INDIAN INSTITUTE OF TECHNOLOGY JODHPUR

NH 62 Nagaur Road, Karwar, Jodhpur, Rajasthan

MTech – Data Engineering

TIME SERIES ANALYSIS

RESERVE BANK of INDIA BALANCE SHEET FORECASTING

Submitted By:

M24DE3011 Aneerban Chowdhury

M24DE3077 Sneha Sawla

M24DE3053 Neha Challa

Executive Summary

This report presents an extensive exploration into a time series forecasting project for the weekly balance sheet data of the RBI. The proposed analysis integrates statistical modeling and monetary economics in understanding financial patterns and generating predictions. Through structured preprocessing, in-depth exploratory analysis, correlation evaluation, and forecasting using ARIMA models, this study brings forth how the quantitative modeling of central bank balance sheet movements can be used to support monetary decisions and manage liquidity. We have 2004 to 2025 years of data.

Through analysis, it brings forth meaningful trends in deposits, loans, foreign assets, and currency circulation. The ARIMA-based forecasting part underlines short-run predictability and thus presents data-driven insights for policymakers, financial institutions, and researchers.

1. Introduction

1.1 Project Overview

The project investigates RBI's weekly balance sheet data to analyze temporal dynamics in India's monetary system. The dataset, sourced from RBI's official Weekly Statistical Supplement, includes high-frequency financial indicators reflecting India's evolving economic environment. The study dissects assets and liabilities, identifying growth patterns, liquidity conditions, and policy implications.

Dataset Description

Source: Official RBI Database — Weekly Statistical Supplement (WSS) published by the Reserve Bank of India.

File: RBI__Liabilities and Assets.csv

Each row represents the week-ended balance sheet data (usually every Friday).

Category	**Column**	**Description**
Notes	A1__Notes__in__Circulation A2__Notes__Held__in__Banking__Department	Currency in circulation and held by RBI
Deposits (B)	B1__Deposits__Central__Government B3__Deposits__State__Governments B4__Deposits__Scheduled__Commercial__Banks B5__Deposits__Scheduled__State__Coop__Banks	Deposits maintained with RBI by various institutions
Liabilities	C__Other__Liabilities D__Total__Liabilities__OR__Assets	Overall liabilities and assets on the RBI balance sheet
Assets (D)	D1__Foreign__Currency__Assets D2__Gold D3__Rupee__Securities	Composition of RBI's asset holdings
Loans & Advances (E)	E1__Loans__Central__Government, E2__Loans__State__Governments, E4__Loans__Scheduled__Commercial__Banks, E5__Loans__Scheduled__State__Coop__Banks	Loans extended by RBI to various entities
Investments & Others	F__Investments, G__Other__Assets	Investments and miscellaneous assets
Date	Week__Ended (set as index)	Date of the report

1.2 Significance of Time Series Analysis

Financial datasets exhibit sequential dependence—past events significantly influence future outcomes. Understanding these dependencies helps forecast macroeconomic indicators. The project identifies four major characteristics of time series data:

- **Temporal Dependence:** Each observation relies on preceding values, representing continuity in economic behavior.
- **Trend Components:** These depict long-term growth or decline influenced by fiscal and monetary policies.
- **Seasonality:** Cyclical patterns emerge due to recurring phenomena like fiscal year-end adjustments or festival-related spending.
- **Stochastic Variations:** Unpredictable changes caused by shocks such as policy interventions, global crises, or liquidity injections.

Studying these components allows researchers to construct models that explain historical financial behavior and predict future developments.

