Forecasting Portugal's GDP: A Comparative Analysis of ARIMA and ETS Models Post-Pandemic Impact

[redacted]

Paper type: Written product (IC)

Date of submission: 13.08.2024

Student number: [redacted] Tomás Gonçalves

Character count (spaces included): 24544

Page count: 10

Abstract

This study investigates the impact of the COVID-19 pandemic on Portugal's GDP by utilizing two prominent time series models: ARIMA and ETS. Leveraging data from the Federal Reserve Economic Data (FRED) database covering the period from 1995 to 2023, the analysis compares actual economic outcomes with hypothetical scenarios that assume the pandemic had not occurred. The research implements a rigorous methodological framework, including data transformations, structural breaks testing, and residual analysis. The findings reveal significant differences between actual GDP figures and the hypothetical non-pandemic scenarios. This study focuses on determining which model more accurately reflects the economic disruptions caused by the pandemic, offering valuable insights for economic strategists and policymakers in enhancing forecasting techniques and formulating resilient economic policies for future crises.

Keywords: *Time Series Forecasting, ARIMA Model, ETS Model, Portugal GDP, Economic Impact of COVID-19, Box-Cox Transformation, Residual Analysis*

Table of Contents

1. Introduction

The COVID-19 pandemic has had profound effects on global economies, disrupting industries and altering economic growth trajectories. Among the many affected nations, Portugal's economy, which relies heavily on sectors such as tourism, experienced significant volatility. Understanding the pandemic's impact on Portugal's GDP and forecasting future economic trends has become crucial for policymakers and economists in planning for recovery and future resilience.

Gross Domestic Product (GDP) is a key economic indicator that measures the total value of goods and services produced within a country over a specific period, usually a quarter or a year. It serves as a broad measure of overall domestic production and functions as a comprehensive scorecard of a country's economic health.

This study aims to address a critical research question: *Which time series model, ARIMA or ETS, offers a more reliable framework for forecasting Portugal's GDP in the post-pandemic context?* The primary objective is to compare the performance of these two models, not only in terms of accuracy but also in understanding the conditions under which each model performs best. This research hypothesizes that the structural changes induced by the pandemic may influence the effectiveness of each model, depending on the specific characteristics of the time series data, such as trends, seasonality, and the presence of structural breaks.

By analyzing historical data from 1995 to 2023, this study seeks to identify which model provides more accurate and reliable forecasts and to understand the underlying reasons for any differences in model performance. The insights gained from this analysis will be valuable for economic strategists and policymakers as they seek to enhance forecasting techniques and formulate resilient economic policies for future crises.

The remainder of this paper is structured as follows: Section 2 describes the dataset used, including its source and characteristics. Section 3 outlines the data exploration and preparation steps, including tests for stationarity and structural breaks. Section 4 presents the forecasting models, their selection criteria, and the results of their application to the dataset. Finally, Section 5 concludes with key findings, implications for economic policy, and suggestions for future research.

2. Dataset Description

This study employs the Portugal Real Gross Domestic Product (GDP) dataset, sourced from the Federal Reserve Economic Data (FRED) database, covering the period from the first quarter of 1995 to the first quarter of 2023. The dataset includes 113 quarterly observations of Real GDP, measured in chained 2010 Euros.

Real GDP is an essential economic metric that adjusts for inflation, allowing for a more accurate reflection of the economy's true growth by accounting for changes in the price level over time. Unlike nominal GDP, which measures economic output based on current prices, Real GDP provides a clearer picture of an economy's performance by isolating the effects of inflation. This makes it particularly useful for long-term economic analysis and for comparing economic performance across different time periods.

The FRED database is a widely respected source for economic data, offering reliable and accurate datasets used in research and forecasting globally. The Portugal GDP series from FRED is part of the OECD's national accounts data, further ensuring the data's credibility and relevance for economic analysis.

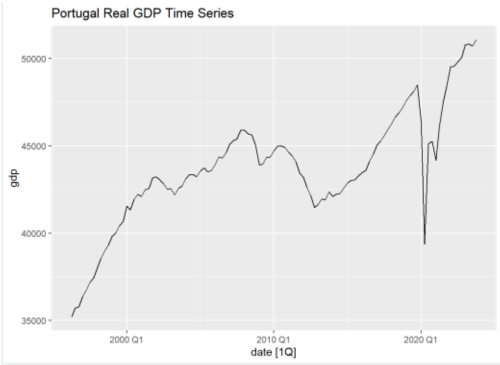


Figure 1: Portugal Real GDP Time Series

3. Data Exploration and Preparation

3.1 Temporal patterns and mathematical transformation

To analyze the Portugal Real GDP data from 1995 to 2023, a comprehensive examination of the data's temporal patterns was conducted, including trends, seasonality, cycles, and potential structural breaks.

This involved visualizing the data through various plots, such as time series charts, autocorrelation functions (ACF), partial autocorrelation functions (PACF), seasonal plots, and decompositions (Appendix A). These visualizations aid in understanding the underlying structures and preparing the data for further analysis.

Initial inspection of the raw data through the time series plot (Figure 1) revealed a clear upward trend with interruptions during the COVID-19 pandemic. The ACF (Appendix A, Figure 1) showed significant positive autocorrelations at early lags, indicating a strong trend component. The PACF (Appendix A, Figure 2) provided further confirmation, with significant values at the initial lag, reflecting the persistence of the trend.

To ensure the data's suitability for advanced time series modeling, mathematical transformations were applied. Specifically, the Box-Cox transformation was utilized to stabilize the variance, a critical step to meet the assumptions required for reliable statistical modeling. The optimal lambda (λ) value of approximately -0.9 was determined using Guerrero's method. This transformation aimed to reduce the impact of heteroskedasticity, making the variance more uniform across the series (Guerrero, 1993).

The STL decomposition of the Box-Cox transformed Portugal real GDP series breaks down the data into trend, seasonal, and residual components. The trend component reflects a steady growth pattern over time, with a notable drop in the first quarter of 2020 due to the COVID-19 pandemic. The seasonal component exhibits a consistent annual pattern, indicating regular fluctuations in GDP throughout the year. The ACF plot shows significant autocorrelations that gradually decline, suggesting a persistence in GDP values over time and indicating that past values continue to influence future values (Hamilton, 1994). The PACF plot identifies significant partial autocorrelations mainly at the first lag, suggesting that the immediate past GDP value is a strong predictor of the current value.

