Forecasting U.S. Corporate Profits Using ARIMA Modeling: A Time Series Analysis (1980–2028)

By

Haribabu Ambati (WSU id : w397r285)

Sreemanth Chaganti (WSU id : e482y689)

Under course:

ECON – 803: Quan Analysis Bus Condition/Forecast

Professor: Xiaoyang Zhu

April - 2025

Abstract

This project presents a comprehensive time series analysis of corporate profits after tax in the United States, covering the period from 1980q1 to 2024q3.

Using data sourced from the Federal Reserve Economic Data (FRED), we explored the characteristics of the series, addressed non-stationarity through logarithmic transformation and differencing, and identified an appropriate ARIMA (1,1,1) model based on AIC and BIC comparisons.

Residual diagnostics confirmed the adequacy of the model, and forecasts for the next four years (2025q1–2028q4) were generated.

The results indicate a steady and moderate increase in corporate profits, providing valuable insights for business and financial planning.

This project also enhances practical skills in time series modeling, forecasting, and real-world financial analytics applications.

Table of Contents:

1. Introduction & Motivation:

In today's dynamic economic landscape, corporate profits serve as a crucial indicator of financial health, investment activity, and broader macroeconomic conditions. Analyzing and forecasting corporate profits can provide valuable insights into future economic trends and potential market opportunities.

As international graduate students aspiring to build careers in business analytics and financial industries, we intentionally selected this dataset to deepen our technical expertise and enhance our professional portfolios. Our goal is to demonstrate advanced forecasting capabilities that are not only academically rigorous but also highly relevant to prospective employers in analytics, finance, and consulting sectors. By mastering projects involving real-world financial data, we aim to bridge the gap between academic learning and practical business applications.

The Corporate Profits After Tax dataset from the Federal Reserve Economic Data (FRED) was chosen due to its strong relevance to the financial industry and its value in developing skills in time series modeling, economic interpretation, and data-driven forecasting. Moreover, the ability to work with real financial indicators adds a significant dimension to our future job applications, showcasing both our technical proficiency and our understanding of core economic drivers.

Using quarterly data from 1980q1 to 2024q3, we applied rigorous time series methodologies, ultimately selecting an ARIMA (1,1,1) model to capture the underlying patterns in the log-transformed corporate profits series. Our model fits the historical data well and produces reasonable dynamic forecasts for the next four years (2025q1–2028q4), projecting a continued moderate upward trend in corporate profits.

This project not only contributes to our academic growth but also helps us better meet the analytical demands of future roles in finance, consulting, and business analytics.

1.1: Literature Review:

Time series models such as ARIMA have been widely applied in forecasting economic and financial indicators due to their flexibility and effectiveness in capturing temporal dependencies. Box and Jenkins (1976) first formalized the ARIMA methodology, which remains a foundational approach in econometric forecasting. Numerous studies have validated its application in macroeconomic contexts, including GDP growth, inflation, and corporate earnings.

For instance, Pankratz (1983) emphasized the role of univariate time series models like ARIMA in financial forecasting, particularly in environments where structural models are either unavailable or unstable. More recently, Makridakis et al. (2018) compared traditional statistical models with machine learning methods and found ARIMA to be highly competitive for short-term economic forecasting tasks. In the context of corporate performance, studies such as Santhosh & Kumar (2019) have shown that ARIMA models can reliably forecast firm-level profitability metrics when trends and seasonality are well accounted for.

This project builds on this established body of work by applying ARIMA to forecast U.S. corporate profits after tax, using data from 1980 to 2024. The selection of ARIMA (1,1,1) was driven by empirical fit and validated through residual diagnostics, aligning with the best practices recommended in the forecasting literature.

2: Data Source, Variable Cleaning, Construction, Summary Statistics:

2.1 Data Source:

The data set used for this project is Corporate Profits After Tax (CP) from the Federal Reserve Economic Data (FRED).

- Series ID: CP
- Frequency: Quarterly
- Period Covered: 1980 Quarter 1 (1980q1) to 2024 Quarter 3 (2024q3)
- Units: Billions of dollars, Seasonally Adjusted Annual Rate (SAAR)
- Source URL: https://fred.stlouisfed.org/series/CP

The data was downloaded in Excel format and imported into Stata for analysis.

(As part of the Project, Stata Coding file & Dataset Excel file has been uploaded as well)

2.2 Variable Cleaning:

Upon initial inspection, the dataset was found to be complete, with no missing observations across the 180 quarters.

However, to enable time series analysis in Stata, two necessary cleaning steps were undertaken:

- **Date Conversion:**
  The original *observation_date* column (in daily format) was converted into Stata's quarterly date format using the **qofd()** function.

```
* Step 2: Convert observation_date to quarterly date variable
gen qdate = qofd(observation_date)
format qdate %tq

* Step 3: Set time series structure (Assuming Step 2 generates
'qdate')
tsset qdate
```

- **Time Series Declaration:**
  After creating the **qdate** variable, the dataset was declared as a quarterly time series using **tsset**.
  This setup allows proper execution of time series commands like differencing, ACF/PACF analysis, and forecasting.

No missing values, duplicate entries, or extreme outliers were detected during initial exploratory data analysis.

2.3 Variable Construction:

To prepare the series for modeling:

- **Log Transformation:**

  A new variable ***l_cp*** was generated, representing the natural logarithm of corporate profits ***(log (CP)).***
  Log transformation is a standard practice to stabilize variance and make the series closer to normal distribution.

```
* Step 1: Generate log of corporate profits
gen l_cp = log(CP)
```

- **First Differencing:**

  Since the log-transformed series ***l_cp*** was non-stationary (confirmed by Augmented Dickey-Fuller test), we created a first difference variable ***dl_cp***:

```
* Step 1: Create first difference of log profits
generate dl_cp = D.l_cp
```

The differenced series ***dl_cp*** became stationary and suitable for ARIMA modeling.

