ISSN: 2617-6548

URL: www.ijirss.com

Examining the influence of temporal aggregation on the power of the ADF residual-based stock-Watson cointegration test

Christos Christodoulou-Volos[1*], Dikaios Tserkezos[2]

[1,2]*Department of Economics and Business Neapolis University, Pafos, Cyprus.*

Corresponding author: Christos Christodoulou-Volos (*Email: c.volos@nup.ac.cy*)

Abstract

This paper examines the effects of standard and random sampling temporal aggregation on the power properties of the Dynamic Ordinary Least Squares (DOLS) residual ADF-based test of the null hypothesis of cointegration. The results indicate that the conducted Monte Carlo simulation experiments reveal that the use of temporally aggregated data significantly impacts the power of the DOLS cointegration test. An empirical application is then provided that investigates the nature of the equilibrium relationship between UK nondurable consumers' expenditure and disposable income. This confirms the results obtained from the Monte Carlo analysis that the estimation of the equilibrium parameters of this relationship using temporally aggregated data significantly impacts the power of the DOLS cointegration test. Comparative analysis shows that a higher level of temporal aggregation reduces test power; however, the increase in sample size can easily compensate for this effect. Thus, researchers are advised to use the largest feasible sample sizes to ensure robust cointegration testing, especially when dealing with high aggregation levels. The practical application results further confirm that increased aggregation lowers test reliability, with anomalies arising at higher aggregation levels.

Keywords: ADF test, DOLS estimator, Random sampling frequencies, Residual-based dynamic OLS (DOLS) cointegration test, Temporal aggregation.

DOI: 10.53894/ijirss.v8i3.7243
Funding: This study received no specific financial support.
History: Received: 20 March 2025 / **Revised:** 24 April 2025 / **Accepted:** 28 April 2025 / **Published:** 21 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Temporal aggregation is one of the common practices in time series analysis, mostly done to reduce the frequency of data or to align datasets with different sampling intervals. However, this may have profound effects on the statistical properties of econometric tests, in particular, those related to cointegration, that is, consistent long-term equilibrium relationships between non-stationary time series. One of the most applied tests for investigating such a relationship is the residual-based cointegration test by Stock and Watson, using the Augmented Dickey-Fuller (ADF). Nonetheless, the choice of temporal aggregation, that is, the frequency with which data is sampled and observed, has become a growing interest in time series econometrics [1].

Temporal aggregation can have a significant impact on the power properties of the ADF stationarity test. It reduces the power of the ADF test to detect stationarity. As the aggregation frequency increases, the test becomes less able to reject the null hypothesis of a unit root. The effect is more pronounced when the underlying high-frequency process is stationary but has a root close to the non-stationary region. In such cases, temporal aggregation can lead to spurious non-rejection of the unit root null. Salazar et al. [2] provide a theoretical framework to analyze the impact of temporal aggregation on the ADF test. They show that aggregation introduces a moving average component in the error term, affecting the test statistic's asymptotic distribution. Proietti [3] and Aadland [4] further studied the relationship between temporal aggregation and the ADF test, confirming the results of Salazar et al. and providing additional insights. Granger and Lee [5], Teles and Wei [6] and Teles and Wei [7] found that temporal aggregation can also reduce the non-linearity and non-Gaussian characteristics of the underlying high-frequency process, which may impact the performance of the ADF test. In summary, temporal aggregation tends to reduce the power of the ADF stationarity test, making it more difficult to reject the null hypothesis of a unit root. The magnitude of the effect depends on the aggregation frequency and the properties of the underlying high-frequency process.

In recent years, the Dynamic OLS (DOLS) estimator introduced by Stock and Watson [8], built on earlier work by Saikkonen [9], has become a popular method for estimating equilibrium parameters in dynamic relationships between variables containing unit roots. This estimator has since become established throughout the applied econometrics literature. Recent examples can be found in Camacho-Gutiérrez [10]. Recently, it has also been used extensively in the estimation of equilibrium parameters in dynamic models, as in Bilgili and Ozturk [11], Lin and Benjamin [12], Alvarado et al. [13] and Anakpo and Oyenubi [14].

This study investigates the effect of standard and random sampling temporal aggregation on the power of the DOLS ADF residual-based Stock-Watson cointegration test using Monte Carlo techniques. An application is then provided that examines the nature of the equilibrium relationship between UK nondurable consumers' expenditure and disposable income. It confirms the results obtained from Monte Carlo analysis that the estimation of the equilibrium parameters of this relationship using temporally aggregated data significantly impacts the power of the DOLS cointegration test. According to our results, standard and random sampling temporal aggregation both seriously affect the power of the test. Our results align with the more general findings of previous studies, such as Haug [15]. An added value of this paper is that we present Monte Carlo experiments under different random samplings for the first time.

