence of nodes (n) in best and worse LSTM performers (in %)

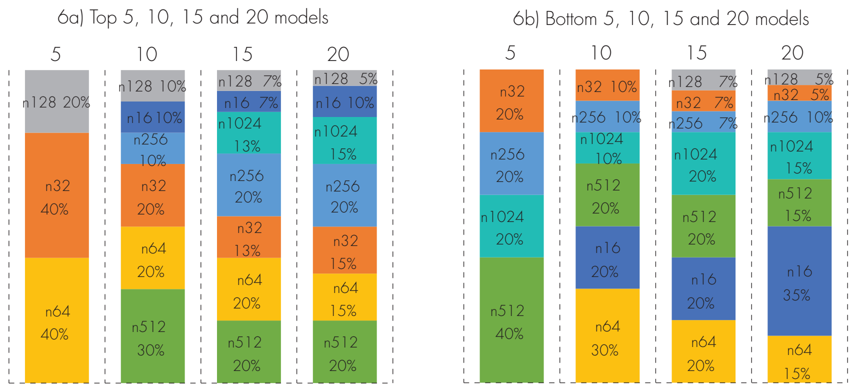


Table 2 lists the top five LSTM models along with their optimal values of the hyperparameters. The quality of house price forecasts in the last out-of-sample 26 quarters is measured by a version of Theil's U-statistic, which compares the loss ratio (RMSE) of each LSTM network with respect to a benchmark forecast derived from a linear autoregressive (AR) model of order p equal to one. A U-statistic below (above) 1 means that neural network models perform better (worse) than the AR(1) predictions. The forecast results show that LSTM networks could be quite competitive against the simpler AR, although a lot of testing is often needed to identify the best neural network(s). As a matter of fact, it is only the top two NNs that could present substantial forecasting gains in comparison to AR, as indicated by their U-statistics of 0.81 and 0.95. Whereas the other neural networks underperform the simpler benchmark by 2 to 9 percent. The relatively low bias and high variance proportions of the selected NNs imply that they may not suffer from underfitting, but perhaps more attention should be paid to achieving the sweet spot, which maintains a state of equilibrium between having a good training fit and the tendency to overfitting.

Interestingly, the best LSTM structure (LSTM(h1; p2; n128)) that makes a significant difference in terms of its relative forecasting power is based on a single hidden layer and 128 nodes, which did not stand out in the discussion on top performers in Figures 3.4.3.a and 3.4.4.a.. This implies that one should strive vigorously by using all possible combinations of these parameters, particularly if his or her main empirical focus is put on forecast accuracy. Nevertheless, the recommended 3 to 4 hidden layers and 32 to 64 nodes (with each having 40 percent occurrence in the top five group) are at least two times more likely to be selected than using h =1 and n = 128 (with a 20 percent occurrence). Consequently, they could be still regarded as a safe play, especially if the NN application is conducted for analytical purposes. As we'll show in the next subsection 3.5, the Shapley values suggest the same explanatory variables with pretty similar contributions on the performance of real house prices in Albania.

Table 2: The Quality of the Best Five LSTM Network Specifications

	U-statistics$)	Bias#)	Variance#)	Epoch no.	Total Params
Benchmark, AR(1)*)	1.00	0.39	0.16		
LSTM(h1; p2; n128)	0.81	0.06	0.64	123	67,592
LSTM(h3; p2; n32)	0.95	0.12	0.70	167	21,256
LSTM(h4; p2; n32)	1.02	0.20	0.68	121	29,576
LSTM(h4; p2; n64)	1.03	0.02	0.49	179	116,488
LSTM(h3; p2; n64)	1.09	0.24	0.49	173	83,464

