1 Data Transformation

The data set consists of daily stock prices for JPMORGAN, APPLE, PEPSICO, and MCDONALDS from 22 August 2000 to 22 August 2024. These prices were aggregated into quarterly averages to create a time series for each company. The data was then transformed into a natural logarithmic scale, with the resulting variables labeled as *lp1q*, *lp2q*, *lpq3*, and *lp4q*, representing the respective log stock prices of JPMORGAN, APPLE, PEPSICO, and MCDONALDS, respectively.

2 Stationarity versus Non-Stationarity of Time Series

2.1 Visual Inspection and Initial Observations:

The analysis begins with examining the level series: *lp1q*, *lp2q*, *lpq3*, and *lp4q*. As shown in **Figure 1: Level Time Series**, all level series exhibit an upward trend, suggesting that the average quarterly log stock prices have increased over time. However, a visible trend does not necessarily imply trend stationarity as this could suggest stochastic non-stationarity – when there is a stochastic trend in the data. To distinguish between trend stationarity and difference stationarity, the time series need to be detrended (**Figure 2: Detrended Time Series**) and differenced (**Figure 3: 1st Order Difference Time Series**).

The residual plot for *lp1q* after detrending (**Figure 2: Detrended Time Series**) shows a similar pattern to the level series, suggesting that *lp1q* may be trend stationary. However, residual plots for *lp2q*, *lp3q* and *lp4q* do not display similar characteristics, indicating these series are not likely to be trend stationary. First-order differencing (**Figure 3: 1st Order Difference Time Series**) effectively removes the trend for all series, with the difference plots exhibiting minimal drift, implying that all level series are difference stationary as difference stationarity always dominates over trend stationarity.

While these observations provide preliminary insights, statistical tests (Unit test) are necessary to confirm the results conclusively.

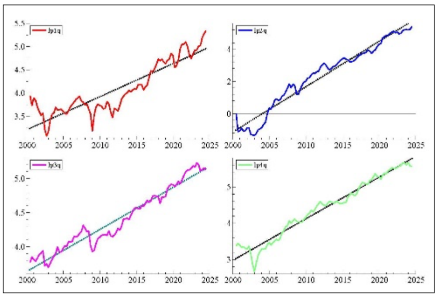


Figure 1: Level Time Series

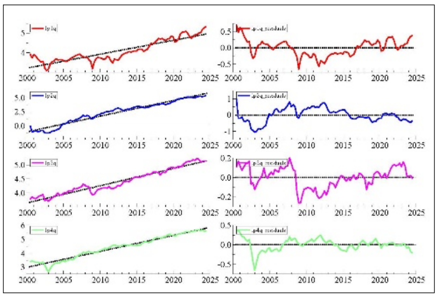


Figure 2: Detrended Time Series

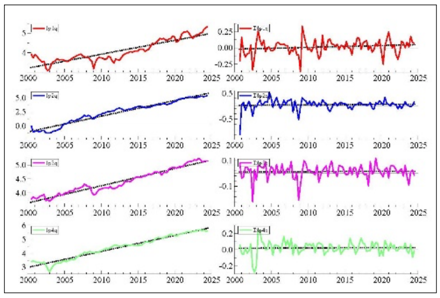


Figure 3: 1st Order Difference Time Series

2.2 Statistical Testing for Stationarity:

The Augmented Dickey-Fuller (ADF) unit root tests were conducted to assess the stationarity of the time series: *lp1q*, *lp2q*, *lp3q* and *lp4q*. The null hypothesis (H_0) states that a unit root is present, indicating non-stationarity $I(1)$ while the alternative hypothesis (H_1) suggests stationarity $I(0)$.

$$\Delta lpiq_t = \alpha + \beta t + \theta \Delta lpiq_{t-1} + \varepsilon_t \; for \; i = 1, \dots, 4 \; (\mathbf{2.1})$$
$$H_0: \theta = 0, lpiq_t \sim I(1), a \; unit \; root \; is \; present \; in \; time \; series - \; does \; not \; possess \; stationarity \; (\mathbf{2.2})$$
$$H_1: \theta \neq 0, lpiq_t \sim I(0), a \; unit \; root \; is \; not \; present \; in \; time \; series - possesses \; stationarity \; (\mathbf{2.3})$$

Variable	t-value (test statistic)	Optimal Lag	Conclusion	Order of Integration	Stationary?
lp1q	-3.219	1	Insufficient Evidence to Reject H_0 (at 5% level).	$I(1)$	No
lp2q	-2.245	1	Insufficient Evidence to Reject H_0 (at 5% level).	$I(1)$	No
lp3q	-2.905	0	Insufficient Evidence to Reject H_0 (at 5% level).	$I(1)$	No
lp4q	-3.801*	1	Insufficient Evidence to Reject H_0 (at 1% level).	$I(1)$	No
Dlp1q	-6.644**	2	Very strong evidence to reject H_0 (at 1% level)	$I(0)$	Yes
Dlp2q	-6.749**	1	Very strong evidence to reject H_0 (at 1% level)	$I(0)$	Yes
Dlp3q	-9.155**	0	Very strong evidence to reject H_0 (at 1% level)	$I(0)$	Yes
Dlp4q	-6.571**	2	Very strong evidence to reject H_0 (at 1% level)	$I(0)$	Yes