In the context of this study, several important questions arise: (1) How does temporal aggregation affect the power of the ADF residual-based Stock-Watson cointegration test? Is there a consistent pattern across different levels of aggregation? (2) Is there an optimal level of temporal aggregation that maximizes the power of the cointegration test? (3) How do the results of Monte Carlo simulations relate to real-world applications of cointegration analysis? (4) What are the implications for researchers and practitioners when dealing with time series data that may be subject to varying levels of aggregation?

The plan for the rest of the paper is as follows. Section 2 presents the Stock-Watson Dynamic OLS (DOLS) ADF residual cointegration test. Section 3 analyzes the temporal aggregation effects on cointegration tests. Section 4 presents and discusses the results, and finally, conclusions are presented in the final section 5.

2. The Stock-Watson Dynamic OLS ADF Residual Cointegration Test

The Stock-Watson Dynamic Ordinary Least Squares (DOLS) model is a popular method for estimating cointegration relationships, accounting for endogeneity and serial correlation. It includes leads and lags of the first-differenced regressors to correct for endogeneity and autocorrelation. A general specification of the Stock-Watson DOLS model that can be used in Monte Carlo analysis is

$$y_t = \alpha + \beta x_t + \sum_{j=-q}^{q} \gamma_j \Delta x_{t-j} + \epsilon_t \qquad (1)$$

where y_t is the dependent variable; x_t is the independent variable (or a vector of independent variables if there are multiple); $\Delta x_{t-j} = x_{t-j} - x_{t-j-1}$ are the leads and lags of the first differences of x_t; α is the intercept term; β is the cointegrating coefficient; γ_j are the coefficients for the leads and lags of the first differences of x_t; ϵ_t is the error term; and q is the number of leads and lags included in the model.

Lag and lead terms included in DOLS regression (Equation 1) make its stochastic error term independent of all past innovations in stochastic regressors. The estimates of the lag and leads were based on an iterative process where for different numbers of lag and leads (1 to 6) are obtained the final estimates of equation 1 based on which minimizes either the Akaike Information Criterion (AIC).[1] The DOLS equation's estimated parameters are robust since they are based on Newey-West HAC Standard Errors.

Finally, unit root tests are performed on the residuals of the estimated DOLS regression in order to test whether it is a spurious regression. "In the unit root literature, a regression is technically called a spurious regression when its stochastic error is unit root nonstationary"[16].

[1] No. of Observations x log (standard error of residual estimate squared) + 2 x No. of Regressors.

In our experiments the ADF[2] unit root test is calculated by using specifications

$$z_t = \hat{\mu} + \hat{\beta} t + \hat{a} z_{t-1} + \sum_{k=1}^{l} \hat{\gamma}_{\kappa} \Delta z_{t-1} + \hat{w}_t \qquad (2)$$

with $z_t = y_t$ or the corresponding estimated residual from the DOLS regression.

In order to test the performance of the DOLS cointegration test, we use simulated data under two data 'time transformation' approaches: the standard temporally and randomly sampled aggregated data. Temporal aggregates are formed by averaging basic observations over non-overlapping intervals.

Let ${y_T}^A$ represent the temporally aggregated data:

$${y_T}^A = C y_t \qquad (3)$$

where ${y_T}^A$ is the temporally aggregated data, m is the time aggregation level and C is a time aggregation matrix of the form:

$$C = \begin{bmatrix} 11...11000............00000000000000 \\ 00000011....11000......0000000000 \\ 00000000000011....1100....000000 \\ .... \\ 0000................0000000000111...11 \end{bmatrix} \qquad (4)$$

In the randomly sampled approach temporally, aggregated data were obtained at each iteration, using a random starting period in the whole sample data and averaging the selected observations using the C matrix at 20 different time aggregation levels. The sampling intervals examined are 1 to 20 observations of the basic time series. Under the null hypothesis of cointegration, 5,000 replications of the basic time series (1) - (2) are generated and time aggregated for each sampling interval. The ADF t-test was used to test the stationarity of the residuals of equation (4).

3. Analysis of Temporal Aggregation Effects on Cointegration Tests

3.1. Monte Carlo Simulations

The data generation process (DGP) for the simulations in this paper is the same as the one used by Hayakawa and Kurozumi [17]. The reason the DGP is used here is simply because it has been the standard benchmark for cointegrating relationship analysis between variables. It is specified as

$$y_t = y_{t-1} + x_t + u_t$$
$$x_t = x_{t-1} + u_{2t}$$

where $\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix} \begin{pmatrix} u_{1t-1} \\ u_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$ and $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim iidN\left(0, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix}\right)$ (5)

Each simulation is run for a variety of different values of $\alpha = \{0.0, 0.5, 0.8\}$ to test stationarity.