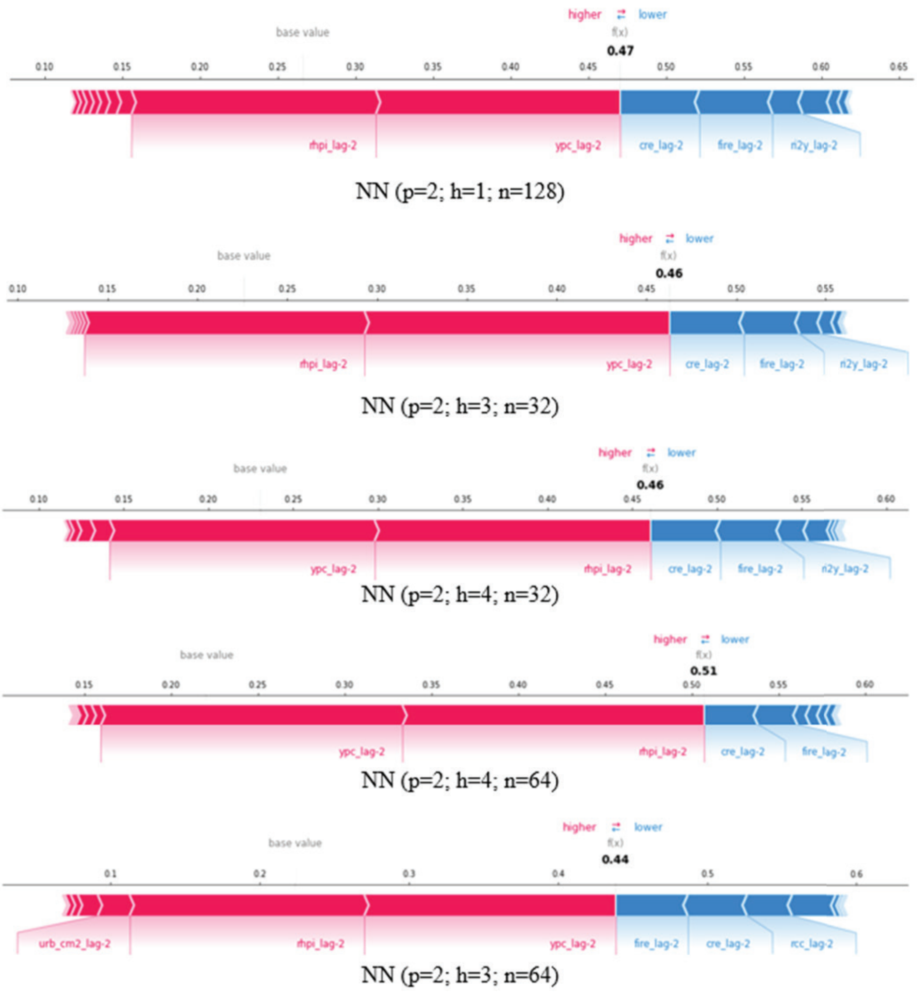
LSTM legend: h = hidden layers; p = input lags; n = nodes in h.
$) Theil's U-statistics is calculated as the ratio of the RMSE of a LSTM forecast to the RMSE of our benchmark AR(1) forecasts.
#) Bias and variance proportions of mean squared forecast errors.
*) The benchmark forecast is derived from a linear autoregressive model of order 1, AR(1). Stationarity tests suggested that RHP is integrated of order I(1), therefore it enters the model in first difference. AR is estimated by the least squares method and the number lags is selected in accordance with Bayesian information criterion. RMSE of the benchmark equals 12.55 for a horizon of 8 quarters.

3.5 WHICH VARIABLES EXPLAIN THE RECENT SURGE IN HOUSE PRICES?

Understanding which variables are the most influential for determining the LSTM model predictions can help us improve upon monitoring and assessing the economic conditions in advance. In search of the theory-driven factors behind the recent run-up in Albania's house prices, we calculate Shapley values which are used to capture the importance of each explanatory variables to the forecasted RHP by quantifying how much each indicator (or feature in the language of machine learning) contributes to the difference between the actual forecast and the mean sample forecast (this is not the difference when a certain variable is removed from the model).

Figure 8 presents the contribution of each input variable including their lags across the top five NN performers over a forecast horizon of two years during the testing set 2017Q1-2023Q2. Each stripe indicates the magnitude of the impact on model prediction (Shapley value). The wider (narrower) a stripe, the more important (less relevant) a feature is. A closer look at the charts reveals that past house price performance together with per capita income are the two essential variables that explain the surge in house prices. Other beneficial indicators could be foreigners' investments in real estate, mortgage loans, construction costs, and the interest rate on 2-year government securities. Perhaps contrary to common perceptions, house rental and urban population growth in relation to the area for new house building permissions are found with only little influence. Last but not least, the second time lag arouses the attention in almost every graph by carrying decisive information for most indicators, while information in the most recent quarter (first lag) generally exhibits little power to determine the outcome.

Figure 7. Identifying the importance of explanatory variables with Shapley values



"Force plotting" the results enables us to look deeper into the impact of each variable in pushing the predicted value farther or closer to average contributions (base value). A red (blue) color stripe indicates that the predictor variable has moved the value towards higher (lower) values. It seems that the two highest contributors – past house price performance and per capita GDP – have put upward pressure on house price forecasts. On the other hand, bank credit, foreign purchases, and government bond rates have contributed on the downward side to the recent house price surge.

4. CONCLUSIONS

The objective of this article was to provide a practical, non-technical introduction to designing a neural network model and demonstrate its application in the field of economics to identify factors behind the recent rapid increase in house prices in Albania. The design procedure was divided in a number of steps, including data preprocessing; training, testing and validation sets; neural network paradigms; and evaluation criteria. Success of neural network applications requires that a researcher must have time, patience and resources to experiments. Despite certain broad rules of thumb that are suggested in empirical neural network literature, much experimentation and imagination is still needed in practice.

We employ Shapley values to distribute a forecast among the selected explanatory indicators in a way that will represent their respective contribution. Our prediction-based analysis demonstrates important role for fundamental indicators in predicting HPI in Albania. Their information content improves upon AR model forecasts in mid-term horizon of two years, in line with most relevant factors identified in Vika and Vika (2023).

ACKNOWLEDGEMENT

The article represents the authors' personal opinions and does not necessarily reflect the views of the institutions where they work. The authors are grateful for the invaluable comments received from the participants at the "17th South-Eastern European Economic Research Workshop" at the Bank of Albania in December 4-5 2023, as well as at the International Scientific Conference on "AI Horizon – A New Era of Creativity and Innovation" organized by the Academy of Sciences of Albania in Tirana, November 8-9, 2024. Special thanks are for Mr Altin Tanku, Head of Research Department at the Bank of Albania, for helpful discussions on house price determinants in Albania.

REFERENCES

Alfaro-Navarro, J.-L.; Cano, E. L.; Alfaro-Cortés, E.; García, N.; Gámez, M.; and Larraz, B. (2020) "A Fully Automated Adjustment of Ensemble Methods in Machine Learning for Modeling Complex Real Estate Systems," Complexity 2020, 5287263.

