PREDICTIVE ANALYSIS FOR DECISION MAKING

Individual Coursework

Versha Sandesh

Student ID: 20395827

Module Code: 7FNCE044W

Module Leader: Dr. Issam Malki

MSc Fintech with Business Analytics

Solution 1.

a. Main statistical properties of data:

The given data belongs to the United States (US) comprising 40 years from January 1959 to December 1998.

The data is time-series because there is single unit (demand for money) observed over time (t). The data includes four variables as below:

$r_t = interest\ rate$

$lm_t = logarithm\ of\ money$

$lp_t = logarithm\ of\ prices$

$lo_t = logarithm\ of\ output$

Table 1

lmt	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
rt	***	0.001	-26.01	0.000	-0.025	-0.022
lpt	0.961	0.014	***	0.000	0.934	0.989
lot	0.439	0.025	***	0.000	0.390	0.488
Constant	0.416	0.056	7.38	0.000	0.305	0.527
URSS			1.754	RRSS		249.548
R-squared			0.993	Number of obs		480.000
F-test			***	Durbin Watson		0.122
Akaike crit. (AIC)			-1323.557	Bayesian crit. (BIC)		-1306.862

The output presents **coefficients** for a linear regression model to analyse the demand for money (lm_t), as dependent variable, in the US using a static long run relationship linking to interest rate, prices, output and a constant. The coefficients represent the estimated effect of each independent/predictor variable on the dependent variable.

Besides, the **standard errors** for each coefficient, as provided in Table 1, indicates the variability of the coefficient estimates in measuring the impact of independent variables on dependent variable. Lower standard errors indicate more precise estimates.

The **t-values** for each coefficient are provided in the table, calculated by dividing the coefficient estimates by their standard errors. The associated **p-values** indicate that all the variables are statistically significant at **95% confidence interval** level (α = 5%).

The **R-squared** is the coefficient of determination, which represents how much variance in dependent variable is explained by the independent variable. The value for R-squared for this dataset is 0.993 (99.3%). However, it indicates to a potential problem of autocorrelation.

Autocorrelation occurs when a variable is correlated with itself in the past or future, a prominent problem with time-series data.

The **F-test** assesses the overall significance of the model, which is calculated in part d.

URSS and **RRSS** refer to Unrestricted Sum of Squared Residuals and Restricted Sum of Squared Residuals, respectively. The RSS of restricted model must always be higher than RSS of unrestricted model.

The **Durbin-Watson Statistic** is used to detect the autocorrelation in the model. As per the rule of thumb in linear equation with time-series data, if R-squared is higher than the Durbin-Watson (DW) test then there is spurious regression, leading to misleading or not statistically sound model. In the given information in Table 1, the value of R-squared (0.993) is significantly higher than the value of DW test (0.122), reflecting to the likelihood of spurious regression.

Akaike Information Criterion **(AIC)** and Bayesian Information Criterion **(BIC)** are measures used to identify the quality of the model. The values of AIC and BIC can be anywhere between $-\infty$ to $+\infty$. The Lower values of AIC and BIC indicate better fitness of model.

The **total number of observations** in the dataset over the 40 years are concluded as 480.

Stationarity:

Stationarity signifies that the statistical properties, such as mean, variance and covariance, of time-series data remain constant over time. The data has stationarity when the data points cross the mean frequently – mean reversion.

Based on the time-series plots in Figure 1, **none of the variables exhibit stationarity** over the sample period as it cannot be determined where the trend ends.

- The lm_t plot shows an upward trending behaviour, indicating non-stationarity in the mean, and violates the assumption of constant mean of data to be stationary.
- The r_t plot illustrates a random walk, which means we cannot predict the time path. This indicates a cluster of variances, meaning that variance is not constant over time.
- The lo_t plot exhibits an upward trend with volatility, implying non-stationarity in the mean and variance.
- The lp_t plot demonstrates an upward deterministic trending pattern, violating the assumption of a constant mean required for stationarity.