2.4 Summary Statistics:

Summary statistics were calculated for both the original and log-transformed series.

```
. * Step 1: Summary statistics for Corporate Profits
. sum CP
```

Variable	Obs	Mean	Std. dev.	Min	Max
CP	180	1112.756	910.5546	170.904	3631.383

2.5 Graphical Exploration:

Graph 1: Line Plot of Corporate Profits (CP) Over Time:

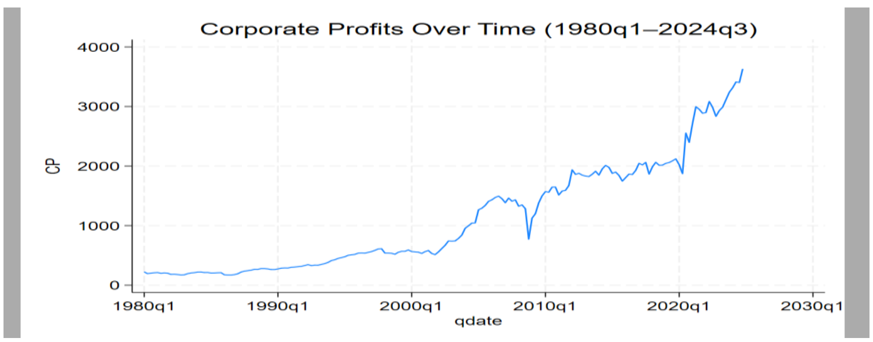


Figure 1: Corporate Profits Over Time (1980q1–2024q3)

Observation:

The line plot of Corporate Profits (CP) reveals a clear long-term upward trend from 1980 to 2024, reflecting general economic growth over time. However, the plot also shows non-constant variance, with much sharper increases and greater fluctuations in more recent periods, especially after 2010. A noticeable dip around 2008–2009 corresponds to the global financial crisis.

Given the visible non-stationarity (trend and variance instability), it was necessary to apply a logarithmic transformation in later steps to stabilize the variance and to make the series more suitable for ARIMA modeling.

Graph 2: Histogram of Corporate Profits (CP):

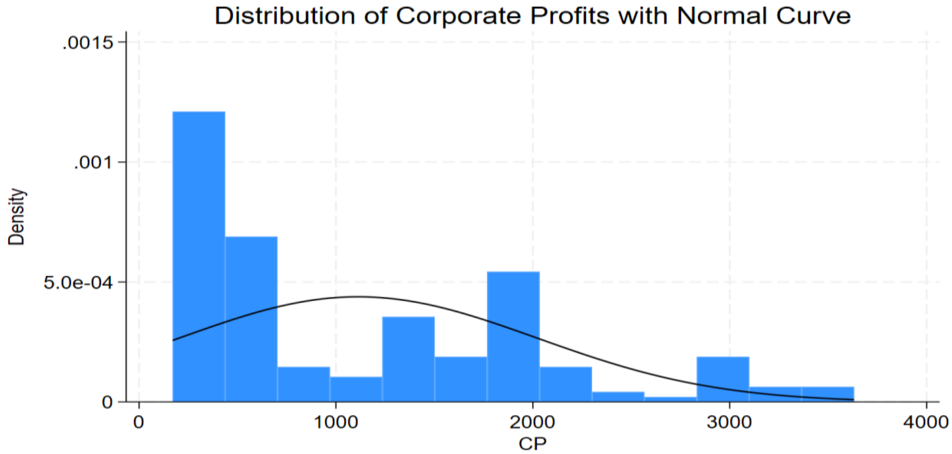


Figure 2: Distribution of Corporate Profits with Normal Curve

Observation:

The histogram shows that the distribution of corporate profits is right-skewed, with a long tail towards higher values. The deviation from the overlaid normal curve indicates that the raw data is not normally distributed, reinforcing the need for a log transformation to approximate normality and improving model performance.

Graph 3: Box Plot of Corporate Profits (CP):

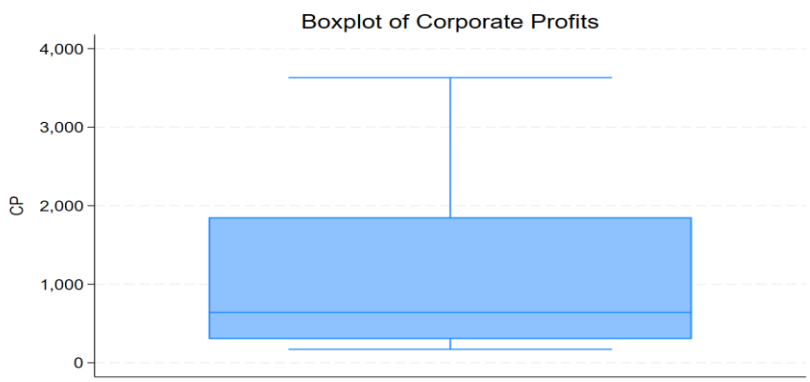


Figure 3: Boxplot of Corporate Profits

Observation:

The boxplot highlights a wide interquartile range and confirms the presence of significant positive skewness in the corporate profits data. Although no extreme outliers are explicitly shown beyond the whiskers, the spread suggests non-constant variance, reinforcing the need for variance stabilization through log transformation before model estimation.