Banerjee, D. and Dutta, S. (2017) "Predicting the housing price direction using machine learning techniques," Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, India, 21–22, pp. 2998–3000, Sept. 2017.

Behr, D. M.; Chen, L.; Goel, A.; Haider, K. T.; Singh, S.; and Zaman, A. (2023) "Estimating house prices in emerging markets and developing economies: a big data approach," World Bank Group, Policy Research Working Paper 10301, February 2023.

Blum, A. (1997). Neural Networks in C++. New York: John Wiley & Sons, 1992, 1st ed.

Boger, Z. and Guterman, H. (1997) "Knowledge extraction from artificial neural networks models," IEEE Systems, Man, and Cybernetics Conference, Orland, FL, USA, 1997.

Ceh, M.; Kilibarda, M.; Lisec, A. and Bajat, B. (2018) "Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments," ISPRS Int. J. Geo-Inf., vol. 7, 168, 2018.

Cerqueira, V.; Torgo, L. and Mozetič, I. (2020) "Evaluating time series forecasting models: an empirical study on performance estimation methods," Mach Learn, vol. 109, p. 1997–2028, 2020.

Chatzidis, I. V. (2019) "Prediction of Housing Prices based on Spatial & Social Parameters using Regression & Deep Learning Methods," Master's Thesis, University of Thessaloniki, Thessaloniki, Greece, 2019.

Fan, C.; Cui, Z. and Zhong, X. (2018) "House Prices Prediction with Machine Learning Algorithms," Proceedings of the 2018 10th International Conference on Machine Learning and Computing, Macau, China, 26–28 February 2018; pp. 6–10.

Feng, C.; Wang, H.; Lu, N.; Chen, T.; He, H.; Lu, Y.; and Tu, X. M. (2014) "Log-transformation and its implications for data analysis," Shanghai Archives of Psychiatry, vol. 26, no. 2, pp. 105-109, 2014.

Hacıevliyagil, N.; Drachal, K. and Eksi, I. H. (2022) "Predicting House Prices Using DMA Method: Evidence from Turkey," Economies vol. 10 (3), 64, 2022. Available online: https://doi.org/10.3390/economies10030064.

Halili, F. (2022) "The Impact of Macroeconomic Factors on the Change of Residential Prices – The Case Study of Albania," International Journal of Applied Research in Management and Economics, 5 (4), pp. 29-44, 2022.

Ho, W. K. O.; Tang, B.-S, and Wong, S. W. (2021) "Predicting property prices with machine learning algorithms," J. Prop. Res., vol. 38, pp. 48–70, 2021.

Hong, J. (2020) "An Application of XGBoost, LightGBM, CatBoost Algorithms on House Price Appraisal System," Hous. Financ. Res., vol. 4, pp. 33–64, 2020.

Joseph, V. R. and Vakayil, A. (2021) "Split: An optimal method for data splitting," Technometrics, vol. 64, no. 2, p. 166–176, 2021.

Kaastra, I. and Boyd, M. (1996) "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10 (3), p. 215-236, 1996.

Koprencka, L. K.; Bundo, Sh. and Qarri, A. (2014) "Factorial analysis of Albanian housing market," Acta Universitatis Danubius, Economica, Vol. 10, No. 4, 2014.

Li, Y.; Wei, C.; and Ma, T. (2019) "Towards explaining the regularization effect of initial large learning rate in training neural networks," Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 2019.

Lleshaj, Ll. and Korbi, A. (2019) "VAR model for short-run endogenous factors in real estate market: case of Albania," American Research Journal Of Humanities Social Science, Vol. 02, Issue 10, pp. 29-36, October 2019.

Lopes, F.; Leal, A.; Pinto, M. F.; Dourado, A.; Schulze-Bonhage, A.; Dümpelmann, M. and Teixeira, C. (2023) "Removing artefacts and periodically retraining improve performance of neural network-based seizure prediction models," Scientific Reports, vol. 13, p. 1-14, 2023.

Marku, E.; Lleshaj, Ll. and Lleshaj, A. (2020) "VECM analysis of house price index. Case of Tirana," European Journal of Marketing and Economics, Vol. 3, Issue 1, January-June 2020.

Mora-Garcia, R.-T.; Cespedes-Lopez, M.-F.; and Perez-Sanchez, V. R. (2022) "Housing Price Prediction Using Machine Learning Al-gorithms in COVID-19 Times," Land 2022, 11, 2100.

Park, B. and Bae, J. K. (2015) "Using machine learning algorithms for housing price prediction: the case of Fairfax County," Virginia housing data. Expert Syst. Appl., vol. 42, pp. 2928-2934, 2015.

Pisa, I.; Santín, I.; Vicario, J. L.; Morell, A.; and Vilanova, R. (2019) "Data preprocessing for ANN-based industrial time-series forecasting with imbalanced data," 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2019, pp. 1-5, Available online, doi: 10.23919/EUSIPCO.2019.8902682.

Rahutomo, R. and Gunawan, F. E. (2018) "Design deep learning neural network for structural health monitoring," Theoretical & Applied Information Technology, vol. 97, no. 5, pp. 1634–1643, 2018.

Song, H.; Kim, M.; Park, D. and Lee, J. (2020) "How does early stopping help generalization against label noise?" Presented at the ICML 2020 Workshop on Uncertainty & Robustness in Deep Learning, July 2020.

