Journal of Empirical Social Science Studies
Published 2020

Application of Machine Learning for Predicting U.S. Bank Deposit Growth: A Univariate and Multivariate Analysis of Temporal Dependencies and Macroeconomic Interrelationships

Hariharan Pappil Kothandapani[1]

[1]**CFA® charterholder, Senior Data Science & Analytics Developer at FHLBC, MS Quantitative Finance @ Washington University in St Louis**
ORCID

RESEARCH ARTICLE

Abstract
This research applies machine learning algorithms to predict the growth rates of bank deposits in the United States using data from 1973 to 2019. The dataset includes weekly deposit records from U.S. commercial banks and key macroeconomic indicators, including GDP, inflation, money supply (M2), recession periods, and interest rates, obtained from the Federal Reserve Economic Data (FRED). The study involved preprocessing steps including date conversion, stationarity testing with the Augmented Dickey-Fuller (ADF) test, and differencing to achieve stationarity. Various models were tested for univariate time series analysis, including SARIMA, Prophet, ETS, LSTM, and Transformer models. LSTM demonstrated the highest predictive accuracy, with the lowest error metrics and the highest R^2 value, proving effective in capturing complex temporal dependencies in deposit data. The study conducted a multivariable analysis incorporating several macroeconomic indicators to explore their relationship with bank deposits. This process included feature scaling, creating lag features, and preserving temporal order during data splitting. Recurrent Neural Networks (RNNs) were evaluated with different lagged periods to assess their impact on model performance. The results indicated that while increasing the number of lags improved the model's fit to the training data, it did not consistently enhance performance on unseen data, highlighting the trade-off between model complexity and generalization. Cointegration analysis confirmed long-term equilibrium relationships between bank deposits and macroeconomic indicators. Further analysis using FMOLS and DOLS revealed that inflation and recessions negatively impacted deposits, while M2 and GDP had positive effects. This study demonstrates the effectiveness of machine learning models, with LSTM proving particularly successful in forecasting bank deposit growth rates. Incorporating multiple macroeconomic variables significantly enhanced predictive accuracy, providing valuable insights into the factors influencing deposit levels. This research contributes to financial forecasting by showcasing the ability of machine learning techniques to integrate economic dynamics into predictive models. This research contributes to the field of financial forecasting by demonstrating the efficacy of machine learning techniques in economic analysis.

OPEN ACCESS
Reproducible Model

Edited by
Associate Editor

Curated by
The Editor-in-Chief

Accepted *January 2020*

Citation
Kothandapani, H. P. (2020) Application of Machine Learning for Predicting U.S. Bank Deposit Growth: A Univariate and Multivariate Analysis of Temporal Dependencies and Macroeconomic Interrelationships

Keywords: Bank deposits, Economic indicators, Financial forecasting, Machine learning, Predictive models, Time series analysis, U.S. GDP

1 Introduction

Bank deposit growth is considered as a fundamental indicator of financial stability and the effectiveness of monetary policy within an economy [1]. A sustained increase in bank deposits typically reflects a favorable economic environment where consumers and businesses are more inclined to save due to stable income levels and positive economic prospects. Conversely, when deposit growth stagnates or declines [2], it may signal underlying economic concerns, such as diminished consumer confidence, reduced disposable income, or expectations of adverse economic conditions. In this context, analyzing trends in deposit growth provides insights into the broader economic sector and can inform policy decisions aimed at maintaining financial stability [3] [4].

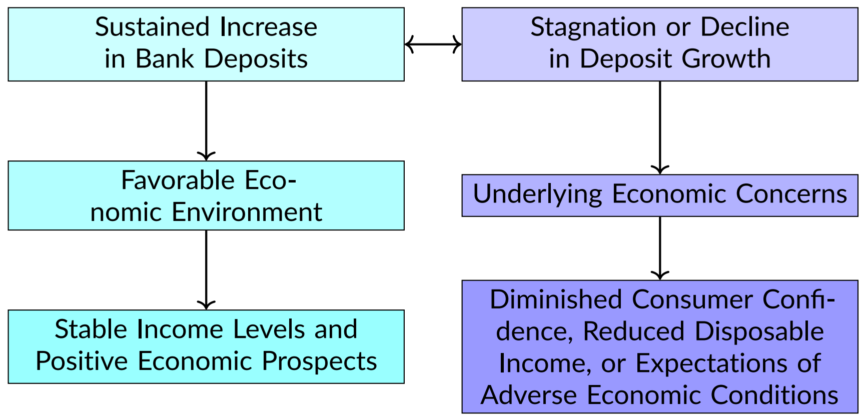

Figure 1. Mechanism depicting the relationship between bank deposit trends and economic conditions.

Interest rates play a pivotal role in influencing bank deposit growth. In a low-interest-rate environment, savers might be disincentivized to keep funds in bank deposits due to the reduced return on savings, leading to slower growth in deposits. Conversely, higher interest rates often encourage savings, as the opportunity cost of holding cash or spending increases, leading to more significant deposit accumulation. Central banks, through their monetary policy tools, can thus indirectly affect bank deposit growth by adjusting interest rates in response to inflationary pressures or economic slowdowns. The interplay between interest rates and deposit growth is complex, as it also depends on the broader economic context, including consumer confidence and expectations regarding future economic conditions [5] [6].

Inflation is another critical factor that can significantly impact bank deposit growth. In periods of high inflation, the real value of money decreases, leading consumers and businesses to seek alternative investment options that may offer better protection against inflationary erosion. This behavior can result in reduced deposit growth as funds are diverted into assets like real estate, commodities, or foreign currencies that are perceived as safer or offering higher returns. On the other hand, in a low inflationary environment, the stability of purchasing power encourages saving in bank deposits, contributing to their growth. Understanding the relationship between inflation and deposit behavior is essential for policymakers to design effective strategies that promote financial stability [7] [8].

Economic growth also exerts a considerable influence on bank deposit trends. In a growing economy, businesses and consumers typically experience increased earnings and profits, which can lead to a higher propensity to save, thus bolstering deposit growth. Economic expansion often leads to job creation, wage increases, and greater disposable income, which all contribute to higher savings rates. Conversely, during periods of economic contraction or recession, deposit growth may slow or decline as individuals and businesses draw on their savings to weather economic hardships. Therefore, tracking bank deposit growth in relation to economic cycles provides a barometer for assessing the broader economic health [5] [9].

Bank deposit growth is also reflective of the public's trust and confidence in the financial system. During times of economic or financial uncertainty, such as a banking crisis or significant market volatility, depositors may withdraw their funds due to fears of bank insolvency or devaluation, leading to a decline in deposit growth [4]. Conversely, in a stable financial environment with