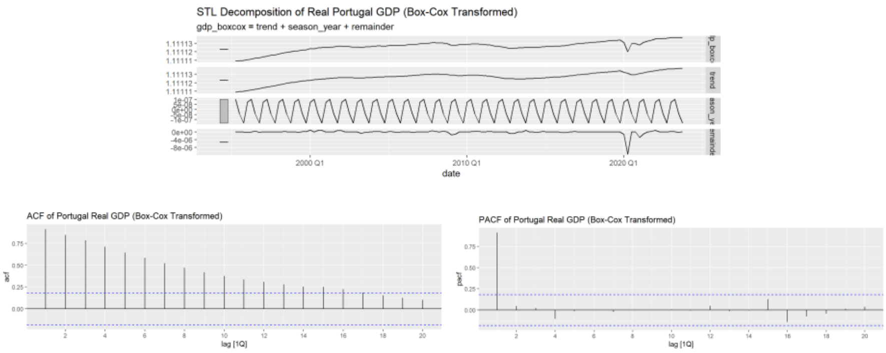


Figure 2: STL Decomposition (top), ACF (left) and PACF (right) of the Box-Cox transformed GDP

3.2 Stationarity & Differencing

For accurate modeling and forecasting using ARIMA, the time series data must be stationary, meaning its statistical properties do not change over time. This is a fundamental assumption of ARIMA models (Box et al., 2015). To assess stationarity, we applied a combination of the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The ADF test examines whether the time series has a unit root, indicating non-stationarity, while the KPSS test checks for stationarity around a deterministic trend (Dickey & Fuller, 1979; Kwiatkowski et al., 1992).

The initial tests indicated non-stationarity in the data, with the ADF test failing to reject the null hypothesis of a unit root and the KPSS test rejecting the null hypothesis of stationarity. These results imply that the time series contains non-stationary components or a trend. As a result, in order to eliminate these trends and stabilize the mean, first-order differencing was applied to the data. Following this transformation, the ADF and KPSS tests were performed again, with both tests confirming stationarity at the 5% significance level.

Differentiating the time series data to confirm stationarity is essential because it guarantees that the data meets the assumptions required for accurate and reliable ARIMA modelling. This approach, supported by Hyndman & Athanasopoulos (2021), allows us to proceed with the modeling process with confidence that the resulting forecasts will be robust and valid. The detailed results, including test statistics and critical values, are presented in appendix B.

3.3 Seasonality Analysis

To assess whether seasonal differencing was necessary, the measure of seasonal strength, F_S, as proposed by Hyndman and Athanasopoulos (2021), was calculated. The formula used for calculating F_S is:

$$F_s = \max\left(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)}\right)$$

where S_t is the seasonal component and R_t is a remainder component.

The calculated F_S value was approximately 0.0299 (Appendix C, Figure 1), which is very low. This low value suggests that the seasonal component in the Box-Cox transformed and differenced data is weak. It implies that the seasonal variations do not significantly contribute to the variance in the data after the transformations. In practical terms, this means that the series does not exhibit strong, regular seasonal patterns and therefore, additional seasonal differencing is not necessary.

3.4 Structural Breaks

Structural breaks can substantially influence the relationships between variables in time series data, potentially compromising the validity of transformations and forecast modeling assumptions (Bai & Perron, 2003). To ensure accurate forecasting and robust model fitting, it is crucial to identify and account for these structural breaks. In this section, we conducted several tests to detect structural breaks: the Quandt Likelihood Ratio (QLR) test, OLS-MOSUM test, OLS-CUSUM test, and the Step Indicator Saturation (SIS) method. Each test offers unique insights into the stability and potential segmentation of the time series data, and all the calculations can be followed in appendix D.

The QLR test, an advanced version of the Chow test, assesses potential breakpoints by comparing model fits before and after suspected break dates within a specified range. This test identifies breakpoints where the F-statistic of the Chow test is significant at the 5% level (Quandt, 1960). In our analysis, the QLR test was performed on the GDP data, identifying a breakpoint at observation number 100, corresponding to the first quarter of 2020 (Figure 3).

The OLS-MOSUM and OLS-CUSUM tests are used to check for stability in a model by analyzing its residuals. The OLS-MOSUM test uses the moving sum of residuals, and the OLS-CUSUM test uses the cumulative sum of recursive residuals. A structural break is detected if the test crosses the critical limits (Zeileis, Kleiber, Krämer, & Hornik, 2003). Our analysis using these tests did not provide evidence for structural breaks, as the test statistics did not cross the critical thresholds (Appendix D, Figure 2 and 3)

The SIS method fits a model with various step indicators to detect significant shifts in the intercept, indicating structural breaks. This method helps in detecting large shifts in the intercept, which are crucial for accurate modeling (Castle, Doornik, & Hendry, 2015). Applying the SIS method to our dataset revealed significant step indicators at several points, suggesting potential break dates (Figure 3).

Combining the results from these tests, we identified a significant structural break in the first quarter of 2020, due to the economic impact of the COVID-19 pandemic. The QLR test indicated this breakpoint with high significance, while the OLS-MOSUM and OLS-CUSUM tests did not show additional breaks. The SIS method corroborated the presence of significant step changes around this period.

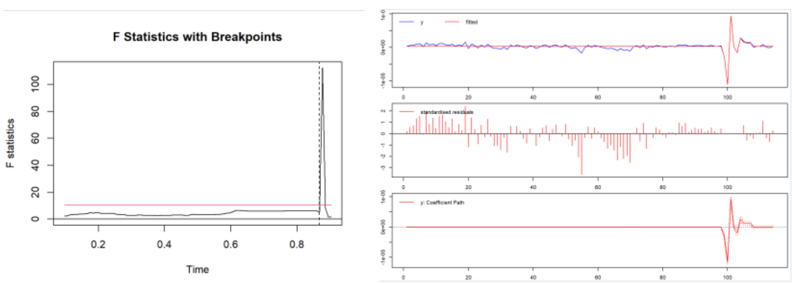


Figure 3: QRL-test (left) and SIS (right)

To proceed with forecasting, the dataset was split into training and test sets according to the structural break detected. The training dataset was then tested for stationarity. Initial tests suggested non-stationarity, necessitating a second differencing. After applying the second differencing, stationarity was confirmed, allowing for robust time series modeling and accurate forecasting. You can follow the calculations in Appendix D.

4. Forecasting

4.1 Forecasting Model Selection

For the task of forecasting Portugal's GDP post-pandemic, the study employs two renowned time series models: ARIMA (AutoRegressive Integrated Moving Average) and ETS (Exponential Smoothing State Space). The selection of these models is based on their proven capabilities in capturing complex economic dynamics and responding to significant structural changes.

ARIMA is highly effective for modeling and forecasting time series because it combines three key components: AutoRegressive (AR), Integrated (I), and Moving Average (MA), which collectively capture the underlying patterns in the data. The model's adaptability enables it to take into account the series' trends, seasonality, and autocorrelations. The 'Integrated' component is especially important as it allows the model to transform non-stationary data into a stationary form through differencing, which is a prerequisite for accurate time series prediction (Box, Jenkins, Reinsel, & Ljung, 2015).