Suljoti, E. (2017) "The implications of financial conditions in the housing market – a cointegrated analysis," Bank of Albania Working Paper 27 (66) 2017.

Tawakuli, A.; Havers, B.; Gulisano, V.; Kaiser, D. and Engel, T. (2024) "Survey: Time-series data preprocessing: A survey and an empirical analysis," Journal of Engineering Research, 2024, ISSN 2307-1877, Available online: https://doi.org/10.1016/j.jer.2024.02.018.

Tian, Y. and Zhang, Y. (2022) "A comprehensive survey on regularization strategies in machine learning," Information Fusion, vol. 80, pp. 146-166, 2022.

Trippi, R. R. and Turban, E. (1993) "Neural networks in finance and investing: using artificial intelligence to improve real-world performance," Probus, Chicago, 1993.

Vika, B. and Tole, Dh. (2023) "Kornizë e të mësuarit të thellë për parashikimin e serive kohore me një apo shumë hapa përpara ," Revista Shqiptare Social Ekonomike, vol. 3 (112), p. 57-70, shtator-dhjetor 2023.

Vika, B. and Vika, I. (2023) "Forecasting house prices in Albania with the deep learning LSTM network," Journal of Economics and Business Issues, vol. 4 (1), pp. 40-55, 2023.

Wang, D. and Li, V. J. (2019) "Mass appraisal models of real estate in the 21st century: a systematic literature review," Sustainability 2019, 11 7006.

Xu, Y. and Goodacre, R. (2018) "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning," Journal of Analysis and Testing, vol. 2, p. 249–262, 2018.

Yzeiraj, E. (2016) "House prices in Albania: deviations from the equilibrium?" Bank of Albania, Working Paper 20 (59) 2016.

Zhang, J.; Yang, Y. and Ding, J. (2023) "Information criteria for model selection," WIREs Computational Statistics, vol. 15, no. 4, 2023.

COMMUNICATIVE TONE OF BANK OF ALBANIA'S MONETARY POLICY REPORT: A SEMANTIC ORIENTATION ESTIMATION

Orion Garo, Research Department, Bank of Albania,

INTRODUCTION

The transmission of a central bank's monetary policy to the wide range of interest rates used in an economy's financial market occurs through a series of interconnected transmission channels. These include: the interest rate channel, the credit channel, the liquidity channel, the asset price channel, the exchange rate channel, as well as the overall stability of the financial system. These channels are directly influenced by the base interest rate set by the central bank.

A channel worth studying is the communication channel that the Bank of Albania (hereinafter referred to as: "the Bank") maintains with the domestic financial and banking market, following the decision-making process for its monetary policy. This communication takes various forms. These include: [1] the Bank's official publications (the annual Monetary Policy Document, the quarterly Monetary Policy Report, Supervisory Council press releases regarding key interest rate updates on the day of meeting, periodic economic analyses covering domestic market macroeconomic developments, etc.); and [2] the unconventional monetary policy instruments at the Bank's disposal (setting required reserve levels, market liquidity management, open market operations, and standing facilities).

Focusing on one of these communication forms: the monetary policy report, we consider important addressing the issue of whether forward guidance is apparent in the Bank's policy-making strategy. Specifically: is there a correlation between the Bank's communication, in terms of informing the domestic financial market, and the subsequent course of monetary policy?
It is worth noting that the publication of the monetary policy report is carried out periodically, on a recurrent basis. The Bank's website stores an archive of these quarterly reports dating back to 2003. Their text content can be utilized to conduct quantitative assessments on the tonality of each report, thereby compiling a quarterly frequency index of a multi-year time span for all archived reports. The purpose of this research is to attain precisely this assessment, in order to determine whether or not the monetary policy report signals, to the domestic economic and banking environment, probable short-term deviations in the key interest rate. The final product of this research is intended to be a quarterly time series of monetary policy's communicative tone, useful for structural analysis and economic forecasting.

As for the methods of this research, the emphasis is placed on obtaining and estimating the data, so as to construct a suitable quantitative indicator of the

Bank's monetary policy communicative tone. The novelty offered in this context relates to a more nuanced reading of the communicative tone within reports, not being satisfied with one-word terms provided through readily-available financial dictionaries, but identifying – through Regex formulations – and calculating the dichotomy of multi-word terms within sentences.

The following content of this research article is structured as follows: The second section presents the data and provides a description of the tools that enable their collection and estimation for the needs of this research. The third section details the results of the index evaluation and the econometric tests for its suitability. The final section lists some concluding remarks. The bibliography can be found at the end of the article.

DATA AND METHODOLOGY

The data were obtained from the contents of the English text version of the Bank's quarterly monetary policy reports. The dataset consists of 86 reports, from beginning of 2003 to mid 2024. The data collection adheres to the principle of dichotomy in semantic analysis (Ko & Chang, 2021), which practically consists of determining the overall semantic orientation (positive or negative) of a written content, factoring in the content's number of words.
For each of sample n's monetary policy reports, a count of one- and multi-word terms, included in the hawkish and dovish categories of monetary policy terminology, was performed.