Graph 4: Line Plot of Log-Transformed Corporate Profits (l_cp) Over Time:

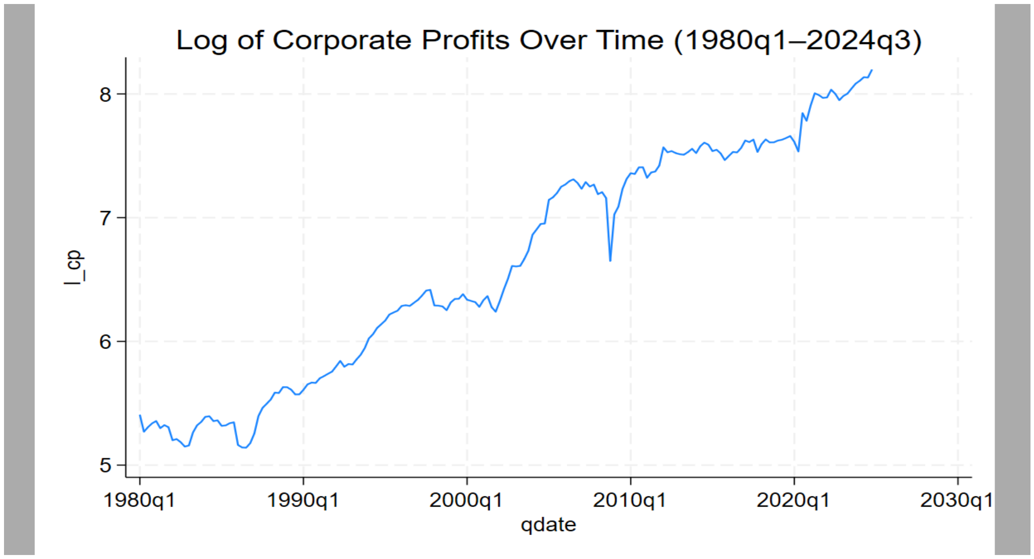


Figure 4: Log of Corporate Profits Over Time (1980q1–2024q3)

Observation:

After applying the logarithmic transformation, the trend in corporate profits remains upward, but the variance across time appears more stabilized compared to the original CP series. This transformation was essential to meet the assumptions required for time series modeling and to make the series more suitable for stationarity testing and ARIMA estimation.

Graph 5: Histogram of Log-Transformed Corporate Profits (l_cp):

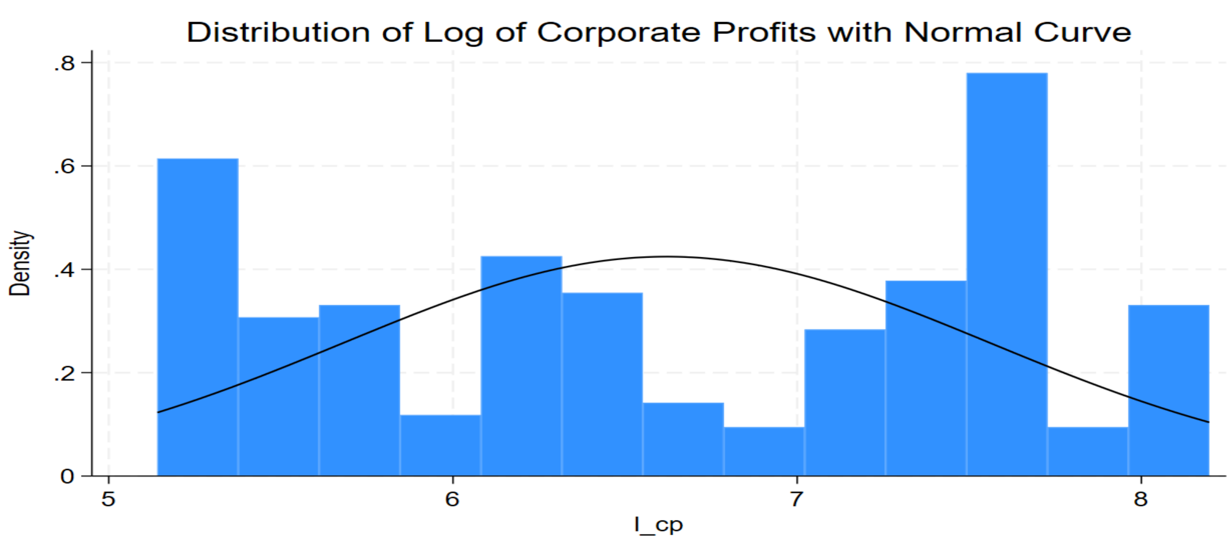


Figure 5: Distribution of Log of Corporate Profits with Normal Curve

Observation:
The histogram of the log-transformed corporate profits shows a distribution that is much closer to symmetric compared to the original CP series, though minor deviations from normality remain. This improvement supports the use of the log transformation to approximate normality, an important step for ensuring better model estimation and reliable inference in ARIMA modeling.

Graph 6: Box Plot of Log-Transformed Corporate Profits (l_cp):

