

Final Presentation

Stationarity and Unit Root Tests

Jordan Andersen

2025-12-09

Introduction and Motivation

A stationary time series, following Hyndman & Athanasopoulos (2018), is one whose statistical properties do not depend on the time at which the series is observed. Intuitively, this implies that any given section of the time series should “look like” any other section in the time series in terms of the joint distribution. While strict stationarity requires that all the joint distributions of a series are time-invariant, weak stationarity requires only that the mean and variance are constant over time, and that the autocovariance is only dependent on the lag between observations, not on time. Weak stationarity is often more applicable in cases with real-world data.

Stationarity is foundational to many time series models, including ARMA and ARIMA. These models assume constant variance, mean and autocovariance in order to make reliable predictions. When a series is non-stationary, often due to trends, shocks, seasonality or volatility, the models parameters become unstable and its predictions become unreliable. However, many real-world time series, especially those in economics and finance, are often non-stationary. This makes it essential to diagnose and address non-stationarity before attempting modeling and forecasting in order to produce reliable estimates.

Unit root tests play a central role in this process. A time series has a unit root when its autoregressive coefficient equals one. This implies that shocks have permanent effects and the process does not necessarily follow a deterministic trend. If a series has a unit root, then it is *difference stationary* meaning the difference must be taken in order to achieve stationarity depending on the order (i.e the presence of one unit root indicates the process is first-difference stationary). Statistical tests such as the Augmented Dickey–Fuller (ADF), Phillips–Perron, and KPSS test evaluate whether a series is difference stationary. In contrast, trend-stationary processes may not have a unit root, but still show non-stationarity. When trend-stationary processes experience a shock, the mean will revert back to the unaffected growing mean, whereas a shock to a unit root process will have a permanent impact on the mean. This difference is important to consider when applying transformations to a non-stationary process since differencing a trend-stationary process is less efficient than de-trending it.

Achieving a stationary series is also crucial when analyzing relationships among multiple time series. Non-stationary series that share similar stochastic trends can exhibit spurious correlations, where high correlation arises solely from the shared stochastic trends rather than any meaningful economic relationship. Granger & Newbold (1974) demonstrated that autocorrelations in the residuals, a sign of non-stationarity, produced spurious correlations and high R^2 values, recommending that methods such as first-differencing be applied to achieve stationarity before producing forecasts. A popular modern example is to consider the time series of ice cream sales and average drownings, which naturally share seasonal peaks in the summertime. If you were to correlate the series, you’ll find apparently strong relationships, even though the two are unrelated apart from their seasonality, leading you to infer a causal relationship. Detecting the non-stationarity through unit root tests and differencing or removing the trend through detrending would help reveal the true variability.

Theoretical Background

This section summarizes the principal methods used in the literature to detect and address non-stationarity in time series data. It outlines key assumptions underlying each method and compares their respective strengths and limitations.

Methods for detecting non-stationarity

Visual inspections in the EDA

Exploratory data analysis is a natural starting point for assessing stationarity since many common forms of non-stationarity can be detected through visual inspection of plots. Visualizations of the raw time series may reveal apparent trends, structural breaks, seasonal patterns, or heteroskedastic variance. Trends look like persistent upward or downward movements without clear reversion to the mean, while seasonality looks like recurring fluctuations at regular intervals (for example, monthly or quarterly). Heteroskedasticity is observed when the magnitude of the variance changes systematically over time.

Autocorrelation function (ACF) plots show the correlation of a time series with itself at different lags and are additional diagnostic tools used to detect non-stationarity. For a stationary series, the ACF shows a rapid decay toward zero, with most correlations falling within the statistical significance bounds after a small number of lags. In contrast, a slowly decaying ACF or repeated significant spikes at regular intervals often indicates non-stationarity through a deterministic trend or seasonality, respectively.

Visual methods are intuitive and easy to implement using modern statistical software. However, they are inherently subjective and may be unreliable when trends are subtle, sample sizes are small, or autocorrelations lie very close to significance thresholds. For these reasons, formal unit root tests are generally required for more rigorous inference.