Table 1: Unit Test Results

For the level series, the results indicate that *lp1q*, *lp2q* and *lp3q* are non-stationary, as there is insufficient evidence to reject H_0 at the 5% significance level. Hence, *lp1q*, *lp2q* and *lp3q* are integrated of order one $I(1)$. However, for *lp4q*, H_0 is rejected at the 5% significance level but not at the (stricter) 1% level. This suggests that *lp4q* may not be stationary with precision. To address this ambiguity, first-order differencing is applied to *lp4q*, which confirmed that it is integrated of order one $I(1)$.

First-order differencing was applied to all series to address non-stationarity. The differenced series (*Dlp1q*, *Dlp2q*, *Dlp3q* and *Dlp4q*) are all stationary at the 1% significance level, which suggests there is strong evidence to reject H_0 at the 1% significance level, confirming that the $I(1)$ series became $I(0)$ after differencing.

In summary, *lp1q*, *lp2q*, *lp3q* and *lp4q* are non-stationary at levels but become stationary after first-order differencing, indicating they are all integrated of order one $I(1)$. These findings ensure the appropriate transformations are applied in subsequent econometric modeling.

3 Model specifications and misspecification tests

3.1 Analysis of the Misspecified ADL (1) Model:

The initial analysis considered an ADL (1) model allowing for up to one lag on both the dependent variable (*lp1q*) and the independent variable (*lp4q*). The general form of the misspecified model is represented as:

$lp1q_t = \alpha_0 + \alpha_1 lp1q_{t-1} + \beta_0 lp4q_t + \beta_1 lp4q_{t-1} + \varepsilon_t$ $\mathbf{(3.1)}$

This model was subjected to a series of restrictions to generate nine nested models, testing specific hypotheses about the significance of variables and the validity of imposed restrictions.

The misspecified ADL (1) model was assessed under the assumption:

1. The Residual error ($\boldsymbol{\varepsilon_t}$) possesses: no serial correlation, no autoregressive conditional heteroskedasticity, Normality [i.e: $\varepsilon_t \sim \mathrm{N}\ (0, \sigma^2)$], and no heteroskedasticity.
2. The Model is well/correctly specified – to obtain this we added dummies.

The results of the hypothesis tests for the misspecified model are summarised in **Table 2: Misspecification Tests Results** below:

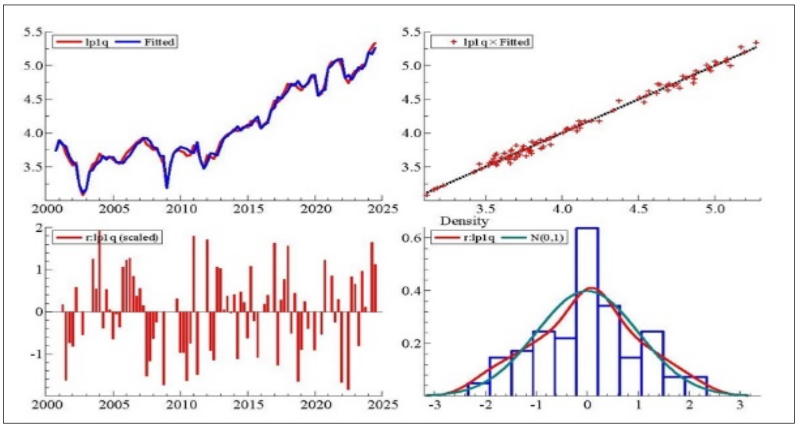


Figure 4: Fitted Model/QQ-plot/ACF/Normality Tests

Nested Model	Restrictions	Test statistic	Reject/Not Reject
Static model **[1]**	$H_0: \alpha_1 = \beta_1 = 0$	F (2,92) = 213.63 [0.0000]**	Reject H_0
Autoregressive model **[2]**	$H_0: \beta_0 = \beta_1 = 0$	F (2,92) = 5.8793 [0.0040]**	Reject H_0
Leading indicator model **[3]**	$H_0: \alpha_1 = \beta_0 = 0$	F (2,92) = 224.51 [0.0000]**	Reject H_0
Difference data model **[4]**	$H_0: \alpha_1 = 1, \beta_0 = -\beta_1$	Chi²(2) = 5.0790 [0.0789]	Do not reject H_0
Distributed lag model **[5]**	$H_0: \alpha_1 = 0$	F (1,92) = 427.12 [0.0000]**	Reject H_0
Partial adjustment model **[6]**	$H_0: \beta_1 = 0$	F (1,92) = 5.3508 [0.0229]*	Reject H_0
Static model with AR (1) errors / AR error **[7]**	$H_0: \alpha_1 * \beta_0 + \beta_1 = 0$	Chi²(1) = 2.9052 [0.0883]	Do not reject H_0
Error correction model **[8]**	$H_0: \alpha_1 + \beta_0 + \beta_1 = 1$	Chi²(1) = 0.58527 [0.4443]	Do not reject H_0
Dead-start model **[9]**	$H_0: \beta_0 = 0$	F (1,92) = 7.7663 [0.0065]**	Reject H_0