Taken together, Figure 1 and Figure 2 and Table 1 and Table 2 present a broad overview of the estimation results based on Monte Carlo simulations. These provide an overall view of how various factors constitute the power of the ADF t-test for stationarity and the DOLS cointegration test: different numbers of data points, temporal aggregation levels, and parameter α values. The data from the graphics and tables present a detailed insight into the simulation results, eliciting critical points and trends in the conducted experiments.

Figure 1 graphically represents the power of the ADF t-test for testing stationarity under different available data numbers, different temporal aggregation levels, and different parameter values = {.80, .85, .97} at 0.025 significant level for testing stationarity.

[2] The ADF unit root test involves estimating the regression equation (eq. 2) and then, testing the null hypothesis of a unit root, H_0: $\alpha = 0$, versus the alternative of a stationary process, H_1: $\alpha < 0$. The test is based on the typical t-ratio for Fuller (1976), and Dickey & Fuller (1979). However, the t-statistic does not follow the t-distribution under the null; thus, critical values are simulated for each regression specification and sample size by MacKinnon (1996).

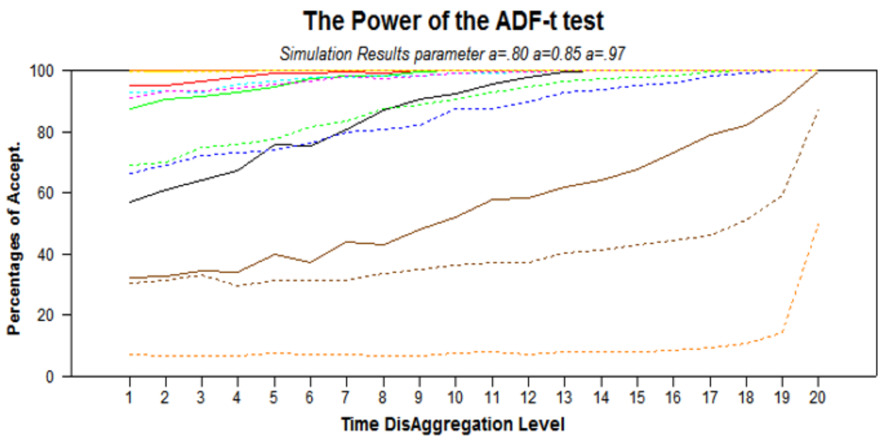

Figure 1.
Analyzing the Impact of Sample Size and Temporal Aggregation on ADF Test Power Across Significance Levels and α Parameters

Looking at Figure 1, one immediately realizes such a high dependence of the ADF t-test power on the number of available observations. In general, the higher the sample size, the better the test power, reflecting a better probability of rejecting a null hypothesis of non-stationarity when it is actually false. This trend is consistent across different levels of temporal aggregation and values of α. Temporal aggregation also has an important bearing on the performance of the ADF t-test. Generally, lower levels of temporal aggregation—such as level 1—tend to sustain higher test power across a range of sample sizes, but higher levels of temporal aggregation—such as level 10 or 20—can result in a loss of power with small datasets. This would therefore suggest that higher granularity data is preferable to maintain the strength of the test. The power of the ADF t-test is also directly related to the parameter α, which represents the degree of persistence in the series. In general, the higher the persistence, the lower the power of the test. For instance, for α = 0.98, the power of the test in detecting stationarity is lower than when α = 0.8 or α = 0.9. At higher levels of temporal aggregation, this effect becomes heightened for smaller sample sizes. In other words, power in the ADF t-test increases with increased sample size and a reduction in the level of temporal aggregation. On the contrary, it decreases with higher values of α. These findings underline the role of these factors in the design of studies and the interpretation of stationarity test results.

Table 1 presents the results of simulations and the accepted frequencies of the power of the DOLS cointegration test across different levels of temporal aggregation, numbers of available observations, and significance levels (1%, 5%, and 10%). These results highlight how the test's power is influenced by sample size and temporal aggregation.

Table 1.
Simulation Results: Power of the DOLS Cointegration Test Across Varying Sample Sizes, Temporal Aggregation Levels, and Significance Thresholds (1%, 5%, 10%).