On the other hand, the ETS model is selected for its strength in modeling data with trends and seasonality through exponential smoothing techniques. This model applies weighted averages of historical observations with exponentially decreasing weights, making it highly adaptable to different time series behaviours. It does not require stationarity in the data, making it ideal for direct modeling of raw GDP data that exhibits non-linear trends and potential seasonal effects, intensified by the pandemic's impact (Hyndman & Athanasopoulos, 2021).

The capacity of the ARIMA and ETS models to reduce forecast errors and capture the underlying economic cycles and trends determines which model is preferred. This includes a thorough analysis of model residuals to ensure they exhibit characteristics of white noise, indicating that the models have successfully captured the information in the data. Given the significant economic impact of the COVID-19 pandemic on Portugal's GDP, achieving ideal residual characteristics might be challenging. The models are rigorously tested on a training dataset, derived from historical GDP data from 1995 to just before the pandemic onset, with validation on subsequent data to assess their predictive accuracy and robustness.

The selection process is guided by several statistical criteria, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and error metrics such as the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These criteria help in identifying the model that best balances fit and complexity, ensuring that the forecasts are both accurate and robust (Hyndman & Athanasopoulos, 2021).

4.2 Results

4.2.1 ARIMA

In developing the ARIMA model for this analysis, the initial model specification was determined by carefully examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the GDP time series after differencing twice to achieve stationarity (Figure 4)

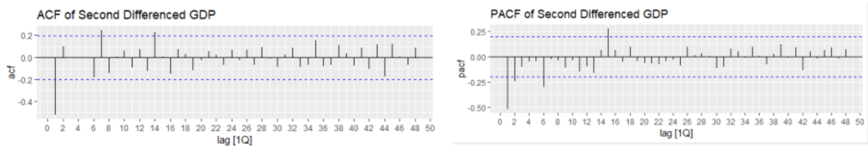


Figure 4: ACF of Second Differenced GDP (left) and PACF of Second Differenced GDP (right)

The ACF plot revealed a significant spike at lag 1, suggesting the inclusion of a moving average (MA) component of order 1 (q=1). Additionally, the PACF plot indicated significant spikes at lags 1 and 2, which supported the inclusion of autoregressive (AR) components of orders 1 and 2 (p=1 and p=2). These observations led to the consideration of the ARIMA(1,2,1)(1,0,1) and ARIMA(2,1,1)(1,0,1) models. Despite the weak seasonal strength (approximately 0.0299), seasonal components were included in the models to ensure even minimal seasonal effects were captured, in line with best practices in time series forecasting.

Additionally, the Auto ARIMA model with drift, specified as ARIMA(1,1,1), was chosen to account for a potential linear trend in the GDP data, a critical factor for accurately modeling economic indicators that typically exhibit steady trends over time. The drift term helps ensure the forecasted values reflect the series' overall direction while balancing model complexity and fit by minimizing AICc.

Every model that was chosen was subjected to residual analysis in order to assess its suitability. The residuals generally centered around zero, though some models exhibited distinct spikes and significant autocorrelations at certain lags. This indicated that while the models captured much of the underlying data structure, certain patterns were not fully accounted for. The residuals' distribution was approximately normal, but with some outliers, suggesting that some models might not fully capture the variability in the data. The residual diagnostics can be observed in Figure 5, and also in Appendix E, including Figures 1 and 2.

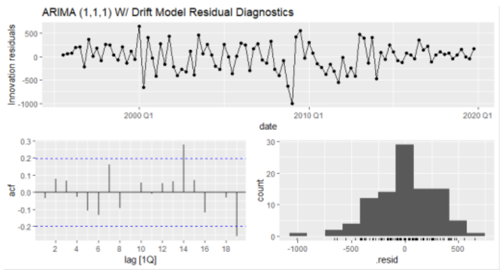


Figure 5: ARIMA (1,1,1) W/ Drift Model Residual Diagnostics

Further diagnostics using the Ljung-Box test indicated that residual autocorrelation was not significant across the models, suggesting that the residuals were independently distributed, a desirable outcome in time series modeling. Additionally, the Shapiro-Wilk test results showed that the residuals were likely normally distributed, with high p-values supporting the adequacy of the ARIMA models and confirming that the error terms behaved as expected under the assumption of normality (Appendix E, Figures 3 and 4).

To conclude, although the ARIMA (2,2,1)(1,0,1) and ARIMA (1,2,1)(1,0,1) models offer slightly better AIC and AICc values, the ARIMA (1,1,1) W/ Drift model stands out with its superior performance on accuracy metrics. Specifically, the ARIMA (1,1,1) W/ Drift model achieves lower RMSE, MAE, and MAPE values compared to the other models, suggesting that it provides more precise forecasts despite its marginally higher AIC, AICc, and BIC. This makes the ARIMA (1,1,1) W/ Drift the most suitable ARIMA choice for forecasting GDP, as it achieves a better balance between predictive accuracy and model efficiency (Table 1).

	AIC	AICc	BIC
ARIMA (1,1,1) W/ Drift - Auto ARIMA	1383.09	1383.52	1393.39
ARIMA (2,2,1)(1,0,1)	1373.92	1374.87	1389.31
ARIMA (1,2,1)(1,0,1)	1373.62	1374.28	1386.44

	RMSE	MAE	MAPE
ARIMA (1,1,1) W/ Drift - Auto ARIMA	3482.22	2528.03	5.67
ARIMA (2,2,1)(1,0,1)	3644.55	2913.46	6.45
ARIMA (1,2,1)(1,0,1)	3692.08	2990.81	6.60

Table 1: Information criteria (left) and test accuracy (right) of ARIMA models.

4.2.2 ETS

The selection of the ETS models was guided by a thorough analysis of the GDP time series characteristics, particularly the trend and seasonality components identified through STL (Seasonal and Trend decomposition using Loess) decomposition (Figure 6).

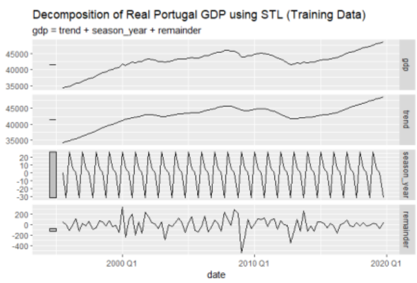


Figure 6: STL Decomposition on the training dataset.

The Auto ETS model was chosen for its systematic selection of the best component combinations, with the ETS(M,Ad,N) configuration proving particularly well-suited given the observed multiplicative error, damped trend, and minimal or no seasonality.

