WORKING PAPER

DEEP LEARNING-BASED FORECASTING OF CURRENCY IN CIRCULATION

Ivan Popović

Available at www.nbs.rs

The views expressed in the papers constituting this series are those of the author(s), and do not necessarily represent the official view of the National Bank of Serbia.

Monetary and FX Operations Department

NATIONAL BANK OF SERBIA

Belgrade, 12 Kralja Petra Street

Telephone: (+381 11) 3027 100

Belgrade, 17 Nemanjina Street

Telephone: (+381 11) 333 8000

www.nbs.rs

Deep Learning-Based Forecasting of Currency in Circulation

Ivan Popović

Abstract: To improve the quality of daily forecasts of currency in circulation (CIC), four deep learning models were developed. The performance of these models was systematically compared with that of the existing ARIMAX (65,1,65,18) model, adjusted through expert judgment. The analysis covers the period from January 2025 to March 2026 and is based on a time series comprising 5,497 daily observations. The results indicate that the ARIMAX model achieved superior forecasting performance over the analysed period. However, the LSTM model with L1+L2 ElasticNet regularisation (Model 3) attained a higher coefficient of determination when January 2025 was excluded from the analysis. This period was characterised by an unforeseen change in pension disbursement dynamics. The paper presents the mathematical foundations of both modelling approaches, provides a detailed description of the developed model architectures, and investigates the phenomenon of overfitting.

Key words: LSTM, ARIMAX, currency in circulation, machine learning, deep learning, overfitting, regularisation, ElasticNet
[**JEL Code]:** C22, C45, C53, E41, E51.

Non-technical summary

The subject of this study is the forecasting of currency in circulation (CIC), one of the key autonomous factors influencing the liquidity of the banking sector. Demand for cash exhibits predictable patterns, typically increasing ahead of weekends, at the beginning of each month, and during holiday periods. However, it is also susceptible to sudden and difficult-to-predict surges, triggered by events such as changes in the timing of pension or social benefit disbursements introduced by government authorities.

The traditional statistical approach to forecasting CIC, based on the ARIMAX model, requires the analyst to specify in advance all factors considered relevant to cash demand dynamics, drawing on prior experience. Such factors typically include public holidays, payment schedules, seasonal patterns, and similar. This approach is well established, transparent, and capable of producing forecasts that are readily interpretable.

Deep learning models based on Long Short-Term Memory (LSTM) neural networks operate differently – they independently learn patterns from large volumes of historical data, without predefined rules. Their principal advantage lies in the ability to capture complex dependencies that may not be readily identifiable by the analyst. However, a key question remains whether such models can maintain robust predictive performance when confronted with previously unseen events – changes they were unable to learn from historical data.

The traditional statistical model incorporates the previous 65 days of currency-in-circulation dynamics and relies on 18 predefined factors identified as significant by the analyst. These include payment dates, public holidays, beginning-of-month effects, and similar.

Four neural network models were designed with progressively more stringent safeguards against overfitting, a common limitation of AI models whereby a model achieves excellent performance on the historical training data but exhibits poor generalisation when exposed to previously unseen situations. Each successive model incorporated additional mechanisms aimed at mitigating this phenomenon, ranging from a baseline to the most complex architecture featuring nine safeguards.

The traditional forecasting model, supplemented by analysts' adjustments, proved to be the most reliable over the evaluation period. The neural network models gravely underperformed in January 2025, when the pension disbursement schedule was modified – an unforeseen event which could not be anticipated by the models. When this period is excluded from the evaluation, the neural network models achieve better predictive performance, suggesting that, given sufficiently representative historical data, they have the potential to outperform the conventional approach.

Particular attention should be given to Model 3, the neural network architecture that incorporated additional mechanisms to prevent overfitting. As a result, the model demonstrated a greater capacity to generalise to previously unseen situations rather than merely fitting patterns observed in historical data.

Contents

1 Introduction

The management of daily banking sector liquidity is one of the core operational responsibilities of every central bank. Within the framework of autonomous factors, CIC is assessed on a daily basis to enable timely responses to changes in banking sector liquidity.

The time series of daily changes in CIC exhibits three structural characteristics that make it particularly challenging to model: (1) pronounced multi-scale seasonality, including weekly, monthly, and annual cycles; (2) the presence of impulse shocks associated with public and religious holidays; and (3) occasional structural changes driven by changes in fiscal and social policies, such as modifications to pension disbursement schedules, unexpected economic shocks, and shifts in household behaviour.

Traditional econometric models of the ARIMA class, and especially their extended ARIMAX variants incorporating exogenous variables, have long represented the standard approach to forecasting cash flows in central banking operations. Advances in machine learning, particularly the development of deep learning architectures for time-series analysis, have raised the question of whether Long Short-Term Memory (LSTM) models can match or outperform conventional econometric approaches in this specific forecasting task.

The LSTM deep learning models were developed in the Python programming language. Artificial intelligence tools, including Claude and ChatGPT, were utilised as coding assistants.

The present study has three main objectives: (1) to develop and systematically evaluate four LSTM-based models tailored to forecasting daily changes in CIC; (2) to compare their performance with that of the ARIMAX (65,1,65,18) model currently employed in operational forecasting; and (3) to formulate practical recommendations regarding model selection for CIC forecasting based on the empirical findings. Particular attention is devoted to the problem of overfitting – a phenomenon in which a model fits the training data extremely well but exhibits poor generalisation to unseen data – which represents a key factor in assessing the practical applicability of LSTM models. The structure of the paper is as follows: Section 2 presents the theoretical framework and a review of the relevant literature; Section 3 describes the mathematical formulations of both approaches; Section 4 presents the developed models and empirical results; Section 5 provides a comparative analysis; and Section 6 concludes the study.

2. Theoretical framework and review of relevant literature

2.1 Liquidity forecasting in central banking

Since the mid-1990s, central banks have intensively developed quantitative models for forecasting autonomous liquidity factors, including currency in circulation. Pioneering studies by Bindseil et al. (2002) and Whitesell (2006) systematised approaches to liquidity forecasting

in the euro area, identifying seasonality and calendar effects as the dominant sources of predictable variability.

In the context of autonomous factors, seasonality in CIC is particularly pronounced. The end of December and the beginning of January, coinciding with pension disbursements, generate fluctuations in CIC that can be an order of magnitude larger than typical daily variations. Modelling this effect requires either explicitly defined artificial variables or model architectures capable of automatically identifying multiple seasonal cycles.

2.2 ARIMA and ARIMAX models in time series forecasting

Box and Jenkins (1970) laid the foundations of the modern approach to time series modelling through the ARIMA (AutoRegressive Integrated Moving Average) framework. The model combines an autoregressive component, differencing to achieve stationarity, and a moving average component, forming a flexible tool for linear modelling of stochastic processes.

The extension to ARIMAX (ARIMA with exogenous variables), formalised by Hamilton (1994), enables the incorporation of deterministic seasonal effects and exogenous shocks directly into the model structure. This feature makes ARIMAX particularly well suited for CIC forecasting, where the influence of holidays and calendar-related factors is dominant and can be precisely quantified through statistical testing of the coefficients associated with exogenous variables.

Studies by Brockwell and Davis (2002) and Lütkepohl (2005) document that ARIMAX models achieve high predictive accuracy on seasonal macroeconomic time series when exogenous variables are appropriately specified. The main limitation of this approach lies in its assumption of linear relationships and the requirement that all relevant variables be manually specified.

2.3 LSTM networks in economic time series forecasting

Hochreiter and Schmidhuber (1997) introduced the LSTM architecture as a solution to the vanishing gradient problem in conventional recurrent neural networks. The gating mechanisms of the LSTM cell enable selective retention of long-term dependencies, which is crucial for economic time series with long memory.

Fischer and Krauss (2018) showed that LSTM networks outperform classical statistical models in forecasting S&P 500 stock returns, particularly during periods of market turbulence. However, Makridakis et al. (2018), in the fourth international forecasting competition (M4), documented that hybrid approaches combining statistical models and machine learning consistently outperform both methods when used individually.