Unit root tests

Augmented Dickey-Fuller test Dickey & Fuller (1981) introduced a formal statistical framework for detecting unit roots in an autoregressive process. They considered the AR(1) process:

$$Y_t = \alpha + pY_{t-1} + \epsilon_t$$

and used maximum likelihood estimation to test the null hypothesis that $p = 1$, indicating a unit root and therefore non-stationarity, against the alternative hypothesis $p < 1$, indicating a stationary series. The Augmented Dickey-Fuller (ADF) test generalizes this framework by including lagged differences of Y_t to account for higher order serial correlations. The ADF test provided one of the first and most widely used standardized test statistics for detecting a unit root, a critical development for economists at the time. However, the test has low power against alternative hypotheses that are very near to the unit root and is sensitive to the deterministic specification choices of intercept, intercept-only or intercept plus trend. It also performs poorly when there is heteroskedasticity in the residuals or structural breaks.

Phillips-Perron test Phillips & Perron (1988) proposed a unit root test with the same null hypothesis as the ADF test, but allowed for serial correlation and heteroskedasticity in the errors. The PP test applies a correction to the test statistic that uses an estimate of the long-run variance of the errors. This made the PP test more robust to a wide class of models, like ARMA and ARIMA, that have serial correlation and heteroskedasticity in the errors. However, the test does depend on choices made for the bandwidth of the kernel that estimates the long-run variance, which can result in low-performance for the test, especially in finite samples.

KPSS test Kwiatkowski et al. (1991) proposed an alternative test in which the null hypothesis is that a series stationary around a deterministic trend and the alternative is the presence of a unit root. The series is decomposed into the sum of a deterministic trend, a random-walk component, and a stationary error term. They implement a Lagrange multiplier test to find whether the random walk component has zero variance. This reversal of the null hypothesis was motivated by the observation that many macroeconomic series fail to reject the unit root null in ADF and PP tests, potentially due to low power rather than true non-stationarity. The KPSS test therefore provides complementary evidence by directly testing whether stationarity can be rejected.

Failing to reject the KPSS null does not prove that a series is stationary, but it indicates that there is insufficient evidence against it. The KPSS test provides a different approach to testing for stationarity and is especially useful when combined with ADF or PP tests to distinguish between difference-stationary, trend-stationary, and ambiguous cases. However, the KPSS test is particularly sensitive to structural breaks and not always robust to small samples, which can lead to spurious rejection of stationarity.

Methods for dealing with non-stationarity

Differencing Differencing transforms a non-stationary series into a stationary one by computing changes between consecutive observations (Hyndman & Athanasopoulos (2018)). Seasonal differencing applies the same principle across seasonal cycles, such as taking differences between observations twelve months apart for monthly data. In practice, the objective is to use the minimum number of differencing required to achieve stationarity. Over-differencing could introduce unnecessary moving-average dynamics, inflate forecast variance, and destroys meaningful long-run relationships among variables by removing too much information. Most economic time series are integrated of order zero or one, and it is rare for higher orders of integration to be required.

De-trending When non-stationarity comes from a deterministic trend rather than a stochastic unit root, detrending may be more appropriate and efficient than differencing. This involves estimating and removing a deterministic trend component so that the remaining series is stationary.

Transformations Transformations to stabilize the variance, usually by taking logarithms, are frequently applied to address heteroskedasticity and exponential growth patterns in time series data (Hyndman & Athanasopoulos (2018)). These transformations can improve the reliability of both the visual diagnostics and unit root tests by making the variance approximately constant over time.

Simulation Study

In their landmark paper on spurious correlations, Granger & Newbold (1974) demonstrate that two independent non-stationary time series, when regressed on each other using OLS, can exhibit highly significant t-statistics and inflated R^2 values, despite being unrelated. This phenomenon arises because standard regression assumptions are violated when the underlying series contain unit roots.

We replicate their simulation to show how two independent random walks can generate misleading regression results. We also extend their original analysis by applying modern unit root tests which were developed after the original paper and allow us to formally diagnose non-stationarity.