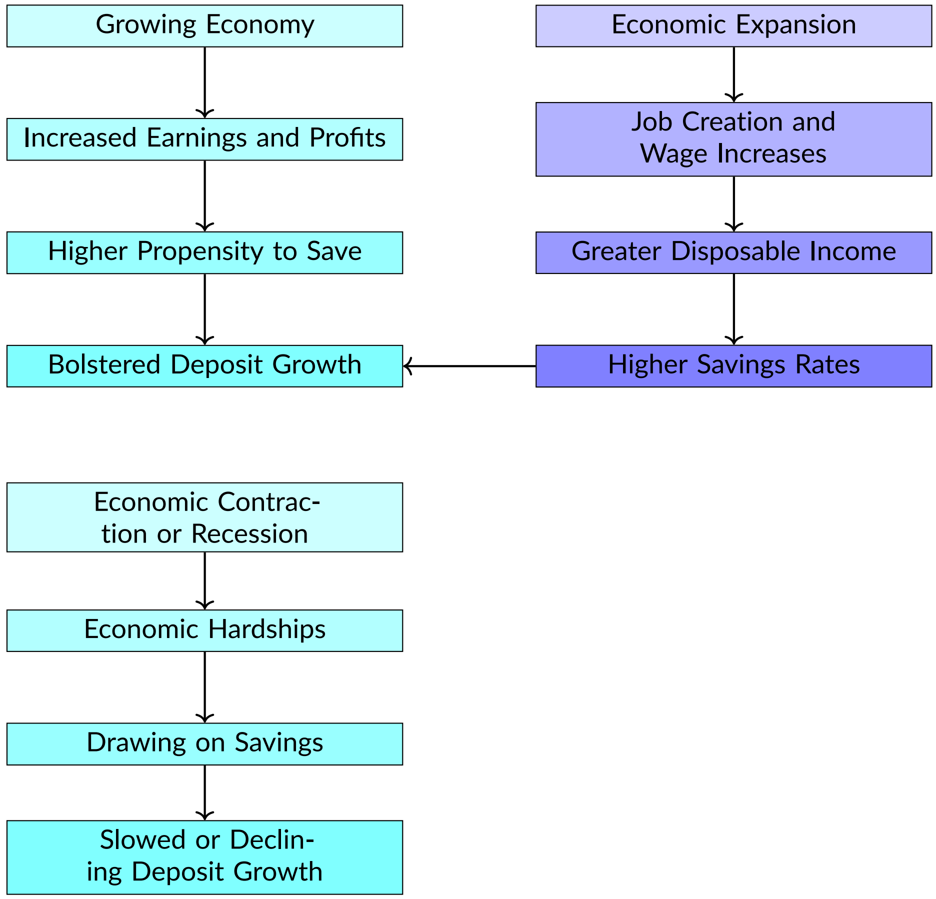

Figure 2. Mechanism illustrating the impact of economic conditions on deposit growth.

robust regulatory frameworks and sound banking practices, deposit growth tends to be stronger as confidence in the financial system is maintained. Hence, deposit growth not only serves as a measure of financial health but also as an indicator of the effectiveness of regulatory oversight and the resilience of the banking sector in the face of external shocks [10].

For financial institutions, the ability to accurately predict deposit growth is critical to maintaining optimal levels of liquidity, ensuring that they have sufficient funds on hand to meet withdrawal demands, while also maximizing the utilization of these funds through lending or investment activities. Liquidity management is a delicate balancing act that requires institutions to anticipate fluctuations in deposit levels and adjust their liquidity buffers accordingly. Inaccurate predictions can lead to either a liquidity shortfall, which could necessitate costly borrowing, or excess liquidity, which might result in suboptimal returns. Thus, sophisticated forecasting models that incorporate macroeconomic indicators, customer behavior analytics, and market trends are essential tools for financial institutions to manage their liquidity effectively and avoid potential risks associated with liquidity mismanagement [11].

Accurate deposit growth predictions also play a significant role in optimizing asset allocation within financial institutions. By understanding expected deposit trends, institutions can better plan their portfolio strategies, ensuring that they allocate assets in a manner that aligns with their liquidity needs and risk appetite. For instance, a forecast of robust deposit growth might enable a bank to allocate more funds to longer-term, higher-yield investments, knowing that liquidity is not an immediate concern. Conversely, if a decline in deposits is anticipated, the institution might shift its focus towards more liquid or short-term assets to maintain financial flexibility. This strategic alignment of asset allocation with deposit trends is crucial for achieving a balance between profitability and risk management.

From a strategic decision-making perspective, deposit growth predictions inform a range of decisions that impact the long-term direction of a financial institution. For example, anticipated deposit increases might support the decision to expand lending activities, launch new financial products, or enter new markets. Conversely, expected stagnation or decline in deposit growth could prompt institutions to focus on cost control measures, improve operational efficiencies, or explore alternative revenue streams. Strategic planning based on accurate deposit growth forecasts helps institutions navigate the complexities of the financial environment and position themselves for sustainable growth. This proactive approach is essential in an industry where changes in market conditions or consumer behavior can have significant implications for a bank's financial health.

For policymakers, understanding deposit trends is integral to the formulation and adjustment of monetary policy. Deposits are a primary component of the money supply, and their growth rates can signal shifts in economic conditions that may require policy interventions. For instance, a surge in deposit growth might indicate increased savings, potentially reflecting subdued consumer spending and a slowdown in economic activity, prompting policymakers to consider measures such as interest rate cuts to stimulate spending. Conversely, sluggish deposit growth might suggest economic overheating, where tightening monetary policy by raising interest rates could be necessary to curb inflationary pressures. Thus, deposit trends are a key data point for central banks in calibrating monetary policy to maintain economic stability.

The impact of deposit growth on money supply and interest rates underscores the interconnectedness of banking operations and macroeconomic policy. Changes in deposit growth can influence the broader financial environment by altering the availability of funds for lending, which in turn affects credit conditions and economic activity. For example, rapid deposit growth can lead to an increase in the supply of loanable funds, exerting downward pressure on interest rates and potentially stimulating economic expansion. On the other hand, slow deposit growth can constrain the availability of credit, leading to higher interest rates and potentially dampening economic activity [12]. Policymakers carefully monitor deposit trends to ensure that their monetary policy settings are aligned with the current economic conditions and to mitigate potential risks to financial stability.

2 Factors Influencing the Deposit Growth

Bank deposit growth rates are significantly influenced by macroeconomic variables such as GDP, inflation, money supply (M2), recession periods, and interest rates. These factors do not operate in isolation but interact within a complex economic framework, guided by various economic theories that help explain their impact on deposit growth. A detailed examination of these interactions, informed by economic theory, provides a clearer understanding of how these variables shape the growth of bank deposits.

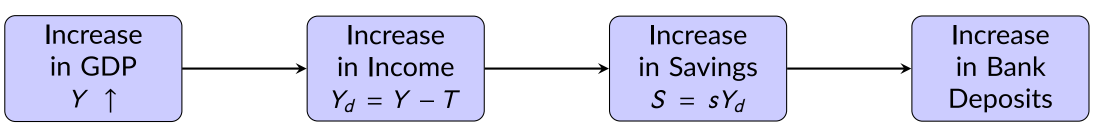

Figure 3. Impact of GDP Growth on Bank Deposit Growth
Theory: Marginal Propensity to Save (MPS), $S = sY_d$ [13]

GDP is a fundamental indicator of economic activity and directly influences bank deposits. According to classical economic theory, an increase in GDP typically leads to higher incomes, which in turn increases the marginal propensity to save. This increase in savings is often reflected in higher bank deposits. The life-cycle hypothesis (LCH), proposed by Franco Modigliani, further elaborates on this by suggesting that individuals plan their savings over their lifetime, saving more during their working years to fund retirement. As GDP grows, individuals may increase their savings to secure their future, leading to a rise in bank deposits.

However, the relationship between GDP and deposit growth is not always straightforward. The

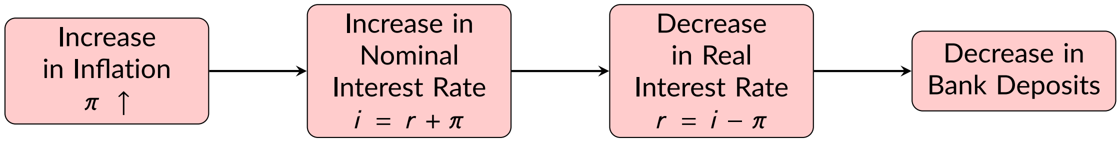

Figure 4. Effect of Inflation on Real Interest Rate and Bank Deposit Growth
Theory: Fisher Equation, $r = i - \pi$ [14]

Keynesian consumption function suggests that as people's incomes rise, their consumption also increases, though not necessarily at the same rate. This could lead to a slower growth rate in savings and, consequently, in bank deposits, especially if the increased income is largely directed toward consumption rather than savings. Additionally, in economies with sophisticated financial markets, individuals might divert a larger portion of their income to investments in stocks, bonds, or other financial instruments, reducing the amount that ends up as bank deposits. This behavior aligns with the portfolio theory, which suggests that individuals allocate their wealth across different assets to maximize returns and minimize risk.