The Auto ETS model was chosen for its methodical selection of the best component combinations, with the ETS(M,Ad,N) configuration proving especially appropriate given the observed multiplicative error, damped trend, and minimal or no seasonality. In addition to the Auto ETS model, models such as ETS(AAN), ETS(AAdN), and ETS(MAN) were included into the analysis. The ETS(AAN) model was selected for its ability to capture the linear trend observed in the data, while the ETS(AAdN) model, with its damped trend, was included to account for a potential slowdown or stabilization in GDP growth, reflecting the leveling off observed in the trend component. The ETS(MAN) model was considered to address variations in the data that might require a multiplicative error model, particularly given any multiplicative effects seen in the remainder component.

Residual analysis across all ETS models indicated that the residuals were mostly centered around zero, suggesting that the models captured much of the fundamental data structure. However, some models exhibited noticeable spikes and significant autocorrelations at certain lags, particularly during periods of economic volatility. Despite these spikes, the residuals' distribution was approximately normal, with some rightward skew, indicating that while the models were effective, there was room for further improvement (Figure 7 and Appendix F).

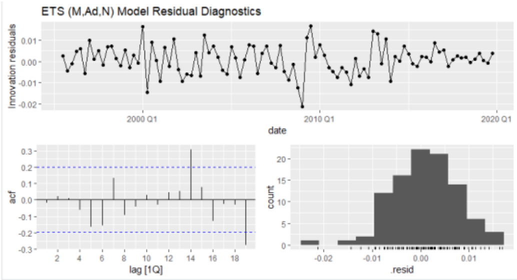


Figure 7: ETS (M,Ad,N) Model Residual Diagnostics

Further diagnostics using the Ljung-Box test indicated that residual autocorrelation was not significant across the models, with high p-values suggesting that the residuals were independently distributed—a desirable outcome in time series modeling. The Shapiro-Wilk test results also showed that the residuals were likely normally distributed, with high p-values across models, reinforcing the adequacy of the ETS models (Appendix F Figures 4 and 5).

The Auto ETS (M,Ad,N) model emerged as the best-performing model, with the lowest AIC (1571.87) and AICc values, suggesting a superior balance between model fit and complexity. The ETS(AAdN) model was a close contender, offering similar performance with the added advantage of modeling a damped trend, which could better reflect future economic conditions if the growth rate is expected to decelerate. The ETS(MAN) and ETS(AAN) models, while robust, showed slightly inferior performance in both information criteria and forecast accuracy, suggesting they may not capture the data's underlying structure as effectively.

	AIC	AICc	BIC
ETS (M,Ad,N) - Auto ETS	1571.87	1572.79	1587.38
ETS (A,A,N)	1572.36	1573.28	1587.87
ETS(A,Ad,N)	1572.75	1573.4	1585.67
ETS(M,A,N)	1574.07	1574.72	1586.99

	RMSE	MAE	MAPE
ETS (M,Ad,N) - Auto ETS	3350.93	2253.67	5.10
ETS (A,A,N)	3964.91	3431.72	7.50
ETS(A,Ad,N)	3325.70	2256.58	5.10
ETS(M,A,N)	3958.59	3422.67	7.48

Table 2: Information criteria (left) and test accuracy (right) of ETS models.

The smoothing parameters of the selected models provided further insights into their behavior. The Auto ETS model, with α=0.9143 and β=0.3970, indicated a strong emphasis on recent observations and a damped trend, making it responsive yet stable for long-term GDP forecasting (Appendix F Figure 6).

4.3 Forecasts Analysis

In this section, the forecasts generated by the best ARIMA and ETS models are compared to assess their predictive performance for the GDP data. The best ARIMA model, ARIMA(1,1,1) with drift, and the best ETS model, ETS(M,Ad,N), were evaluated using several accuracy metrics, including RMSE, MAE, and MAPE, as shown in Table [3] and visually represented in Figure [8].

	RMSE	MAE	MAPE
ARIMA (1,1,1) W/ Drift - Auto ARIMA	3482.22	2528.03	5.67
ETS (M,Ad,N) - Auto ETS	3350.93	2253.67	5.10

Table 3: Test Accuracy of ARIMA (1,1,1) W/ Drift anf ETS (M,Ad,N) on test dataset.

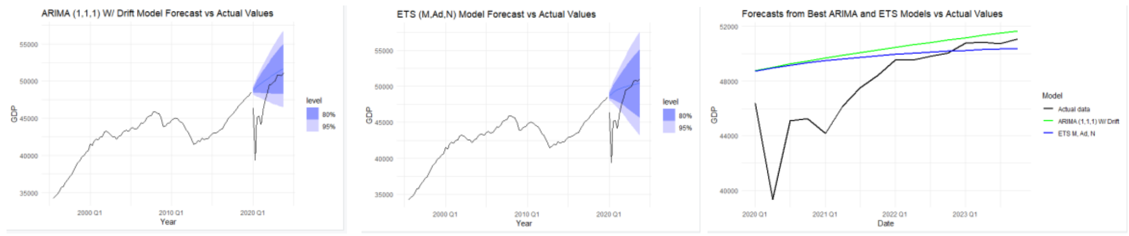


Figure 8: ARIMA (1,1,1) W/ Drift and ETS (M,Ad,N) forecast of test dataset.

The unprecedented global disruption caused by the COVID-19 pandemic had a profound impact on economic indicators worldwide, including Portugal's GDP. The sharp and unexpected economic contraction followed by a period of recovery introduced significant volatility into the GDP data. This unpredictability, particularly during 2020 and 2021, has heavily influenced the accuracy of forecasting models like ARIMA and ETS, making it challenging to predict GDP movements with precision. According to the Bank of Portugal, the country's central bank, during the peak of the pandemic's economic impact, Portugal's GDP experienced one of its most severe contractions, with a decline of over 16% year-on-year in the second quarter of 2020. This period saw GDP fall to approximately 39 billion euros, a level not seen since 1999, highlighting the magnitude of the economic impact (Banco de Portugal, 2020). The subsequent recovery was gradual and uneven, further complicating the task of forecasting models, which rely on historical data patterns to predict future trends.

As referred previously, given the significant structural break in the GDP data caused by the COVID-19 pandemic, particularly in Q1 2020, I proceeded to split the dataset at this critical point. This split allows for a more nuanced analysis by isolating the pre-pandemic period from the turbulent economic conditions that followed. By focusing on the data prior to the pandemic, the ARIMA and ETS models can provide predictions that reflect a hypothetical scenario where the GDP trajectory continued without the disruption caused by COVID-19. This approach offers an alternative view, essentially forecasting what the GDP could have been if the pandemic had not occurred.

Recognizing the volatility introduced by the pandemic, especially during 2020 and 2021, and its impact on the forecasting models and the accuracy metrics (Table 3 and Figure 8), I decided to also evaluate the models' performance based on the period from 2022-Q1 to 2023-Q4 (Figure 9). This period represents a more stable phase for GDP, as the economy had begun to stabilize post-pandemic. By calculating the accuracy metrics specifically for this period, I aimed to determine which model provided more reliable predictions under less volatile conditions.