Table 2: Misspecification Tests Results

Key Findings:

Strong evidence (at the *1%* significance level) was found to reject the null hypothesis (H_0) for five models: Static model **[1]**, Autoregressive model **[2]**, Leading indicator model **[3]**, Distributed lag model **[5]**, and Dead-start model **[9]**, indicating that excluded variables or restrictions in these models are statistically significant. Furthermore, at the 5% significance level, the Partial adjustment model **[6]** provided sufficient evidence to reject H_0, suggesting that excluded variable ($lp4q_{t-1}$) was significant.

There was insufficient evidence to reject H_0 for three models: Difference data model **[4]**, Static model with AR (1) errors **[7]**, and Error correction model **[8]**. This indicates that the respective restrictions could not be statistically invalidated. However, the results highlight that the Error correction model [8] stands out as a best candidate for further analysis, given its consistency with the data and absence of evidence to reject H_0 (largest p-value: 0.4443).

3.2 Analysis of the Misspecified ADL (1) Model:

To address the identified misspecifications, the ADL (1) model was re-specified by incorporating 19 dummy variables. The specified model is expressed as: $\boldsymbol{lp1q_t = \alpha_0 + \alpha_1 lp1q_{t-1} + \beta_0 lp4q_t + \beta_1 lp4q_{t-1} + \varepsilon_t + dummies}$. The specification tests for the revised model were conducted to reassess the nested hypotheses, with the results presented in **Table 3: Specification Tests Results** below:

Nested Model	Restrictions	Test statistic	Conclusion
Static model **[1]**	$H_0: \alpha_1 = \beta_1 = 0$	F (2,73) = 448.19 [0.0000]**	**Reject H_0**
Autoregressive model **[2]**	$H_0: \beta_0 = \beta_1 = 0$	F (2,73) = 11.708 [0.0000]**	**Reject H_0**
Leading indicator model **[3]**	$H_0: \alpha_1 = \beta_0 = 0$	F (2,73) = 482.92 [0.0000]**	**Reject H_0**
Difference data model **[4]**	$H_0: \alpha_1 = 1, \beta_0 = -\beta_1$	$Chi^2(2)$ = 13.549 [0.0011]**	**Reject H_0**
Distributed lag model **[5]**	$H_0: \alpha_1 = 0$	F (1,73) = 886.39 [0.0000]**	**Reject H_0**
Partial adjustment model **[6]**	$H_0: \beta_1 = 0$	F (1,73) = 9.4480 [0.0030]**	**Reject H_0**
Static model with AR (1) errors / AR error **[7]**	$H_0: \alpha_1 * \beta_0 + \beta_1 = 0$	$Chi^2(1)$ = 7.1287 [0.0076]**	**Reject H_0**
Error correction model **[8]**	$H_0: \alpha_1 + \beta_0 + \beta_1 = 1$	$Chi^2(1)$ = 1.3806 [0.2400]	**Do not reject H_0**
Dead-start model **[9]**	$H_0: \beta_0 = 0$	F (1, 73) = 13.851[0.0004]**	**Reject H_0**

Table 3: Specification Tests Results

Key Findings:

Strong evidence was found to reject H_0 at the 1% significance level for eight models: Static model **[1]**, Autoregressive model **[2]**, Leading indicator model **[3]**, Difference data model **[4]**, Distributed lag model **[5]**, Partial adjustment model **[6]**, Static model with AR (1) errors / AR error **[7]**, and Dead-start model **[9]**. This indicates that excluded variables or restrictions are statistically significant.

The **Error correction model [8]** emerges as the most robust and well-specified option, supported by a p-value of 0.2400, which indicates insufficient evidence to reject the null hypothesis (H_0) at the 5% significance level. Hence, the model is consistent with the restriction $\alpha_1 + \beta_0 + \beta_1 = 1$, confirming its ability to capture short-run dynamics while maintaining compatibility with potential long-run equilibrium relationships.

The results emphasise the importance of addressing misspecifications in econometric modeling to ensure reliable and coherent outcomes. This conclusion is based on both misspecification and specification tests, which further validate the robustness of the error correction model in this context.