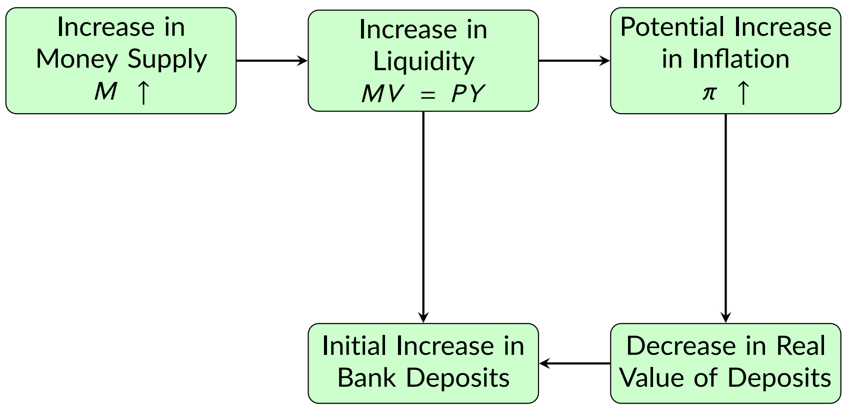

Figure 5. Impact of Money Supply (M2) on Bank Deposit Growth
Theory: Quantity Theory of Money, $MV = PY$ [15]

Inflation, which measures the general rise in price levels, has a complex impact on bank deposit growth, mediated through its effect on real and nominal interest rates. According to the Fisher effect, nominal interest rates are adjusted to compensate for expected inflation, meaning that higher inflation typically leads to higher nominal interest rates. This adjustment is intended to preserve the real rate of return on deposits, thereby maintaining the incentive to save. However, if inflation rises faster than nominal interest rates, the real interest rate (nominal rate minus inflation) could become negative, discouraging savings and reducing bank deposit growth. The theory of rational expectations also plays a role here, as individuals form expectations about future inflation and adjust their saving and investment behavior accordingly. If they expect inflation to persist, they might shift their savings from bank deposits to assets that are perceived as better hedges against inflation, such as real estate or commodities.

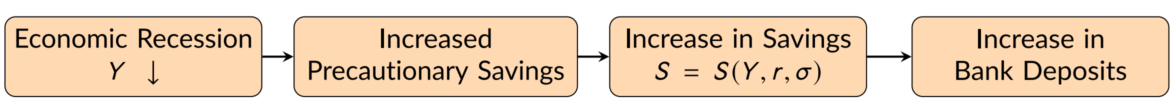

Figure 6. Effect of Recession on Bank Deposit Growth Through Precautionary Savings
Theory: Precautionary Savings Function, $S = S(Y, r, \sigma)$ [16]

In extreme cases, such as hyperinflation, confidence in the currency can collapse, leading to a sharp reduction in bank deposits as individuals seek to protect their wealth through foreign currency holdings or tangible assets. This phenomenon is consistent with the quantity theory of money, which posits that a rapid increase in the money supply, if not matched by an increase in real output, leads to inflation. If inflation expectations become unanchored, people might withdraw their deposits en masse, leading to a contraction in the deposit base.

The money supply, specifically M2, which includes cash, checking deposits, and easily convertible near money, is another critical factor influencing bank deposit growth. The quantity theory of money, articulated by economists like Milton Friedman, suggests that an increase in the money supply, holding velocity and output constant, should lead to a proportional increase in nominal GDP and price levels. When the central bank expands the money supply, typically through open market operations or other forms of monetary stimulus, the immediate effect is often an increase in bank reserves and deposits.

However, the long-term impact on bank deposits depends on how this increased money supply is utilized within the economy. If the expansion in M2 leads to higher inflation without corresponding real GDP growth, the real value of deposits could decline, reducing the incentive to hold money in banks. This outcome aligns with the expectations-augmented Phillips curve, which suggests that if monetary expansion leads to inflation without reducing unemployment, inflationary pressures will erode real returns, discouraging savings. Conversely, if the increase in M2 corresponds with genuine economic growth, deposit growth may be sustained as the economy expands and savings increase in tandem.

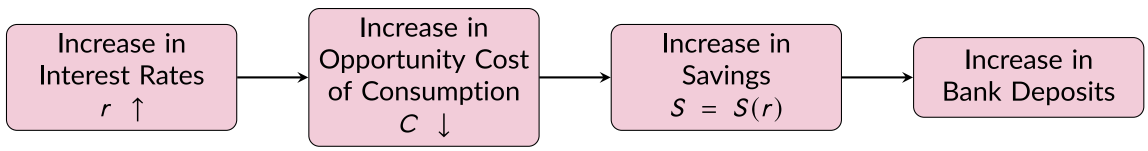

Figure 7. Impact of Interest Rates on Bank Deposit Growth
Theory: Loanable Funds Theory, $S = S(r)$ [17]

Recession periods present a unique context for bank deposit growth, often leading to seemingly paradoxical effects. During recessions, economic activity contracts, leading to lower incomes and reduced consumption. However, the uncertainty and fear associated with recessions often lead to an increase in precautionary savings. The permanent income hypothesis (PIH), developed by Milton Friedman, suggests that individuals base their consumption on expected lifetime income rather than current income. During a recession, when future income prospects are uncertain, individuals might increase their savings to smooth consumption over time, leading to an increase in bank deposits.

This behavior is further explained by the concept of the precautionary motive for saving, where individuals hold more liquid assets, such as bank deposits, during periods of economic uncertainty to hedge against future income shocks. The paradox of thrift, a concept introduced by Keynes, also plays a role here. As individuals and businesses collectively increase their savings during a recession, overall demand in the economy may decrease, potentially prolonging the recession. This increased saving, while stabilizing individual finances, can lead to higher bank deposits in the short term.

The impact of recessions on bank deposits also depends on the type of recession and the accompanying policy responses. In a financial crisis-induced recession, such as the 2007-2008 Global Financial Crisis, there might initially be a sharp decline in deposits as financial institutions face liquidity issues and depositor confidence wanes. However, aggressive monetary interventions by central banks, such as lowering interest rates and implementing quantitative easing, can restore confidence and lead to a recovery in deposits as the economy stabilizes.

Interest rates, both nominal and real, are crucial in determining bank deposit growth. The classical loanable funds theory posits that the interest rate is determined by the supply of and demand for loanable funds. When interest rates are high, saving becomes more attractive, leading to an increase in bank deposits. This relationship is straightforward: higher interest rates increase the opportunity cost of consuming today rather than saving, thus encouraging individuals to defer consumption and increase savings.

However, in a low-interest-rate environment, the situation becomes more complex. If nominal interest rates are low and inflation is moderate, the real return on bank deposits might be close to zero or even negative. According to the liquidity preference theory, developed by Keynes, individuals prefer to hold money (or liquid assets) when interest rates are low because the opportunity cost of holding money is minimal. In such cases, deposit growth might still be sustained if banks are perceived as safe havens during economic uncertainty, even though the returns on deposits are low.

The Taylor rule, a principle guiding central banks on setting interest rates based on economic conditions, provides further insight. According to the Taylor rule, central banks adjust nominal interest rates based on deviations of actual inflation from target inflation and actual GDP from potential GDP. In periods of low inflation and below-potential GDP, central banks might lower interest rates to stimulate economic activity. While this could reduce the attractiveness of bank deposits due to lower returns, the overall impact on deposit growth depends on whether the lower interest rates succeed in stimulating economic activity and boosting income levels, which could eventually lead to higher deposit growth.

The theory of financial repression, where governments keep interest rates artificially low to reduce the cost of public debt servicing, also impacts deposit growth. In such environments, savers might receive negative real returns on their deposits, discouraging savings in traditional bank accounts. However, if there are capital controls or other restrictions that limit the ability to move funds out of the banking system, deposits might still grow, driven more by necessity than by attractive returns.

3 Challenges in Traditional Predictive Models

Traditional models for predicting bank deposit growth, such as autoregressive integrated moving average (ARIMA) and vector autoregression (VAR), have relied heavily on historical data and linear relationships [18]. These models often fall short in capturing the complexities of financial data, which are influenced by a multitude of factors that exhibit nonlinear and dynamic behaviors. The increasing volatility in financial markets and the intricate interdependencies among macroeconomic variables have highlighted the limitations of these traditional approaches. As a result, there is a growing need for more sophisticated models that can handle large datasets, uncover hidden patterns, and improve predictive accuracy [19].