In the context of central bank cash flow forecasting, Papadimitriou et al. (2019) applied an LSTM model to daily liquidity flows of the European Central Bank and reported improvements over ARIMA models for forecasting horizons shorter than five days. Huang et al. (2020) showed that regularisation techniques – particularly Dropout and L2 regularisation

– have a critical impact on the generalisation performance of LSTM models when applied to short financial time series.

The key open research question is the extent to which LSTM models can robustly generalise in the presence of structural breaks in the time series – a setting in which classical econometric models with expert adjustments have demonstrated advantage in empirical studies.

2.4 Overfitting and regularisation in deep learning

Overfitting – a situation in which a model fits the training data too closely, including random noise, while performing poorly on new observations – is a fundamental challenge in applying deep neural networks to economic time series with limited numbers of observations. Srivastava et al. (2014) introduced Dropout regularisation as an effective mechanism for mitigating overfitting, while Ioffe and Szegedy (2015) developed Batch Normalisation, which stabilises the distribution of activations across layers.

Zou and Hastie (2005) formalised Elastic Net regularisation as a combination of L1 (Lasso) and L2 (Ridge) penalties. L1 regularisation promotes sparse solutions by eliminating irrelevant parameters, whereas L2 regularisation prevents parameter explosion. The combination yields a model that is simultaneously parsimonious, stable, and robust to multicollinearity in the input variables.

3. Mathematical formulation of the model

3.1 ARIMAX model

3.1.1 Basic structure

Let y_t denote the daily change in CIC at time t. The ARIMAX (p,d,q) model with exogenous variables X_t is defined as follows:

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t + \beta' X_t \quad (3.1)$$

where:

$\Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p$ – is the autoregressive (AR) polynomial of order p in the lag operator L, representing the dependence of the process on its own lagged values;

$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$ – is the moving average (MA) polynomial of order q, which captures dependence on past forecast errors;

$(1-L)^d$ – is the differencing operator of order (d), which transforms the time series into a stationary process;

$\varepsilon_t \sim WN(0, \sigma^2)$ – is a white noise process with zero mean and constant variance;

X_t – denotes the vector of exogenous variables (artificial seasonal variables, holidays);

β' – the vector of coefficients associated with the exogenous variables.

3.1.2 Stationarity and identification

The stationarity condition of the ARIMAX model requires that all inverse roots of the polynomial $\Phi(z)$ lie inside the unit circle in the complex plane, i.e. $|z| < 1$ for all z such that $\Phi(z) = 0$. Similarly, invertibility requires that all inverse roots of the polynomial $\Theta(z)$ lie inside the unit circle.

The ARIMAX (65,1,65,18) model was selected based on a comparative analysis of the coefficient of determination and the AIC/BIC criteria across a range of competing models with different orders of the AR and MA components. For the ARIMAX (65,1,65,18) model analysed in this study, d=1 implies that daily changes in CIC are modelled (first differences), rather than the level of the series. This is consistent with the empirical finding that CIC is an integrated process of order one, I(1), as confirmed by the Augmented Dickey-Fuller (ADF) test. The autoregressive component (p=65) and the moving average component (q=65), together with 18 exogenous variables, provide the model with sufficient capacity to capture seasonal dynamics extending up to a three-month horizon.

$$\Delta y_t = (1-L)y_t = y_t - y_{t-1} \quad (3.2)$$

where Δy_t represents the daily change in CIC, measured in millions of dinars.

3.1.3 Exogenous variables

In the ARIMAX (65,1,65,18) model, the following categories of exogenous variables are included:

- Weekly effects (week-of-year, four binary variables): W1, W6, W10 and W48, where $W\_j = 1$ if it is the first week of the year;
- Monthly effects (one binary variable): Secondw, indicating the second week of the month;
- Holiday effects (13 variables): binary variables for national and religious holidays;

$$X_t = [W1, W6, W10, W48, SECONDW, NY\_1A, NY\_4A, NY\_5A, CHR\_2A, EH\_1A, EH\_2A, TW1, TW4, WT1, WT4, MT2, MT3, MT4] \quad (3.3)$$

A total of 18 statistically significant exogenous variables were selected based on a p-value threshold of ≤ 0.10 (10% significance level), ensuring that each variable contributes to explaining the variance of the dependent variable.

The selection of exogenous variables was performed in multiple iterations. In the first step, each candidate variable was tested for statistical significance within a simple regression model. In the second step, variables were included sequentially according to their level of statistical significance, with multicollinearity assessed using the Variance Inflation Factor (VIF) test. In the third step, the final specification was evaluated using the Akaike Information Criterion (AIC) and the Bayesian Schwarz Criterion (BIC):

$$AIC = -2\ln(L) + 2k \quad (3.4)$$

$$BIC = -2\ln(L) + k\ln(n) \quad (3.5)$$

where L is the maximum value of the likelihood function, k is the number of parameters, and n is the number of observations. Minimisation of the AIC/BIC criteria ensures an optimal balance between model fit and model simplicity (parsimony).

3.1.4 Residual diagnostics of the ARIMAX model

The validity of the ARIMAX model is assessed through a series of diagnostic tests applied to the residuals $\hat{\varepsilon}_t = y_t - \hat{y}_t$:

Durbin-Watson test: This test examines the presence of first-order autocorrelation in the residuals. A value of DW $\approx$ 2 indicates no autocorrelation; DW < 2 suggests positive autocorrelation, and DW > 2 indicates negative autocorrelation. For the ARIMAX (65,1,65,18) model, DW = 2.02, indicating satisfactory residual properties.

Ljung-Box Q test: This tests H_0: $\rho_1 = \rho_2 = ... = \rho_m = 0$ for all lags up to m. Failure to reject H_0 confirms that the residual white noise and in fact all seasonal structure are adequately captured by the model.

Jarque-Bera test: The normality of the residual distribution is assessed based on skewness and kurtosis coefficients. Mild deviations from normality are typical for high-frequency financial time series.

ARCH test: The ARCH test examines the presence of conditional heteroskedasticity, a situation in which the variance of the error terms is not constant over time. The presence of ARCH effects may require extending the model with a GARCH component.

All diagnostic tests confirm the adequacy of the ARIMAX (65,1,65,18) model specification: DW = 2.02 (no autocorrelation), the Q test fails to reject the null hypothesis (H_0) up to lag 100, and the residual distribution is approximately symmetric. These results suggest that the model adequately captures the seasonal structure of CIC.

The ARCH test confirms statistically significant presence of conditional heteroskedasticity in the residuals (F – statistic = 2.39, p = 0.0049), although its practical impact on model reliability is limited. This is further supported by the sample size of 711 observations, which provides sufficient asymptotic justification for absorbing the observed form of conditional heteroskedasticity.

3.2 LSTM model

3.2.1 Architecture of the LSTM cell

The LSTM cell is the fundamental building block of recurrent deep neural networks equipped with gating mechanisms. At time step t, given the input vector x_t and the previous hidden state h_{t-1}, the LSTM cell computes the following (see Diagram 1):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.4)$$

Input gate determines which new information is written into the cell state.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.5)$$

Forget gate: determines the proportion of the previous cell state that is retained.

$$g_t = \tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \quad (3.6)$$

New cell state: new candidate values used for updating the memory.

$$C_t = f_t \odot C_{t-1} + i_t \odot g_t \quad (3.7)$$

Cell state update: a combination of retained past information and newly proposed candidate values.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.8)$$