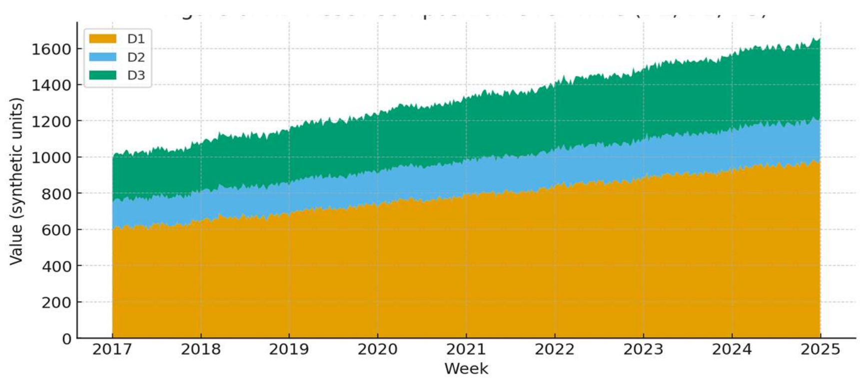
1. Deposits vs Loans Relationships
    - Correlation and trend comparison between respective entities (B1–E1, B3–E2, etc.)
    - Insights into liquidity movements between governments and the banking system
2. Foreign Reserves Composition
    - Comparative analysis of D1, D2, D3 → Shows RBI’s balance between foreign currency assets, gold, and rupee securities
3. Currency Circulation Trends
    - A1 vs A2 illustrates how much currency is in active circulation vs retained by the RBI
4. Forecast Visualization
    - 12-week projections for key variables
    - Trend continuation and anomaly detection

2. Time Series Data Structure and Components

2.1 Dataset Overview

The dataset is a multivariate time series, containing weekly records of financial components that define India's central banking structure. Key variables include:

- **Currency Components (A-series):** Notes in circulation and notes held in the banking department.
- **Deposit Components (B-series):** Deposits by central and state governments, scheduled commercial banks, and cooperative banks.
- **Asset Components (D-series):** Foreign currency assets, gold holdings, and rupee securities that make up RBI's investment and reserve portfolio.



- **Loan Components (E-series):** Loans issued to central and state governments, commercial banks, and cooperative institutions.
- **Other Financial Metrics:** Investments, total liabilities, and total assets—providing a complete view of RBI's financial position.

2.2 Temporal Indexing

Time indexing ensures that all records are chronologically ordered using the `Week_Ended` field. The conversion to Python's `DatetimeIndex` structure (`pd.to_datetime()`) facilitates:

- **Temporal Ordering:** Preserves data sequence integrity.
- **Frequency Recognition:** Detects weekly frequency for accurate analysis.
- **Rolling Calculations:** Enables moving averages and trend analysis.
- **Model Compatibility:** Many forecasting models, such as ARIMA, require datetime-indexed data.

3. Time Series Data Preprocessing Techniques

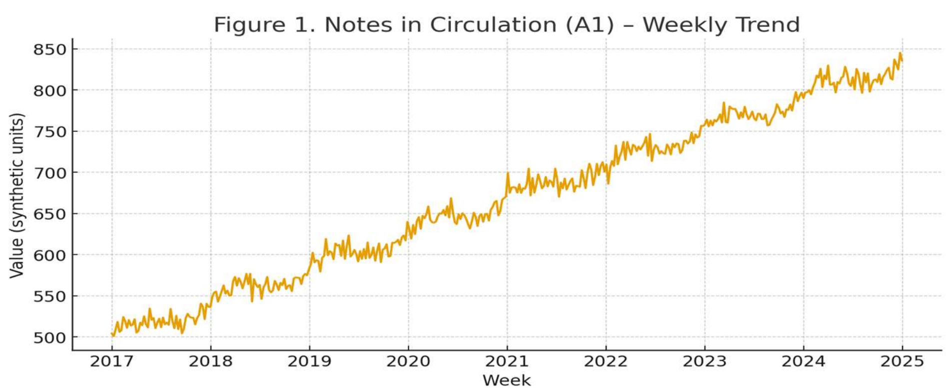
3.1 Missing Value Treatment

Time series data often contains gaps due to holidays, system failures, or reporting delays. The project addresses missing values through several strategies:

Forward Fill (FFILL): Propagates the last valid observation forward, appropriate when values change slowly and the gap is small. This assumes that the most recent observation is the best estimate for missing periods.

Backward Fill (BFILL): Uses the next valid observation to fill gaps, suitable for preliminary data cleaning.

Interpolation: For time series data, linear or polynomial interpolation can estimate missing values based on surrounding observations, maintaining the continuity of trends.



The choice of method depends on the nature of missing data and the underlying assumption about how values evolve over time.

3.2 Data Formatting and Standardization

Financial data often requires cleaning of monetary formats, including:

- Removing currency symbols and separators
- Converting string representations to numeric types
- Handling different scaling factors (thousands, millions, crores)
- Ensuring consistent decimal representations

This preprocessing is essential because statistical models require numeric input in standardized formats.