4 Cointegration vs spurious regression

4.1 Do lp3q and lp4q cointegrate?

As *lp3q* and *lp4q* are of the same integrated order, $I\ (1)$, we were able to check if cointegration exists. From the analysis of *lp3q* (PEPSICO) and lp4q (MCDONALDS) it suggests that their observed relationship is likely a spurious correlation rather than evidence of cointegration. This was obtained from running a regression of *lp4q* with *lp3q* and checking if the residuals obtained from the regression is of integrated order zero, $I(0)$. Our OLS regression of $lp3q_t = \beta_0 + \beta_1 lp4q_t + \varepsilon_t$, yielded a high R^2 of 0.92, indicating a seemingly strong fit, diagnostic tests revealed significant residual autocorrelation ($AR\ test\ p = 0.0000$) and model misspecification ($RESET\ test\ p = 0.0000$), which are classic symptoms of spurious regressions. Additionally, the ADF test on the residuals failed to reject the null hypothesis of a unit root ($t - adf = -2.033, critical\ value = -2.89$), confirming that the residuals are non-stationary and thus invalidating the notion of a long-term equilibrium relationship between *lp3q* and *lp4q*. These results, combined with issues of heteroskedasticity ($ARCH\ test\ p = 0.0000$) and non-normal residuals, strongly suggest that the observed correlation arises from shared trends (i.e: as seen in **Figure 5: Time Series plots of lp3q and lp4q**) rather than a genuine strong cointegrating relationship

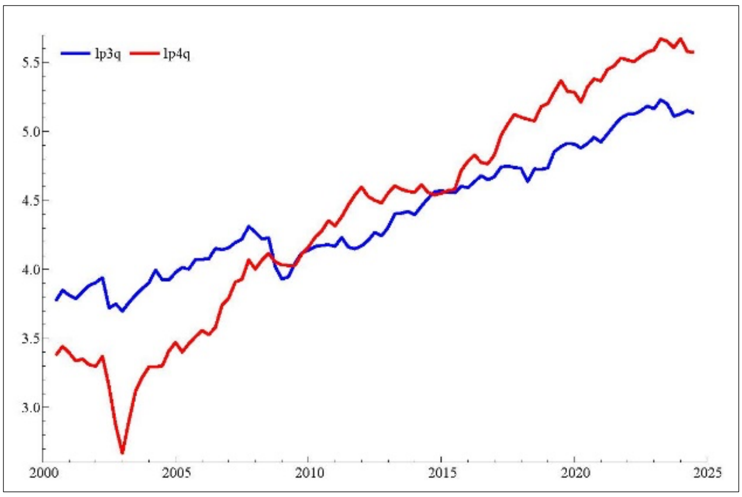


Figure 5: Time Series plots of lp3q and lp4q

4.2 Empirical Findings:

Variable	Coefficient	Standard Error	t-value	t-prob	Part.R²
Constant	0.158495	0.09050	1.75	0.0832	0.0323
lp3q_1	-0.0730628	0.04072	-1.79	0.0761	0.0338
lp4q_1	0.0386780	0.02205	1.75	0.0827	0.0324
Dlp4q	0.308397	0.06558	4.70	0.0000	0.1938

Table 4: UECM Tests Results

The Unrestricted Error Correction Model (UECM) reveals a significant $\lambda = -0.0730628$ which is negative and different from zero. This suggests the presence of a weak cointegration relationship between *lp3q* (the dependent variable) and *lp4q* (the regressor). The long-run parameters are estimated as:

$$\frac{\alpha_0}{1-\alpha_1} = -\frac{\alpha *}{\lambda} = -\frac{0.158495}{-0.0730628} = 2.169 \ (\mathbf{4.1}) \qquad \frac{\beta_0}{1-\ \beta_1} = -\frac{\beta *}{\lambda} = -\frac{0.0386780}{-0.0730628} = 0.529 \ (\mathbf{4.2})$$

These findings indicate that, in the long run, a 1-unit increase in *lp4q* is associated with a 0.529-unit increase in *lp3q*. However, the relatively low magnitude of λ suggests a slow adjustment process back to equilibrium of nearly 14 years.

The Engle-Granger two-step procedure provides additional insights. In the first step, the residuals from the static regression of *lp3q* on *lp4q* are stationary at the 5% significance level but not at the 1% level, indicating weak cointegration.

Variable	Coefficient	Standard Error	t-value	t-prob	Part.R²
residuals_1	-0.0719868	0.04080	-1.76	0.0809	0.0324
Dlp4q	0.327399	0.06302	5.19	0.0000	0.2249

Table 5: Engle and Granger 2nd step Results

In the second step, the restricted error correction model shows that the coefficient of the lagged residuals aligns closely with the UECM estimate. This coefficient implies that approximately 7% of the disequilibrium from the previous period is corrected annually.