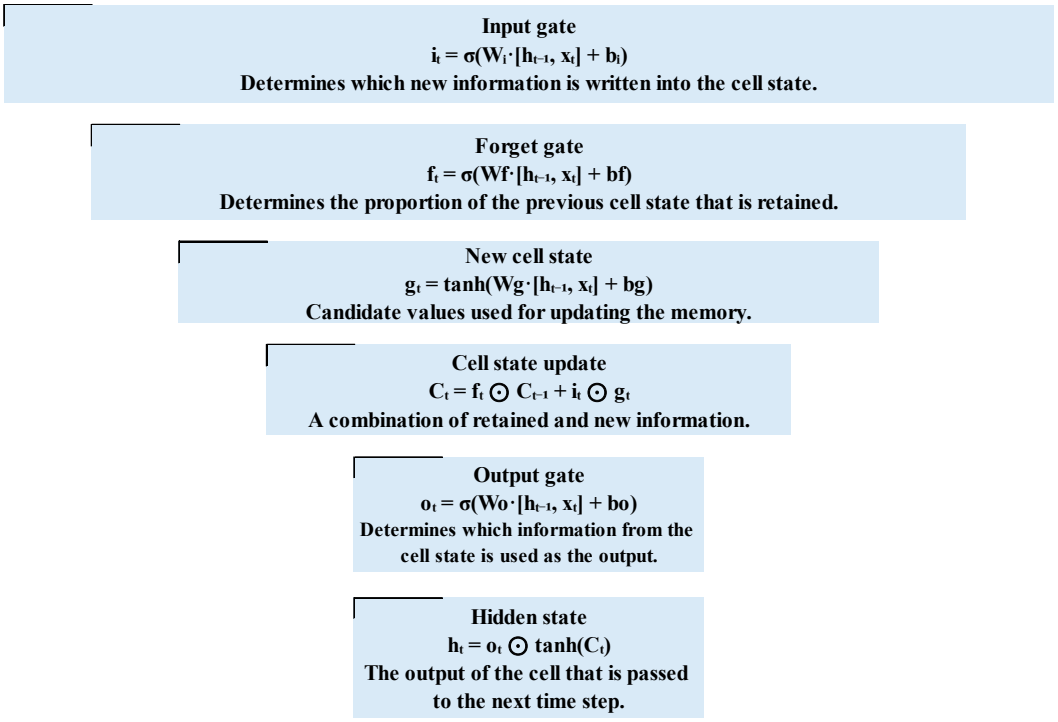
Output gate: determines which information from the cell state is used as the output.

$$h_t = o_t \odot \tanh(C_t) \quad (3.9)$$

Hidden state: the output of the cell that is passed to the next time step.

where: **σ(·)** – denotes the sigmoid function; **tanh(·)** – the hyperbolic tangent function; ⊙ – element-wise (Hadamard) multiplication; **W_f, W_i, W_g, W_o** – the weight matrices of the gates; **b_f, b_i, b_g, b_o** – bias vectors.

Figure 1 **Architecture of the LSTM cell**



3.2.2 Dense layers and activation functions

The LSTM layers are followed by fully connected (Dense) layers, which perform non-linear transformations of the extracted representations:

$$a^{(l)} = f^{(l)}(W^{(l)} \cdot a^{(l-1)} + b^{(l)}) \quad (3.10)$$

where **f^(l)** denotes the activation function of layer l. For the hidden Dense layers, the ReLU (Rectified Linear Unit) is used:

$$\mathrm{ReLU}(z) = \max(0, z) \quad (3.11)$$

The ReLU activation introduces non-linearity while avoiding the vanishing-gradient problem for positive input values. For the final output layer, which forecasts changes in CIC – a variable that is not restricted in sign – a linear activation function is employed:

$$f^{(\mathrm{exit})}(z) = z \quad (3.12)$$

3.2.3 ElasticNet regularisation

L1+L2 ElasticNet regularisation modifies the loss function by adding penalty terms on the network weights:

$$L\_total = L\_MSE + \lambda_1 \Sigma|w\_j| + \lambda_2 \Sigma w\_j^2 \quad (3.13)$$

where $\mathbf{L\_MSE = (1/n)\Sigma(y_t - \hat{y}_t)^2}$ is the Mean Squared Error (MSE), $\boldsymbol{\lambda_1 = 0.0001}$ is L1 (Lasso penalty coefficient) and $\boldsymbol{\lambda_2 = 0.001}$ is the L2 (Ridge penalty coefficient). The L1 penalty promotes sparse solutions by driving unimportant parameters to zero, whereas the L2 penalty reduces the magnitudes of all parameters, thereby improving model stability.

3.2.4 Sliding window mechanism and iterative forecasting

LSTM models employ a sliding window of length T_w to structure the input data:

$$X\_t^{(\mathrm{window})} = [x\_\{t-T\_w+1\}, x\_\{t-T\_w+2\}, \ldots, x\_t] \quad (3.14)$$

where T_w = 30 business days for Models 1–3, and T_w = 20 for Model 4. Each sample in the training set consists of a sequence of T_w consecutive observations, with the corresponding output being a one-step-ahead forecast.

For long-term forecasting, an iterative (recursive) forecasting strategy is employed:

$$\hat{y}\_\{t+k\} = f\_LSTM([\hat{y}\_\{t+1\}, \ldots, \hat{y}\_\{t+k-1\}, x\_\{t+k\}]) \quad (3.15)$$

where the forecast generated at the previous step is fed back as the input to the subsequent step. This strategy is necessary for producing forecasts up to the end of 2035, but it is susceptible to the accumulation of forecast errors over long horizons.

3.2.5 Performance metrics

The performance of the models was evaluated using the following standard metrics:

- The coefficient of determination R^2, which indicates the percentage of the variation in changes in CIC explained by the model.

$$R^2 = 1 - \Sigma(y\_t - \hat{y}\_t)^2 \,/\, \Sigma(y\_t - \bar{y})^2 \quad (3.16)$$

- Mean Absolute Error (MAE) = $(1/n)\,\Sigma|y\_t - \hat{y}\_t|$ *(3.17)*

- Root Mean Squared Error (RMSE) = $\sqrt{[(1/n)\,\Sigma(y\_t - \hat{y}\_t)^2]}$ *(3.18)*

- Direction Accuracy of CIC Change Forecasts $= (1/n)\ \Sigma\ 1[\mathrm{sgn}(y_t - y_{t-1}) = \mathrm{sgn}(\hat{y}_t - y_{t-1})]$ *(3.19)*

The evaluation metrics are computed during both the model training and model testing phases. During the training phase, the model automatically identifies 80 candidate time windows that may be statistically significant for forecasting daily changes in CIC. During the testing phase, model selection is performed by choosing the time window that yields the lowest forecast errors. The training period spans from 31 December 2004 to 31 December 2024, while the testing period covers the year 2025.

The MAE gap and RMSE gap are indicators of how well the model generalises to the testing period. As measures of overfitting, the MAE gap and RMSE gap are defined as the relative differences between the training and testing metrics:

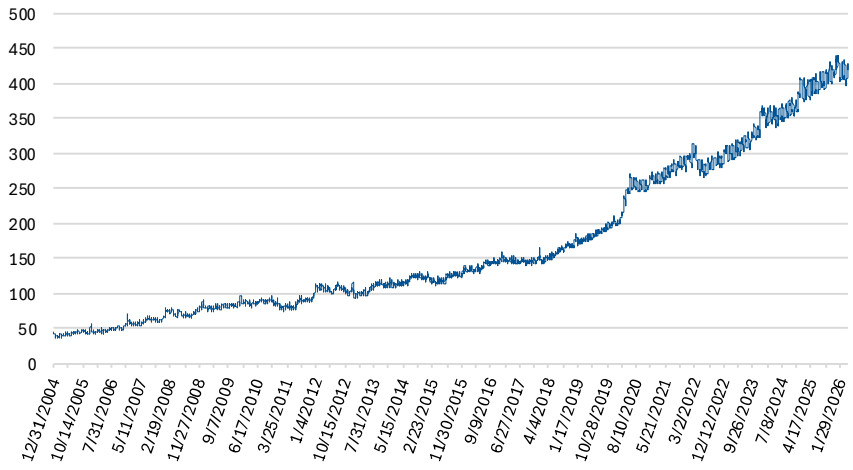
$$\mathrm{MAE\_gap} = (\mathrm{MAE\_test} - \mathrm{MAE\_train}) \,/\, \mathrm{MAE\_train} \times 100\% \quad (3.20)$$

4. Developed models and empirical results

4.1 Data preparation for analysis

The analysis is based on a time series of daily changes in CIC covering the period from 31 December 2004 to 26 January 2026, comprising a total of 5,497 business days (observations).