	RMSE	MAE	MAPE
ARIMA (1,1,1) W/ Drift - Auto ARIMA	826.01	789.05	1.57
ETS (M,Ad,N) - Auto ETS	464.62	438.23	0.87

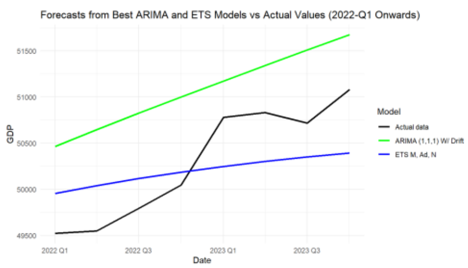


Figure 9: Test Accuracy of ARIMA (1,1,1) W/ Drift anf ETS (M,Ad,N) on filtered test dataset and visual representation.

The results from the more stable period of 2022-Q1 to 2023-Q4 indicated that the ETS model slightly outperformed the ARIMA model across all accuracy metrics. The ETS model's lower RMSE, MAE, and MAPE values suggest it captured the underlying GDP patterns more effectively during this period. Furthermore, the visual comparison of the forecast plots revealed that the ETS model's predictions were more closely aligned with the actual GDP values, particularly as the economy began to stabilize from the pandemic's impact.

The superior performance of the ETS(M,Ad,N) model can be attributed to its ability to effectively capture the characteristics of GDP data through its multiplicative error structure and damped trend. The multiplicative error allows the model to adapt to variations where errors scale with the level of the series, while the damped trend component ensures that the model realistically reflects slowing long-term trends. This makes the ETS(M,Ad,N) model particularly well-suited for modeling Portugal's GDP as it stabilizes post-pandemic.

5. Conclusion

The purpose of this project was to analyze the impact of the COVID-19 pandemic on Portugal's GDP and to identify the most accurate forecasting model for predicting future GDP trends. The study evaluated various time series models, ultimately determining that the ETS (M,Ad,N) model emerged as the most accurate and reliable for forecasting Portugal’s GDP in the post-pandemic context.

The findings indicated that the structural changes brought about by the pandemic posed significant challenges to traditional models like ARIMA, which struggled to capture the abrupt shifts in economic activity. In contrast, the ETS model proved more adaptable, capturing the multiplicative errors and damped trends that characterized the post-pandemic GDP data.

Reflecting on the broader economic context, it is clear that while the COVID-19 pandemic affected economies worldwide, Portugal was particularly vulnerable due to its heavy reliance on tourism. In 2023, tourism accounted for 19.6% of Portugal’s GDP, underscoring the sector's significant role in the national economy. Looking forward, the World Travel & Tourism Council (WTTC) projects that by 2034, the tourism sector will contribute an estimated 22.4% to the economy. While this highlights the sector's strength, it also poses a risk in times of global uncertainty (WTTC, 2023).

Given this reliance, it is crucial for Portugal to complement its tourism sector with strategic investments in other industries. By fostering growth in areas such as technology, green energy, and digital services, Portugal could build a more resilient and diversified economic base. This approach would not only mitigate the impact of future global shocks but also position the country for sustainable long-term growth, ensuring that it remains competitive in a rapidly changing global economy.

6. References

Bai, J., & Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics, 18*(1), 1-22.

Banco de Portugal. (2020). *Economic Bulletin, October 2020*. Bank of Portugal. Retrieved from https://www.bportugal.pt/en/publication/economic-bulletin-october-2020.

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

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2015). Step Indicator Saturation. *Journal of Econometrics, 184*(2), 396-408.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427-431.

Federal Reserve Bank of St. Louis. (2023). Portugal Real GDP, FRED Economic Data. Retrieved from https://fred.stlouisfed.org/series/CLVMNACSCAB1GQPT.

Guerrero, V. M. (1993). Time series analysis using the Box-Cox transformation. *Journal of Time Series Analysis, 14*(3), 235-249.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts: Melbourne, Australia. OTexts.com/fpp3.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1-3), 159-178.

OECD. (2023). OECD National Accounts Data. Retrieved from https://data.oecd.org.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association, 55*(290), 324-330.

World Travel & Tourism Council. (2023). Portugal Travel and Tourism Poised for Historic Year. Retrieved from WTTC.

Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis, 44*(1-2), 109-123.

7. Appendix

Appendix A - Temporal patterns and mathematical transformation

Figure 1 – ACF of Portugal Real GDP

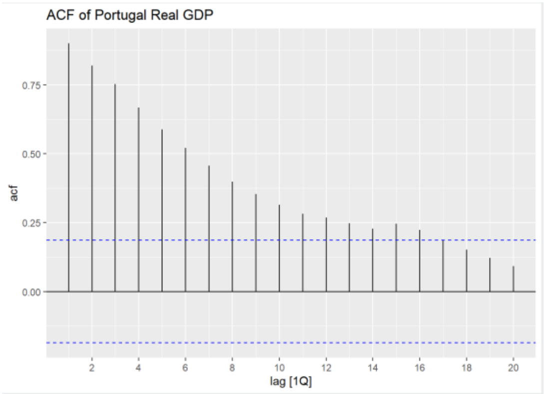


Figure 2 – PACF of Portugal Real GDP

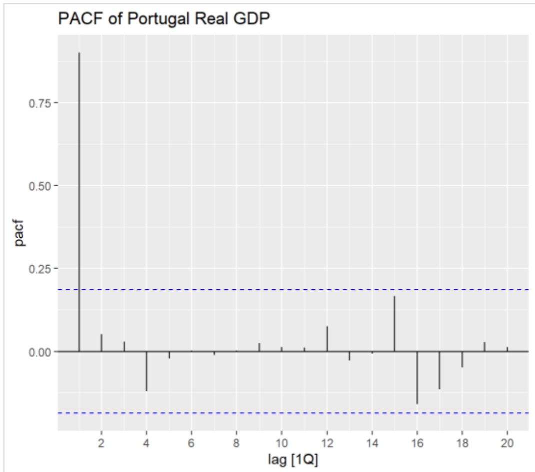
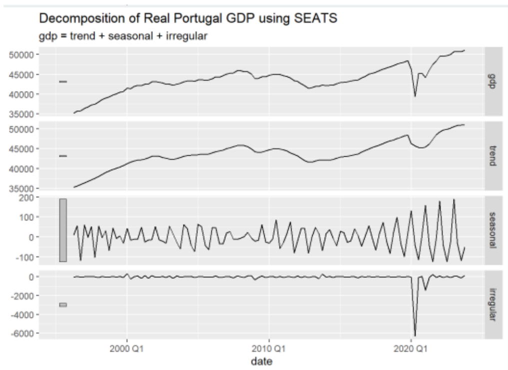