Let Y_t and X_t be two independent random walks of length $T = 100$:

$$Y_t = Y_{t-1} + W_t, W_t \sim N(0, 1)$$

$$X_t = X_{t-1} + U_t, U_t \sim N(0, 1)$$

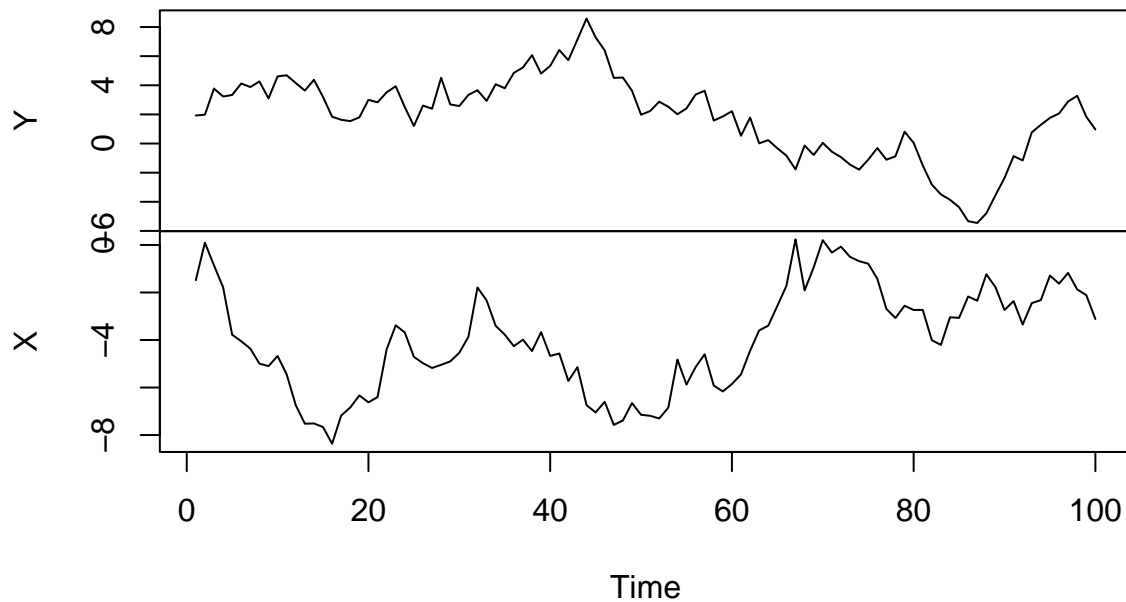
where W_t and U_t are mutually independent Gaussian white-noise error terms.

```
set.seed(12)

# Simulate two random walks with 100 time steps
N <- 100

X <- cumsum(rnorm(N))
Y <- cumsum(rnorm(N))
```

Two Independent Random Walks: X and Y



Despite the fact that we know the two series are independent, regressing Y_t on X_t produces a highly significant coefficient (p-value < 0.001) and an inflated R^2 value, suggesting a strong relationship between the two series.

Spurious Correlation between X and Y

```
=====
                Dependent variable:
                -----
                        Y
-----
X                      -0.677***
                        (0.113)

Constant                -0.836
                        (0.510)

-----
Observations              100
R2                        0.268
Adjusted R2              0.261
Residual Std. Error      2.451 (df = 98)
```

F Statistic 35.940*** (df = 1; 98)
=====

Note: *p<0.1; **p<0.05; ***p<0.01

Consistent with Granger & Newbold (1974), our Monte Carlo simulations indicate a Type I error rate exceeding 70%, meaning that the null hypothesis of no relationship is falsely rejected in more than two-thirds of replications. Moreover, the average R^2 remains high across simulations, further reinforcing that the non-stationarity of two series can lead to false relationships where none exist.