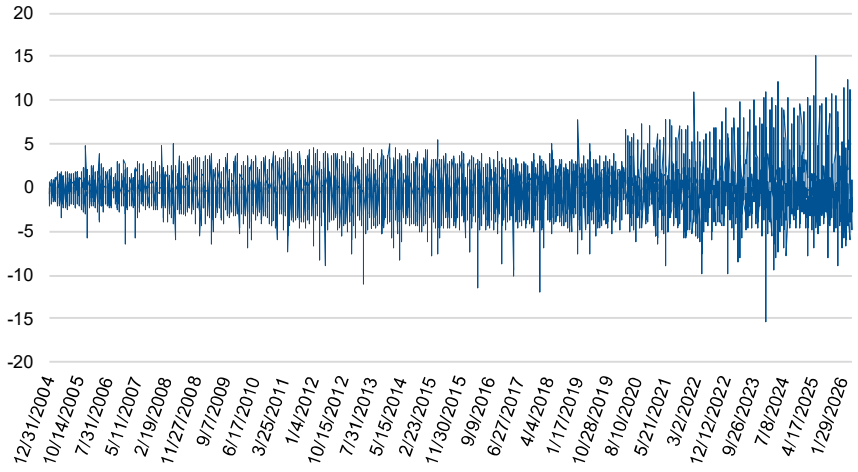
Chart 1 **Cash in circulation (daily volumes)**
in RSD bn



Source: NBS.

The CIC series exhibits the following statistical properties: a mean daily change close to zero, a standard deviation on the order of 2–3 billion Serbian dinars, pronounced positive skewness resulting from asymmetric December and January effects, and the presence of multiple seasonal cycles (5-day, ~22-day, ~65-day, and 260-day cycles).

Chart 2 **Cash in circulation (daily changes)**
in RSD bn



Source: NBS.

The results of the ADF test confirmed that the daily changes in CIC (Δy_t) are stationary ($p < 0.01$), whereas the level of CIC follows an integrated process of order one, I(1). In all LSTM models, the dependent variable was standardised (mean=0, std=1) prior to training and transformed back to its original scale (denormalised) for model evaluation.

4.2 Model 1: LSTM (128→64→16→*1*)

4.2.1 Architecture and parameters

Model 1 is a deep LSTM neural network with a four-layer funnel architecture, comprising a total of 200,500 trainable parameters. The input space consists of 164 variables: 129 exogenous variables, 34 automatically generated seasonal variables, and lagged values of the target variable. The 30-business-day sliding window implies that each observation in the training set is represented by an input tensor – i.e. a data structure passed to the model for processing and learning – of dimension [30 × 164] (Table 1).

Table 1 **Model 1 Architecture and hyperparameters**

Parameter	Value	Note
Layer Architecture	128→64→16→1	Funnel architecture; LSTM+Dense
Number of parameters	200,500	74.8% in the first layer
Window mechanism	30 business days	Sliding window
Dropout	0.30	Between LSTM layers
Optimizer	Adam	lr = 0.001 (default)
Batch size	16	Lower value → better generalization
EarlyStopping patience	10 epochs	Monitoring: val_loss, indicating model error on new data

Parameter	Value	Note
Training period (optimal)	2.3.2006 – 31.12.2024	4,914 observations
Regularisation	Dropout only	Without L1/L2

During the training process, using a 30-business-day sliding window, the model automatically identified 80 candidate periods that could be statistically significant for forecasting changes in CIC. Among all analysed periods, the model selected as optimal the one in which forecasting errors on the test set (year 2025) were the lowest, i.e. where the coefficient of determination (R^2) was the highest. Model 1 identifies the period starting on 2 March 2006 as optimal, with $R^2 = 0.9019$ on the training set.

4.2.2 Results and diagnostics

On the test set (261 observations, 2025), Model 1 achieves an R^2 of 0.7745 and an MAE of RSD 1,329 mn, with a Direction Accuracy of 82.3% – demonstrating that the model accurately predicts the direction of daily CIC changes in more than four out of five cases (82% of cases). This is also the highest level of Direction Accuracy for daily CIC changes among all tested models.

However, the overfitting analysis reveals a significant gap between training and test performance: the MAE gap amounts to +162.5%, which by far exceeds the threshold level of 20%. The model has memorised historical patterns, including the specific noise characteristics in the training set, which makes it potentially unstable in long-term projections.

A detailed look at monthly performance reveals an interesting heterogeneity: in eight out of the 15 analysed months, Model 1 achieves an R^2 above 0.75, while in the remaining seven months (mainly January and December), it records a significant drop. This is consistent with a known characteristic of LSTM models – good performance under regular conditions, but increased sensitivity to structural anomalies.

The finding related to direction error is also interesting: out of 261 forecasted days, Model 1 incorrectly identifies the direction of change in only 47 cases (17.7%). Of those 47 cases, 31 (66%) occur in January and December – which once again confirms that holiday and seasonal periods are the key source of errors for all models (Table 2).

Table 2 **Model 1 – Key parameters**

TRAINING vs TEST — OVERFITTING ANALYSIS						
Parameter	**Training value**	**Test value**	**Difference (Test–Train)**	**Difference %**	**Status**	**Interpretation**
R^2	0.9019	0.7745	-0.1273	-14.12	**OVERFIT**	Lower gap = better model
MAE	506.2555	1328.7652	822.5097	162.47	**OVERFIT**	Higher MAE test = normal below 20%
RMSE	709.5949	1764.4323	1054.8375	148.65	**OVERFIT**	Higher RMSE test = normal below 20%
Direction Accuracy (%)	84.5177	82.3077	-2.21	-2.61	**OK**	Difference < 5pp = OK
Durbin-Watson	1.9565	1.7673	-0.1892	-9.67	**OK**	Both should be close to 2.0
Mean bias	35.037	-213.7663	-248.8033	-710.12	**CAUTION**	Both should be close to 0
Residual Std	708.7294	1751.4353	1042.7059	147.12	**CAUTION**	Std test higher by max 30%

TRAINING vs TEST — OVERFITTING ANALYSIS						
Parameter	**Training value**	**Test value**	**Difference (Test–Train)**	**Difference %**	**Status**	**Interpretation**
Pearson r	0.9531	0.8827	-0.0704	-7.39	**CAUTION**	Difference < 0.05 = OK
Theil U	0.1672	0.2566	0.0895	53.51	**OK**	Both should be < 0.5
CONCLUSION: ✗ SIGNIFICANT OVERFITTING – the model is too adjusted to the training set						

4.3 Model 2: LSTM (128→64→16→1) – second overfitting criterion

Model 2 shares an identical architecture with Model 1 and achieves almost the same numerical performance. The only difference lies in the criterion for overfitting classification: instead of the MAE gap ≥ 20% which categorically classifies the model as overfitted, Model 2 applies a threshold of $\Delta R^2 < 0.15$ (Table 3).

Table 3 **Model 2 – Overfitting criterion**

Criterion (ΔR^2)	Lower threshold	Upper threshold	Overfitting level	Description	Result (0.127)
$\Delta R^2 < 0.05$	0.00	< 0.05	**Mild overfitting**	The model generalises well, with a minimal difference between the training and test sets.	—
$\Delta R^2 < 0.15$	0.05	< 0.15	**Moderate overfitting**	Acceptable difference; the model has a certain tendency towards overfitting.	**✓ 0.127**
$\Delta R^2 \geq 0.15$	0.15	1.00	**Significant overfitting**	Large difference; the model overfits the training data excessively, resulting in poor generalisation.	—

CONCLUSION
$\Delta R^2 = 0.127$ → Moderate overfitting ($0.05 \leq \Delta R^2 < 0.15$) → The model is acceptable, but there is room for improvement through regularisation.

Since $\Delta R^2 = 0.127 < 0.15$, Model 2 is classified as moderate overfitting – which indicates that the choice of parameters for measuring overfitting has a substantial impact on the interpretation of the results (Table 4).