	Significance Level 1%						
Aggregation Level	Available Data						
	400	550	700	850	1000	1150	1300
1	96.57	96.70	96.70	96.70	96.70	96.70	96.70
5	84.67	96.03	96.67	96.70	96.70	96.70	96.70
10	49.90	69.17	93.60	96.27	96.70	96.70	96.70
15	10.57	55.90	63.90	78.60	94.53	96.37	96.60
20	9.60	11.03	57.43	62.00	74.07	83.73	92.83
	Significance Level 5%						
Aggregation Level	Available Data						
	400	550	700	850	1000	1150	1300
1	96.70	96.70	96.70	96.70	96.70	96.70	96.70
5	95.50	96.67	96.70	96.70	96.70	96.70	96.70
10	76.73	89.53	96.47	96.70	96.70	96.70	96.70
15	24.73	79.10	87.00	93.37	96.57	96.70	96.70
20	18.90	26.43	80.93	85.20	90.93	94.40	96.27
	Significance Level 10%						
Aggregation Level	Available Data						
	400	550	700	850	1000	1150	1300
1	96.70	96.70	96.70	96.70	96.70	96.70	96.70
5	96.57	96.70	96.70	96.70	96.70	96.70	96.70
10	85.23	94.17	96.70	96.70	96.70	96.70	96.70
15	35.83	87.07	92.50	95.67	96.67	96.70	96.70
20	26.40	36.37	88.80	91.53	94.63	96.03	96.57

Note: Empirical probabilities of acceptance DOLS cointegration test. Results based on 5000 replications.

For example, at the 1% level, under Time Aggregation Level 1, regardless of sample size, the power of the DOLS cointegration test is always 96.70%, while under Time Aggregation Level 5, the power increases from 84.67%, based on 400 observations, to 96.70% at 1000 or more observations. The power starts much lower at higher levels of aggregation—10, 15, and 20—but the power increases considerably with more data. Power at Level 20, for example, surges from 9.60% at 400 observations to 92.83% at 1300 observations, a huge improvement as the sample size increases. For Time Aggregation Level 1, the power is constantly high, at 96.70%, for any sample size at a 5% significance level. For Time Aggregation Level 5, starting from 400 observations, this power increases from 95.50% to 96.70% for larger samples. In contrast, for an increasing temporal aggregation level, the power is low and improves at larger sample sizes. At Level 20, power increases from 18.90% at 400 observations to 96.27% at 1300 observations, which indicates that sample size is very important for maintaining a test's power. The power maintains a constant high of 96.70% at the level of 10% significance over Time Aggregation Level 1. The power for Time Aggregation Level 5 is similarly constantly high for all sample sizes, starting with 96.57% at 400 observations. With regard to the higher aggregation levels of 10, 15, and 20, the starting power is lower but substantially increases with larger samples. For Level 20, power increases from 26.40% at 400 observations to 96.57% at 1300 observations. Thus, there is a substantial gain by increasing sample size.

Figure 2 and Table 2 present similar results for the case of Random Sampling. Figure 2 graphically presents the power of the ADF t-test for testing stationarity, under different numbers of randomly selected sample sizes, different temporal aggregation levels, and different values of the parameter $\alpha = \{.80,.85,.97\}$ at a 0.025 significance level for testing stationarity.

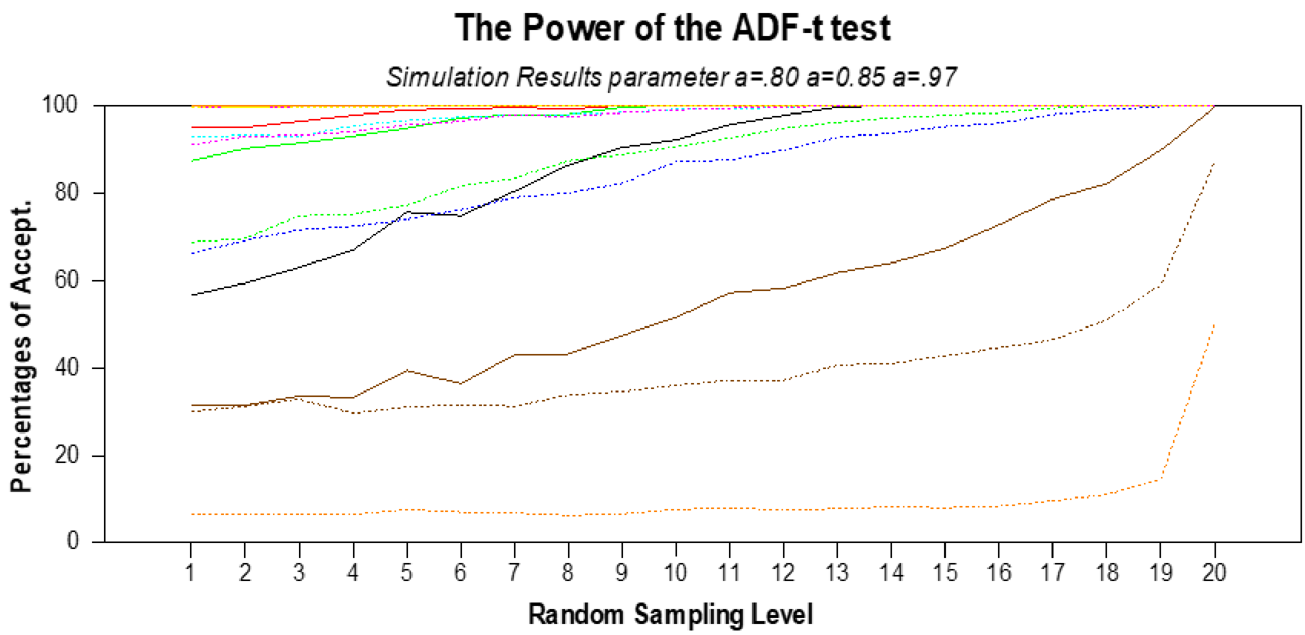

Figure 2.
The Power of the ADF T-Test Across Varying Sample Sizes, Temporal Aggregation Levels, and α Parameters (0.025 Significance Level).