Simulations	Error_Rate	Average_R_Squared
1000	0.763	0.242012

If unit root tests are applied prior to estimation, both series are found to be non-stationary. The ADF test fails to reject the null hypothesis of a unit root for both X_t and Y_t , while the KPSS test strongly rejects the null hypothesis of stationarity around a deterministic trend ($p < 0.05$), which is aligned with the generation of the data as a random walk. It also weakly rejects at the 10% level the null hypothesis of a unit root. The results of these tests provide strong evidence that both series are non-stationary and do not have a deterministic trend, meaning they require differencing before attempting modeling and forecasting.

Series	ADF_Statistic	ADF_P_Value	Method
X	-1.6428	0.7246	Augmented Dickey-Fuller Test
Y	-1.5093	0.7799	Augmented Dickey-Fuller Test

Series	KPSS_Statistic	KPSS_P_Value	Method
X	0.4519	0.0548	KPSS Test for Level Stationarity
Y	0.4327	0.0631	KPSS Test for Level Stationarity
X	0.2931	0.0100	KPSS Test for Trend Stationarity
Y	0.3506	0.0100	KPSS Test for Trend Stationarity

After taking the first differences of both series, we no longer see a significant p-value and the R^2 decreases dramatically towards zero, indicating that there is not sufficient evidence of correlation between both of these series and that the previous regression produced a spurious correlation.

Differenced Regression of X and Y

Dependent variable:	
diff(Y)	
diff(X)	0.023 (0.093)
Constant	-0.106 (0.091)
Observations	99
R2	0.001

Adjusted R2	-0.010
Residual Std. Error	0.907 (df = 97)
F Statistic	0.058 (df = 1; 97)
=====	
Note:	*p<0.1; **p<0.05; ***p<0.01

This simulation illustrates how spurious regression arises shared stochastic trends rather than genuine economic or statistical relationships. Without looking for non-stationarity and testing for unit roots, we could wrongly assume strong predictive or causal relationships where they do not actually exist.

Practical Case Study

In this case study, we examine annual constant Gross Domestic Product (GDP) data from the International Monetary Fund (IMF) to assess whether national GDP time series are stationary. We apply both the visual diagnostic techniques as described above and formal unit root tests to evaluate the time-series properties of GDP, and we assess the implications of these findings for transformation choices and forecasting.

The data are drawn from the IMF World Economic Outlook (WEO) database (*World Economic Outlook (WEO)* (2025)) and consist of annual nominal constant GDP values, which have been adjusted to account for inflation, from 196 countries between 1980-2025. The figures are reported in the domestic currency in billions.