Simultaneously, a count of the report's total of words was needed, for the hawkish and dovish terms to be reported as numerators of s constant number of words K (as for the case of most studies in the related literature, research kept K at 100 words). The index consisted of n values, each of which was calculated as the ratio of hawkish terms to dovish terms for the constant K. The figure below presents a table of some key descriptive statistics of the dataset:

Figure 1 Tabular presentation of some of the data's descriptive statistics:

	n:	Hawkish terms:	Dovish terms:	No. of words:	TSOI:
Sample:	86				
Sum:		8,171.00	7,230.00	551,486.00	-
Minimum:		11.00	7.00	681.00	-0.38
Maximum:		185.00	185.00	10,564.00	0.53
Mean:		95.01	84.07	6,412.63	0.11
Median:		95.00	73.50	6,320.50	0.04

Source: Publications of the monetary policy report, Bank of Albania

It is worth noting that the calculation of each report's total of words excluded stop words.[1]

[1] *The most frequent words in the text are common words such as determiners, auxiliary verbs, pronouns, conjunctions, etc. These words, known as stop words, usually do not contain much information and also obscure interesting content in the text, due to their high frequency of use. For this reason, these words are usually removed before data analysis is performed. (Albrecht, et al., 2020).*

The main data collection and estimation tool to obtain the above-tabulated statistics was the Python environment and programming language (after the data was collected and TSOI index was estimated, related econometric tests were performed with the SAS program). The data collection steps, automatically followed based on the Python code, are listed below:

1. The corpus[2] of this search was downloaded from the Bank's web page and saved locally. As mentioned above, it consists of a collection of monetary policy reports.
2. Based on the instructions of Albrecht, et al. (2020, pp. 3-10), for ease of access and following the standard grouped data model, the entire corpus, divided by n quarterly reports, was moved to n separate elements of a dataframe.[3]
3. The cleaning of data for the entire corpus was carried out. This step consisted of a text filtering, to strip out of it unnecessary elements and characters such as titles, subtitles, new lines, bullet points, brackets, apostrophes, footnotes, stop words, as well as the written contents of prefaces, tables, figures, graphs, and informative spaces reserved for economic analyses and forecasts. Additionally, parts of the text unrelated to the local economy (regional and global economic developments) were removed.
4. The text was divided into fragments (known in literature as tokens), based on its sentences. This step also includes the machine learning component of this research's methods, and it was necessary to perform the data estimation. In order for any semantic orientation phrases to be identified in the sentence, the report text must first be fragmented according to the sentences. This is because the algorithm may identify phrases that start in one sentence and end in the next, consequently, calculations of the terms' totals may result in errors.

The execution of these steps required the use of the Pdfplumber, Pandas, Regex, and NLTK libraries. The automatic procedure continued with steps for estimating the collected data, which consists of the text filtered as per the aforementioned data collection steps. The data estimation steps are listed below:

1. The total number of words in the text was calculated.
2. The total number of hawkish and dovish terms was calculated. It is worth clarifying that a "term" refers to: [1] any word having a positive or negative connotation and, [2] any phrase categorized as hawkish or dovish. Lists of hawkish and dovish words were provided in the form of dictionaries, published by Loughran & McDonald (2011), Correa, et al. (2021), and Gonzalez & Tadle (2021).[4] The novelty this research brings

[2] *A term borrowed from linguistics; it represents the collection of texts of all reports used for this research. From a statistical viewpoint, it is simultaneously the sample and the population of our data.*

[3] *Equivalent to a MS Excel spreadsheet.*

[4] *In line with the methodology used by Chong and Ho (2022), from the dictionaries of Loughran and McDonald (2011), Correa, et al. (2021), Gonzalez dhe Tadle (2021), the words with positive semantic orientation were categorized under hawkish terms, while those with negative semantic orientation were categorized under dovish terms.*

to the existing domestic literature is the identification and inclusion of hawkish and dovish phrases in the text. To determine the total number of hawkish terms, the total number of hawkish words was added to the total number of hawkish phrases. The same was done to calculate the total number of dovish terms.

3. The total number of false-positive and false-negative terms was calculated. These are one-word hawkish and dovish terms preceded by negative prefixes such as: 'can not', 'unlikely to', 'despite', 'neither / nor', etc. After the calculation, the total of false-positive terms was subtracted from the total of corresponding hawkish terms, while the total of false-negative terms was considered zero.[5]
4. The estimation of the TSOI index was carried out adhering to the theoretical approaches of Correa, et al. (2021) and Rutkowska & Szyszko (2024), as specified below:

$$IOST_t = \frac{K * \left(\frac{P_t}{(P_t + N_t)} - \frac{N_t}{(P_t + N_t)} \right)}{T_t} \quad (1)$$

where, for the written content of each report t:

- P is the total of hawkish terms;
- N is the total of dovish terms;
- T is the words' total, and;
- K is a constant.