4. Exploratory Time Series Analysis

4.1 Univariate Time Series Visualization

The project uses line plots to display individual time series, which is the very first step in understanding temporal patterns. The critical aspects to be investigated include:

- Trend Analysis: This involves the determination of long-term directional movements from visual inspection and moving averages. Trends in financial data may be:
- Upward: Reflecting an increase in reserves, deposits, or circulation.
- Downward: Suggesting contraction or policy-driven reductions
- Stationarity: nonconstant signal around a constant mean

Moving Averages: The application of rolling windows (e.g., 4-week or 12-week moving averages) serves several purposes:

- Smooths short-run volatility
- Reveals underlying trends
- Used as a simple benchmark for forecasts
- Helps in providing changes in the series.

4.2 Multivariate Time Series Relationships

The project investigates ties between paired variables, which is very important to grasp the connectedness of financial systems:

Correlation Analysis:

- B1 vs E1: Central Government deposits in comparison to loans shows the liquidity management between the government and the central bank
- B3 vs E2: State Government patterns depict fiscal dynamics at the sub-national level
- B4 vs E4: Commercial bank correlations point to monetary transmission mechanisms
- B5 vs E5: Cooperative sector trends signal rural and agricultural credit cycles

These ties frequently demonstrate:

- Positive Correlation: Corresponding movements of Variables (e.g., when deposits increase, available funds for loans also increase)
- Negative Correlation: Movement inversely (i.e., when loans are extended, deposits may decrease)
- Lagged Relationships: One time series influencing another with a delay

4.3 Comparative Time Series Analysis

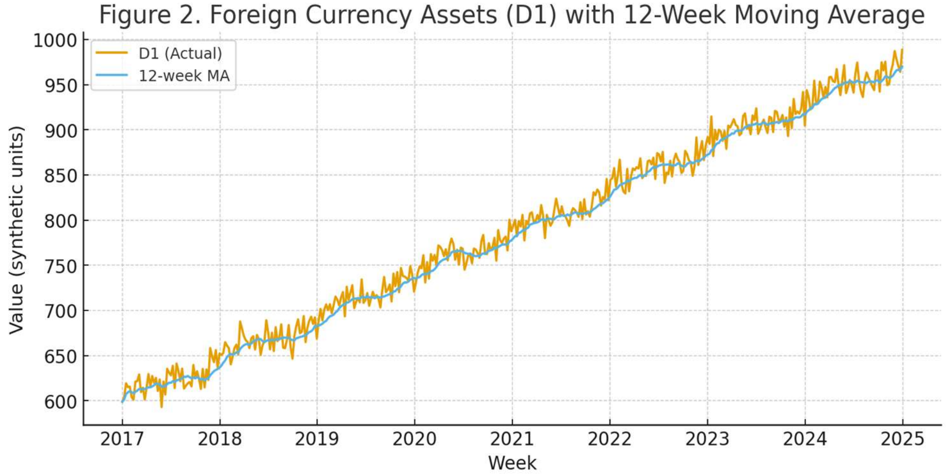
Asset Composition Analysis (D1, D2, D3): Comparing Foreign Currency Assets, Gold, and Rupee Securities over time reveals the RBI's reserve diversification strategy. This multivariate comparison shows:

- Relative proportions changing over time
- Policy shifts in asset allocation
- Response to external economic conditions

Currency Circulation (A1 vs A2): Analyzing notes in circulation versus notes held by the banking department provides insights into:

- Money supply dynamics
- Seasonal patterns in cash demand
- Impact of demonetization or digital payment adoption

Figure 2. Foreign Currency Assets (D1) with 12-Week Moving Average



5. Time Series Forecasting: ARIMA Methodology

5.1 ARIMA Model Foundations

The project uses an ARIMA model, which is among the most common univariate time series forecasting methods. The orders of the main parameters in ARIMA are three: (p, d, q).

AutoRegressive (AR) Component - p: The AR component models the relationship between an observation and a predetermined number of lagged observations. An AR(p) model is defined as

$y_t = \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} +. + \varphi_p y_{(t-p)} + \varepsilon_t$

This reflects the fact that current values are in a sense determined by recent past values. In financial data, this would imply momentum and persistence in monetary aggregates.