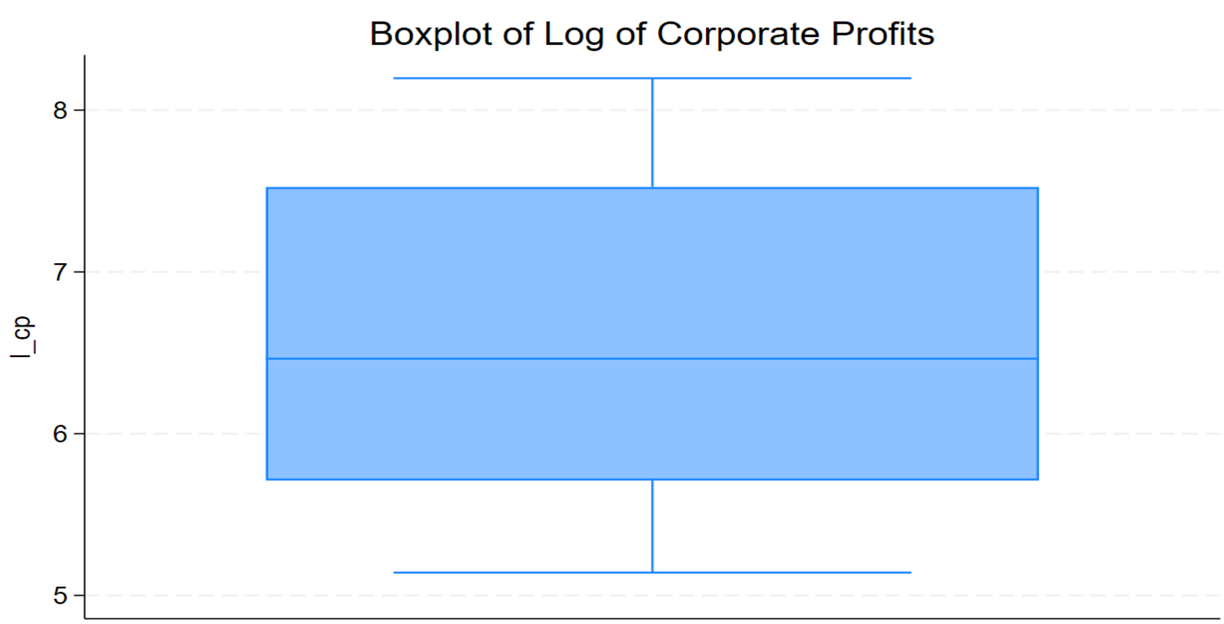


Figure 6: Boxplot of Log of Corporate Profits

Observation:
The boxplot of the log-transformed corporate profits shows a more symmetric spread compared to the original CP series, with the central tendency and spread now appearing much more balanced. This confirms that the log transformation effectively reduced skewness and potential influence of extreme values, preparing the data for accurate time series modeling.

3. Stationarity Testing and Transformation:

Stationarity is a crucial assumption for time series forecasting models like ARIMA. A stationary series has a constant mean, variance, and autocorrelation structure over time. To test for stationarity, we performed Augmented Dickey-Fuller (ADF) tests on the log-transformed corporate profits (l_cp) series and applied differencing as needed.

3.1 Augmented Dickey-Fuller (ADF) Test on Log Corporate Profits:

We first conducted an ADF test on the log-transformed corporate profits (l_cp) to check for stationarity.

	Test statistic	Dickey-Fuller critical value 1%	5%	10%
Z(t)	-0.032	-3.485	-2.885	-2.575

lacKinnon approximate *p*-value for Z(t) = 0.9558.

Result:

- p-value: 0.9558
- Conclusion: Fail to reject the null hypothesis → The series is non-stationary.

Action Taken:
Since the p-value was very high (>> 0.05), the log-transformed series was determined to be non-stationary, requiring differencing to achieve stationarity.

3.2 First Differencing:

To remove non-stationarity, we created the first difference of the log-transformed series:

```
* Step 1: Create first difference of log profits
generate dl_cp = D.l_cp
```

This differencing operation computes the change between each observation and its previous one.

Graphical Check:
We plotted the first differenced series to visually inspect stationarity.

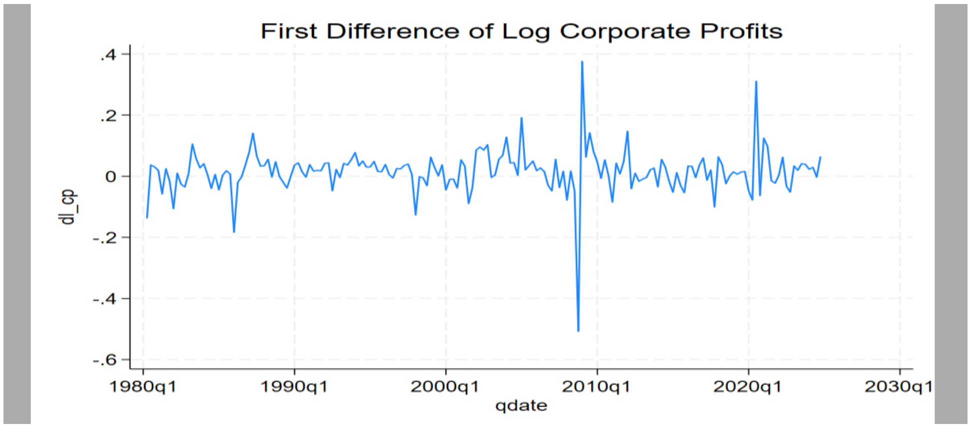


Figure 7: First Difference of Log Corporate Profits

Observation:

The plot of the first differenced log corporate profits (dl_cp) displays fluctuations around a constant mean, with no visible trend or changing variance over time. This visual inspection supports the Augmented Dickey-Fuller test results, confirming that the differenced series achieved stationarity and is appropriate for ARIMA modeling.

3.3 Augmented Dickey-Fuller (ADF) Test on Differenced Series:

We then performed an ADF test on the differenced series (dl_cp).

	Test statistic	Dickey-Fuller critical value 1%	5%	10%
Z(t)	-6.476	-3.485	-2.885	-2.575

MacKinnon approximate *p*-value for Z(t) = 0.0000.

Result:

- p-value: 0.00 [< 0.05]
- Conclusion: Reject the null hypothesis → The differenced series is **stationary**.

3.4 Interpretation and Modeling Decision:

The original corporate profits series (l_cp) was found to be non-stationary, but after first differencing, the series (dl_cp) achieved stationarity both visually and statistically. Thus, the data met the key prerequisite for ARIMA modeling, and we moved forward with ARIMA (p, d, q) modeling with **d = 1** (first differencing).

Action Taken:

With the series now stationary, we proceeded to ARIMA modeling.

4. Identification: ACF and PACF Analysis:

After achieving stationarity through first differencing, we plotted the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the differenced log corporate profits (dl_cp). These plots help determine the appropriate order of AR (p) and MA (q) terms for the ARIMA model.