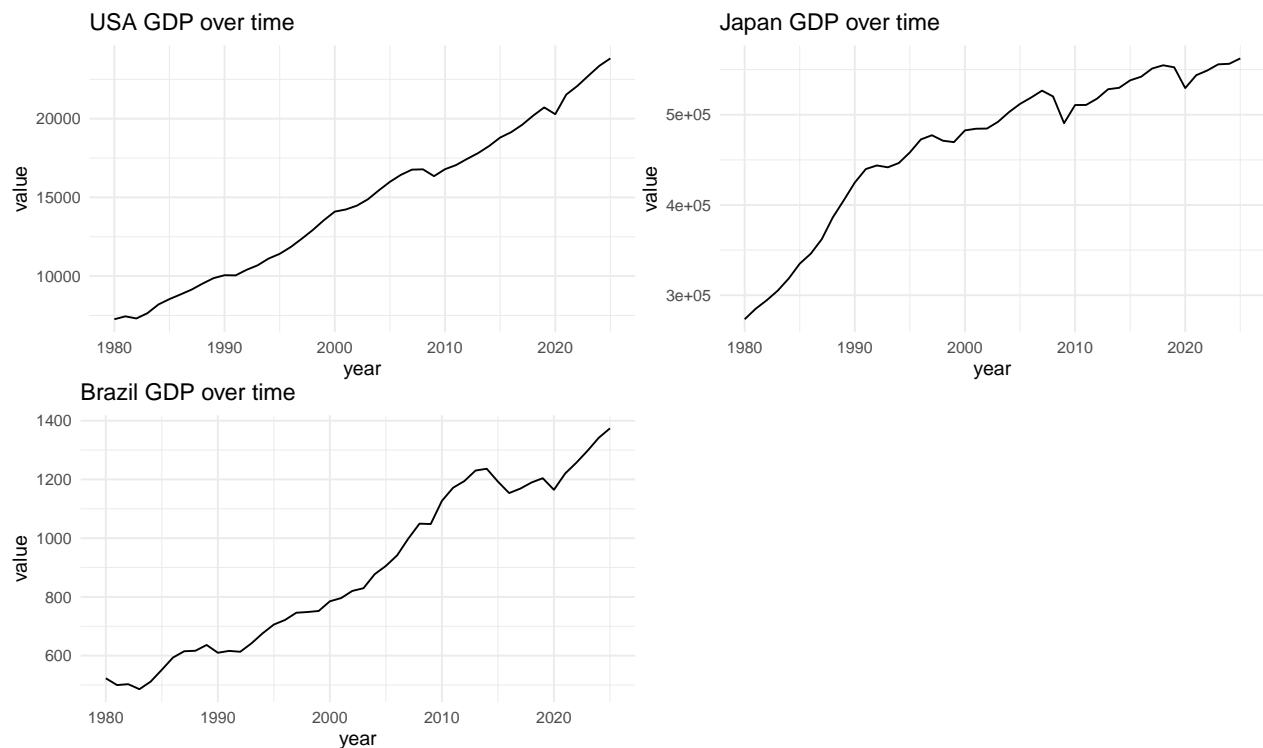
To illustrate the concept of stationarity and the behavior of unit roots and trends across different macroeconomic contexts, this case study will focus on three countries: USA, Japan and Brazil. The United States serves as a benchmark advanced economy with a deterministic trend and multiple shocks, including the Global Financial Crisis and the COVID-19 recession. Japan represents an economy characterized by stagnated growth since the early 1990s, which provides an example of near unit root behavior and low test power. Brazil represents a high-volatility emerging economy that experienced extreme inflation and a major stabilization shift with the Real Plan in 1994, making it a good case of unit root misclassifications due to structural breaks. Together, these three countries provide a comprehensive case study for evaluating the importance of stationarity and unit root tests, as well as the limitations and reliability of the tests under varying circumstances.

Exploratory Data Analysis

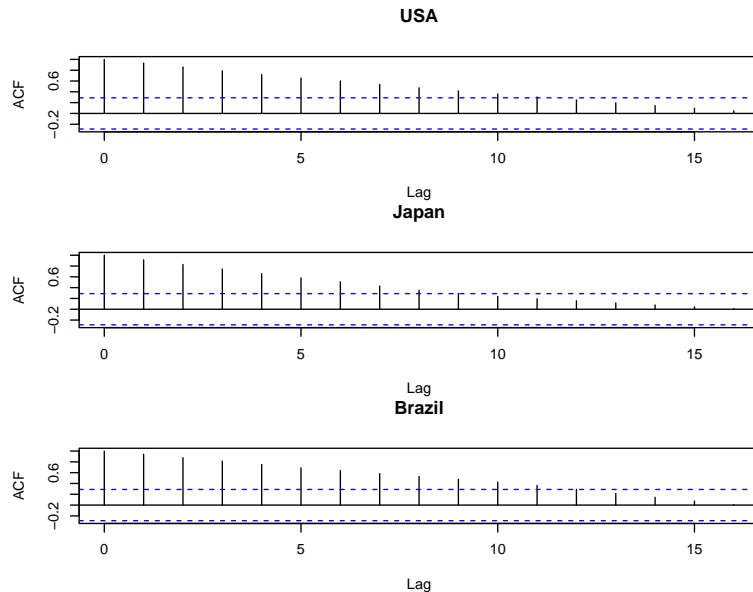
It's important to begin by exploring our data visually to see if there are any key indicators of non-stationarity, such as deterministic trends, heteroskedastic variance, structural breaks or seasonal elements. Visual inspection will provide us with insight into whether a series may require transformations, differencing, or de-trending prior to formal modeling and forecasting.