Table 4 **Model 2 – Key parameters**

TRAINING vs TEST — OVERFITTING ANALYSIS						
Parameter	**Training value**	**Test value**	**Difference (Test–Train)**	**Difference %**	**Status**	**Interpretation**
R^2	0.9019	0.7745	-0.1273	-14.12	**OVERFIT**	Lower gap = better model
MAE	506.2555	1328.7652	822.5097	162.47	**OVERFIT**	Higher MAE test = normal below 20%
RMSE	709.5949	1764.4323	1054.8375	148.65	**OVERFIT**	Higher RMSE test = normal below 20%
Direction Accuracy (%)	84.5177	82.3077	-2.21	-2.61	**OK**	Difference < 5pp = OK
Durbin-Watson	1.9565	1.7673	-0.1892	-9.67	**OK**	Both should be close to 2.0
Mean bias	35.037	-213.7663	-248.8033	-710.12	**CAUTION**	Both should be close to 0
Residual Std	708.7294	1751.4353	1042.7059	147.12	**CAUTION**	Std test higher by max 30%
Pearson r	0.9531	0.8827	-0.0704	-7.39	**CAUTION**	Difference < 0.05 = OK
Theil U	0.1672	0.2566	0.0895	53.51	**OK**	Both should be < 0.5
CONCLUSION: ✗ MODERATE OVERFITTING – the model is more adapted to the training set						

4.4 Model 3: LSTM (128→64→16→1) with ElasticNet regularisation

Model 3 introduces L1+L2 ElasticNet regularisation ($\lambda_1 = 0.0001$, $\lambda_2 = 0.001$) on the densely connected layers (dense layers), shares the identical architecture of Models 1 and 2, and the same sliding window mechanism of 30 business days. The regularisation reduces the weight amplitudes, particularly on dense layers, which directly suppresses the tendency towards overfitting.

The regularisation alters the optimal training period length: Model 3 identifies the period from 12 June 2012 (3,276 observations) instead of 2006 to 31 December 2024. This is consistent with theoretical expectations – a regularised model can learn effectively even from a shorter but richer period, without excessive reliance on distant history.

The effects of regularisation are measurable and statistically significant: ΔR^2 falls from 0.127 (Model 1) to 0.056 (Model 3), while the MAE gap decreases from 162.5% to 73.9%. Model 3 achieves an $R^2 = 0.7344$ and Direction Accuracy = 78.9% on the test set, along with an improved Durbin-Watson coefficient of 1.829 and a more symmetric residual distribution (skewness = -0.100 compared to +0.283 for Model 1), indicating higher-quality modelling of residual noise (Table 5).

Table 5 **Model 3 – Key parameters**

TRAINING vs TEST — OVERFITTING ANALYSIS						
Parameter	**Training value**	**Test value**	**Difference (Test–Train)**	**Difference %**	**Status**	**Interpretation**
R^2	0.7899	0.7344	-0.0556	-7.03	**CAUTION**	Lower gap = better model
MAE	804.5368	1399.2087	594.6719	73.91	**OVERFIT**	Higher MAE test = normal below 20%
RMSE	1145.2979	1915.1021	769.8043	67.21	**OVERFIT**	Higher RMSE test = normal below 20%
Direction Accuracy (%)	83.1433	78.8462	-4.2971	-5.17	**OK**	Difference < 5pp = OK
Durbin-Watson	1.9466	1.8291	-0.1175	-6.03	**OK**	Both should be close to 2.0
Mean bias	-317.6571	-438.2904	-120.6333	-37.98	**OK**	Both should be close to 0
Residual Std	1100.3641	1864.2741	763.91	69.42	**CAUTION**	Std test higher by max 30%
Pearson r	0.8978	0.8655	-0.0323	-3.6	**OK**	Difference < 0.05 = OK
Theil U	0.2396	0.2787	0.0391	16.33	**OK**	Both should be < 0.5
CONCLUSION: ✗SIGNIFICANT OVERFITTING – the model is too adapted to the training set						

A particularly valuable finding is that Model 3, despite its slightly lower overall R^2 compared to Models 1 and 2, achieves a higher average $R^2 = 0.805$ during the February–March 2026 period. This indicates that regularisation improves generalisation precisely where it matters most – on the latest, previously unseen data. This characteristic makes Model 3 the most suitable LSTM candidate for future operational testing (Table 10).

4.5 Model 4: LSTM (64→32→8→1) – anti-overfitting architecture

Model 4 represents a fundamental architectural redesign aimed at eliminating overfitting through the application of nine concurrent anti-overfitting strategies. The reduced network (64→32→8→1) has only 71,300 parameters – three times fewer than the previous models.

Each of the nine strategies contributes to reducing overfitting at a different level (Table 6): (1) reduced architecture 64→32→8→1, (2) Batch Normalisation after both LSTM layers, (3) increased Dropout to 0.45 (standard) and 0.10 (recurrent), (4) Dense Dropout of 0.20, (5) enhanced L1+L2 regularisation (ten times stronger than the previous model), (6) Gradient Clipping (clipnorm=1.0), (7) Weight Decay in the Adam optimiser (AdamW effect), (8) larger batch size (32 neurons instead of 16 neurons), and (9) smaller window size of 20 instead of 30 days.

Model selection introduces a filter: out of 80 tested configurations, only 13 pass the MAE gap ≤ 20% filter. The optimal training period length is from 19 May 2021 to 31 December 2024.

Table 6 **Model 4 – Anti-overfitting strategies**

9 Anti-Overfitting Strategies		
1 ARCHITECTURE	**2 NORMALISATION**	**3 REGULARISATION**
Reduced Architecture	Batch Normalization	Increased Dropout
64 → 32 → 8 → 1	After both LSTM layers	0.45 std · 0.10 rec
Reducing network capacity limits its ability to memorise noise.	BatchNorm returns values to the normal range (~0). More stable training, with a lower risk of overfitting.	Standard and recurrent dropout for the LSTM layers.
4 REGULARISATION	**5 REGULARISATION**	**6 OPTIMISATION**
Dense Dropout	L1 + L2 Regularisation	Gradient Clipping
p = 0.20	10× stronger than before	clipnorm = 1.0
Additional regularisation of the dense layers. More aggressive dropout would destroy the weak signal.	Stronger penalisation of large weights. Stronger regularisation → a simpler model.	Prevents exploding gradients. Never change weights by more than 1.0 in a single step.
7 OPTIMISATION	**8 DATA**	**9 DATA**
Weight Decay (AdamW)	Larger Batch Size	Smaller Window Size
Adam + decay	32 instead of 16	20 instead of 30 days
Directly reduces weights. The model remains simple, improving generalisation.	Larger batch → a more stable gradient and less noise. Effect: the model fluctuates less → less overfitting.	Smaller window → less information → less chance of memorising specific noise.

The selected configuration achieves $R^2 = 0.6651$ with $\Delta R^2 = +0.033$ and an MAE gap = +26.2% – confirming that the anti-overfitting strategy achieves its goal (Table 7).

Table 7 **Model 4 – Key parameters**

TRAINING vs TEST — OVERFITTING ANALYSIS						
Parameter	**Training value**	**Test value**	**Difference (Test–Train)**	**Difference %**	**Status**	**Interpretation**
R^2	0.6979	0.6651	-0.0328	-4.7	**OK**	Gap < 0.05 = excellent
MAE	1251.4888	1579.8329	328.3441	26.24	**CAUTION**	MAE test ≤ 20% > Train = OK
RMSE	1758.9828	2150.2113	391.2285	22.24	**CAUTION**	RMSE test ≤ 20% > Train = OK
Pearson r	0.857	0.8366	-0.0203	-2.37	**OK**	Difference < 0.05
Dir. acc. (%)	74.026	70.7692	-3.2567	-4.4	**OK**	Difference < 5pp

TRAINING vs TEST — OVERFITTING ANALYSIS						
Parameter	**Training value**	**Test value**	**Difference (Test–Train)**	**Difference %**	**Status**	**Interpretation**
Residual Std	1758.9741	2131.5099	372.5358	21.18	**OK**	Std test ≤ 30% > Train
Durbin-Watson	2.0616	1.6716	-0.39	-18.92	**OK**	Both ≈ 2.0
Theil U	0.3295	0.3454	0.016	4.85	**OK**	Both < 0.5
CONCLUSION: ⚠ MILD OVERFITTING – the model works somewhat better on the training set						

Unfortunately, reducing the network capacity degraded predictive performance: Model 4 is consistently the weakest across all forecast accuracy parameters (Table 10).