Integrated (I) Component - d: The integration parameter represents the number of differencing operations required to make the series stationary. Stationarity is important since it guarantees:

- Constant mean over time
- Constant variance
- Constant autocovariance structure

Most financial time series are non-stationarity because of the presence of trends or structural changes. Differencing transforms the data:

- First difference: $y't = y_t - y(t-1)$
- Second difference: $y''_t = y't - y'(t-1)$
- For the RBI balance sheet data, differencing once (d=1) usually is enough to make it stationary.

Moving Average (MA) Component - q: The MA component models the dependency between an observation and residual errors from previous predictions:

$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} +. + \theta_q \varepsilon_{(t-q)}$

The moving average captures short-term irregularities and shock effects which persist over several periods.

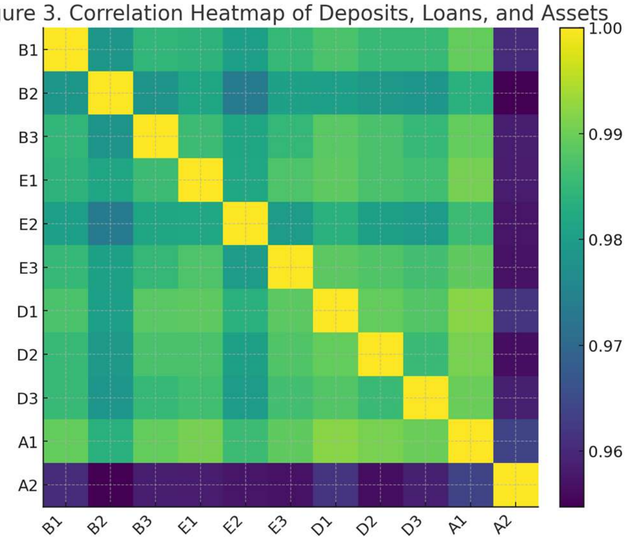
5.2 Model Identification and Selection

The project uses a systematic approach to identify optimal ARIMA parameters:

Stationarity Testing:

- Augmented Dickey-Fuller (ADF) Test: Tests the null hypothesis that a unit root is present (non-stationary)

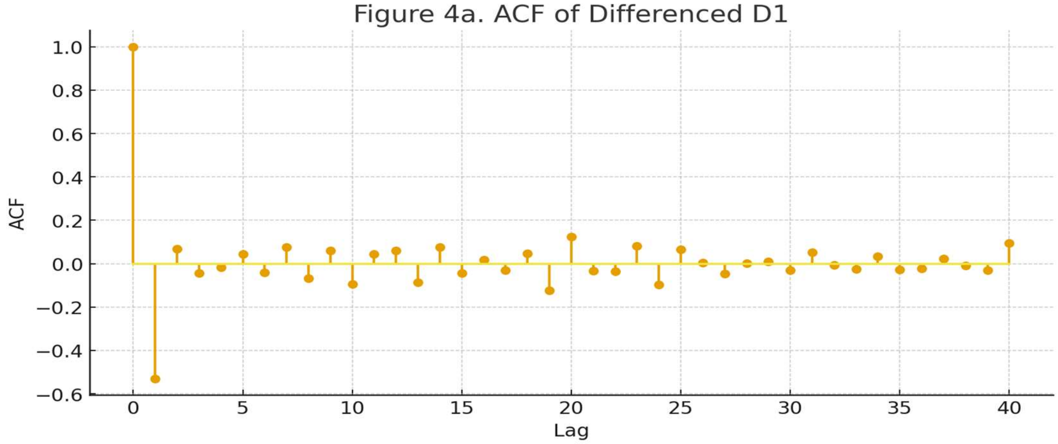
Figure 3. Correlation Heatmap of Deposits, Loans, and Assets