In the time series plots, the United States shows a strong and persistent upward trend in GDP over the sample period, indicating long-run economic growth. There are two significant declines that are visible around 2008, corresponding to the Global Financial Crisis, and 2020, corresponding to the COVID-19 pandemic, indicating major structural shocks to the underlying trend. Japan's GDP shows rapid growth leading up into 1990, followed by a prolonged period of stagnation with only modest increases afterwards. Brazil's GDP appears relatively stagnant at low levels prior to the early 1990s, followed by a sustained growth phase, and then a noticeable flattening around 2010 before sharply increasing again in 2020, suggesting multiple structural changes in the distribution throughout the sample.

Each of these visual patterns suggest that all three series may be non-stationary. These observations indicate that we should apply formal unit root tests in the analysis to distinguish between difference-stationary and trend-stationary dynamics, and assess what types of transformations are required before modeling and forecasting.



Visual inspection of the ACF plots for all three countries shows a slow decay in the autocorrelation across many lags. This pattern indicates either a unit root process or a strongly trending series and, in agreement with the above time series plots, suggests that the GDP series are likely non-stationary.



Unit Root Tests

Now that we have visually inspected our data and found evidence of possible non-stationarity, the next step is to apply formal unit root tests to see if our speculations are supported. Looking at the results of the unit root tests for the USA GDP data, both the ADF and Phillips-Perron tests produce insignificant p-values, failing to reject the null hypothesis of a unit root. The KPSS level-stationary test, in agreement with the

ADF and PP tests, rejects the null hypothesis of weak stationarity. However, the KPSS trend-stationary test fails to reject the null hypothesis of trend-stationarity, suggesting that the non-stationarity in the USA data is due to a deterministic trend. This tells us that applying a de-trending method could be more efficient than first-differencing.

Table 4: USA Unit Root and Stationarity Tests

	Test	Statistic	P_Value	H0
Dickey-Fuller	ADF	-1.4613697	0.7881234	Unit root (nonstationary)
Dickey-Fuller Z(alpha)	Phillips-Perron	-8.3447886	0.6110551	Unit root (nonstationary)
KPSS Level	KPSS (Level)	1.2461842	0.0100000	Stationary (Level)
KPSS Trend	KPSS (Trend)	0.1127233	0.1000000	Stationary (Trend)

Now taking a look at the unit root tests for Japan's GDP, we see some similar results. The ADF test fails to reject the null hypothesis of a unit root, as does the Phillips-Perron. The KPSS level test also supports these findings by rejecting the null hypothesis of weak stationarity. However, the KPSS trend test rejects the null hypothesis of trend-stationarity as well. This suggests that Japan's GDP is not stationary around a deterministic trend, suggesting that shocks to the series will be persistent throughout time. The combination of these test results suggests that Japan's GDP is a near-unit-root series, rather than a clearly trend-stationary or difference-stationary series. This result highlights the difficulty that can arise when trying to distinguish between true unit roots and highly persistent stationary dynamics, an issue that was highlighted in the paper by Kwiatkowski et al. (1991), where they discussed issues with standard unit root tests failing to reject the null if the process is highly persistent over time.

Table 5: Japan Unit Root and Stationarity Tests

	Test	Statistic	P_Value	H0
Dickey-Fuller	ADF	-2.586601	0.3407153	Unit root (nonstationary)
Dickey-Fuller Z(alpha)	Phillips-Perron	-3.704918	0.8996941	Unit root (nonstationary)
KPSS Level	KPSS (Level)	1.115804	0.0100000	Stationary (Level)
KPSS Trend	KPSS (Trend)	0.272580	0.0100000	Stationary (Trend)

Finally, looking at the results from the unit root tests for Brazil's GDP, we see that the ADF and PP tests again fail to reject the null hypothesis of a unit root. The KPSS level test rejects the null hypothesis of stationarity, also indicating the presence of a unit root. However, the KPSS trend test produces a p-value > 0.05 but less than 0.10. This suggests that Brazil's GDP *could* be trend stationary, but the evidence is weak. This is likely due to the structural shocks that we observed earlier in the data for Brazil, which can produce a series that looks like it has trending behavior, but it is not well captured from a deterministic trend. Recall from the theoretical background section that KPSS tests can be sensitive to structural breaks, leading sometimes to spurious rejections of the null hypothesis.