4.6 ARIMAX (65,1,65,18) model

The ARIMAX model was developed based on the series of daily CIC changes from 3 January 2022 to 23 September 2024 period (711 observations), which has the highest statistical significance in forecasting daily CIC changes. The choice of this period reflects the assessment that CIC dynamics since 2022 are structurally more stable and more relevant for short-term forecasting compared to older data that cover the COVID period (Table 3 in the annex).

The model specification is the result of an iterative testing process. The starting point was the identification of stationarity – the ADF test confirms that daily CIC changes (first difference) are stationary. Autocorrelation (ACF) and partial autocorrelation (PACF) plots suggest the presence of significant autocorrelations at lags up to 65 working days. By specifying AR(65) and MA(65) components with first-order differencing, the ARIMAX (65,1,65,18) model was obtained.

The final model achieves Adjusted $R^2 = 0.76$, a Durbin–Watson statistic of 2.02, and a standard error of regression significantly below the standard deviation of the dependent variable – confirming high predictive power (Table 8).

Table 8 **ARIMAX (65,1,65,18) – Summary indicators of model quality**

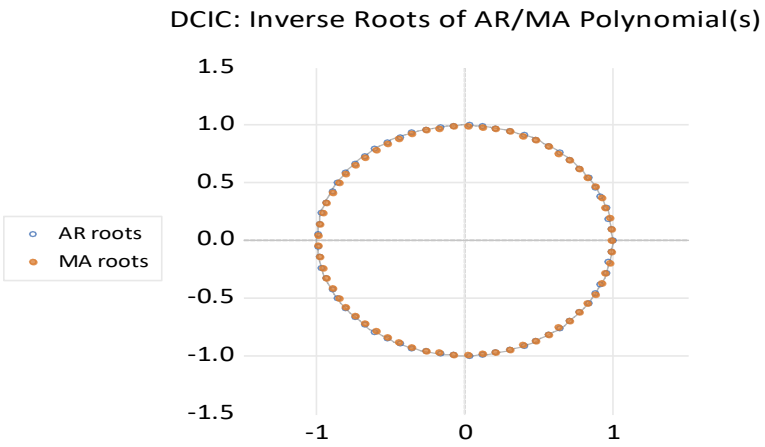
Statistics	Value	Interpretation
Adjusted R^2	0.76	76% of CIC variance explained
Standard error of regression	< S.D. of dependent variable	Good predictive power
Durbin–Watson	2.02	No residual autocorrelation
Stationarity (AR inverse roots)	Within the unit circle	Model is stable
Invertibility (MA inverse roots)	Within the unit circle	Model is invertible
Number of parameters	18 exogenous + AR/MA	Simple model

Although the ARCH test confirms the statistically significant presence of conditional heteroskedasticity in the residuals (F-statistic = 2.39, p = 0.0049), its practical impact on the model's reliability is limited in scope. This is indicated by the low coefficient of determination of the ARCH test regression ($R^2 = 0.04$), which shows that only 4% of the variability in the squared residuals can be statistically explained, while the remaining part represents a random

component that the model objectively cannot capture. Given that the primary goal of the model is forecasting, not statistical inference (i.e. drawing conclusions about the entire population based on a sample), the presence of heteroskedasticity does not cause the parameter estimates obtained by the maximum likelihood method to become systematically biased; as the number of observations increases, they converge to the true parameter values. The sample size of 711 observations provides a sufficient asymptotic basis to absorb the observed form of conditional heteroskedasticity (Table 4 in the annex).

All AR and MA inverse roots lie within the unit circle, guaranteeing the stationarity and invertibility of the model (Chart 3).

Chart 3 **ARIMAX (65,1,65,18) – Plot of AR and MA inverse roots within the unit circle**



5. Comparative performance analysis

5.1 Metrics

The performance of all models is compared based on monthly R^2, MAE and RMSE values for the period January 2025 – March 2026. Each model is also assigned an overfitting category based on ΔR^2 and the MAE gap, which quantify the difference between training and test performance (Table 9).

Table 9 **Summary performance metrics of all models – test period (2025)**

Model	R^2 test	MAE test (RSD bn)	ΔR^2	MAE gap	Dir. acc.	Overfitting
ARIMAX (65,1,65,18) + expert	0.790	1.353	N/A	N/A	N/A	N/A
Model 1: LSTM 128→64→16→1	0.7745	1.329	0.127	+162.5%	82.3%	High
Model 2: LSTM 128→64→16→1 (alt.)	≈ Model 1	≈ Model 1	0.127	+162.5%	≈ Model 1	Moderate

Model	R^2 test	MAE test (RSD bn)	ΔR^2	MAE gap	Dir. acc.	Overfitting
Model 3: LSTM 128→64→16→1 + ElasticNet	0.7344	1.399	0.056	+73.9%	78.9%	Moderate
Model 4: LSTM 64→32→8→1 (anti-OF)	0.6651	Highest	0.033	+26.2%	Lowest	Low

Note: N/A – not applicable in the same manner; Dir. Acc. – Direction Accuracy; OF – overfitting.

The ARIMAX (65,1,65,18) model, adjusted by expert judgement, gives the best CIC forecast results for the period January 2025 – March 2026 compared to the other tested models. The coefficient of determination R^2 is the highest, and the mean absolute error (MAE) is the lowest compared to all models, indicating that the forecasted values are close to the actual ones.

Table 10 **A comparative overview of the performance of all models**

Month	ARIMAX*			Model 1			Model 2			Model 3			Model 4		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Jan 25	**0.698**	**1,634**	**1,263**	0.000	3,119	2,013	0.014	3,048	1,968	0.045	3,162	1,980	0.007	2,937	2,122
Feb 25	0.766	2,107	1,666	**0.851**	**1,723**	1,505	**0.851**	1,727	1,507	0.841	1,740	**1,457**	0.789	2,219	1,579
Mar 25	0.812	1,601	1,173	0.832	1,539	1,260	0.832	1,540	1,258	**0.882**	**1,198**	**877**	0.791	1,672	1,249
Apr 25	0.828	1,658	1,297	0.865	1,546	**1,203**	0.864	1,549	1,205	**0.868**	**1,483**	1,237	0.839	1,754	1,322
May 25	0.851	1,978	1,470	0.919	1,523	1,194	**0.920**	**1,519**	**1,191**	0.830	1,994	1,300	0.771	2,795	1,801
Jun 25	0.748	1,906	1,564	0.874	1,397	1,145	**0.875**	**1,395**	**1,142**	0.830	1,632	1,301	0.825	1,793	1,449
Jul 25	0.757	1,859	1,319	0.854	1,461	1,171	0.854	1,462	1,173	0.880	**1,321**	**1,066**	**0.884**	1,513	1,256
Aug 25	0.811	1,615	1,239	**0.820**	**1,457**	**1,107**	0.795	1,490	1,151	0.799	1,549	1,220	0.772	1,696	1,318
Sep 25	0.685	2,130	1,364	**0.854**	1,504	**1,045**	**0.854**	**1,503**	**1,045**	0.846	1,683	1,186	0.819	1,859	1,245
Oct 25	0.828	**1,696**	1,375	0.844	1,746	**1,358**	0.844	1,746	**1,358**	0.844	1,810	1,531	**0.847**	1,947	1,643
Nov 25	**0.921**	**1,130**	**874**	0.774	1,836	1,449	0.774	1,836	1,449	0.700	2,095	1,642	0.725	2,040	1,549
Dec 25	0.779	**2,005**	**1,632**	0.781	2,145	1,779	0.781	2,145	1,779	**0.838**	2,042	1,645	0.771	2,363	1,839
Jan 26	0.343	2,524	1,692	0.249	2,705	1,887	0.249	2,705	1,887	**0.536**	**2,296**	**1,636**	0.202	2,660	2,052
Feb 26	**0.862**	**1,714**	**1,363**	0.727	2,282	1,747	0.727	2,282	1,747	0.755	2,171	1,668	0.748	2,314	1,685
Mar 26	0.774	2,120	1,531	**0.908**	**1,424**	**1,130**	**0.908**	**1,424**	1,131	0.825	1,899	1,408	0.727	2,498	1,901
Average Feb 25–Mar 26	**0.769**	**1,860**	**1,397**	**0.797**	**1,735**	**1,356**	**0.795**	**1,737**	**1,359**	**0.805**	**1,780**	**1,370**	**0.751**	**2,080**	**1,563**
Average Jan 25–Mar 26	**0.764**	**1,845**	**1,388**	**0.744**	**1,827**	**1,400**	**0.743**	**1,825**	**1,399**	**0.755**	**1,872**	**1,410**	**0.701**	**2,137**	**1,601**

**ARIMAX model adjusted by expert judgement Green = best value in the row Red = weakest value in the row*

The LSTM models record weaker performance, primarily due to the unforeseen change in the pension payment schedule in January 2025, since this change was not represented in the training set and the models could not learn it (Table 10).