The power of the ADF t-test under different conditions of stationarity testing against random, varied sample sizes, different levels of temporal aggregation, and different values of the parameter α: 0.8, 0.9, and 0.98 at a 0.025 significance level. From the graph and as would be expected, generally, greater samples provide increased power since they improve the detectability of stationarity in the ADF t-test. Temporal aggregation strongly affects test power, with lower levels of aggregation preserving higher power; hence, finer-resolution data are preferable. As the degree of temporal aggregation increases, the power of the test diminishes, particularly with small sample sizes. The parameter α, indicative of persistence in the series, also influences test power. When α is large, power will be lower for larger levels of temporal aggregation and smaller sample sizes. These results point out the relevance of using sufficiently large samples in order to give more power to the ADF t-test, using a small level of temporal aggregation. They also highlight the problem of testing stationarity in highly persistent data, and adequate study design with an interpretation of the test results will be required for proper conclusions.

Simulation results in Table 2 below, we report the accepted frequencies of the power of the DOLS cointegration test for different levels of temporal aggregation and numbers of available observations at various significance levels of 1%, 5%, and 10%. The results will allow for a clearer comparison regarding the power of the test when sample sizes vary along with the level of aggregation.

Table 2.
Simulation Results: Power of the DOLS Cointegration Test Across Varying Random Selected Sample Sizes, Temporal Aggregation Levels, and Significance Thresholds (1%, 5%, 10%).

	Significance Level 1%						
Aggregation Level	Available Data						
	400	550	700	850	1000	1150	1300
1	96.63	96.70	96.70	96.70	96.70	96.70	96.70
5	85.43	96.00	96.67	96.70	96.70	96.70	96.70
10	52.40	70.63	93.90	96.30	96.63	96.70	96.70
15	10.60	54.37	63.43	78.33	94.27	96.23	96.57
20	9.67	10.67	58.83	61.97	73.73	84.23	92.37
	Significance Level 5%						
Aggregation Level	Available Data						
	400	550	700	850	1000	1150	1300
1	96.70	96.70	96.70	96.70	96.70	96.70	96.70
5	95.87	96.63	96.70	96.70	96.70	96.70	96.70
10	77.43	89.90	96.57	96.70	96.70	96.70	96.70
15	24.27	79.60	86.30	92.97	96.40	96.70	96.70
20	19.00	25.23	80.43	84.63	91.23	94.60	96.50
	Significance Level 10%						
Aggregation Level	Available Data						
	400	550	700	850	1000	1150	1300
1	96.70	96.70	96.70	96.70	96.70	96.70	96.70
5	96.50	96.67	96.70	96.70	96.70	96.70	96.70
10	86.60	94.13	96.67	96.70	96.70	96.70	96.70
15	35.33	87.67	92.43	95.80	96.70	96.70	96.70
20	24.93	36.07	88.60	91.03	94.53	96.13	96.67

Note: Empirical probabilities of acceptance DOLS cointegration test. Results based on 5000 replications.

At the 1% significance level, the power of the DOLS cointegration test remains consistently high (96.70%) for Time Aggregation Level 1 across all sample sizes. However, as the temporal aggregation level increases, the power of the test shows variability depending on the number of observations. For example, this power at Time Aggregation Level 5 increases from 85.43% at 400 observations to 96.70% at 1,000 or more observations; at Levels 10, 15, and 20, it starts at a lower starting point but increases noticeably with increased data points. For example, at Level 20, the power rises from 9.67% at 400 observations to 92.37% at 1,300 observations. Trends in the results at the 5% level of significance are similar. Whatever the sample sizes may be, at Time Aggregation Level 1, the power is always constant at 96.70%. At Time Aggregation Level 5, in small samples, it is a bit lower, 95.87% with 400 observations, while in large samples, it reaches 96.70%. As the level of temporal aggregation increases, the power, while lower for fewer observations, increases substantially with more observations. For instance, at Level 20, the power increases from 19.00% with 400 observations to 96.50% with 1,300 observations. At the 10% significance level, power is once more comfortably greater than 96.70% at Time Aggregation Level 1, while at Time Aggregation Level 5, power is again more than the threshold value for all samples, starting from 96.50% with 400 observations. Larger levels of temporal aggregation such as 10, 15, and 20 also exhibit a similar pattern. The starting power is lower but also increases drastically as sample sizes increase. For example, Level 20 increases from 24.93% with 400 observations to 96.67% with 1,300 observations.