Ferson et al. (2003a) and Ferson et al. (2003b) explore the risks of spurious regression in financial contexts, particularly when highly autocorrelated variables are involved. They emphasise how persistence in predictor variables, such as dividend yields, can exaggerate the appearance of predictability. Similarly, the results for *lp3q* and *lp4q* may reflect shared stochastic trends rather than a genuine equilibrium relationship, which was shown from our results. As shown in Hendry's (2004) Section II (pp. 189–194), spurious regressions often arise when non-stationary series are regressed on one another without proper adjustments, such as first-differencing or cointegration tests. In our results *lp4q* had only 1 lag rejecting the null hypothesis of a unit root, compared to 4 lags accepting the null of unit root, which proved an inconsistency in its integrated order compared to *lp3q* which was all $I\ (1)$. This potentially affected our results, which led to a weak cointegration.

While Hendry (2004) Section IV (pp. 198–204) highlights the power of equilibrium correction models in capturing both long and short-run dynamics, it also warns against overinterpreting weak or inconsistent results. In this case, the UECM results fail to provide conclusive evidence of cointegration and instead suggest a relationship driven by short-term fluctuations. This is consistent with the theory of spurious regressions, as non-stationary series can produce artificially significant relationships when modeled without accounting for their underlying stochastic properties.

Weak cointegration in this case may stem from omitted variables or external factors influencing both lp3q and lp4q beyond their direct relationship. For example, PEPSICO's stock price (*lp3q*) may depend not only on MCDONALDS' stock price (*lp4q*) but also on broader market conditions, industry-specific trends, or macroeconomic dynamics. The findings support the hypothesis of weak cointegration rather than a purely spurious correlation.

The "drunkard and the dog" analogy, as explained by Hendry (2004), provides a relatable way to describe cointegration in the context of our analysis. Imagine a drunkard wandering randomly and their dog also moving randomly but tethered to the drunkard. Despite their independent short-term movements, the tether ensures they remain connected, reflecting a long-term equilibrium. This can be likened to the relationship between *lp3q* (PEPSICO) and *lp4q* (MCDONALDS) in our analysis. The UECM results showed weak evidence of cointegration, suggesting a loose tether that maintains some equilibrium connection between the two variables. This may be from MCDONALDS presumably selling PEPSICO products, or other fast-food demands, affecting sales at MCDONALDS. However, the dominance of short-term dynamics implies that this tether is not particularly strong, and the relationship could still be influenced by shared trends rather than a robust long-term adjustment. This highlights the fragility of the connection and the need for cautious interpretation of the results.

4.3 Engle & Granger 2-Step Procedure:

The implementation of the 2-step Engle-Granger procedure and the unrestricted equilibrium correction model (UECM) provides complementary insights into the relationship between *lp3q* (PepsiCo) and *lp4q* (McDonald's). The Engle-Granger procedure involves:

1. Estimating the long-run equilibrium relationship between the two variables.
2. Testing the residuals for stationarity.

In our case, the ADF test on the residuals failed to strongly reject the null hypothesis of a unit root (t=−2.033, critical value = -2.89 at 5%), suggesting weak evidence for cointegration. This implies that while a long-term equilibrium might exist, it is not robust or stable enough to confirm strong cointegration.

The UECM, on the other hand, tests both short-term dynamics and long-term adjustments. The results indicated marginal significance of the lagged level term for *lp3q* (p=0.0761), suggesting some degree of equilibrium correction. However, the short-term dynamics dominated, with *Δlp4q* showing strong significance, indicating that short-term factors play a major role in driving the relationship. This aligns with Hendry's (2004) discussion in Section III (pp. 194–198), where the equilibrium correction model balances long-run equilibrium with short-term adjustments, capturing dynamic processes more effectively than static models.

Overall, the weak evidence for cointegration in the Engle-Granger procedure, combined with the dominance of short-term dynamics in the UECM, suggests that any equilibrium relationship between *lp3q* and *lp4q* is fragile and heavily influenced by short-term effects. Hendry's emphasis on dynamic modeling underscores the importance of using UECMs to account for both short-run changes and long-run equilibria, providing a more nuanced understanding of relationships like this one.

5 Granger causality

5.1 Theoretical Framework:

Granger causality is a statistical method used to determine whether one time series (Xt) can help predict another (Y_t). If the past values of X_t significantly improve the prediction of Y_t, X_t is said to Granger-cause Y_t.

Granger causality is particularly valuable because it provides a formal, testable framework to assess causation. A classic example is N. Thurman and Mark E. Fisher humorous exploration of the "chicken or egg" problem. In this study, Granger explored the long-debated question of whether chicken or egg came first. While the example is humorous and ironic, it underscores the practicality of Granger causality in providing mathematical and empirical results to address questions that were previously speculative or philosophical. This example demonstrates how the methodology can be adapted to economic and financial research, enabling analysts to uncover the real causal relationships between variables.