Challenges in Traditional Predictive Models	**Machine Learning as an Advanced Analytical Tool**
Traditional models such as ARIMA and VAR rely heavily on historical data and linear relationships.	Machine learning can analyze large datasets and identify complex patterns not apparent with traditional methods.
These models often fail to capture the complexities of financial data, influenced by nonlinear and dynamic behaviors.	Machine learning models nonlinear relationships and adapts to changing data distributions, making it suitable for financial forecasting.
Increasing volatility in financial markets and intricate interdependencies among macroeconomic variables limit the effectiveness of traditional approaches.	Machine learning can integrate temporal dependencies with macroeconomic variables, offering a holistic view of factors driving bank deposit growth.
There is a need for more sophisticated models to handle large datasets, uncover hidden patterns, and improve predictive accuracy.	The flexibility of machine learning algorithms, from simple regressions to complex neural networks, allows for tailored approaches to specific data challenges.

Table 1. Comparison of Traditional Predictive Models and Machine Learning in Financial Forecasting

Machine Learning as an Advanced Analytical Tool Machine learning has emerged as a powerful tool for analyzing large datasets and identifying complex patterns that are not readily apparent with traditional statistical methods. Its ability to model nonlinear relationships and adapt to changing data distributions makes it suitable for financial forecasting. In the context of bank deposit growth, machine learning can integrate temporal dependencies with macroeconomic variables, providing a more holistic view of the factors driving deposit trends. The flexibility of machine learning algorithms, ranging from simple regression models to complex neural networks, allows for tailored approaches that can address specific challenges in the data.

4 Research objective

The banking sector relies heavily on accurate forecasting to maintain stability and foster growth. Predicting deposit growth is crucial as it directly influences liquidity management, credit allocation, and overall financial health. Traditional statistical methods, while useful, often fall short in capturing the complex, nonlinear relationships inherent in economic data. Machine learning, with its advanced computational capabilities and adaptability, offers a promising alternative. This paper investigates the efficacy of ML algorithms in predicting U.S. bank deposit growth through both univariate and bivariate analyses, aiming to enhance understanding of temporal dependencies and macroeconomic interrelationships.

5 Methods

Recurrent Neural Networks (RNNs) are a class of artificial neural networks where connections between nodes form a directed graph along a sequence, allowing the model to exhibit dynamic temporal behavior [20]. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs, making them well-suited for tasks where data is sequential, such as time series analysis, language modeling, and speech recognition. The architecture of an RNN allows it to maintain a hidden state that captures information about previous inputs, which is updated at each time step based on both the current input and the previous hidden state. This recursive process can be mathematically described as:

$$h_t = \sigma(W_h \cdot h_{t-1} + W_x \cdot x_t + b_h)$$

$$y_t = \sigma(W_y \cdot h_t + b_y)$$

where h_t represents the hidden state at time t, x_t is the input at time t, W_h and W_x are weight matrices, b_h and b_y are bias terms, and σ is an activation function, typically a sigmoid or hyperbolic tangent. Despite their theoretical appeal, standard RNNs suffer from several practical issues, notably the vanishing gradient problem, where gradients of the loss function with respect to the weights diminish exponentially as they are propagated back through time. This issue severely hampers the learning of long-term dependencies. Various extensions of RNNs, such as Long Short-Term Memory (LSTM) [21] and Gated Recurrent Units (GRUs), have been developed to mitigate these problems by introducing mechanisms that better capture long-term dependencies. Cointegration tests, often employed in econometrics, are crucial when analyzing non-stationary time series data to determine whether a linear combination of two or more series can result in a stationary series [22]. Cointegration implies a long-run equilibrium relationship between the series, even though the individual series themselves may be non-stationary. The Engle-Granger test, Johansen test, and Phillips-Ouliaris test are among the most commonly used cointegration tests. The Engle-Granger test involves first estimating the long-term relationship using ordinary least squares (OLS) and then applying an Augmented Dickey-Fuller (ADF) test on the residuals to check for stationarity. The Johansen test, on the other hand, is a more comprehensive method that allows for multiple cointegrating relationships and is based on a Vector Autoregressive (VAR) model [23]. The test statistic for the Johansen test is derived from the eigenvalues of the stochastic matrix formed by the VAR process, and the number of cointegrating vectors is determined based on trace and maximum eigenvalue statistics. The mathematical expression for the Engle-Granger cointegration test is as follows [24]. Suppose we have two non-stationary series y_t and x_t. The cointegration relationship can be modeled as:

$$y_t = \alpha + \beta x_t + u_t$$

where u_t represents the residuals. If u_t is stationary, then y_t and x_t are cointegrated. The ADF test is then applied to u_t to determine whether it is stationary. The null hypothesis H_0 of the Engle-Granger test is that u_t has a unit root (i.e., it is non-stationary), which implies no cointegration. Parameter instability is another critical concern in time series analysis when dealing with models that are estimated over long periods where structural breaks or changes in the underlying data-generating process might occur. The Hansen Parameter Instability test is a statistical method used to assess whether the parameters of a model are stable over time. This test is based on the recursive estimation of the model and examines the stability of the parameters by comparing estimates from different sub-samples of the data. The test statistic, denoted as λ_{max}, is computed from the largest eigenvalue of the covariance matrix of the recursive residuals. The test's null hypothesis is that the parameters are stable over time, and rejection of this hypothesis suggests the presence of structural breaks or non-constant parameters.

The Dynamic Ordinary Least Squares (DOLS) estimator is a technique designed to provide robust estimates in cointegrated systems by addressing endogeneity and serial correlation issues. DOLS extends the standard OLS by including leads and lags of the differenced explanatory variables, thereby removing endogeneity in the cointegrating equation. The inclusion of these leads and lags serves to account for the feedback effects from the dependent variable to the explanatory variables, which are typical in cointegrated systems. The DOLS estimator can be expressed as:

$$y_t = \alpha + \beta x_t + \sum_{j=-p}^{q} \gamma_j \Delta x_{t+j} + u_t$$

where Δx_{t+j} represents the lead or lag of the differenced explanatory variable, and p and q denote the number of leads and lags, respectively. The resulting estimator is asymptotically efficient and has desirable small-sample properties, making it a popular choice for estimating cointegrating relationships.

Fully-Modified Ordinary Least Squares (FM-OLS) is another method for estimating cointegrating vectors in the presence of non-stationarity. FM-OLS improves upon standard OLS by correcting for the effects of serial correlation and endogeneity in the regressors. This correction is achieved by modifying the OLS estimator to account for the long-run covariance between the regressors and the error term. The FM-OLS estimator involves two main adjustments: first, adjusting the OLS residuals for serial correlation by using non-parametric estimates of the long-run variance, and second, correcting the OLS estimates for endogeneity by incorporating a correction term based on the covariance between the regressors and the error term. The FM-OLS estimator is useful in situations where the cointegrating relationship involves multiple variables and the regressors exhibit endogeneity or serial correlation.

The mathematical representation of the FM-OLS estimator can be described as follows. Consider a cointegrated system represented by:

$$y_t = \alpha + \beta x_t + u_t$$

where u_t is the error term, which may be serially correlated and correlated with x_t. The FM-OLS estimator modifies the standard OLS estimator by adjusting the residuals and incorporating a correction term $\hat{\beta}_{FM-OLS}$ to account for the bias induced by endogeneity and serial correlation:

$$\hat{\beta}_{FM-OLS} = \hat{\beta}_{OLS} - \hat{\Gamma}_{xy}\hat{\Omega}_{xx}^{-1}$$

where $\hat{\Gamma}_{xy}$ is the covariance matrix between the residuals and the regressors, and $\hat{\Omega}_{xx}$ is the long-run variance of the regressors. This adjustment ensures that the FM-OLS estimator is consistent and asymptotically efficient, even in the presence of endogeneity and serial correlation.

6 Data

The study utilizes data on deposits at all commercial banks in the United States from 1973 to 2019, recorded in billions of U.S. dollars. This dataset is seasonally adjusted and measured on a weekly basis, providing a detailed and continuous view of the trends and fluctuations in bank deposits over an extensive historical period. In addition to deposit data, several key macroeconomic indicators are included to enhance the analysis: *Gross Domestic Product (GDP), inflation rates, money supply (M2), recession periods, and interest rates.*

GDP data is measured quarterly in billions of dollars and adjusted for seasonal variations on an annual basis, offering insights into the overall economic activity and growth within the country. Inflation rates are derived from the Consumer Price Index (CPI) and reflect the rate at which the general level of prices for goods and services rises, eroding purchasing power. The money supply (M2) includes cash, checking deposits, and easily convertible near money, which provides a broader view of the liquidity available in the economy. Recession periods are identified based on the National Bureau of Economic Research (NBER) business cycle dating, and interest rates are tracked to understand their impact on both deposit levels and broader economic conditions.