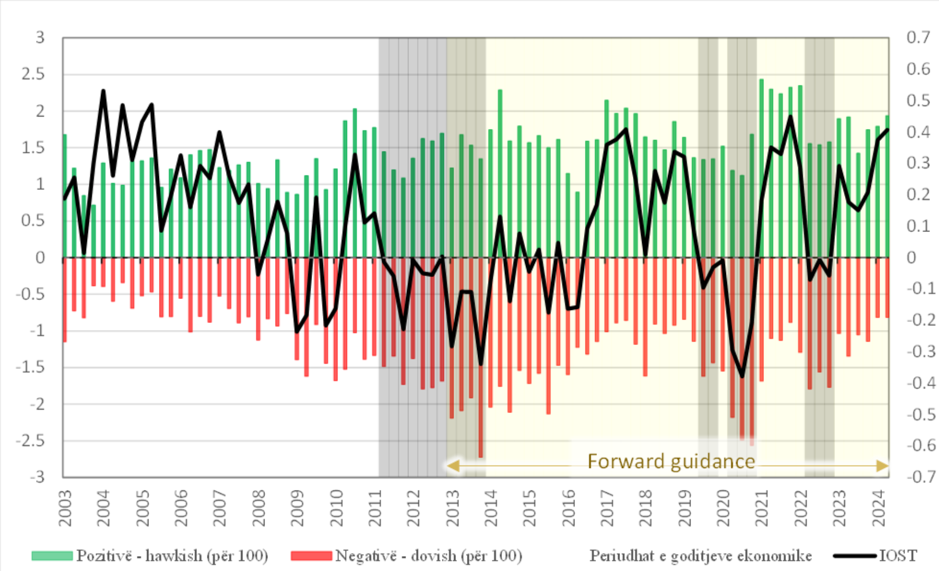
As mentioned above, the constant term K was kept at 100 words. K allows the formula to maintain proportionality between voluminous-content and compact-content monetary policy reports.

RESEARCH RESULTS

Based on the above methodology, specifically the estimations according to specification (1), we have constructed the TSOI index for the period 2023-2024, as presented in Figure 2. TSOI reports the net percentage of the hawkish semantic tone of the report and varies between the values -1 and 1 (Rutkowska & Szyszko, 2024). It was chosen to be scaled this way for ease of graphical interpretation, as when the index takes a value above zero, it means that the hawkish communication tone dominates the content of that particular quarterly report, and vice versa. The figure also details the number of hawkish terms (green bars) and dovish terms (red bars) – for K = 100 words – in the the report's content.

[5] *It was acted upon in this manner because negations of positive terms take on a negative semantic orientation, while negations of negative terms take on a neutral semantic orientation. (Correa, et al., 2021).*

Figure 2: Graphical presentation of TSOI's trend over time, and of the dichotomous terms.



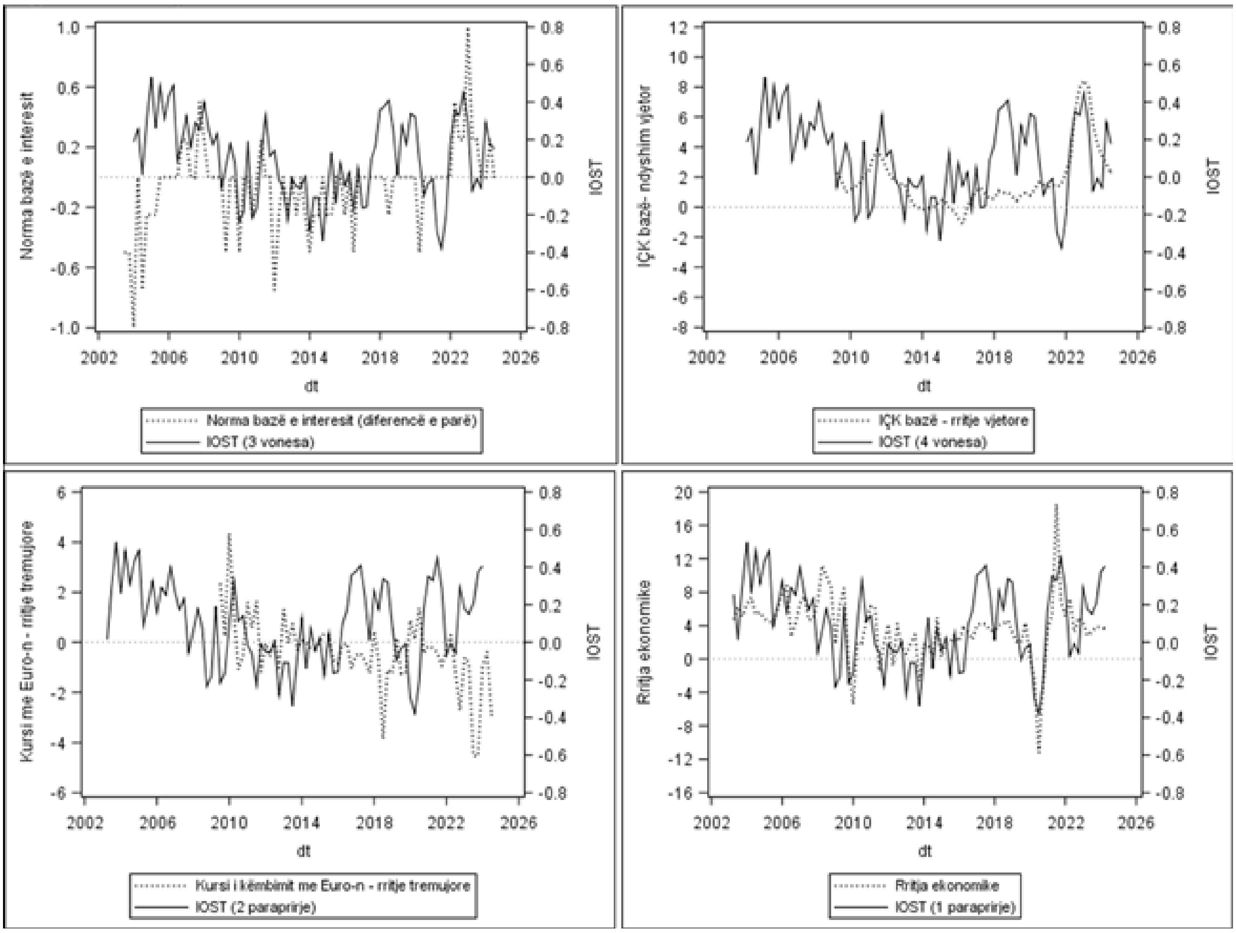
Source: Publications of the monetary policy report, Bank of Albania