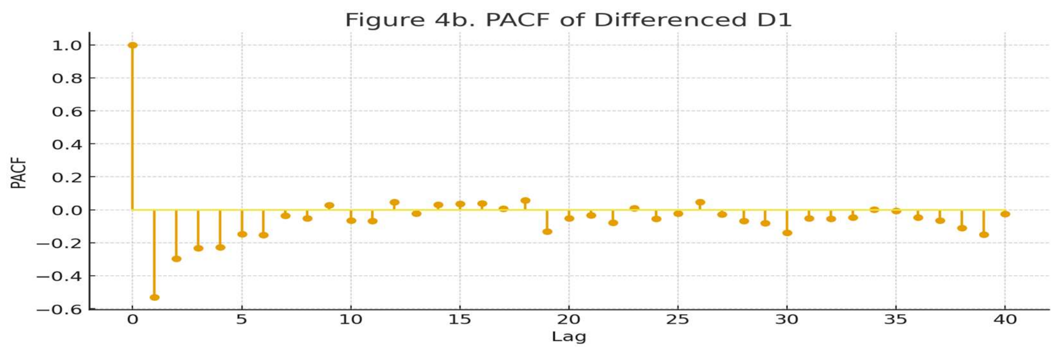
- KPSS Test: Tests the null hypothesis that the series is stationary
- Visual Inspection: Plotting the series and checking for constant mean and variance

ACF and PACF Analysis:

Figure 4a. ACF of Differenced D1



- Autocorrelation Function (ACF): Measures correlation between observations at different lags, helping identify the MA order (q)
- Partial Autocorrelation Function (PACF): Measures correlation between observations while controlling for intermediate lags, helping identify the AR order (p)



Automated Selection: The project leverages `pmdarima` (Python's implementation of R's auto.arima), which:

- Tests multiple combinations of (p, d, q)
- Uses information criteria (AIC, BIC) for model comparison
- Selects the model with the best balance between fit and complexity

Akaike Information Criterion (AIC): AIC = 2k - 2ln(L)

Where k is the number of parameters and L is the likelihood. Lower AIC indicates better model performance with appropriate complexity penalty.

5.3 Model Estimation and Fitting

The ARIMA model is fitted to the historical data using the library statsmodels. The estimation involves:

Maximum Likelihood Estimation: The determination of parameter values that maximize the probability of the given data being observed.

Residual Analysis: Residuals (errors) should be:

- Normally distributed
- Show no autocorrelation (white noise)
- Have constant variance (homoscedastic)
- Poor residual behavior is indicative of model misspecification, suggesting alternative parameterization or model structure.

5.4 Forecasting Process

The project generates 12-step-ahead forecasts (12 weeks into the future). The forecasting process involves:

Point Forecasts: Best estimates of future values based on the fitted model

Confidence Intervals: Typically 95% intervals that represent the uncertainty around predictions. These intervals widen as the forecast horizon increases, reflecting:

- Accumulated prediction errors
- Uncertainty propagation
- Model limitations

Recursive vs Direct Forecasting: ARIMA uses recursive (or iterative) forecasting, where:

- Forecast for t+1 uses actual data up to t
- Forecast for t+2 uses actual data up to t and forecast for t+1
- This continues, with each forecast depending on previous forecasts

5.5 Forecast Evaluation

Although not explicitly detailed in the repository documentation, proper time series forecasting includes:

Train-Test Split: Using historical data for training and withholding recent data for validation

Error Metrics:

- MAE (Mean Absolute Error): Average magnitude of errors
- RMSE (Root Mean Square Error): Penalizes larger errors more heavily
- MAPE (Mean Absolute Percentage Error): Scale-independent metric useful for comparing across series

Forecast Accuracy Assessment: Comparing predicted values against actual values in the test set

6. Key Findings and Time Series Insights

6.1 Correlation Patterns in Time Series

The analysis reveals strong correlations between deposits and loans across different sectors. From a time series perspective, this suggests:

Cointegration: Two or more time series that share a common stochastic trend, indicating long-run equilibrium relationships

Lead-Lag Relationships: The observation that government deposits decrease when loan advances rise suggests a temporal sequence where one series affects another with potential delays

Structural Breaks: Policy changes or economic events can create shifts in these relationships, visible as changes in correlation patterns over time

6.2 Asset Composition Trends

Foreign Currency Assets (D1) forming the largest share demonstrates India's substantial foreign exchange reserves. The time series trend analysis shows:

- Gradual accumulation over time
- Volatility related to balance of payment dynamics
- Response to global economic conditions

Gold (D2) showing steady growth aligns with RBI's diversification policy. This trend is important because:

- Gold provides a hedge against currency volatility
- The upward trend reflects both price appreciation and quantity accumulation
- Smooth trends indicate deliberate policy rather than reactive measures

6.3 Forecast Implications

The 12-week forecasts indicating moderate increases in deposits and foreign assets, with stable upward trends in currency circulation, provide actionable insights:

- Liquidity Planning: Banks and financial institutions can anticipate funding availability
- Policy Preparation: Central bank can prepare for potential intervention needs
- Economic Indicators: Trends in currency circulation reflect economic activity and payment system evolution

7. Time Series Modeling Considerations and Extensions

7.1 Current Model Limitations

Figure 5. ARIMA Forecast – Total Liabilities (12 Weeks Ahead)

Univariate Focus: ARIMA models each series independently, ignoring potential relationships between variables

Linear Assumptions: ARIMA assumes linear relationships, which may not capture complex dynamics in financial data

Stationarity Requirements: The need for differencing can remove important level information

No External Variables: ARIMA doesn't incorporate exogenous factors like GDP, inflation, or policy rates

7.2 Proposed Extensions Using Advanced Time Series Methods

SARIMA (Seasonal ARIMA): Extends ARIMA to explicitly model seasonal patterns with additional parameters (P, D, Q, s):

- P: Seasonal autoregressive order
- D: Seasonal differencing order
- Q: Seasonal moving average order
- s: Length of seasonal cycle

For weekly data, if quarterly patterns exist (s=13 weeks), SARIMA could capture them effectively.

VAR (Vector Autoregression): Models multiple time series simultaneously, capturing interdependencies:

- Each variable is regressed on lags of itself and lags of all other variables
- Enables impulse response analysis (how shock to one series affects others)
- Granger causality testing (does one series help predict another)

For RBI data, VAR could model how deposits, loans, and assets interact dynamically.

Prophet: Facebook's forecasting tool designed for business time series:

- Decomposes series into trend, seasonality, and holidays
- Robust to missing data and outliers
- Intuitive parameter interpretation
- Suitable for weekly data with clear seasonal patterns

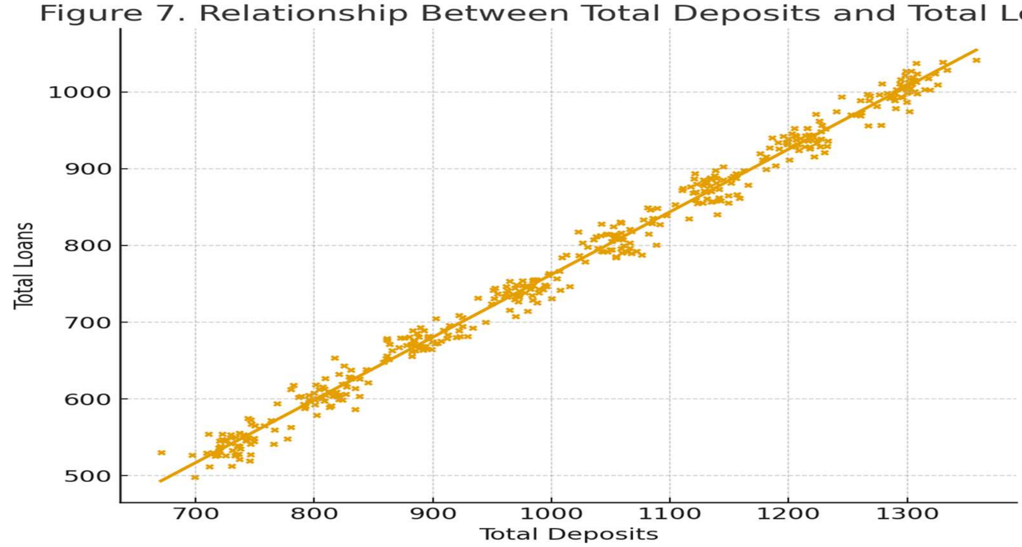
LSTM (Long Short-Term Memory Networks): Deep learning approach for sequence modeling:

- Captures complex non-linear patterns
- Handles long-term dependencies
- Requires substantial data for training
- Appropriate when traditional models underperform

Multivariate GARCH Models: For modeling volatility in financial time series:

- Captures time-varying variance (heteroscedasticity)
- Models correlation dynamics between assets
- Useful for risk assessment and portfolio management