All of these datasets are sourced from the Federal Reserve Economic Data (FRED) provided by the Federal Reserve Bank of St. Louis, a reliable and widely used source for economic and financial data.

7 Preprocessing

Preprocessing steps include converting the relevant date column to a datetime format to facilitate its use in time series analysis. The study then checks for missing values in the target column, and if any are found, their count is output for further handling if necessary.

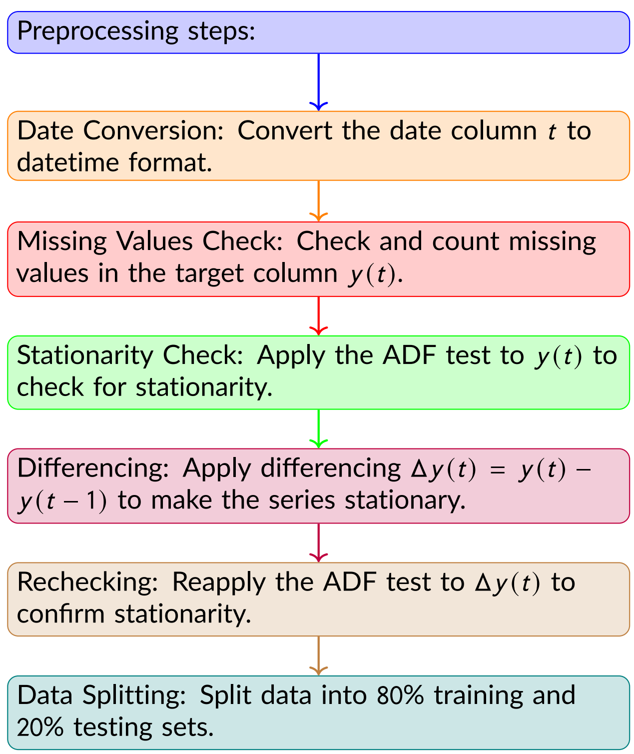

Figure 8. Preprocessing Steps

Stationarity is a fundamental concept in time series analysis, referring to a time series whose statistical properties, such as mean, variance, and autocorrelation, are constant over time. Mathematically, a time series $\{y_t\}$ is stationary if for all t, the joint probability distribution does not change when shifted in time. This implies that $\mathbb{E}[y_t] = \mu$, $\text{Var}(y_t) = \sigma^2$, and the covariance $\text{Cov}(y_t, y_{t+h})$ depends only on the lag h and not on time t. Stationarity is a crucial assumption in many statistical and econometric models, such as those used for forecasting, as it ensures that the model's parameters remain consistent over time.

The Augmented Dickey-Fuller (ADF) test is a widely used statistical test for stationarity, specifically for detecting the presence of a unit root in a time series. The presence of a unit root indicates that the series is non-stationary and exhibits a stochastic trend. The ADF test is an extension of the Dickey-Fuller test and includes lagged differences of the series to account for higher-order autocorrelations. The test is based on estimating the following regression model:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \epsilon_t,$$

where $\Delta y_t = y_t - y_{t-1}$ is the first difference of the series, α is a constant, βt is a time trend, and p is the number of lagged differences included in the model. The null hypothesis $H_0 : \gamma = 0$ suggests that the series has a unit root, implying non-stationarity.

The test statistic for the ADF test is compared to critical values from the Dickey-Fuller distribution to determine whether to reject the null hypothesis. If the test statistic is less than the critical value, the null hypothesis of a unit root is rejected, indicating that the series is stationary. However, if the series is non-stationary, differencing can be applied to transform it into a stationary series. Differencing involves subtracting the previous observation from the current observation, mathematically expressed as $\Delta y_t = y_t - y_{t-1}$. If the series is differenced d times to achieve stationarity, it is said to be integrated of order d, denoted as $I(d)$.

The process of differencing is a common technique to make a time series stationary. By differencing the series, we remove trends and seasonal structures, thereby stabilizing the mean of the series. For example, first differencing a series involves transforming y_t to $\Delta y_t = y_t - y_{t-1}$, while second differencing, if necessary, involves differencing the first differences: $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$. If a series is differenced once and achieves stationarity, it is said to be integrated of order one, or $I(1)$. In practice, most non-stationary time series can be made stationary through first or second differencing.

After differencing, it is standard practice to reapply the ADF test to confirm the stationarity of the differenced series. If the test confirms that the differenced series is stationary, we can proceed with modeling or forecasting using methods that assume stationarity. The importance of ensuring stationarity lies in the fact that most time series models, such as ARIMA (AutoRegressive Integrated Moving Average) models, require the input series to be stationary for the results to be valid and reliable. Therefore, the combination of differencing and the ADF test is a powerful approach for preparing non-stationary time series data for analysis.

To assess the stationarity of the time series, the Augmented Dickey-Fuller (ADF) test is applied to the target column, with the ADF statistic and p-value printed for evaluation. If the series is determined to be non-stationary, differencing is applied to make it stationary, and the resulting differenced series is stored in a new column. The stationarity of this differenced series is rechecked by applying the ADF test again to confirm that the series has become stationary. Subsequently, the dataset is split into training and testing sets, with 80% of the data allocated to the training set and 20% to the testing set, and the sizes of these sets are printed. In this specific case, no missing values were found. The initial ADF statistic was 2.6176 with a p-value of 0.9991, indicating non-stationarity, while after differencing, the ADF statistic was -5.9634 with a p-value of 2.01e-07, confirming that the series became stationary.

8 Results

The results of this study present a detailed evaluation of various predictive modeling techniques applied to forecast the growth rates of bank deposits in the United States, based on historical data from 1973 to 2019. The analysis is divided into two main parts: univariate time series analysis and bivariate analysis incorporating GDP data.

Description	Value
No missing values found	—
ADF Statistic	2.6176313640743274
p-value	0.9990782084455896
ADF Statistic after differencing	-5.963421789813776
p-value after differencing	2.0138275665144432e-07

Table 2. ADF Test Results

8.1 Univariate temporal dependencies analysis results

In the univariate time series analysis, five different models were applied: SARIMA, Prophet, ETS, LSTM, and Transformer. Each model's performance was evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). The SARIMA model, commonly used for time series forecasting, showed moderate predictive capability with a relatively high error (MAE of 2179.148 and RMSE of 2925.733) and a negative R^2 value of -0.0381. The negative R^2 indicates that the model does not explain the variability in the deposit data well and performs worse than a simple mean-based prediction.

Model	MAE	RMSE	MAPE	R^2
SARIMA	2179.1480	2925.7333	13.6149	-0.0381
Prophet	1596.4204	2210.0113	nan	0.4077
ETS	3931.9758	4868.9082	25.4012	-1.8749
LSTM	767.6600	987.4566	21.1641	0.8797
Transformer	1970.7211	2810.7283	11.9358	0.0268

Table 3. Performance Metrics for Different Forecasting Models

The Prophet model, which is designed for time series with strong seasonal effects and missing data, performed better than SARIMA, with lower error metrics (MAE of 1596.420 and RMSE of 2210.011) and a positive R^2 value of 0.4077, indicating a better fit to the data. However, the model's inability to compute MAPE suggests potential issues with handling certain data characteristics. The ETS model, which is based on exponential smoothing, performed poorly, with the highest error metrics (MAE of 3931.976 and RMSE of 4868.908) and a negative R^2 value of -1.8749. This performance indicates that ETS was not well-suited to this time series data, possibly due to its inability to capture the underlying patterns effectively.