Granger causality aligns with the idea that "causation does not imply correlation," as the two concepts are distinct. Correlation measures the degree of association between variables but does not consider directionality or temporal precedence. For instance, two variables may be highly correlated due to a shared underlying factor (spurious correlation), yet one does not cause the other. Granger causality, in contrast, focuses on whether one variable's history improves the prediction of another, providing a more meaningful framework for understanding causation.

The broader context of the Granger causality has been extensively discussed in econometrics. Hendry (2004) highlights its role in identifying dynamic relationships between time series, noting that it is a critical tool for understanding causality in econometric models. Granger causality is also closely linked to concepts like cointegration, where long-term equilibrium relationships imply Granger causality in at least one direction. Thurman and Fisher (1988) provide a practical example of Granger causality in their analysis of the relationship between chicken and egg production. Their study used Granger causality to empirically determine whether changes in egg production predict changes in chicken production or vice versa, demonstrating how the method can disentangle complex temporal dynamics in real-world data.

Mathematically is tested by comparing the prediction error (variance) of a model that uses only Y's past values versus one that uses both Y's and X's past values. If adding X's history improves the prediction, Granger causality is said to exist.
In brief is just estimating this model:

$$y_t = \alpha + \sum_{i=1}^{n} \beta_{1,i} y_{t-1} + \sum_{i=1}^{n} \beta_{2,i} x_{t-1} + u_t \quad u_t \sim WN$$

And testing the two hypotheses:

$$H_0: \sum_{i=1}^{n} \beta_{2,i} = 0 \quad H_1: \sum_{i=1}^{n} \beta_{2,i} \neq 0$$

If we fail to reject the null hypothesis, x_t has no tendency to lead movements in y_t. If we reject the null hypothesis, the forecasting error will be bigger if we omit lagged values of x_t.

5.2 Empirical Findings

In our use case the results of the Granger-causality tests to determine whether movements in ***lp4q*** Granger-cause movements in the other three stock prices (***lp1q, lp2q*** and ***lp3q)*** are presented in the table below:

Variables	Results
Lp1q	Subset F (4,84) = 1.9429 [0.1108]
Lp2q	Subset F (4,84) = 1.4081 [0.2384]
Lp31	Subset F (4,84) = 1.6269 [0.1751]

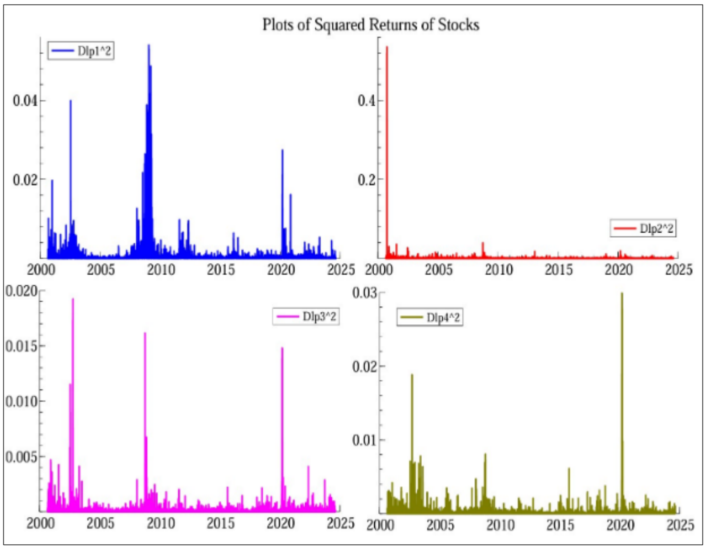
Table 6: Granger-causality Tests Results

In all cases, the *p-values* exceed the 5% significance level, indicating that we fail to reject the null hypothesis that lagged values of ***lp4q*** have no predictive power for the other three stock prices.

6 (G)ARCH Models:

6.1 Test for ARCH effects using a Lagrange multiplier test:

ARCH effects explain the occurrence of serial correlation (or autoregressive conditional) of heterosckedacity in variance returns, in our case the squares of Dlp1, Dlp2, Dlp3, and Dlp4. To test for the presence of ARCH effects we use the F-test as a proxy for the Lagrange multiplier tests on the squared of the first differences of our series with the hypothesis below.



From the plot we can observe the few numbers of spikes in Dlp2 (red) showing the low number of volatility clusters, more at the start and a decline towards the end, which can be explained by ARCH1-2 tests where there was presence of ARCH effects but not for the higher orders, of 5, and 10, compared to the higher volatility clustering in the other series.

$\boldsymbol{H_0}$: No Serial correlation between variances; no presence of ARCH effects.
$\boldsymbol{H_1}$: Serial correlation between variances; presence of ARCH effects.

Assumptions:

1. The relationship between squared residuals and their lags is linear.
2. Squared residuals are independent over time after model fitting.
3. Squared residuals show no serial correlation.
4. The time series is stationary (stable mean and variance).
5. The ARCH-LM test assumes normally distributed residuals.