Figure 3 – SEATS Decomposition of Portugal Real GDP



Appendix B - Stationarity & Differencing

Test 1 - ADF Test with Trend after-differentiation:

```
> # Test 1: Augmented Dickey-Fuller test with a trend
> summary(ur.df(as.ts(pt$d_gdp_boxcox), type = 'trend', lag = 24, selectlags
= 'AIC'))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
       Min         1Q     Median         3Q        Max
-1.227e-05 -1.791e-07  1.959e-07  4.220e-07  4.953e-06

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.031e-07  5.196e-07  -0.583   0.5612
z.lag.1     -1.541e+00  1.692e-01  -9.109 4.37e-14 ***
tt           7.289e-09  7.255e-09   1.005   0.3180
z.diff.lag   2.386e-01  1.073e-01   2.224   0.0289 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.659e-06 on 82 degrees of freedom
Multiple R-squared:  0.6437,	Adjusted R-squared:  0.6307
F-statistic: 49.39 on 3 and 82 DF,  p-value: < 2.2e-16


Value of test-statistic is: -9.109 27.6586 41.4871

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

Test 2-KPSS Test with Trend after-differentiation:

```
> # Test 2: KPSS test with a trend
> summary(ur.kpss(as.ts(pt$d_gdp_boxcox), type = 'tau'))

#######################
# KPSS Unit Root Test #
#######################

Test is of type: tau with 4 lags.

Value of test-statistic is: 0.135

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

Test 3 - ADF Test with Drift after-differentiation:

```
> # Test 3: Augmented Dickey-Fuller test with a drift
> summary(ur.df(as.ts(pt$d_gdp_boxcox), type = 'drift', lag = 24, selectlags
= 'AIC'))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
       Min         1Q     Median         3Q        Max
-1.204e-05 -3.071e-07  2.049e-07  4.901e-07  5.304e-06

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.866e-07  1.800e-07   1.036   0.3030
z.lag.1     -1.522e+00  1.681e-01  -9.053 5.12e-14 ***
z.diff.lag   2.288e-01  1.068e-01   2.142   0.0351 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.659e-06 on 83 degrees of freedom
Multiple R-squared:  0.6394,	Adjusted R-squared:  0.6307
F-statistic: 73.57 on 2 and 83 DF,  p-value: < 2.2e-16


Value of test-statistic is: -9.0529 40.9785

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

Test 4 - KPSS Test with Level after-differentiation:

```
> # Test 4: KPSS test with a level
> summary(ur.kpss(as.ts(pt$d_gdp_boxcox), type = 'mu'))

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.1891

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Test 5 - ADF Test with None after-differentiation:

```
> # Test 5: Augmented Dickey-Fuller test with none
> summary(ur.df(as.ts(pt$d_gdp_boxcox), type = 'none', lag = 24, selectlags =
'AIC'))

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
       Min         1Q     Median         3Q        Max
-1.183e-05 -1.270e-07  3.891e-07  6.708e-07  5.624e-06

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1     -1.5028     0.1672  -8.990 6.23e-14 ***
z.diff.lag   0.2191     0.1065   2.058   0.0427 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.66e-06 on 84 degrees of freedom
Multiple R-squared:  0.6347,	Adjusted R-squared:  0.626
F-statistic: 72.97 on 2 and 84 DF,  p-value: < 2.2e-16


Value of test-statistic is: -8.9895

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Appendix C - Seasonality Analysis

Figure 1 - Seasonal Strength Calculation (R Code)

```
> ### 9 - Seasonality Analysis
> #STL Decomposition on Box-Cox transformed and differenced data
> stl_decomp <- pt %>% model(STL(d_gdp_boxcox ~ season(window = "periodic")))
> components <- components(stl_decomp)
> # Extract seasonal and remainder components
> S_t <- components$season_year
> R_t <- components$remainder
> # Calculate the variance of seasonal and remainder components
> var_Rt <- var(R_t, na.rm = TRUE)
> var_St_Rt <- var(S_t + R_t, na.rm = TRUE)
> # Calculate seasonal strength F_s
> F_s <- max(0, 1 - var_Rt / var_St_Rt)
> #Print
> print(F_s)
[1] 0.02987415
```

Appendix D – Structural Breaks

Figure 1 – QLR Test (R Code)

```
> ### 10.1 - QLR Test (Chow Test)
> qlr <- Fstats(Lag0 ~ 1 + Lag1, data = pt.lag, from = 0.10)
> plot(qlr, alpha = 0.1, main = "F Statistics")
> test <- sctest(qlr, type = "supF")
> print(test)

        supF test

data:  qlr
sup.F = 112.34, p-value < 2.2e-16

> breaks <- breakpoints(qlr, alpha = 0.01)
> print(breaks)

         Optimal 2-segment partition:

Call:
breakpoints.Fstats(obj = qlr, alpha = 0.01)

Breakpoints at observation number:
100

Corresponding to breakdates:
0.8684211
> breakpoint_dates <- pt$date[breaks$breakpoints]
> print(breakpoint_dates)
<yearquarter[1]>
[1] "2020 Q1"
# Year starts on: January
```

Figure 2 – OLS-based MOSUM Test

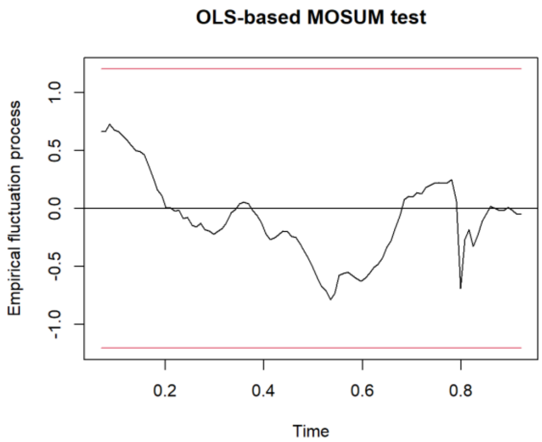


Figure 3 – OLS-based CUSUM Test

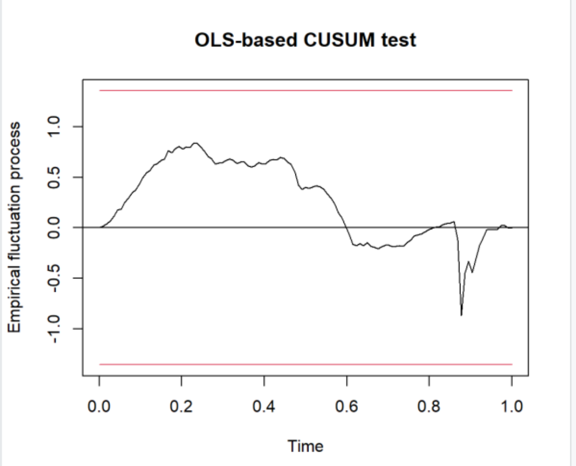


Figure 4 – SIS Test (R Code)