The LSTM model demonstrated the strongest performance among the univariate models, with the lowest error metrics (MAE of 767.660 and RMSE of 987.457) and a high R^2 value of 0.8797. This result indicates that LSTM, a type of recurrent neural network capable of learning long-term dependencies, was effective in modeling the complex temporal patterns in the bank deposit data. The Transformer model, another deep learning approach, showed moderate performance with an MAE of 1970.721, RMSE of 2810.728, and an R^2 value of 0.0268. While it outperformed SARIMA and ETS, its performance was still inferior to LSTM, possibly due to its architectural differences in handling sequential data.

8.2 Multivariable Macroeconomic Interrelationship analysis results

```
Epoch 49/50
5/5  0s 19ms/step - loss: 6.6809e-06 - val_loss: 0.0048
Epoch 50/50
5/5  0s 21ms/step - loss: 7.2623e-06 - val_loss: 0.0048
5/5  1s 68ms/step
2/2  0s 6ms/step
Training MSE: 6.544580421788998e-06
Training MAE: 0.0016377903880896445
```

```
Training RMSE: 0.0025582377570876787
Testing MSE: 0.00480853216587751
Testing MAE: 0.04463006866046173
Testing RMSE: 0.0693435805671838
```

	Hansen Parameter Instability Cointegration Test			
Statistic	Stochastic Trends (m)	Deterministic Trends (k)	Excluded Trends (p2)	Probability
Lc statistic	0.005430	4	0	> 0.2

Table 4. Results of the Hansen Parameter Instability Cointegration Test for the series DEPOSIT, GDP, INFLATION, M2, and RECESSION, testing the null hypothesis that the series are cointegrated. The cointegrating equation includes a constant (C). The p-values are based on Hansen (1992b) Lc(m2=4, k=0) distribution, where m2=m-p2 represents the number of stochastic trends in the asymptotic distribution.

Table 5. Dynamic Least Squares (DOLS) Results

Dependent Variable:	DEPOSIT			
Method:	Dynamic Least Squares (DOLS)			
Sample (adjusted):	1973Q3 2019Q4			
Cointegrating equation deterministics:	C			
Fixed leads and lags specification:	lead=1, lag=1			
Long-run variance estimate	Bartlett kernel, Newey-West fixed bandwidth = 5.0000			
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
GDP	0.086350	0.031011	2.784516	0.0059
INFLATION	-191.5082	90.70594	-12.111308	0.0361
M2	0.793206	0.038194	20.76788	0.0000
RECESSION	-129.1248	168.1273	-10.768018	0.0435
C	-681.7733	169.2787	-4.027519	0.0001
R-squared	0.997471	**Mean dependent var**		5339.653
Adjusted R-squared	0.997251	**S.D. dependent var**		4762.715
S.E. of regression	249.7143	**Sum squared resid**		11473730
Long-run variance	273033.1			

Table 6. Fully Modified Least Squares (FMOLS) Results

Dependent Variable:	DEPOSIT			
Method:	Fully Modified Least Squares (FMOLS)			
Sample (adjusted):	1973Q2 2019Q4			
Cointegrating equation deterministics:	C			
Long-run covariance estimate	Bartlett kernel, Newey-West fixed bandwidth = 5.0000			
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
GDP	0.061361	0.027335	2.244762	0.0259
INFLATION	-55.67781	56.49245	-22.985580	0.0255
M2	0.811519	0.033803	24.00733	0.0000
RECESSION	-140.4528	109.3293	-11.284677	0.0004
C	-527.3452	135.3257	-3.896858	0.0001
R-squared	0.995559	**Mean dependent var**		5375.492
Adjusted R-squared	0.995469	**S.D. dependent var**		4825.022
S.E. of regression	324.7814	**Sum squared resid**		20885625
Long-run variance	307742.1			

In the Multivariable analysis, the performances of Recurrent Neural Networks (RNNs) across different lagged periods show how lag selection affects the model's predictive accuracy. Starting with a single lagged period, the model showed a steady improvement in loss metrics throughout

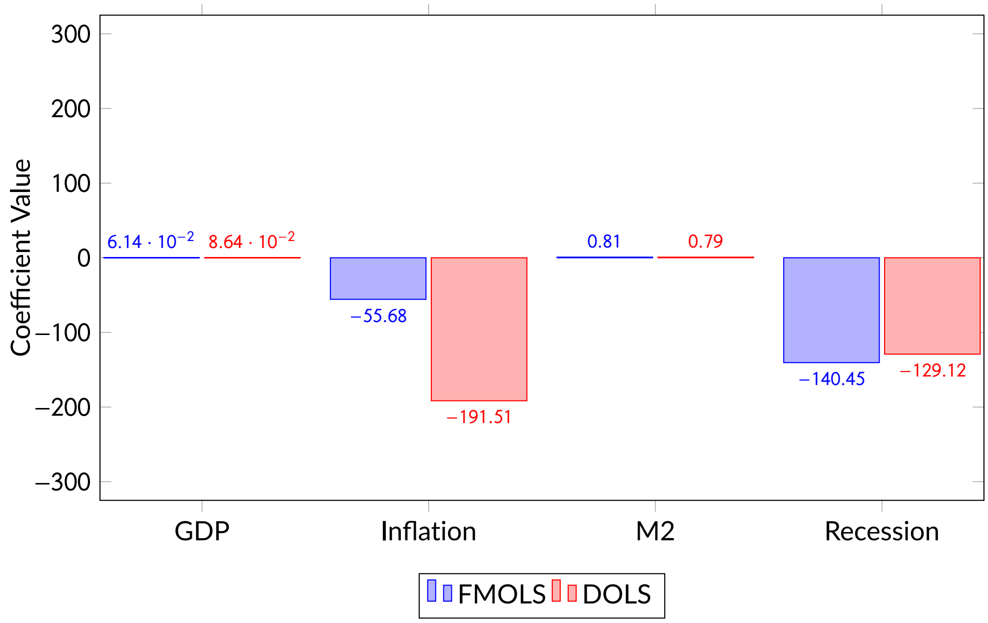

Figure 9. Comparison of Coefficients from FMOLS and DOLS Models

the training process. By the final epoch, the training loss was 0.0606, with a validation loss of 0.0455. The training Mean Squared Error (MSE) was 0.0598, and the Root Mean Squared Error (RMSE) was 0.2446, indicating a decent level of accuracy. The model's performance on the testing data was slightly better, with a testing MSE of 0.0455, suggesting that the model generalizes well with this lag configuration.

Increasing the lagged period to two introduced some changes in the model's performance. The final training loss decreased to 0.0538, but the validation loss increased slightly to 0.0574, indicating potential overfitting. The training MSE for this configuration was 0.0546, with an RMSE of 0.2337. The testing phase showed a slight increase in errors, with a testing MSE of 0.0574 and an RMSE of 0.2396. These results suggest that while the model's training performance improved slightly, its generalization to unseen data did not improve and may have worsened.

When extending the lagged period to three, the model's performance exhibited further nuances. The training loss continued to decrease, reaching 0.0475 by the final epoch, but the validation loss remained relatively stable at 0.0545. The training MSE was 0.0458, with an RMSE of 0.2139. However, the testing performance did not improve significantly, as indicated by a testing MSE of 0.0545 and an RMSE of 0.2335. This suggests that increasing the lagged period further did not lead to a substantial improvement in model performance, and the benefit of additional lags might be diminishing.

With a four-lagged period, the model's performance metrics showed a different trend. The training loss dropped to 0.0301, the lowest observed across all configurations, suggesting that the model was fitting the training data very well. However, the validation loss increased to 0.0709, indicating a higher level of overfitting. The training MSE was 0.0259, with an RMSE of 0.1610, which is the lowest among all models. Despite the low training error, the testing MSE increased to 0.0709, and the RMSE rose to 0.2663, indicating that the model's ability to generalize to new data had decreased as the lagged period increased.

These results indicate that while increasing the number of lagged periods in an RNN can improve the model's fit to the training data, it does not necessarily translate to better performance on unseen data. There is a trade-off between fitting the training data and generalizing to new data, and the optimal lag configuration appears to be a balance between these two factors.

In addition to these RNN results, the cointegration analysis using Hansen's parameter instability test for the series DEPOSIT, GDP, INFLATION, M2, and RECESSION was conducted. The null hypothesis, which posits that the series are cointegrated, was not rejected given the Lc statistic of 0.005430 with a probability value greater than 0.2. This suggests that the variables share a long-term equilibrium relationship, despite short-term deviations.