If January 2025 is excluded from the forecast period and the period February 2025 – March 2026 is observed, all LSTM models except Model 4 have better performance compared to ARIMAX (65,1,65,18). Model 3 (LSTM 128→64→16→1 with ElasticNet L1+L2 regularisation) has the highest coefficient of determination R^2. This shows that the additional L1+L2 ElasticNet regularisation succeeded in reducing overfitting and improving generalisation, i.e. the model's ability to perform well on new, unseen data, not just on the data it was trained on (Table 10).

5.2 Monthly performance analysis

An analysis of monthly R^2 values (Table 10) reveals significant performance variations across all models during the tested period. Three characteristic behavioural patterns can be observed:

January 2025 (structural break): All LSTM models record a drastic drop in performance ($R^2 \approx 0.3–0.5$) due to the unforeseen change in the dynamics of pension payments. The ARIMAX model, thanks to expert adjustment, shows greater resilience ($R^2 = 0.698$), although it also records weaker performance than in other months.

February 2025 – October 2025 (stable period): LSTM models 1, 2 and 3 achieve performance comparable to the ARIMAX model adjusted by expert judgement, and in certain months (March, June, September) even better.

November 2025 (special period): The ARIMAX model adjusted by expert judgement achieves by far the best coefficient of determination, $R^2 = 0.921$. LSTM models ($R^2 \approx 0.75–0.80$) lag behind because it is more difficult for them to forecast without explicit exogenous variables.

5.3 Error analysis

Across all models, RMSE is significantly higher than MAE, which indicates the presence of occasional large errors that are disproportionately penalised by the RMSE metric (Table 10). These extremes are primarily related to:

– Pension and salary payment days – when cash flow can be several times higher than the average;

– Extended holiday periods – especially the Christmas celebrations;

– Unexpected changes in fiscal policy – such as the alteration of the pension payment calendar in January 2025.

A common feature across all models is a skewed distribution of residuals with positive skewness coefficients – the models more frequently underestimate CIC growth than overestimate it.

For all models, the autocorrelation structure of the residuals was also analysed using the Durbin–Watson statistic. The ARIMAX model adjusted by expert judgement achieves DW = 2.02 – a practically ideal result. The LSTM models record DW values between 1.77 and 1.83, which indicates mild positive autocorrelation of residuals (Table 2 in the annex).

Table 11 **Advantages and disadvantages of ARIMAX and LSTM models**

Aspect	ARIMAX model	LSTM model
Interpretability	High: coefficients on exogenous variables directly quantify the effects of holidays and seasonality	Low: “black box”; relationships between variables are implicitly encoded in the weight matrices
Seasonality	Explicit: requires all relevant dummy variables to be defined in advance	Automatic: the model itself identifies seasonal patterns from the data

Aspect	ARIMAX model	LSTM model
Structural changes	Requires manual adjustment and expert judgement; responds faster with intervention	Responds slowly – a new change must be in the training set for the model to learn it
Long-term forecasting	Limited – parameters do not evolve automatically; periodic re-estimation is required	Risky for horizons > 2 years due to the accumulation of errors in iterative forecasting
Computational requirements	Minimal – rapid estimation and evaluation, suitable for operational use	Substantial – training requires GPU resources; evaluation is fast
Overfitting risk	Low – the model is simple by construction; statistical variable selection	High on short series – requires careful regularisation and a validated training process
Development and maintenance	Standardised process; available literature; supported in Eviews/R/Python	Requires specialised knowledge in deep learning; more complex adjustment process

6. Conclusion

This paper presents a systematic comparative analysis of four LSTM deep machine learning models and the ARIMAX (65,1,65,18) econometric model, adjusted by expert judgement, for forecasting daily changes in CIC. The analysis is based on a time series of 5,497 daily observations from 2004 to 2026, with the analysed period being January 2025 – March 2026. The research generated five key findings.

Finding 1 – Superiority of the ARIMAX model over the entire period. The ARIMAX model adjusted by expert judgement achieves superior performance in the observed period: the lowest MAE, the highest overall R^2, and robustness, with good forecasting in January 2025 – a period marked by a structural change in the pension payment dynamics.

Finding 2 – Competitiveness of regularised LSTM in stable periods. LSTM Model 3 with ElasticNet regularisation achieves an average $R^2 = 0.805$ in the period February 2025 – March 2026 – higher than the coefficient of determination (R^2) of the ARIMAX model. In the absence of structural changes, regularised LSTM Model 3 provides very good forecasting without expert adjustment. Recommendation: Model 3 is the most promising candidate for operational testing in the parallel monitoring phase.

Finding 3 – Effectiveness of regularisation in controlling overfitting. ElasticNet regularisation reduces ΔR^2 from 0.127 (Model 1) to 0.056 (Model 3) and improves the Durbin-Watson coefficient from 1.77 to 1.83. The aggressive anti-overfitting architecture (Model 4) eliminates overfitting ($\Delta R^2 = 0.033$), but at the cost of predictive power ($R^2 = 0.6651$). There is an optimal level of regularisation between these two extremes that should be systematically explored.

Finding 4 – Criticality of defining overfitting parameters. Models 1 and 2, identical in architecture and numerical performance, are classified differently depending on whether overfitting is measured through the MAE gap (Model 1: high overfitting) or ΔR^2 (Model 2:

moderate overfitting). This classification difference can be crucial for the decision on the operational application of the model.

Finding 5 – Complementarity of approaches and hybrid potential. ARIMAX and LSTM models are not competitors – they are complementary approaches with different strengths and weaknesses. The ARIMAX model provides interpretability, transparency, and resilience to structural changes. The LSTM model provides automatic learning of non-linear patterns and potential for adaptability, i.e. the ability to adjust to new conditions, changes, or different situations.

Increasing the number of observations, assuming the absence of significant structural changes, could further improve the performance of LSTM models.

Annex: Detailed comparison of model architectures

The annex provides a more detailed technical overview of all five models, with complete hyperparameter specifications, the results of the optimal training period search, and summarised diagnostic indicators. This information is intended for technical users who wish to reproduce or adapt the developed models.