```
> print(sis)

Date: Thu Aug  8 12:12:28 2024
Dependent var.: y
Method: Ordinary Least Squares (OLS)
Variance-Covariance: Ordinary
No. of observations (mean eq.): 114
Sample: 1 to 114

SPECIFIC mean equation:

              coef   std.error   t-stat   p-value
mconst  2.5543e-07  5.3853e-08   4.7432 6.657e-06 ***
sis99  -3.0297e-06  5.3583e-07  -5.6543 1.359e-07 ***
sis100 -8.3803e-06  7.5394e-07 -11.1155 < 2.2e-16 ***
sis101  2.0551e-05  7.5394e-07  27.2577 < 2.2e-16 ***
sis102 -9.1694e-06  7.5394e-07 -12.1620 < 2.2e-16 ***
sis103 -1.8368e-06  7.5394e-07  -2.4362  0.016524 *
sis104  4.4302e-06  7.5394e-07   5.8761 5.007e-08 ***
sis105 -1.3644e-06  6.1559e-07  -2.2164  0.028818 *
sis108 -1.1952e-06  3.6788e-07  -3.2489  0.001556 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Diagnostics and fit:

                 Chi-sq df   p-value
Ljung-Box AR(1)   19.765  1 8.757e-06 ***
Ljung-Box ARCH(1)  6.659  1  0.009866 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

SE of regression    0.00000
R-squared           0.88829
Log-lik.(n=114)  1489.41777
```

Test 1 - ADF Test with Trend after second differentiation:

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
       Min         1Q     Median         3Q        Max
-1.236e-06 -2.381e-07 -3.239e-08  3.058e-07  1.079e-06

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.070e-07  1.691e-07  -0.633  0.52928
z.lag.1     -2.775e+00  4.749e-01  -5.843  1.8e-07 ***
tt           1.783e-09  2.621e-09   0.680  0.49867
z.diff.lag1  1.253e+00  4.288e-01   2.922  0.00478 **
z.diff.lag2  9.276e-01  3.636e-01   2.551  0.01310 *
z.diff.lag3  7.548e-01  2.912e-01   2.592  0.01178 *
z.diff.lag4  5.681e-01  2.040e-01   2.785  0.00701 **
z.diff.lag5  3.132e-01  1.054e-01   2.970  0.00417 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.672e-07 on 65 degrees of freedom
Multiple R-squared:  0.751,     Adjusted R-squared:  0.7242
F-statistic: 28.01 on 7 and 65 DF,  p-value: < 2.2e-16


Value of test-statistic is: -5.8432 11.3839 17.0736

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2  6.50  4.88  4.16
phi3  8.73  6.49  5.47
```

Test 2-KPSS Test with Trend after second differentiation:

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: tau with 3 lags.

Value of test-statistic is: 0.0209

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

Test 3 - ADF Test with Drift after second differentiation:

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
       Min         1Q     Median         3Q        Max
-1.240e-06 -2.297e-07 -2.967e-08  2.974e-07  1.099e-06

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.910e-09  5.449e-08   0.035  0.97214
z.lag.1     -2.734e+00  4.692e-01  -5.828 1.84e-07 ***
z.diff.lag1  1.216e+00  4.236e-01   2.871  0.00550 **
z.diff.lag2  8.962e-01  3.592e-01   2.495  0.01511 *
z.diff.lag3  7.313e-01  2.880e-01   2.539  0.01347 *
z.diff.lag4  5.543e-01  2.022e-01   2.742  0.00785 **
z.diff.lag5  3.062e-01  1.045e-01   2.930  0.00465 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.653e-07 on 66 degrees of freedom
Multiple R-squared:  0.7493,    Adjusted R-squared:  0.7265
F-statistic: 32.87 on 6 and 66 DF,  p-value: < 2.2e-16


Value of test-statistic is: -5.8276 16.9826

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1  6.70  4.71  3.86
```

Test 4 - KPSS Test with Level after second differentiation:

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 3 lags.

Value of test-statistic is: 0.0497

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Test 5 - ADF Test with None second differentiation:

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
       Min         1Q     Median         3Q        Max
-1.238e-06 -2.278e-07 -2.776e-08  2.992e-07  1.101e-06

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1      -2.7341     0.4656  -5.872 1.48e-07 ***
z.diff.lag1   1.2162     0.4204   2.893  0.00515 **
z.diff.lag2   0.8963     0.3565   2.514  0.01434 *
z.diff.lag3   0.7315     0.2858   2.559  0.01275 *
z.diff.lag4   0.5544     0.2006   2.763  0.00738 **
z.diff.lag5   0.3063     0.1037   2.953  0.00433 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.618e-07 on 67 degrees of freedom
Multiple R-squared:  0.7493,    Adjusted R-squared:  0.7268
F-statistic: 33.37 on 6 and 67 DF,  p-value: < 2.2e-16


Value of test-statistic is: -5.8718

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.6 -1.95 -1.61
```