Model	ARCH Tests	F-value	Decision
$\boldsymbol{R1 = Dlp1^2}$	ARCH (1-2)	F (2,6257) = 504.97 [0.0000]	Very strong evidence to reject $\boldsymbol{H_0}$ at 1% significance level.
	ARCH (1-5)	F (5,6251) = 271.65 [0.0000]	Very strong evidence to reject $\boldsymbol{H_0}$ at 1% significance level.
	ARCH (1-10)	F (10,6241) = 157.64 [0.0000]	Very strong evidence to reject $\boldsymbol{H_0}$ at 1% significance level.
$\boldsymbol{R2 = Dlp2^2}$	ARCH (1-2)	F (2,6257) = 3.2313 [0.0396] *	Strong evidence to reject $\boldsymbol{H_0}$ at 5% significance level.
	ARCH (1-5)	F (5,6251) = 1.8613 [0.0976]	Insufficient evidence to reject $\boldsymbol{H_0}$ at 5% significance level.
	ARCH (1-10)	F (10,6241) = 1.6146 [0.0958]	Insufficient evidence to reject $\boldsymbol{H_0}$ at 5% significance level.

$R3 = Dlp3^2$	ARCH (1-2)	F (2,6257) = 593.15 [0.0000]	Very strong evidence to reject H_0 at 1% significance level.
	ARCH (1-5)	F (5,6251) = 316.74 [0.0000]	Very strong evidence to reject H_0 at 1% significance level.
	ARCH (1-10)	F (10,6241) = 169.08 [0.0000]	Very strong evidence to reject H_0 at 1% significance level.
$R4 = Dlp4^2$	ARCH (1-2)	F (2,6257) = 213.06 [0.0000]	Very strong evidence to reject H_0 at 1% significance level.
	ARCH (1-5)	F (5,6251) = 138.01 [0.0000]	Very strong evidence to reject H_0 at 1% significance level.
	ARCH (1-10)	F (10,6241) = 116.32 [0.0000]	Very strong evidence to reject H_0 at 1% significance level.

From our results ARCH effects is not present in Dlp2 hence GARCH (p, q) might not be appropriate for the modelling of its conditional volatility. After running this process in OxMetrics all 4 variables Dlp1, Dlp2, Dlp3 and Dlp4 were all misspecified according to normality, Hetero test, RESET, AR, and ARCH tests.

6.2 Best Univariate GARCH Representation:

Our model selection was:

1. Dlp1: ARMA(0,0)-GJR-GARCH(1,1) - Normal Distribution
2. Dlp2: ARMA(0,0)-GARCH(1,1) - Normal Distribution
3. Dlp3: ARMA(1,0)-GARCH(1,1) - Normal Distribution
4. Dlp4: ARMA(1,1)-EGARCH(1,1) - Student distribution

$$GARCH(p,q)Representation: h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, where\ \alpha + \beta < 1$$
$$GJR - GARCH(1,1)\ Representation: h_t = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\alpha\varepsilon_{t-1}^2 I(\varepsilon_{t-1}^2 < 0) + \beta h_{t-1}$$
$$EGARCH(1,1)\ Representation: ln(h_t) = \omega + \alpha\left(\left|\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right| - \sqrt{\frac{2}{\pi}}\right) + \gamma\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta ln(h_{t-1})$$

Where, h_t = conditional variance, ω = unconditional variance component and ε_t = errors and shocks

The model selection process was carried out on Dlp1, Dlp2, Dlp3, Dlp4, where we chose based on:

1. A model with the least number of misspecifications is preferred
2. A model with lowest Schwarz (BIC) is preferred, more of a check for parsimonious model

Dlp1: The analysis of the GJR-GARCH (1,1) and GARCH (1,1) models reveals important insights into the volatility dynamics of Dlp1. Both models are generally well-specified, as diagnostic tests indicate no significant serial correlation in residuals or squared residuals. The GJR-GARCH (1,1) model provides a better fit, with a higher log-likelihood (16860.3) and lower AIC, and effectively captures asymmetry in volatility through a significant Gamma term (p < 0.001), reflecting the greater impact of negative shocks on volatility. However, the model fails the fourth-moment condition (1.00341 > 1), raising concerns about its robustness under extreme leptokurtic behaviour, which is evident in the high excess kurtosis (18.86) and rejection of normality (p = 0.0000) in both models. The standard GARCH (1,1) model, while slightly less effective in capturing volatility asymmetry and with a lower log-likelihood (16788.7), satisfies both the second- and fourth-moment conditions, making it more robust under heavy-tailed distributions. Overall, the GJR-GARCH (1,1) model is preferred for its superior fit and ability to model asymmetry, but the GARCH (1,1) model is a safer choice when stability under extreme events is a priority. To note both models showed strong convergence to the normal distribution.