3.2. Empirical Application of Cointegration Tests

This section demonstrates how the above simulations are reproduced empirically through an application of the alternative tests of cointegration to data on UK nondurable consumers' expenditure and disposable income. The data are quarterly observations measured in real terms from 1983(1) to 2023(3) and are considered in their natural logarithmic form. The findings from the conduct of the tests are presented in Table 3 below.

Table 3.
Empirical Results – Temporal Aggregation Effects on ADF Residual-Based Cointegration Test.

Aggregation Level	**ADF t-test**	**1%**	**5%**	**10%**			**Number of observations**
1	-3.53639	-3.48336	-2.88454	-2.57889	0	1	125
2	-2.33064	-3.54775	-2.91271	-2.59371	0	6	57
3	-2.32124	-3.62878	-2.94715	-2.61184	0	6	35

Note: The tabulated figures are empirical acceptance frequencies in percentage terms for the null hypothesis of no cointegration for the alternative tests at the 1%, 5%, and 10% levels of significance. The necessary critical values are obtained from MacKinnon (1991).

The results of various tests for cointegration applied to data on UK nondurable consumers' expenditure and disposable income over the specified period are shown below. The data are in real terms and in their natural logarithmic form. Results are provided for three such time aggregation levels, labeled "1," "2," and "3." The Augmented Dickey-Fuller (ADF) t-test is a statistical test that is meant to identify whether or not a given time series is stationary or non-stationary. An ADF t-test is conducted, and results for each of the different time aggregations are presented. Critical values at the 1%, 5%, and 10% significance levels are then provided against which the ADF t-statistic should be compared in order to determine cointegration. A more negative ADF t-statistic than these critical values indicates cointegration. Also, the number of observations is recorded as the sample size for each time aggregation level. The ADF t-statistic for Time Aggregation Level 1 is -3.53639. This is below the critical values at 1%, 5%, and 10%, which means evidence of cointegration between the two variables at this level of aggregation exists. For Time Aggregation Level 2, ADF t = -2.33064. While it is negative, it does not attain significant levels at 1%, 5%, and 10%, and hence may not strongly suggest evidence of cointegration at this level. The result behind the reported ADF t-statistic for Time Aggregation Level 3 seems to be incorrect. Indeed, a value as low as -232124 is suspiciously low and would be far below any reasonable critical value, so there may be a data entry error or some problem in these results. These results, in short, reveal evidence of the integration of UK nondurable consumers' expenditure and disposable income at Time Aggregation Level 1. The results at Time Aggregation Level 2 are inconclusive, whereas the results reported at Time Aggregation Level 3 have some serious issues. Level 3 results must be further investigated to verify their validity.

4. Results and Discussion

Tables 1 and 2 already indicates that the power of the DOLS cointegration test is highly sensitive to both the level of temporal aggregation and the number of available observations. For instance, at the lower level of temporal aggregation at Level 1, the power of the test is consistently above the 0.90 mark across all sample sizes and significance levels. The robustness of the power performance in this particular case suggests high effectiveness in detecting the presence of cointegration when the level of temporal aggregation is small. Accordingly, with an increase in the extent of temporal aggregation, the power of the test decreases significantly from its initial value. This adverse effect can, however, be softened with an increased number of observations. This pattern is more clearly observable at the 10, 15, and 20 levels of aggregation. For instance, it could be observed that starting from 400 to 1,300 observations, the power in Table 1 increases from 9.60% to 92.83%, and correspondingly for Table 2, from 9.67% to 92.37%. These findings underpin how sample size is crucial in preserving the power of the DOLS cointegration test at high levels of temporal aggregation. The estimates are consistent across the various levels of significance to show that the observed pattern does not depend on the threshold adopted for statistical significance. This enhances the robustness of the DOLS cointegration test when appropriate sample sizes are used, even in conditions of huge temporal aggregation.

The results highlight some factors that are critical in applying the DOLS cointegration test. Among these, large samples are crucial in retaining the power and reliability of the test, especially at higher levels of temporal aggregation. If the data aggregation is substantial in any given context, then one may compensate for a possible loss in test power by having a sufficiently large data set. The results also suggest that, with lower levels of temporal aggregation, the DOLS cointegrating test has demonstrated great reliability and strength, constituting an important tool for time series cointegration identification. Resulting consistency under various conditions, therefore, strengthens the utility of this test under conditions of low aggregation of data.