Further econometric analysis was conducted using Fully Modified Least Squares (FMOLS) and Dynamic Least Squares (DOLS) methods to estimate the long-run relationships among these variables. The FMOLS results indicate that GDP has a positive coefficient of 0.0614, suggesting a direct but modest impact on DEPOSIT levels. Inflation has a significant negative impact, with a coefficient of -55.6778, reflecting its adverse effect on deposits. M2, representing money supply, has a strong positive coefficient of 0.8115, indicating its crucial role in determining deposit levels. The RECESSION variable has a significant negative effect, with a coefficient of -140.4528, indicating that economic downturns significantly reduce deposit levels. The model's R-squared value of 0.9956 suggests that the explanatory variables account for nearly all the variance in DEPOSIT levels, with a high degree of precision in the long-run variance estimate.

The DOLS estimation, with a lag of one period, produced similar results, albeit with some differences in the magnitude of coefficients. GDP's impact was slightly higher at 0.0864, while the negative effect of inflation increased significantly to -191.5082. M2's influence remained strong, with a coefficient of 0.7932, and the recession's negative impact was slightly less severe at -129.1248. The model's R-squared was slightly higher at 0.9975, further supporting the strong explanatory power of the selected variables.

The LSTM model emerged as the most effective univariate model for predicting bank deposit growth rates, demonstrating its ability to capture complex temporal dependencies in the data. in the multivariable analysis, the RNN models demonstrated varying degrees of effectiveness based on the number of lagged periods, with the model performance generally decreasing as the lag increased beyond one or two periods, suggesting potential overfitting with more complex configurations. The cointegration analysis confirmed the long-term equilibrium relationship between DEPOSIT and the other economic indicators, with both FMOLS and DOLS methods highlighting the significant impact of these variables on DEPOSIT, albeit with some differences in the magnitude of their influence. These findings provide a robust understanding of the dynamic relationships and suggest the importance of selecting lagged periods in time series models to balance model complexity and generalization performance.

9 Policy implications

9.1 Regulatory Oversight and Policy Formulation

The results of the study highlight the need for robust regulatory oversight. The differing performances of models like SARIMA, Prophet, ETS, LSTM, and Transformer indicate that not all models are equally effective in capturing the complexities of financial data. Regulatory frameworks must ensure that financial institutions use models that are both reliable and appropriate for the data they analyze.

LSTM models, which outperformed traditional statistical models, exemplify the shift toward machine learning in financial forecasting. However, the poor performance of models like SARIMA and ETS suggests that outdated approaches still in use may not effectively capture financial trends. This discrepancy calls for regulatory bodies to update guidelines and standards, ensuring that institutions adopt advanced models where appropriate while maintaining rigorous standards for model validation, governance, and transparency.

The study also emphasizes the importance of integrating macroeconomic factors like GDP, inflation, and money supply into regulatory oversight. Traditional economic theories, such as the Quantity Theory of Money, suggest a direct link between money supply and nominal deposits, which aligns with the study's findings. However, the adverse impact of inflation on deposits complicates this relationship, underscoring the need for regulatory policies that account for both short-term fluctuations and long-term trends. Regulators, including the Federal Reserve and the Office of the Comptroller of the Currency (OCC), should incorporate the findings from advanced predictive

Table 7. Regulatory Oversight and Policy Formulation

Aspect	**Requirement**	**Description**
Model Risk Management	Advanced Model Adoption	Financial institutions should adopt models like LSTM for improved accuracy in forecasting.
Governance	Continuous Model Validation	Ongoing validation and review of models to ensure compliance with regulatory standards.
Macroeconomic Integration	Incorporate Economic Indicators	Regulatory policies should integrate indicators like GDP, inflation, and M2 for better oversight.
Stress Testing	Enhanced Scenario Analysis	Use advanced models in stress testing to evaluate resilience under adverse economic conditions.

models into their macroprudential oversight. The study's identification of long-term relationships between deposits and key economic indicators suggests that these variables should be closely monitored to manage systemic risks effectively. Systemic risk management, aimed at preventing the collapse of financial systems, can benefit from the predictive power of models like LSTM and RNNs, which can anticipate the effects of macroeconomic shocks on deposits.

Stress-testing models that incorporate advanced findings should be a regulatory priority. Such models can help financial institutions evaluate their resilience under adverse conditions. By adopting advanced machine learning models in stress testing, regulators can better assess the vulnerabilities within the banking system and take steps to mitigate risks. Financial institutions must align with evolving regulatory standards by adopting advanced predictive modeling techniques. The findings indicate that traditional models may no longer suffice in understanding modern financial markets. Therefore, institutions should invest in developing and deploying machine learning models like LSTM to enhance forecasting accuracy and improve risk management. Model validation and governance, as outlined in regulatory guidelines like BCBS 239, are essential for the successful adoption of these models. Financial institutions should implement robust validation frameworks that evaluate model accuracy, robustness, and interpretability, ensuring compliance with regulatory standards. Governance structures must support ongoing review and documentation of these models to manage potential risks effectively.

9.2 Implications for Monetary Policy and Economic Stability

The study's findings have significant implications for monetary policy, particularly in managing the relationship between bank deposits and macroeconomic factors like GDP, inflation, and money supply. Understanding these relationships is crucial for formulating policies that ensure economic stability.

Table 8. Implications for Monetary Policy and Economic Stability

Variable	**Effect**	**Monetary Policy Implication**
Money Supply (M2)	Positive Influence on Deposits	Expansion of money supply can increase deposit levels; central banks should manage M2 to ensure liquidity.
Inflation	Erosion of Deposit Value	High inflation negatively impacts the real value of deposits; monetary policy should focus on maintaining price stability.
Economic Cycles	Cyclical Impact on Deposits	Recessions reduce deposit levels; countercyclical policies can help stabilize deposit growth during downturns.

The positive relationship between money supply (M2) and bank deposits, as revealed by the study, underscores the importance of liquidity management in monetary policy. According to the Quantity Theory of Money, increasing the money supply should boost nominal deposits, enhancing liquidity within the banking system. However, the study also reveals that inflation negatively impacts deposit levels, consistent with the Fisher Equation, which relates nominal interest rates to real interest rates and inflation. This suggests that central banks must balance stimulating economic growth through monetary expansion with maintaining price stability to prevent deposit erosion.

Recessions significantly impact deposit levels, reflecting the cyclical nature of economic activity. The study's findings align with the financial accelerator theory, which suggests that economic shocks are amplified by financial market frictions, leading to deeper downturns. This insight emphasizes the need for countercyclical monetary policies that stabilize deposit levels during downturns, helping to maintain overall economic stability.

Monetary policy can mitigate the cyclical effects of economic fluctuations by adjusting interest rates to support deposit levels during downturns and control inflation during expansions. This approach helps to manage the economy's cyclical nature and protect financial stability.

9.3 Strategic Implications for Financial Institutions

The study provides actionable insights for financial institutions in managing deposit growth and mitigating economic risks. The superior performance of the LSTM model suggests that banks should prioritize advanced predictive modeling techniques to improve decision-making.

Table 9. Strategic Implications for Financial Institutions

Area	Strategic Action	Implication
Liquidity Management	Adoption of Predictive Models	Utilize advanced models to forecast deposit trends and manage liquidity buffers effectively.
Interest Rate Strategy	Adjust for Inflation	Modify interest rates to counteract inflationary pressures and maintain competitive deposit offerings.
Capital Planning	Enhanced Stress Testing	Integrate advanced models into stress testing to ensure adequate capital levels and prepare for economic shocks.

Effective liquidity management is critical for financial stability. The study's findings on the relationship between money supply and deposits indicate that banks should monitor macroeconomic indicators like M2 closely. Anticipating changes in these indicators allows banks to adjust liquidity buffers and asset-liability management strategies, ensuring they can meet customer demands.

Inflation's negative impact on deposits highlights the importance of managing interest rate risk in an inflationary environment. Banks must consider how inflation affects the real value of deposits and adjust interest rates to remain competitive and prevent outflows. Advanced predictive models, such as LSTM and RNNs, provide banks with the tools needed to forecast inflation trends and adjust interest rate strategies accordingly.