Dlp2: The comparison of GARCH (1,1) and APARCH (1,1) models for (Dlp2) reveals important insights into their performance in capturing volatility dynamics. However, the ARCH-LM test results indicate no significant ARCH effects (p > 0.05) in (Dlp2), suggesting that volatility modelling may not be strictly necessary. Despite this, volatility models are often applied to financial series to explore hidden dynamics or to maintain consistency with other analyses.

The GARCH (1,1) model provides a log-likelihood of 15530.4 with significant parameters for ARCH (alpha_1 = 0.103) and GARCH (beta_1 = 0.888) and satisfies the second-moment condition (alpha_1 + beta_1 = 0.99154 < 1). However, it fails the fourth-moment condition (1.00443 > 1), raising concerns about its robustness under heavy-tailed distributions. Diagnostic tests show no significant serial correlation in residuals or squared residuals (p > 0.05), confirming a well-specified model. Nonetheless, the series exhibits significant leptokurtosis (Kurtosis = 138.87) and rejects normality (p = 0.0000).

The APARCH (1,1) model improves upon the GARCH model, with a higher log-likelihood of 15596.8 and better AIC values, reflecting superior fit. This model captures asymmetry through the Gamma parameter (gamma_1 = 0.344), (p < 0.001), and the Delta parameter (delta = 1.09), (p < 0.001) allows for greater flexibility in modelling volatility clustering. The APARCH model satisfies both the second- and fourth-moment conditions, ensuring greater robustness under extreme events. However, given the absence of significant ARCH effects in (Dlp2), the marginal improvement in fit may not justify the added complexity of the APARCH model.

In conclusion, while both models are technically well-specified, the lack of significant ARCH effects in (Dlp2) suggests that volatility modelling may not be necessary. If a volatility model is required, the GARCH (1,1) model is simpler and sufficient, while the APARCH (1,1) model offers a better fit and greater flexibility at the cost of added complexity. The choice depends on the goals of the analysis: parsimony and simplicity favour the GARCH (1,1), while the APARCH (1,1) is better suited for capturing asymmetry and robustness under extreme events. To note both models showed strong convergence to the normal distribution.

Dlp3: The analysis of the ARMA (1,0) - GARCH (1,1) or APARCH (1,1) models for (Dlp3) evaluates their performance in capturing volatility dynamics. The GARCH (1,1) model achieves a log-likelihood of 19891.6 and satisfies both the second moment (alpha_1 + beta_1 = 0.97693 < 1) and fourth-moment conditions, ensuring stability under heavy-tailed distributions. Residual diagnostics show no significant serial correlation in residuals or squared residuals, with (p > 0.05) for most lags, confirming that the model is well-specified. The model's simplicity results in a lower parameter count (5 parameters) and a more parsimonious fit, as reflected in its BIC value of -6.346. However, the data's leptokurtic nature (Kurtosis = 2.93) and rejection of normality (p = 0.0000) highlight limitations in capturing extreme events.

The APARCH (1,1) model achieves a higher log-likelihood of 19947.5 and accounts for asymmetry and non-linear effects through significant Gamma (gamma_1 = 0.521) and Delta (delta = 1.283) parameters. Despite these advantages, the APARCH model has a higher parameter count (7 parameters), and its BIC value of -6.361 is slightly less parsimonious than that of the GARCH (1,1) model. Furthermore, the model shows residual ARCH effects at short lags (p = 0.0052) for ARCH 1-2), indicating some unexplained volatility clustering.

In conclusion, based on model specification and the BIC criterion, the GARCH (1,1) model is preferred for (Dlp3). It offers a simpler and more parsimonious approach while remaining well-specified and stable. While the APARCH (1,1) model provides additional flexibility and a slightly better fit, the improvements are marginal and do not outweigh the benefits of simplicity and better penalization of model complexity provided by the GARCH (1,1) model. To note both models showed strong convergence to the normal distribution.

Dlp4: The analysis of (Dlp4) using $ARMA(0,0) - GARCH\ (1,1)$ following a Normal distribution or $ARMA(0,0) - EGARCH\ (1,1)$ following a student distribution reveals that both are well-specified with insignificant specification issues and no normality. EGARCH model had a weak convergence to the student distributions whiles the GARCH model had a strong convergence to the Normal distribution. The GARCH (1,1) model achieves a log-likelihood of 18884.8, satisfies stability conditions alpha_1 + beta_1 = 0.98588 < 1), and has a BIC of -6.026, making it more parsimonious with only 4 parameters. In contrast, the EGARCH (1,1) model improves the log-likelihood to 19139.1 and effectively captures asymmetry and heavy tails using the student t-distribution, but its higher parameter counts results in a BIC of -6.103, but still better than the BIC of the GARCH model. While the EGARCH model provides a better fit and greater flexibility, the GARCH (1,1) model balances simplicity and specification, hence we choose the EGARCH model choice based on the BIC criterion.