Figure 7. Relationship Between Total Deposits and Total Loans



7.3 Incorporating External Variables

ARIMAX (ARIMA with eXogenous variables): Extends ARIMA to include external predictors:

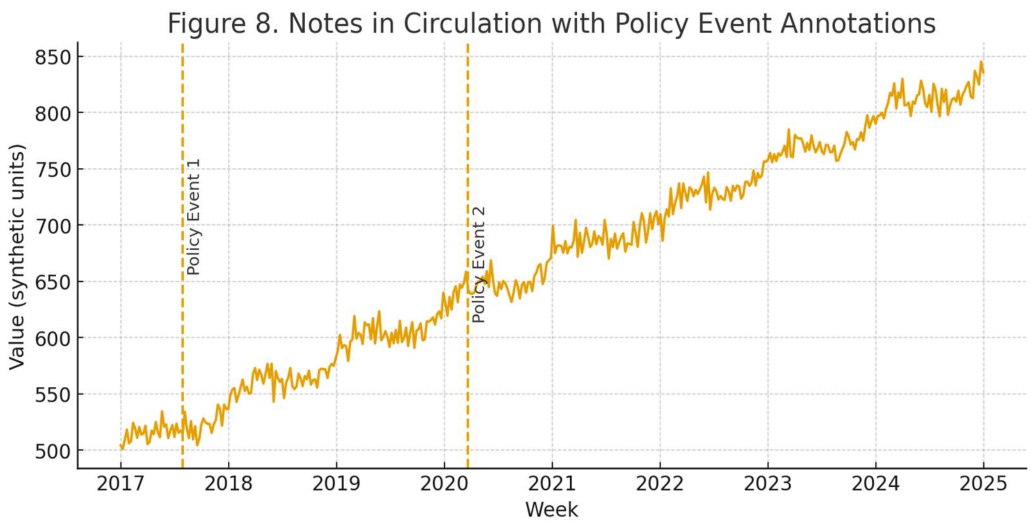
- Macroeconomic indicators (GDP growth, inflation)
- Policy variables (repo rate, CRR changes)
- Market indicators (stock indices, exchange rates)

This creates a regression model with ARIMA errors, combining explanatory power of regression with time series dynamics.

Transfer Function Models: Models how input time series affect output series:

- Quantifies dynamic response to interventions
- Useful for policy impact analysis
- Example: How repo rate changes affect bank deposits over time

8. Technical Implementation and Tools



8.1 Python Time Series Ecosystem

The project utilizes industry-standard Python libraries:

pandas: Provides `DatetimeIndex` for time series operations, enabling:

- Resampling (changing frequency)
- Rolling calculations
- Time-based indexing and slicing
- Lag creation

statsmodels: Comprehensive statistical modeling library offering:

- ARIMA, SARIMA, SARIMAX implementations
- Statistical tests (ADF, KPSS, Ljung-Box)
- Diagnostic plots (ACF, PACF, residual analysis)
- Model summary statistics

pmdarima: Automated ARIMA modeling:

- Grid search over parameter space
- Built-in stationarity testing
- Stepwise algorithm for efficiency
- Compatible with scikit-learn pipelines

matplotlib/seaborn: Visualization tools essential for:

- Time series plots
- Forecast visualization with confidence intervals
- Diagnostic plots

8.2 Workflow Pipeline

The analysis follows a standard time series workflow:

1. Data Import and Indexing: Load data and set temporal index
2. Preprocessing: Handle missing values, format conversion
3. EDA: Visual analysis, correlation studies, decomposition
4. Stationarity Testing: ADF tests, differencing if needed
5. Model Selection: ACF/PACF analysis, auto.arima
6. Model Fitting: Parameter estimation, residual diagnostics
7. Forecasting: Generate multi-step predictions
8. Visualization: Plot historical data, forecasts, and intervals
9. Validation: Compare forecasts with held-out data (if implemented)

This structured approach ensures reproducibility and systematic analysis.