The vertical gray stripes in the graph represent the strongest shocks that the Albanian economy has suffered during the 20-year period under analysis. When the country's economic conditions are stable and favorable, TSOI generally fluctuates above the value of 0, but it is observed to consistently fall below 0 when these conditions deteriorate. After conducting the relevant tests, it is worth noting that this series results stationary and does not display any seasonality. In the graph, a structural break of the index's trend around 2013 is immediately noticeable, an episode coinciding with the consequences faced by the Albanian economy after the global financial crisis and the European public debt crisis. At the beginning of the series, compared to its average, TSOI fluctuates in relatively high positive values. This coincides with the suitable conditions, and favorable economic environment, prevailing in the years 2003-2007. However, by 2013, this index is observed to show a pronounced downward trend. The 2011-2013 period results in the lowest partial of this index: an average of -0.13. 2013 marked the beginning of the implementation of forward guidance by the Bank of Albania (Hoxholli, 2018). The graph shows that, starting from this year, TSOI's trend gradually moves upward, and during the last two quarters, the index shows values comparable to those at the beginning of the series.

A significant part of the results is the carryout of tests aimed at establishing (the presence or absence of) correlations between TSOI and a set of financial and economic aggregates. The presence of correlations would provide evidence that monetary policy communication does signal the domestic economic and banking environment on possible future changes of the key interest rate. Figure 3 presents graphically how TSOI correlates with some of the key economic indicators, which include: [1] the key interest rate, [2] core inflation, [3] the Euro exchange rate,[6] and [4] economic growth.

[6] *It is worth clarifying that for the key interest rate, core inflation, and the Euro exchange rate, we have performed monthly data averaging to transform them into quarterly data, with the aim of adapting to TSOI's quarterly frequence.*

Figure 3: Graphs of TSOI's correlation with a set of financial and real indicators.



Source: INSTAT; publications of the monetary policy report, Bank of Albania.

Figure 4 provides further details of these relationships, presenting statistical outputs of IOST's correlation to these indicators.

Figure 4:
A. Tabular presentation of the results of Pearson's correlation tests between IOST and the core interest rate

Variable:	Transformation:	TSOI's lag/lead with which the variable displays the strongest correlation:	Timespan: 2003-2024 (full) Vrojtime:	Koeficienti:	P-value:
Key interest rate	first difference:	3rd lag	83	0.262	0.016 (**)
	annual growth (yoy):	4th lag	82	0.236	0.033 (**)

Variable:	Transformation:	Timespan: 2003-2012 Data points:	Coefficient:	P-value:	Timespan: 2013-2024 (forward guidance) Data points:	Coefficient:	P-value:
Key interest rate	first difference:	39	0.248	0.138	46	0.362	0.013 (**)
	annual growth (yoy):	36	0.209	0.22	46	0.416	0.004 (***)

B. Tabular presentation of the results of Pearson's correlation tests between IOST and a selection of financial and real economic indicators of the domestic economy.

Variable:	Transformation:	TSOI's lag/lead with which the variable displays the strongest correlation:	Timespan 2003-2024		
			Data points:	Coefficient:	P-value:
Core CPI,	annual growth (yoy):	4th lag	62	0.359	0.004 (***)
Euro exchange rate,	quarterly growth (qoq):	2nd lead	59	-0.402	0.002 (***)
Gross domestic product	annual growth (yoy):	1st lead	85	0.544	<0.001 (***)

Source: Publications of the monetary policy report, Bank of Albania

Specifically, Figure 4/A presents statistical outputs for two tests between TSOI and the core interest rate (which represents the reaction of the domestic banking market), in transforming the latter as: [1] first difference and as [2] annual growth. The tests were conducted for the full timespan (2003 – 2024) analyzed, as well as for each of its two segments: the one before (2003 – 2012), and the one after the Bank's implementation of forward guidance (2013 - 2024). Several features of the correlations stand out:

1. For the full timespan taken into analysis, it is observed that the correlations result in being statistically significant. For the time segment before the implementation of forward guidance, the correlations are absent (do not fall within the 5% significance threshold). Meanwhile, for the time segment coinciding with the Bank's implementation of forward guidance (2013-2024), these two correlations are within the 5% significance threshold. This finding shows that the correlation between signaling—by the Bank—about the future direction of its monetary policy and the key interest rate is present and coincides in time with the beginning and continuation of the implementation of forward guidance by the Bank.
2. The core interest rate shows the strongest link to TSOI's lags, precisely to the lags of 3-4 quarters. This result implies that the reaction of the domestic banking market to the Bank's signals regarding the key interest rate dynamics occurs 3 quarters to 1 year later.
3. The above correlations have positive signs. This shows that the short-term changes in TSOI and those in the key interest rate exhibit the same direction. Therefore, the dominance of hawkish tones in the monetary policy report signals positive short-term changes in the key interest rate 3-4 months later, and vice versa.