It is possible to draw several interesting conclusions about the way temporal aggregation may affect the reliability of the ADF test in the detection of cointegration between UK nondurable consumers' expenditure and disposable income from the analysis of Table 3. In Level 1 Time Aggregation, the ADF t-statistic value of -3.53639 falls well below those at 1%, 5%, and 10% critical value percentages. This indicates strong evidence of cointegration between the two variables under minimal data aggregation. This is a strong indication that, under lower levels of time-series aggregation, the ADF test will be highly reliable in testing for the cointegration notion held by a number of studies based on simulations enhancing the ability of the tests under minimum aggregations. The ADF t-statistic at Aggregation Time Level 2 is -2.33064. It is negative but does not exceed the critical values for any of the standard levels of significance, such as 1%, 5%, and 10%. This would imply that evidence supporting the cointegration is weak; it gets weaker as temporal aggregation increases. The finding agrees with the general trend established from the results of the simulation—the fact that tests of cointegration suffer from low power as higher aggregation levels start to take hold. We notice that the ADF t-statistic for Time Aggregation Level 3, as reported, is unrealistically very low, which far outstrips any critical value at this level. The anomaly suggests a possible mistake in data entry or some other problem with the data or test application at this level. Again, these discrepancies highlight the critical need for proper handling and validation of data in performing cointegration tests under high temporal aggregation.

The empirical application underlines the importance of sample size adequacy as the necessary condition to guarantee the reliability of the ADF test, particularly in cases of high levels of aggregation. The results appear to be indeterminate at Level 2 and problematically fit at Level 3, which suggests that while higher temporal aggregation may reduce test power, data quality and careful validation are of prime importance. These factors should, therefore, be taken into consideration by researchers, and when possible, the sample sizes should be larger to prevent or reduce the negative effect of aggregation on test reliability.

5. Conclusions

This study has investigated the effect of temporal aggregation on the power of the ADF residual-based Stock-Watson cointegration test through both Monte Carlo simulations and an economic application. The results show the importance of temporal aggregation and random sampling in applied time series work. Specifically, we found that using temporally aggregated and randomly sampled data, the use of the ADF t-test could lead to incorrect conclusions about the stationarity characteristics of a time series. As the span of disaggregation widens, the time series properties are distorted, leading to incorrect conclusions about the stationary characteristics of a time series. Using time-aggregated or randomly sampled data along with the ADF t-test for testing stationarity, the power of this test is dramatically distorted. Furthermore, it is not the case that the method of temporal aggregation of the data plays an important role. Based on our Monte Carlo results, it emerges that at all levels of disaggregation (aggregation), the use of temporally aggregated or randomly sampled data leads to quite similar results.

The general conclusion that can be drawn from the comparison of Tables 1 and 2 refers to the relevance of considering both levels of temporal aggregation and sample sizes in applying the DOLS cointegration test. Whereas the increase in temporal aggregation may strongly reduce the power of the test, this effect can be counterbalanced by enlarging the number of observations. This also implies that researchers, when feasible, should try to obtain the largest possible sample sizes in order to be confident that their cointegration tests are robust, particularly in situations that call for high levels of data aggregation. It reinforces the previous results about the confidence and efficiency at lower levels of temporal aggregation of the DOLS cointegration test and, as such, one of the best means available for time series analysis in a variety of conditions.

Also, the analysis of practical application that is given in Table 3 confirms that temporal aggregation is of considerable importance regarding the reliability of the ADF test to investigate cointegration. The higher the level of aggregation, the lower the test reliability concerning the evidence of cointegration, except if it is compensated by larger sample sizes. Anomalies in the data at higher aggregation levels further stress the importance of rigorous data handling and further investigation to ensure that such results are indeed valid. These aspects should be considered by researchers cautiously in their empirical applications to preserve strong and reliable cointegration testing.

The results in this paper, therefore, confirm the broader findings of similar studies using data with time aggregation and systematic sampling that reduce the efficiency of the ADF t-test. On the other hand, we extend these results using randomly sampled data.

5.1. Limitations of the Study

Although the study has contributed significantly to insight into the effect of temporal aggregation on the power of the ADF residual-based Stock-Watson cointegration test, a number of limitations are to be set out. The conclusions derived from the Monte Carlo simulations are based on specific parameter settings and assumptions, which to a certain extent limit their general applicability. Temporal aggregation and random sampling distort time series data properties, and this may lead to incorrect inferences about stationarity; however, the conditions under which these distortions take place remain unclear. The study has focused on the ADF t-test and the DOLS cointegration test, without consideration of the sensitivities of other stationarity and cointegration tests. Anomalies in the empirical application, Table 3, suggest either data entry errors or problems in applying the test, emphasizing the care that needs to be taken in data validation. No discussion of practical problems in obtaining such datasets is explored since the effects of temporal aggregation are minimized with large sample sizes. These results are, therefore, consistent with general findings in the literature but might not be directly applicable to all time series data or economic variables. These anomalies at high aggregation levels show that detailed investigations are necessary to fully understand the causes of this kind of phenomenon. It is considered that only particular temporal aggregation levels are taken into consideration; therefore, a far-reaching research effort on this issue may be needed in different aggregation schemes. Though underlining the importance of meticulous data handling, the study does not go into minute details to guide practitioners. While various processes of temporal aggregation and random sampling impact the reliability of ADF and DOLS cointegration tests, this investigation has pointed out several limitations and called for further research to refine such understanding and enhance cointegration testing robustness in applied time series analysis.