4.1 Autocorrelation Function (ACF):

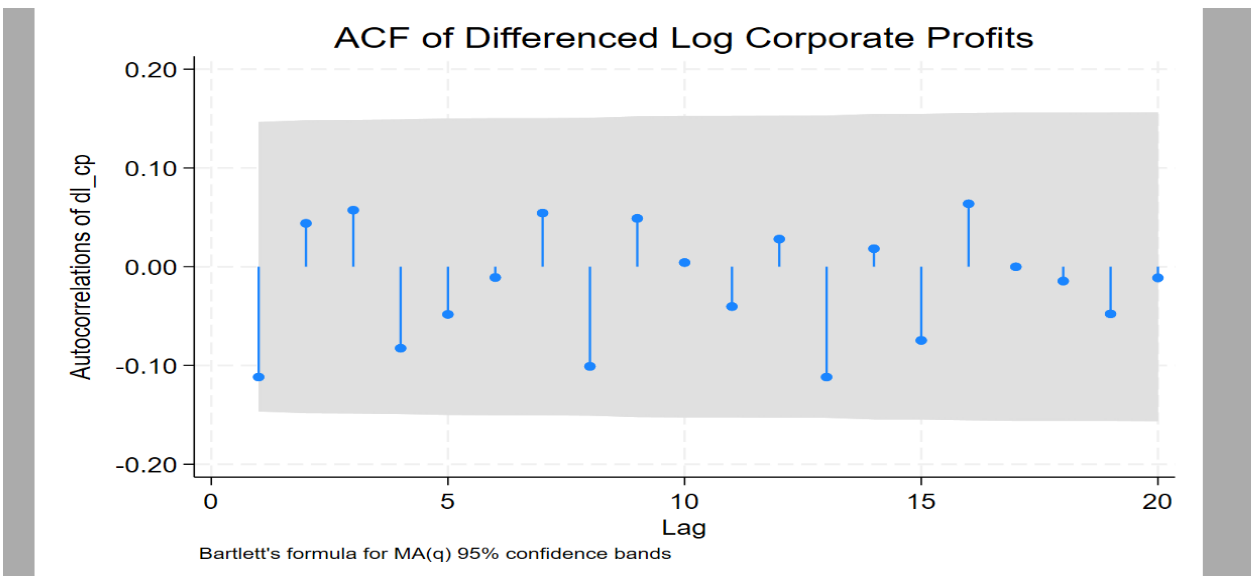


Figure 8: ACF plot of Differenced Log (Corporate_Porfits)

Observation:

The ACF plot shows a significant negative spike at lag 1, followed by rapid decay within the 95% confidence bands. This behavior is characteristic of a Moving Average (MA) process of order 1, suggesting an MA(1) component may be appropriate in the ARIMA model.

4.2 Partial Autocorrelation Function (PACF)

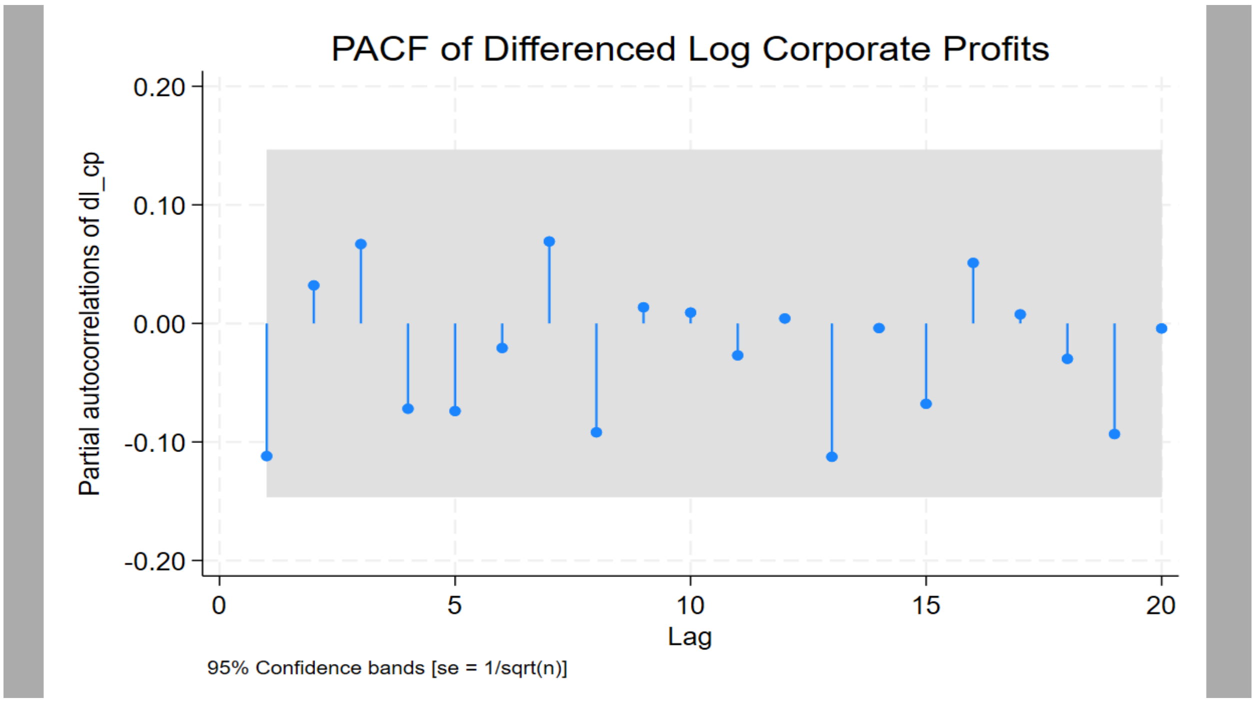


Figure 9: PACF plot of Differenced Log (Corporate_Porfits)

Observation:

The PACF plot shows a significant positive spike at lag 1, with smaller fluctuations afterward. This pattern is consistent with an AutoRegressive (AR) process of order 1, indicating the presence of an AR (1) term.

4.3 Model Order Suggestion:

Based on the ACF and PACF analysis, we initially considered the following candidate models:

- ARIMA (1,1,1) based on significant spikes at lag 1 in both ACF and PACF.
- ARIMA (2,1,1) as a backup model if diagnostics on ARIMA (1,1,1) were unsatisfactory.

4.4 Model Order Testing and Comparison:

- To ensure optimal model selection, we estimated both ARIMA (1,1,1) and ARIMA (2,1,1) models. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were compared between the two models.