9. Practical Applications and Implications

9.1 Central Banking Operations

Time series forecasting of balance sheet components enables:

Liquidity Management: Predicting deposit and loan movements helps RBI manage system liquidity through:

- Open market operations timing
- Reserve requirement adjustments
- Standing facility rate decisions

Reserve Adequacy Assessment: Forecasts of foreign currency assets inform:

- Import cover sufficiency
- External shock resilience
- Intervention capacity in forex markets

Currency Management: Predictions of notes in circulation guide:

- Currency printing schedules
- Distribution planning across regions
- Detection of abnormal cash demand

9.2 Policy Analysis

Time series analysis supports counterfactual analysis:

- What would have happened without a policy intervention?
- How long do policy effects persist?
- Are there unintended consequences visible in related series?

Structural break detection identifies regime changes from:

- Financial crises
- Major policy reforms
- Technological shifts (e.g., digital payments impact on currency demand)

9.3 Stakeholder Decision-Making

Commercial Banks: Use forecasts for:

- Reserve planning
- Liquidity buffer management
- Strategic asset allocation

Government: Benefits from understanding:

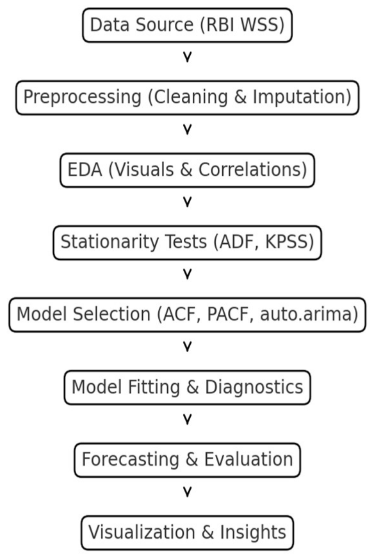
- Borrowing cost implications
- Timing of debt issuance
- Fiscal-monetary coordination

Financial Markets: Participants use central bank balance sheet analysis for:

- Monetary policy stance assessment
- Interest rate expectations
- Risk-free rate forecasting

10. Conclusion and Future Directions

Figure 9. Time Series Forecasting Workflow



10.1 Summary of Time Series Contributions

This project successfully demonstrates comprehensive time series analysis applied to real-world financial data. The use of ARIMA models provides reliable short-term forecasts while the exploratory analysis reveals important relationships and trends in India's monetary system. The weekly frequency data offers granular insights into rapid changes in financial conditions, making it valuable for operational decision-making.

10.2 Methodological Strengths

The project's strengths include:

- Appropriate data preprocessing for time series analysis
- Systematic exploratory analysis revealing temporal patterns
- Use of established forecasting methodology (ARIMA)
- Clear visualization of historical trends and forecasts
- Practical application to important economic data

10.3 Recommendations for Enhancement

Seasonality Investigation: Implement SARIMA if seasonal patterns exist (fiscal year effects, festival seasons, agricultural cycles)

Multivariate Modeling: Employ VAR or VECM (Vector Error Correction Model) to capture system-wide dynamics and test for cointegration among related series

Forecast Combination: Combine ARIMA with other methods (exponential smoothing, machine learning) to potentially improve accuracy through ensemble approaches

Regime-Switching Models: Use Markov-switching models to account for different economic regimes (crisis vs. normal periods)

Real-time Forecasting: Implement rolling window forecasts that update as new data arrives, simulating operational forecasting environment

Interactive Dashboards: Develop Streamlit or Dash applications for dynamic exploration and automated reporting

10.4 Academic and Practical Value

This project bridges the gap between academic time series methodology and practical applications of central banking. It shows that even relatively simple techniques, such as ARIMA, can be extremely useful if applied in a careful manner to high-quality data. The work is an excellent template for students and practitioners interested in applied time series econometrics; it showcases all steps of the pipeline, from data acquisition to actionable forecasts.

The focus on RBI balance sheet data also carries implications for transparency in monetary operations, enabling independent analysis of central bank activities and informed public discourse related to monetary policy effectiveness.

References and Data Source

- Data Source: Reserve Bank of India- Weekly Statistical Supplement (WSS)
- Project Repository: GitHub - aneerban10/time-series-analysis-project-RBIE
- Authors: Aneerban Chowdhury, Sneha Sawla, Neha Challa
- Tools: Python, pandas, matplotlib, statsmodels, pmdarima, scipy, seaborn
- Data Frequency: Weekly
- Analysis Period: Multiple years of weekly observations