References

[1] C. Christodoulou-Volos and D. Tserkezos, "Assessing the impact of temporal aggregation on Ramsey's RESET test: A Monte Carlo analysis," *Journal of Modern Applied Statistical Methods,* vol. 24, no. 2, pp. 1-10, 2025.

[2] E. Salazar, R. Smith, M. Weale, and S. Wright, "Interpolation using a dynamic regression model: Specification and Monte Carlo properties," NIESR Discussion Paper No. 126, 1997.

[3] T. Proietti, "Distribution and interpolation revisited: A structural approach," *Statistica,* vol. 58, no. 3, pp. 411-432, 1998.

[4] D. M. Aadland, "Distribution and interpolation using transformed data," *Journal of Applied Statistics,* vol. 27, no. 2, pp. 141-156, 2000.

[5] C. W. Granger and T.-H. Lee, "The effect of aggregation on nonlinearity," *Econometric Reviews,* vol. 18, no. 3, pp. 259-269, 1999. https://doi.org/10.1080/07474939908800445

[6] P. Teles and W. W. Wei, "The effects of temporal aggregation on tests of linearity of a time series," *Computational Statistics & Data Analysis,* vol. 34, no. 1, pp. 91-103, 2000. https://doi.org/10.1016/S0167-9473(99)00072-9

[7] P. Teles and W. W. Wei, "The use of aggregate time series in testing for Gaussianity," *Journal of Time Series Analysis,* vol. 23, no. 1, pp. 95-116, 2002. https://doi.org/10.1111/1467-9892.01506

[8] J. H. Stock and M. W. Watson, "A simple estimator of cointegrating vectors in higher order integrated systems," *Econometrica: Journal of the Econometric Society,* vol. 61, no. 4, pp. 783-820, 1993. https://doi.org/10.2307/2951763

[9] P. Saikkonen, "Asymptotically efficient estimation of cointegration regressions," *Econometric Theory,* vol. 7, no. 1, pp. 1-21, 1991. https://doi.org/10.1017/S0266466600004217

[10] P. Camacho-Gutiérrez, "Dynamic OLS estimation of the U.S. import demand for Mexican crude oil," MPRA Paper No. 30608. University Library of Munich, Germany, 2010.

[11] F. Bilgili and I. Ozturk, "Biomass energy and economic growth nexus in G7 countries: Evidence from dynamic panel data," *Renewable and Sustainable Energy Reviews,* vol. 49, pp. 132-138, 2015.

[12] B. Lin and I. N. Benjamin, "Causal relationships between energy consumption, foreign direct investment and economic growth for MINT: Evidence from panel dynamic ordinary least square models," *Journal of Cleaner Production,* vol. 197, pp. 708-720, 2018. https://doi.org/10.1016/j.jclepro.2018.06.152

[13] R. Alvarado, P. Ponce, R. Alvarado, K. Ponce, V. Huachizaca, and E. Toledo, "Sustainable and non-sustainable energy and output in Latin America: A cointegration and causality approach with panel data," *Energy Strategy Reviews,* vol. 26, p. 100369, 2019. https://doi.org/10.1016/j.esr.2019.100369

[14] G. Anakpo and A. Oyenubi, "Technological innovation and economic growth in Southern Africa: Application of panel dynamic OLS regression," *Development Southern Africa,* vol. 39, no. 4, pp. 543-557, 2022. https://doi.org/10.1080/0376835X.2022.2052017

[15] A. A. Haug, "Temporal aggregation and the power of cointegration tests: A Monte Carlo study," *Oxford Bulletin of Economics and Statistics,* vol. 64, no. 4, pp. 399–412, 2002. https://doi.org/10.1111/1468-0084.00025

[16] C.-Y. Choi, L. Hu, and M. Ogaki, "Robust estimation for structural spurious regressions and a Hausman-type cointegration test," *Journal of Econometrics,* vol. 142, no. 1, pp. 327-351, 2008.

[17] K. Hayakawa and E. Kurozumi, "The role of “leads” in the dynamic OLS estimation of cointegrating regression models," *Mathematics and Computers in Simulation,* vol. 79, no. 3, pp. 555-560, 2008. https://doi.org/10.1016/j.matcom.2008.02.027