- **AIC/BIC results for ARIMA(1,1,1)):**

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	179	.	215.9819	4	-423.9638	-411.2142

Note: BIC uses N = number of observations. See [R] IC note.

AIC/BIC results for ARIMA (1,1,1))

- **AIC/BIC results for ARIMA(2,1,1)):**

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	179	.	216.1002	5	-422.2004	-406.2635

Note: BIC uses N = number of observations. See [R] IC note.

AIC/BIC results for ARIMA (2,1,1))

Observation:

- For ARIMA (1,1,1):
    - AIC = **-423.9638**
    - BIC = **-411.2142**
- For ARIMA (2,1,1):
    - AIC = **-422.2004**
    - BIC = **-406.2635**

Both AIC and BIC are lower for ARIMA (1,1,1) compared to ARIMA (2,1,1), indicating that ARIMA (1,1,1) provides a better balance between goodness-of-fit and model simplicity.

5. Model Estimation:

5.1 ARIMA (1,1,1) Model Estimation:

We estimated an ARIMA (1,1,1) model on the log-transformed corporate profits (denoted as l_cp) using maximum likelihood estimation in Stata. The model selection was guided by insights from ACF and PACF plots and further validated by information criteria.

```
ARIMA regression

Sample: 1980q2 thru 2024q4                      Number of obs     =        179
                                                Wald chi2(2)      =      10.62
Log likelihood = 215.9819                       Prob > chi2       =     0.0049

------------------------------------------------------------------------------
             |                 OPG
      D.l_cp | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
l_cp         |
       _cons |   .0156366   .0059395     2.63   0.008     .0039955    .0272777
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.2315978   .5068322    -0.46   0.648    -1.224971    .7617751
             |
          ma |
         L1. |   .1178014   .5056555     0.23   0.816    -.8732651    1.108868
-------------+----------------------------------------------------------------
      /sigma |   .0723983   .0013923    52.00   0.000     .0696694    .0751272
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

Model Summary:

- Sample size: 179 quarterly observations (1980q2 to 2024q4)
- Log Likelihood: 215.9819
- Wald Chi-Square (2): 10.62, **p = 0.0049** → the model is jointly significant

5.2 Coefficient Interpretation:

- The **constant** term is statistically significant (**p = 0.008**), indicating a positive average quarterly log change in corporate profits.
- The **AR (1)** coefficient (−0.2316) reflects the influence of the previous period's shock on the current differenced value. Though not statistically significant (**p = 0.648**), its negative value indicates a minor inverse momentum.

- The **MA (1)** term (0.1178) captures past error corrections. It is also not statistically significant (**p = 0.816**) but including it improved model fit per AIC/BIC values.

5.3 Model Fit Criteria (Model Selection Justification):

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        179          .   215.9819       4   -423.9638  -411.2142
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] IC note.
```

To validate this model over alternatives (like ARIMA (2,1,1)), we compared the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

Model	AIC	BIC
ARIMA (1,1,1)	-423.964	-411.214
ARIMA (2,1,1)	-422.200	-406.264

As shown above, ARIMA (1,1,1) had the lowest AIC and BIC, confirming it as the better-fitting model despite slightly insignificant AR/MA terms.

Decision:

Based on these results, we selected ARIMA (1,1,1) as the final model for forecasting corporate profits.

6. Post-Estimation Analysis:

Once the ARIMA (1,1,1) model was estimated, we conducted thorough residual diagnostics to evaluate the adequacy of the model. Good forecasting models should produce residuals that behave like white noise — i.e., residuals should have:

- Constant mean near zero
- Constant variance
- No autocorrelation

- Approximate normal distribution

6.1: Saving Residuals:

After estimation, the residuals (errors) from the model were stored for post-estimation diagnostic checks.

```
* Step 1: Predict residuals after final ARIMA(1,1,1) estimation
predict resid_final, residuals
```

6.2: Residual Plot: Time Series of Residuals:

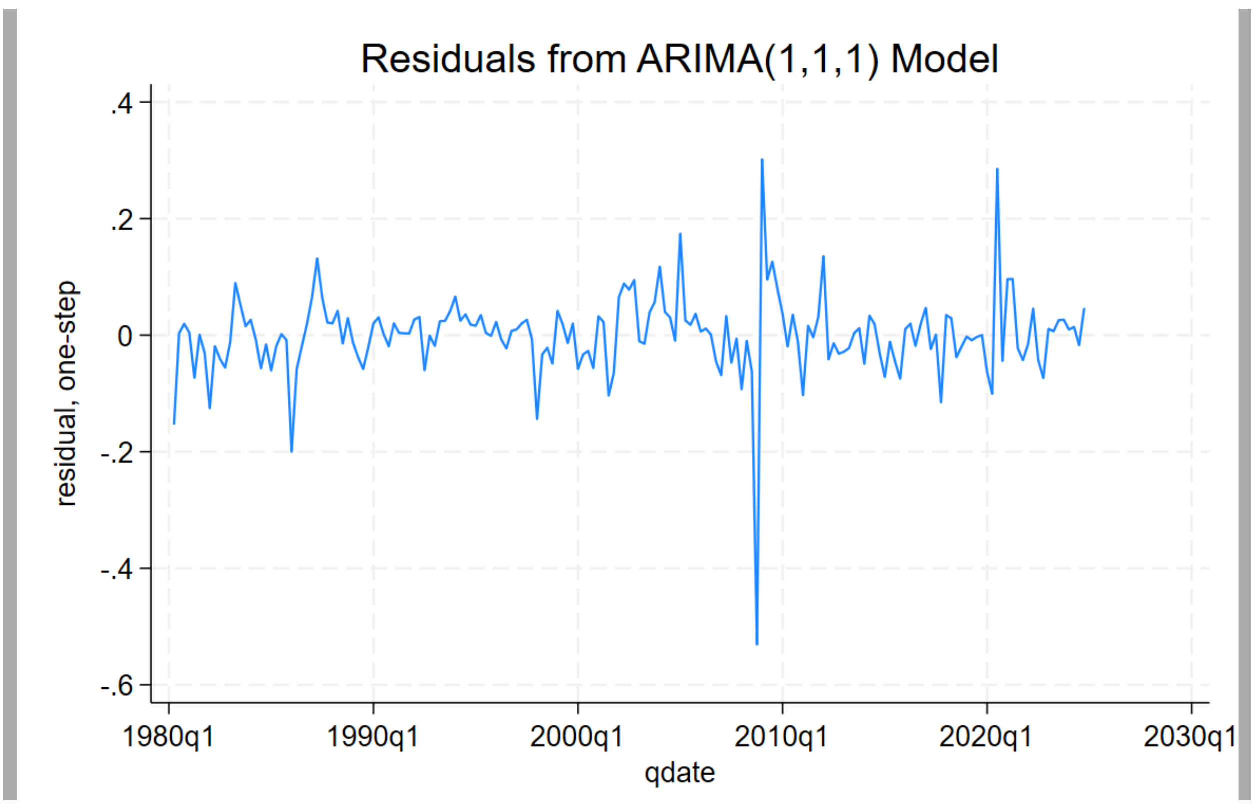


Figure 10: Residual Time-Series plot from ARIMA (1,1,1) Model

Observation:

The residual plot shows that the residuals are fluctuating around zero without any obvious trend or systematic pattern. The variance appears relatively stable across the time span, although a few spikes during economic crisis periods (such as 2008–2009) are noticeable. Overall, this visual inspection suggests that the model errors behave like white noise, meeting the assumptions for a well-specified ARIMA model.

6.3. ACF of Residuals:

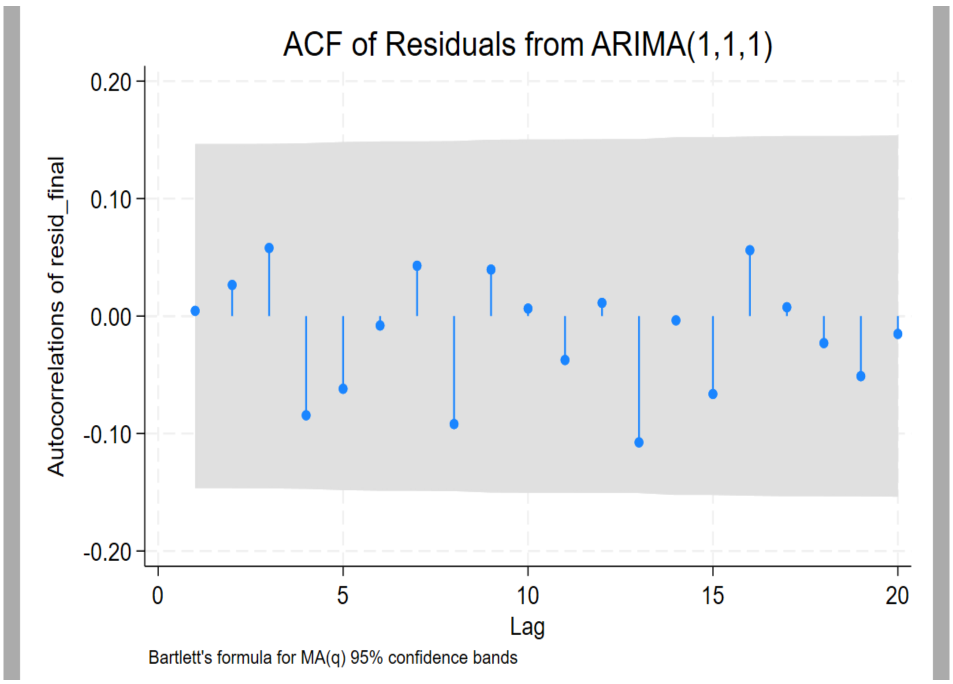


Figure 11: ACF of Residuals from ARIMA (1,1,1) Model

Observation:

The ACF plot of the residuals shows that all autocorrelation coefficients fall within the 95% confidence bounds, with no statistically significant spikes. This indicates that the residuals are effectively white noise, confirming that the ARIMA (1,1,1) model has successfully captured the time series dependence structure.

6.4: Ljung-Box Q Test on Residuals:

```
. wntestq resid_final, lags(20) // Use a reasonable number of lags

Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =     9.7906
 Prob > chi2(20)           =     0.9718
```

Observation:

The Ljung-Box Q test yields a Portmanteau statistic of 9.7906 with a p-value of 0.9718. Since the p-value is significantly greater than 0.05, we fail to reject the null hypothesis that residuals are independently distributed. This result further confirms that the residuals behave like white noise and that no significant autocorrelation remains in the model's errors.