Table 1 **Complete hyperparameter specification of all LSTM models**

Hyperparameter	Model 1	Model 2	Model 3	Model 4
Architecture (LSTM→Dense→out)	128→64→16→1	128→64→16→1	128→64→16→1	64→32→8→1
Total parameters	200,500	200,500	200,500	71,300
Dropout (standard)	0.30	0.30	0.30	0.45
Recurrent dropout	–	–	–	0.10
Dense dropout	–	–	–	0.20
L1 regularisation	–	–	0.0001	0.001
L2 regularisation	–	–	0.001	0.01
Batch Normalisation	No	No	No	Yes
Gradient Clipping	–	–	–	1.0
Window size (days)	30	30	30	20
Batch size	16	16	16	32
Optimiser	Adam	Adam	Adam	AdamW
EarlyStopping patience	10	10	10	10
OF filters at selection	MAE gap ≤ 20%	$\Delta R^2 < 0.15$	MAE gap ≤ 20%	MAE gap ≤ 20%
Optimal training start	2.3.2006	2.3.2006	12.6.2012	19.5.2021
Number of training observations	4,914	4,914	3,276	945

Table 2 **Results of the optimal training period search – summary statistics**

Search statistics	Model 1	Model 2	Model 3	Model 4
Number of tested configurations	80	80	80	80
Passed through the OF filter	N/A	53 ($\Delta R^2 < 0.15$)	N/A	13 (MAE ≤ 20%)
R^2 training (selected model)	0.9019	0.9019	0.8618	0.6984
R^2 test (selected model)	0.7745	0.7745	0.7344	0.6651
ΔR^2 (training – test)	0.127	0.127	0.056	0.033
MAE gap (%)	+162.5%	+162.5%	+73.9%	+26.2%

Search statistics	Model 1	Model 2	Model 3	Model 4
Durbin-Watson (residual test)	1.774	1.774	1.829	1.801
Residual skewness	+0.283	+0.283	-0.100	+0.147

Table 3 **ARIMAX (65,1,65,18)**

Dependent Variable: DCIC
Method: ARMA Maximum Likelihood (BFGS)
Date: 02/21/25 Time: 13:46
Sample: 1/03/2022 9/23/2024
Included observations: 711
Convergence achieved after 31 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
W10	-622.8686	259.1914	-2.403122	0.0165
TW1	10127.95	1636.893	6.187302	0.0000
MT3	1948.424	755.0318	2.580585	0.0101
MT2	4584.982	837.0957	5.477250	0.0000
EH_2A	-2478.763	571.2319	-4.339329	0.0000
W48	3460.174	561.9205	6.157764	0.0000
EH_1A	-6688.144	2641.319	-2.532123	0.0116
MT4	-9082.940	1516.119	-5.990914	0.0000
W1	-1141.472	550.1693	-2.074765	0.0384
SECONDW	572.8555	348.7151	1.642761	0.1009
NY_1A	-7287.528	949.1455	-7.677988	0.0000
WT1	2641.636	1284.037	2.057289	0.0400
WT4	-3694.829	1686.075	-2.191378	0.0288
TW4	-6157.223	3678.150	-1.674000	0.0946
NY_4A	-2478.705	769.3149	-3.221965	0.0013
NY_5A	-3411.392	1460.454	-2.335844	0.0198
CHR_2A	3358.348	1512.940	2.219749	0.0268
W6	1052.604	277.8336	3.788614	0.0002
AR(45)	0.169553	0.015234	11.12981	0.0000
AR(65)	0.777811	0.017361	44.80256	0.0000
AR(1)	0.077832	0.015160	5.134141	0.0000
AR(2)	-0.109737	0.014551	-7.541568	0.0000
AR(22)	0.081253	0.013115	6.195303	0.0000
MA(64)	-0.073675	0.030963	-2.379487	0.0176
MA(65)	-0.586921	0.034954	-16.79116	0.0000
MA(43)	0.076906	0.028518	2.696732	0.0072
MA(44)	0.130228	0.029151	4.467349	0.0000
SIGMASQ	2539056.	102895.7	24.67600	0.0000

R-squared	0.765310	Mean dependent var	101.6203
Adjusted R-squared	0.756032	S.D. dependent var	3291.504
S.E. of regression	1625.775	Akaike info criterion	17.72566
Sum squared resid	1.81E+09	Schwarz criterion	17.90551
Log likelihood	-6273.474	Hannan-Quinn criter.	17.79513
Durbin-Watson stat	2.021595		

Table 4 **Heteroskedasticity test: ARCH**

Heteroskedasticity Test: ARCH

F-statistic	2.392522	Prob. F(12,686)	0.0049
Obs*R-squared	28.07917	Prob. Chi-Square(12)	0.0054

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 06/20/26 Time: 17:01
Sample (adjusted): 1/19/2022 9/23/2024
Included observations: 699 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2314059.	342482.5	6.756722	0.0000
RESID^2(-1)	0.145720	0.038164	3.818236	0.0001
RESID^2(-2)	0.075450	0.038568	1.956270	0.0508
RESID^2(-3)	-0.032200	0.038671	-0.832683	0.4053
RESID^2(-4)	-0.031701	0.038596	-0.821363	0.4117
RESID^2(-5)	-0.007807	0.038605	-0.202229	0.8398
RESID^2(-6)	0.034927	0.038599	0.904868	0.3659
RESID^2(-7)	0.019319	0.038600	0.500481	0.6169
RESID^2(-8)	0.021279	0.038609	0.551134	0.5817
RESID^2(-9)	-0.069998	0.038598	-1.813531	0.0702
RESID^2(-10)	-0.012271	0.038666	-0.317374	0.7511
RESID^2(-11)	-0.016229	0.038564	-0.420839	0.6740
RESID^2(-12)	-0.029133	0.038162	-0.763406	0.4455

R-squared	0.040170	Mean dependent var	2564340.
Adjusted R-squared	0.023380	S.D. dependent var	5099925.
S.E. of regression	5039953.	Akaike info criterion	33.72211
Sum squared resid	1.74E+16	Schwarz criterion	33.80673
Log likelihood	-11772.88	Hannan-Quinn criter.	33.75483
F-statistic	2.392522	Durbin-Watson stat	1.997415
Prob(F-statistic)	0.004949		

Literature

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016), 265–283

Alex Martelli, Anna Martelli Ravenscroft, Steve Holden, Paula McGuire (2023). Python за програмере

Sebastian Raschka i Vahid Mirjalili (2020). Python машинско учење

Sebastian Raschka, Yuxi (Hayden) Liu i Vahid Mirjalili (2022). Машинско учење PyTorch и Scikit-Learn

Bindseil, U., Camba-Mendez, G., Hirsch, A., Weller, B. (2004). Excess reserves and the implementation of monetary policy of the ECB. ECB Working Paper Series, No. 361. European Central Bank

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control, 5th edition. John Wiley & Sons

Chollet, F. (2018). Deep Learning with Python. Manning Publications

Diebold, F.X., Mariano, R.S. (1991). Comparing predictive accuracy

Fischer, T., Krauss, C. (2017). Deep learning with long short-term memory networks for financial market predictions

Gal, Y., Ghahramani, Z. (2016). Dropout as a Bayesian approximation

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press

Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780

Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift

Kingma, D.P., Ba, J. (2015). Adam: A method for stochastic optimization

Lim, B., Arık, S.Ö., Loeff, N., Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLOS ONE, 13(3), e0194889

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929–1958

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2023). Attention is all you need

Whitesell, W. (2006). Interest rate corridors and reserves. Journal of Monetary Economics, 53(6), 1177–1195

Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175

Zou, H., Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net

Abbreviations

Abbreviation/term	Definition
CIC	Currency in circulation – the total value of banknotes and coins in circulation
ARIMAX	*AutoRegressive Integrated Moving Average with exogenous variables – an extended ARIMA model incorporating exogenous variables*
LSTM	*Long Short-Term Memory – a type of recurrent neural network using gating mechanisms to retain long-term dependencies*
ReLU	*Rectified Linear Unit – activation function: f(x) = max(0,x)*
Dropout	Regularisation technique: randomly disabling neurons during training to prevent overfitting
ElasticNet	Combined L1 + L2 regularisation that simultaneously promotes sparse solutions and reduces parameter magnitudes
Overfitting	Overfitting of the model to the training data; the model describes historical data well but generalises poorly to new data
R^2	Coefficient of determination: the proportion of variance in the dependent variable that is explained ($0 \leq R^2 \leq 1$)
MAE	*Mean Absolute Error – the mean absolute forecast error*
RMSE	*Root Mean Squared Error – the square root of the mean squared error; it penalises larger errors, meaning that larger errors have a greater impact, i.e. large deviations are penalised more heavily*
Direction Accuracy	The proportion of cases in which the model correctly forecasts the direction of change (increase/decrease)
Batch Normalization	Normalisation of activations by mini-batch (a smaller group of data); stabilises the training of deep neural networks
EarlyStopping	A mechanism for stopping training early when the error on the validation set does not improve over a specified number of patience epochs. It indicates how many epochs the model may go without improvement before training is stopped.
Window size	Length of the sliding window (number of consecutive days) used as input to the LSTM model
Dense layer	Densely connected neural-network layer (Fully Connected Layer) This means that each neuron is connected to all neurons in the preceding layer.