Figure 4/B shows the statistical outputs for TSOI's correlations with the core inflation, the Euro exchange rate, and economic growth. These results bring into focus a broader framework of the reciprocal relationship between the Bank's communication channel and domestic economic developments in general. The tests were conducted for the full timespan (2003–2024). It is observed that all TSOI's relationships to these indicators result within the 1% statistical significance threshold, which is evidence of the co-dynamics of the Bank's communicative tone with the domestic financial and real economic activity.

These correlations of TSOI with the economy aggregates highlight the role of the Bank's communication channel in informing the domestic financial market about the future course of monetary policy.

CONCLUDING REMARKS

In conclusion, we emphasize that this research was dedicated to analyzing the communicative tone present in the Bank's quarterly monetary policy reports, with the objective of estimating a time series of the semantic orientation of this tone, useful for economic analysis and forecasts in the future.

The research data consisted on English version text contents of the quarterly monetary policy reports published by the Bank; they constitute a sample of 86 data-points, from the beginning of 2003 to mid-2024. The methodology belongs to semantic orientation analysis, where the specific approach consists of the dichotomous division of terms. Through it, an assessment of the overall sentiment present in the text—hawkish or dovish—is made possible by calculating the totals of words and phrases.

The main tools for data collection and analysis consisted of the Python programming language and environment. With the aim of ensuring estimation accuracy, the methodology included a process for eliminating stop words from the total number of words, and used a constant comparative base of 100 words to calculate the identified terms and their ratio in each text content. The product of the analysis was an index calculated from the ratio of positive (hawkish) terms to negative (dovish) terms identified in the text content.

The process of collecting the data was a crucial part in this research. It involved the steps of: downloading and locally saving monetary policy reports, organizing the data into a dataframe by quarter, cleaning the content from unnecessary text elements, and segmenting it according to sentences. This process was succeeded by the estimation of the index, which consisted of: calculating the total number of dichotomous terms for each text content, identifying positive and negative terms (including words and phrases), identifying false-positive and false-negative terms and subtracting them from the respective totals, and calculating the final index based on specification (1), adapted according to the theoretical approaches of the referenced literature.

Regarding the communicative tone of the Bank's monetary policy, the attained results highlighted several attributes.

- Firstly, the TSOI series is characterized by stationarity. This is because, during the dynamics of monetary policy communication over time, no long-term increasing or decreasing trends are expected in the use of positive or negative tones present in the terminology of reports. These tones are relatively uniform, and their polarity is entirely dependent on the time positioning and the gravity of economic developments in the country. Consequently, analyses of the potential correlations of this series with other economic indicators will focus on their common short-term changes.
- Secondly, the grafically presented data show that the negative tones of monetary policy communication dominate during periods of decline and economic tension, while positive tones are present during periods

of growth. This dynamic precedes in time and aligns with the short-term trends of interest rates in the domestic banking market. The tests confirmed a one-way Granger causality from TSOI to the base interest rate. These findings on TSOI's behavior confirm that the communication of the Bank of Albania does signal the future course of monetary policy.

- Thirdly, the tests of the TSOI series were carried out for two time segments separately: from the beginning of the series until 2012, and from 2013 onwards. For the first time segment, it resulted that the communicative tone of monetary policy did not show a statistically significant correlation with the key interest rate. For the second time segment, which coincides with the policy decision by the Bank to apply forward guidance, it resulted as expected: the communicative tone of monetary policy showed a correlated and leading dynamic with respect to the key interest rate.
- Fourthly, the dynamics of the communicative tone of monetary policy aligns with some of the key economic indicators. Specifically, a positive correlation is observed between: TSOI's 1-quarter delay and economic growth, as well as; TSOI's 2-quarter delay and core inflation. Meanwhile, the relationship between TSOI and the exchange rate with the Euro results in a negative correlation. TSOI's dynamics traces that of the exchange rate with a 2-quarter delay. This finding highlights the synchronicity of monetary policy communication with the overall economic developments.

REFERENCES:

Albrecht, et al. (2020). Blueprints for Text Analytics Using Python. Authors: Albrecht, Jens; Ramachandran, Sidharth; Winkler, Christian. Sebastopol: O'Reilly Media, Inc.

Chong, E., & Ho, S.-J. (2022). Measuring Text-Based Sentiments from Monetary Policy Statements. Në Bank of Malaysia (Re.), Data Science in Central Banking (fv. 1-7). Rome: Bank of Italy.

Correa, et al. (2021). Sentiment in Central Banks' Financial Stability Reports. Review of Finance, 85-120. Authors: Correa, Ricardo; Garud, Keshav; Londono, Juan M.; Mislang Nathan. doi:10.1093/rof/rfaa014

Gonzalez, M., & Tadle, R. C. (2021). Sentiment comparisons on monetary policy releases from inflation-targeting regimes. 5th BIS-CCA Ressearch Network Conference (fv. 1-37). Bank for International Settlements.

Hoxholli, R. (2018). Challenges to monetary policy conduct in Albania. Gjetur në Central Bank of Bosnia website.

Ko, C.-R., & Chang, H.-T. (2021). LSTM-based sentiment analysis for stock price forecast. PeerJ Computer Science, 7(408), 1-23. doi:http://dx.doi.org/10.7717/peerj-cs.408

Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66, 1. doi:10.1111/j.1540-6261.2010.01625.x

Rutkowska, A., & Szyszko, M. (2024, May 17). Dictionary-based sentiment analysis of monetary policy communication. Qual Quant, 1-24. doi:10.1007/s11135-024-01896-9