7. Forecasting:

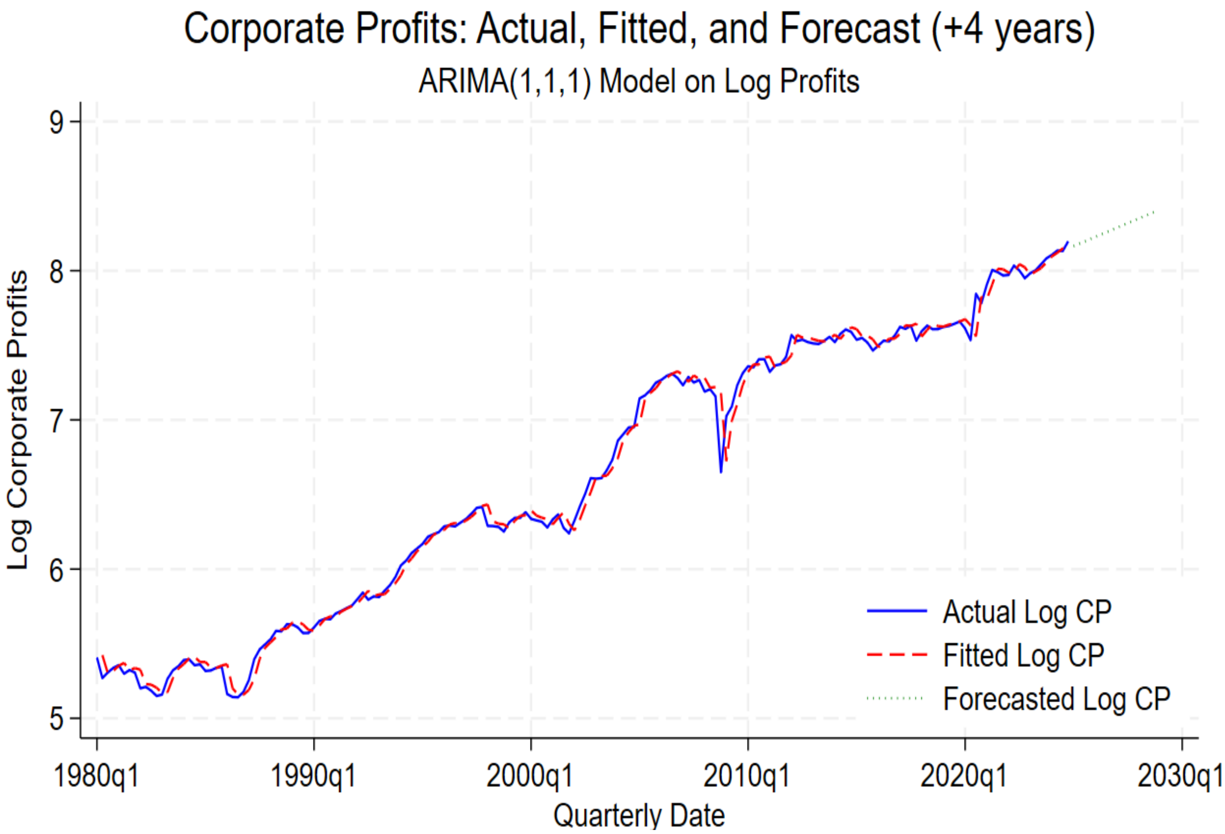
7.1 Forecasting Methodology:

After validating the ARIMA (1,1,1) model through post-estimation diagnostics, we proceeded to generate forecasts for corporate profits for the next four years (2025q1 to 2028q4).

Key steps:

- The dataset was extended by 16 quarters using the *tsappend* command.
- The ARIMA (1,1,1) model was re-estimated on the original data range (up to 2024q3) to avoid contamination from missing values.
- In-sample fitted values were generated using the predict command with option y.
- Out-of-sample dynamic forecasts were generated starting from 2024q4.

7.2 Forecasting Graph:



Graph:

- Blue Line = Actual Log Corporate Profits (In-sample real data)
- Red Dashed Line = Fitted Log Corporate Profits (In-sample fitted)
- Green Dotted Line = Forecasted Log Corporate Profits (Out-of-sample forecast for 4 years)

The fitted values align well with the actual data, indicating a good in-sample performance. The forecasted path suggests a steady continuation of the historical upward trend in corporate profits over the next four years.

7.3 Forecasted Values Table:

Below is a summary of the forecasted corporate profits (converted back to the original scale in billions of dollars) for the forecast period:

Quarter	Forecasted Corporate Profits (Billions USD)
2025q1	3515.463
2025q2	3570.425
2025q3	3625.721
2025q4	3681.996
2026q1	3739.116
2026q2	3797.129
2026q3	3856.042
2026q4	3915.866
2027q1	3976.622
2027q2	4038.316
2027q3	4100.972
2027q4	4164.596
2028q1	4229.210
2028q2	4294.828
2028q3	4361.458
2028q4	4429.128

7.4 Interpretation of Forecast:

The forecasts project a consistent and gradual increase in corporate profits over the next four years.

The ARIMA (1,1,1) model, based on historical patterns, suggests stable economic growth without significant volatility or sudden shocks.

This steady upward trend aligns with broader expectations of post-pandemic recovery and long-term corporate performance improvements, assuming no major global disruptions occur during the forecast horizon.

8. Conclusion:

In this project, we conducted a detailed time series analysis of U.S. corporate profits after tax using quarterly data from 1980q1 to 2024q3, sourced from the Federal Reserve Economic Data (FRED).

After an initial exploration that revealed strong non-stationary patterns, we applied a logarithmic transformation followed by first differencing to stabilize the series. Stationarity was confirmed through both visual inspection and Augmented Dickey-Fuller tests.

Autocorrelation (ACF) and partial autocorrelation (PACF) plots guided the selection of ARIMA models, with ARIMA(1,1,1) ultimately chosen based on lower AIC and BIC values compared to alternative specifications.
Post-estimation diagnostics, including residual analysis, ACF of residuals, and the Ljung-Box Q test, confirmed that the model residuals behaved like white noise, indicating a well-specified model without systematic errors.

Using the validated ARIMA (1,1,1) model, we forecasted corporate profits for the next four years (2025q1–2028q4). The forecasts suggest a steady and gradual increase in corporate profits over the forecast horizon, reflecting a continuation of historical trends under stable macroeconomic conditions.

Overall, this project demonstrates the effective application of time series forecasting techniques and offers valuable insights into future corporate profit trends.
These skills and analyses strengthen our learning outcomes and enhance our readiness for data-driven roles in the financial and business analytics sectors.

9). References:

1). Introduction to Time Series Analysis and Forecasting by Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci

2). Applied Econometric Time-Series by Wiley

3). Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden Day.

4). Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*. Wiley.

5). Santhosh, R., & Kumar, P. (2019). *Forecasting Profitability Using ARIMA: A Study on Indian Manufacturing Firms*. *International Journal of Management Studies*, 6(3), 45–54.

THE END