Appendix E – ARIMA

Figure 1 – ARIMA (2,2,1)(1,0,1)[4] Residual Diagnosis

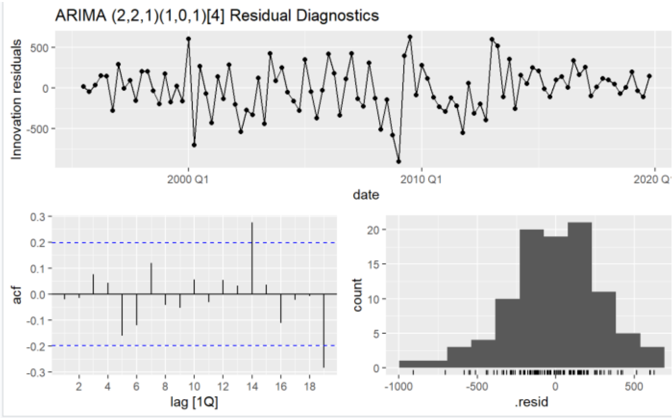


Figure 2 – ARIMA (1,2,1)(1,0,1)[4] Residual Diagnosis

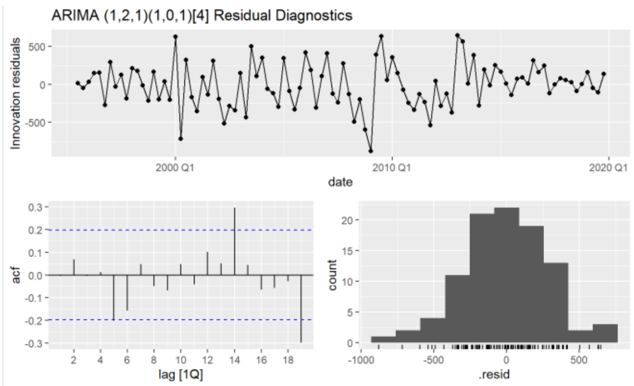


Figure 3 – ARIMA Ljung-Box Test

```
# Display Ljung-Box test results
print(lb_test_auto_arima_drift)
A tibble: 1 × 3
.model                  lb_stat lb_pvalue
<chr>                     <dbl>     <dbl>
auto_arima_model_drift     8.40     0.395
print(lb_test_guessed_arima_1)
A tibble: 1 × 3
.model          lb_stat lb_pvalue
<chr>             <dbl>     <dbl>
guessed_arima_1    7.45     0.489
print(lb_test_guessed_arima_2)
A tibble: 1 × 3
.model          lb_stat lb_pvalue
<chr>             <dbl>     <dbl>
guessed_arima_2    8.66     0.372
```

Figure 4 – ARIMA Shapiro-Wilk Test

```
> print(shapiro_test_auto_arima_drift)

        Shapiro-Wilk normality test

data:  modelsg %>% select(auto_arima_model_drift) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.98539, p-value = 0.3525

> print(shapiro_test_guessed_arima_1)

        Shapiro-Wilk normality test

data:  modelsg %>% select(guessed_arima_1) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.99114, p-value = 0.7673

> print(shapiro_test_guessed_arima_2)

        Shapiro-Wilk normality test

data:  modelsg %>% select(guessed_arima_2) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.99188, p-value = 0.8222
```

Appendix F – ETS

Figure 1 – ETS (A,A,N) Residual Diagnosis

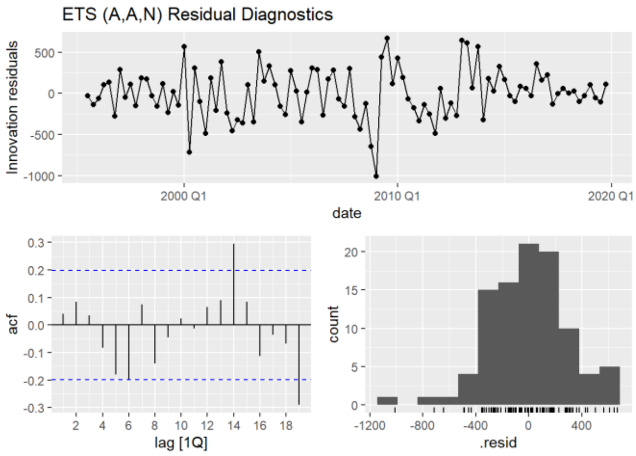


Figure 2 – ETS (A,Ad,N) Residual Diagnosis

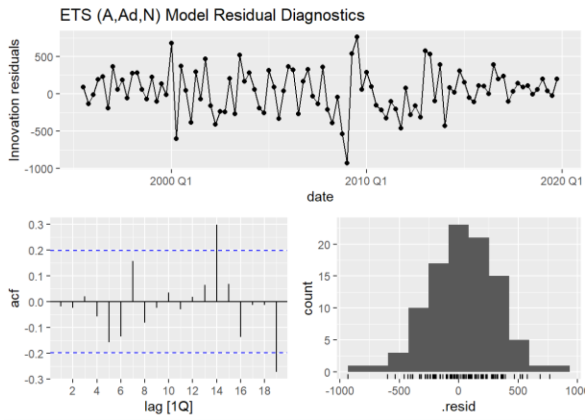


Figure 3 – ETS (M,A,N) Residual Diagnosis

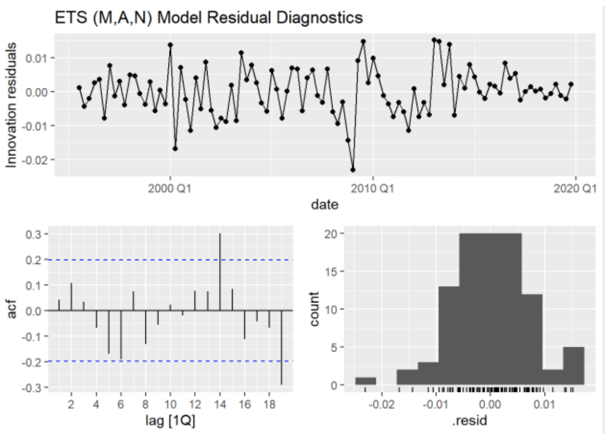


Figure 4 – ETS Ljung-Box Test

```
> print(lb_test_auto_ets)
# A tibble: 1 × 3
  .model    lb_stat lb_pvalue
  <chr>       <dbl>     <dbl>
1 auto_ets     9.53     0.300
> print(lb_test_ets_AAN)
# A tibble: 1 × 3
  .model  lb_stat lb_pvalue
  <chr>     <dbl>     <dbl>
1 ets_AAN    12.4     0.135
> print(lb_test_ets_AAdN)
# A tibble: 1 × 3
  .model    lb_stat lb_pvalue
  <chr>       <dbl>     <dbl>
1 ets_AAdN     8.61     0.376
> print(lb_test_ets_MAN)
# A tibble: 1 × 3
  .model  lb_stat lb_pvalue
  <chr>     <dbl>     <dbl>
1 ets_MAN    12.6     0.128
```

Figure 5 – ETS Shapiro-Wilk Test

```
	Shapiro-Wilk normality test

data:  ets_models %>% select(auto_ets) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.99296, p-value = 0.8924

> print(shapiro_test_ets_AAN)

	Shapiro-Wilk normality test

data:  ets_models %>% select(ets_AAN) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.98664, p-value = 0.4287

> print(shapiro_test_ets_AAdN)

	Shapiro-Wilk normality test

data:  ets_models %>% select(ets_AAdN) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.99256, p-value = 0.8681

> print(shapiro_test_ets_MAN)

	Shapiro-Wilk normality test

data:  ets_models %>% select(ets_MAN) %>% residuals() %>% select(.resid) %>% as.ts()
W = 0.98738, p-value = 0.4788
```

Figure 6 – Smoothing Parameters (M,Ad,N) Model

```
> ### 26 - Smoothing Parameters ETS
> # Auto ETS Model
> report(ets_models %>%
+          select(auto_ets))
Series: gdp
Model: ETS(M,Ad,N)
  Smoothing parameters:
    alpha = 0.9143283
    beta  = 0.3969956
    phi   = 0.8882289

  Initial states:
     l[0]      b[0]
 33799.42 459.3424

  sigma^2:  0

     AIC     AICc      BIC
1571.866 1572.790 1587.376
>
```