Table 6: Brazil Unit Root and Stationarity Tests

	Test	Statistic	P_Value	H0
Dickey-Fuller	ADF	-2.324430	0.4449583	Unit root (nonstationary)
Dickey-Fuller Z(alpha)	Phillips-Perron	-9.800383	0.5205049	Unit root (nonstationary)
KPSS Level	KPSS (Level)	1.230161	0.0100000	Stationary (Level)
KPSS Trend	KPSS (Trend)	0.123278	0.0920777	Stationary (Trend)

Implications for Transformations and Modeling

The unit root test results have direct and important implications for how each GDP series should be transformed prior to econometric modeling and forecasting. Because ARMA and ARIMA-type models require weak stationarity, failing to appropriately transform non-stationary series would lead to unstable parameter estimates, spurious relationships, and unreliable forecasts.

For the United States, the KPSS trend-stationary test tells us that de-trending the series or including a time trend explicitly in the modeling equation, would be an appropriate transformation. This transformation preserves the long-run growth information while ensuring weak stationarity in the residuals. For Japan's GDP series, first differencing is required to achieve stationarity before estimation. The results of the KPSS tests suggest that it may also be useful to create an ARIMA model based on growth rates rather than GDP levels, ensuring that forecasts reflect stochastic rather than deterministic long-run dynamics. For Brazil, simple detrending may be insufficient, while differencing risks discarding meaningful dynamics that result from the different structures in the series. A practical approach would involve first differencing combined with modeling the structural breaks. Future work could implement break-robust unit root tests to further improve the model's accuracy.

From a forecasting perspective, misclassifying these integration properties would have serious consequences. Treating a difference-stationary series as trend-stationary would lead to systematically biased forecasts and underestimated uncertainty, while treating a trend-stationary series as difference-stationary would unnecessarily eliminate valuable long-run information. The appropriate transformations identified here by examining the stationarity of the series and implementing unit root tests ensure that forecast models produce statistically valid estimates.

Conclusions and Future Directions

This report demonstrates that stationarity and unit root testing are not just diagnostic formalities, but foundational steps that determine valid modeling strategies, inference, and forecasting performance. Through theoretical discussion, Monte Carlo simulation, and an applied macroeconomic case study using IMF GDP data, we showed how non-stationarity produces spurious regression relationships, inflates goodness-of-fit measures, and undermines statistical validity when left unaddressed. The visual inspection and test results for the United States, Japan, and Brazil further illustrate the practical challenges of distinguishing difference-stationary, trend-stationary, and near-unit-root processes, especially in the presence of structural breaks. Together, these findings reinforce the necessity of combining visual diagnostics, complementary unit root tests, and economic context when analyzing real-world time series. Future work should incorporate break-robust testing procedures and explore cointegration among integrated macroeconomic variables to preserve long-run equilibrium relationships while maintaining statistical validity.

References

- Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4), 1057–1072. <https://doi.org/10.2307/1912517>
- Granger, & Newbold. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2, 111–120. [https://doi.org/10.1016/0304-4076\(74\)90034-7](https://doi.org/10.1016/0304-4076(74)90034-7)
- Hyndman, & Athanasopoulos. (2018). ARIMA models. In *Forecasting: Principles and practice* (3rd ed., pp. Chapter 9). OTexts.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1991). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54, 159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. <https://doi.org/10.2307/2336182>

World economic outlook (WEO). (2025). Accessed December 8, 2025; International Monetary Fund. [https://data.imf.org/en/Data-Explorer?datasetUrn=IMF.RES:WEO\(9.0.0\)%7B.uri%7D](https://data.imf.org/en/Data-Explorer?datasetUrn=IMF.RES:WEO(9.0.0)%7B.uri%7D)