The effects of macroeconomic variables like GDP, inflation, and recessions on deposits underscore the need to integrate these factors into capital planning and stress-testing frameworks. Regulatory bodies like the Federal Reserve mandate stress testing under the Dodd-Frank Act Stress Test (DFAST) and the Comprehensive Capital Analysis and Review (CCAR) to evaluate capital adequacy under adverse scenarios.

Integrating advanced predictive models into stress testing can enhance accuracy and reliability, ensuring that banks are better prepared for economic shocks. Understanding how macroeconomic variables influence deposits allows banks to anticipate changes and plan for potential challenges, improving overall stability.

This study's findings highlight the need for updated regulatory frameworks, advanced predictive modeling, and strategic planning within financial institutions. The superior performance of machine learning models like LSTM underscores the importance of adopting these techniques in financial forecasting. Regulatory bodies must ensure that institutions use validated, transparent, and robust models, while monetary authorities should leverage these insights to balance economic growth and price stability. Financial institutions, in turn, should align their strategies with these findings, focusing on liquidity management, risk mitigation, and capital adequacy to navigate the complexities of modern financial markets effectively.

10 Conclusion

This study offers a comprehensive analysis of predictive modeling techniques applied to forecast the growth rates of bank deposits in the United States, covering the period from 1973 to 2019. The evaluation spanned both univariate time series analysis and multivariate analysis incorporating macroeconomic indicators such as GDP. Through this investigation, several key insights have emerged.

The univariate time series analysis involved five models: SARIMA, Prophet, ETS, LSTM, and Transformer. Among these, the LSTM model demonstrated the highest predictive accuracy, characterized by the lowest error metrics and the highest R^2 value, making it the most effective model for capturing the complex temporal dependencies within the bank deposit data. In contrast, traditional models like SARIMA and ETS showed limitations, with high error metrics and, in some cases, negative R^2 values, indicating poor fit and predictive capability. The performance of the Prophet model, while better than SARIMA and ETS, was still inferior to the LSTM, primarily due to its difficulty in handling certain data characteristics, such as missing values or atypical seasonal patterns.

The multivariate analysis explored the impact of incorporating macroeconomic variables, using Recurrent Neural Networks (RNNs) with varying lagged periods. The results showed that while increasing the lagged periods initially improved the model's fit to the training data, it did not consistently enhance performance on unseen data. This suggests a trade-off between model complexity and generalization, highlighting the need for careful selection of lag parameters to avoid overfitting.

Further econometric analysis through Fully Modified Least Squares (FMOLS) and Dynamic Least Squares (DOLS) confirmed the significant long-term relationships between bank deposits and key economic indicators. The positive influence of GDP and money supply (M2) on deposit levels aligns with established economic theories, while the negative impact of inflation underscores the sensitivity of deposit growth to price stability. The analysis also revealed that recessions have a profound negative effect on deposit levels, emphasizing the cyclical vulnerability of the banking sector.

The complexity of these models, which rely on deep neural networks with multiple layers and numerous parameters, makes it difficult to interpret how specific inputs are transformed into outputs. This lack of transparency poses challenges for financial institutions that need to justify their model predictions to regulators, stakeholders, and decision-makers.

The issue of model interpretability is especially critical in regulatory oversight and policy formulation. Regulators require transparency in the models used by financial institutions to ensure that decisions based on these models are sound and reliable. The opaque nature of advanced models like LSTM and Transformer can hinder this requirement, as it becomes challenging to explain why a model made a particular prediction or to identify potential biases within the model. Consequently, there is a trade-off between the accuracy and the interpretability of the models used for forecasting bank deposit growth rates, which must be carefully managed by financial institutions and regulators alike.

The dependency on historical data also raises concerns about the models' ability to adapt to structural changes in the economy. Economic relationships that held true in the past may evolve or break down due to changes in regulatory frameworks, technological advancements, or shifts in consumer behavior. If the models are not continuously updated with new data and re-calibrated to reflect these changes, their predictions may become outdated or inaccurate. This limitation highlights the importance of not only relying on historical data but also incorporating mechanisms for model updating and adaptation to ensure that predictions remain relevant and reliable in a changing economic sector.

In the study, increasing the number of lagged periods initially improved the fit to the training data, as evidenced by decreasing training loss. However, this improvement did not consistently translate to better performance on the testing data, which indicated potential overfitting. As the lagged periods increased, the models showed signs of diminishing returns, where the added complexity did not yield corresponding gains in predictive accuracy. This trade-off between model complexity and generalization highlights the challenges of finding the optimal configuration for predictive models. Overly complex models risk capturing idiosyncratic patterns that do not generalize, thus reducing their practical utility in real-world forecasting applications.

References

[1] Diamond DW, Dybvig PH. Banking theory, deposit insurance, and bank regulation. The Journal of Business. 1986;59(1):55-68.

[2] Anginer D, Demirguc-Kunt A, Zhu M. How does deposit insurance affect bank risk? Evidence from the recent crisis. Journal of Banking & finance. 2014;48:312-21.

[3] Buser SA, Chen AH, Kane EJ. Federal deposit insurance, regulatory policy, and optimal bank capital. The journal of Finance. 1981;36(1):51-60.

[4] Goldstein I, Pauzner A. Demand–deposit contracts and the probability of bank runs. the Journal of Finance. 2005;60(3):1293-327.

[5] Han R, Melecky M. Financial inclusion for financial stability: Access to bank deposits and the growth of deposits in the global financial crisis. World bank policy research working paper. 2013;(6577).

[6] Demirgüç-Kunt A, Kane E, Karacaovali B, Laeven L. Deposit insurance around the world: a comprehensive database. 2008.

[7] Calomiris CW. Deposit insurance: lessons from the record. Economic Perspectives. 1989;13(3):10-30.

[8] Hannan TH, Hanweck GA. Bank insolvency risk and the market for large certificates of deposit. Journal of money, credit and banking. 1988;20(2):203-11.

[9] Mondschean TS, Opiela TP. Bank time deposit rates and market discipline in Poland: the impact of state ownership and deposit insurance reform. Journal of Financial Services Research. 1999;15:179-96.

[10] Merton RC. On the cost of deposit insurance when there are surveillance costs. Journal of Business. 1978:439-52.

[11] Kareken JH, Wallace N. Deposit insurance and bank regulation: A partial-equilibrium exposition. Journal of business. 1978:413-38.

[12] Jacewitz S, Pogach J. Deposit rate advantages at the largest banks. Journal of Financial Services Research. 2018;53:1-35.

[13] Mikesell RF, Zinser JE. The nature of the savings function in developing countries: a survey of the theoretical and empirical literature. Journal of Economic Literature. 1973;11(1):1-26.

[14] Sun Y, Phillips PC. Understanding the Fisher equation. Journal of Applied Econometrics. 2004;19(7):869-86.

[15] Friedman M. Quantity theory of money. In: Money. Springer; 1989. p. 1-40.

[16] Carroll CD, Kimball MS. Precautionary saving and precautionary wealth. CFS Working Paper; 2006.

[17] Lindner F. Does saving increase the supply of credit? A critique of loanable funds theory. IMK Working Paper; 2013.

[18] Nelson BK. Time series analysis using autoregressive integrated moving average (ARIMA) models. Academic emergency medicine. 1998;5(7):739-44.

[19] Kinney Jr WR. ARIMA and regression in analytical review: An empirical test. Accounting Review. 1978:48-60.

[20] Schmidt RM. Recurrent neural networks (rnns): A gentle introduction and overview. arXiv preprint arXiv:191205911. 2019.

[21] Graves A, Graves A. Long short-term memory. Supervised sequence labelling with recurrent neural networks. 2012:37-45.

[22] Hubrich K, Lütkepohl H, Saikkonen P. A review of systems cointegration tests. Econometric Reviews. 2001;20(3):247-318.

[23] MacKinnon JG. Critical values for cointegration tests. Queen's Economics Department Working Paper; 2010.

[24] Bayer C, Hanck C. Combining non-cointegration tests. Journal of Time series analysis. 2013;34(1):83-95.