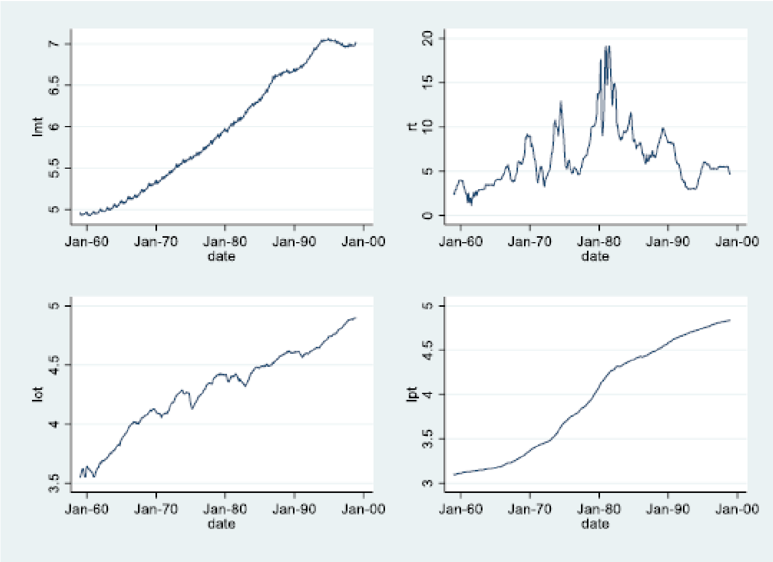

Figure 1

The unit root tests, such as Augmented Dickey-Fuller (ADF) test, which test the null hypothesis of non-stationarity against the alternative stationarity, or the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, that examines whether a time series remains within certain bands over time, can be further conducted to assess for the stationarity.

b. Mathematical expression of the model:

The mathematical model for the given data can be expressed as:

$$lm_t = \beta_0 + \beta_1 r_t + \beta_2 lp_t + \beta_3 lo_t + u_t$$

Where:

$\beta_0, \beta_1, \beta_2, \beta_3$ are unknown coefficients,

u_t is unknown error term.

The key assumptions of the linear model including explanatory variables (r_t, lp_t, lo_t) and error term u_t are the following:

1. Y (logarithm of money) is a linear function of explanatory variables.
2. The explanatory variables are linearly independent from each other (i.e. no multicollinearity).
3. Explanatory variables are fixed (non-stochastic/non-random) with rank k (full column rank) → E[x | u] = E[x]
4. Explanatory variables are strictly exogenous. This implies that the explanatory variables are not correlated with the error term (u_t), i.e. corr[x,u] = 0.
5. The errors are homoscedastic, meaning the variance of the error term (u_t) is constant across all observations (no heteroskedasticity).

6. The errors are serially uncorrelated with each other over time (no autocorrelation).
7. The distribution between dependent variable and error term should be normal distribution, and that error term is a vector with E[u] = 0, meaning the expected value or mean of error term is 0.

The validity of these assumptions is must for the OLS to be a reliable and best linear unbiased estimator (BLUE). Violating these assumptions would result in following implications for the estimated model:

1. If the true relationship between dependent variable and explanatory variables is non-linear then the linear model (OLS) may provide biased and inconsistent estimates of the coefficients.
2. If two or more explanatory variables are highly correlated with each other, then it implies multicollinearity, and the variables inflate the results, making it difficult to assess the individual effects of the variables.
3. If explanatory variables are not fixed leads to the problem of endogeneity, implying that explanatory variables are correlated with errors, i.e. E[x|u] ≠ E[x], which causes OLS estimator to be inefficient.
4. Violating the assumption of exogeneity results in endogeneity, meaning that a variable in model is affected by the variables that are not in the model, implying that the OLS estimator is no longer unbiased nor consistent.
5. If the variance in error term is not constant, then it results in heteroskedasticity, leading to invalid statistical inferences.
6. The presence of autocorrelation in the errors will result in inefficient estimates and biased standard errors.
7. If the assumption E[u]=0 is violated, it indicates that the error term has a non-zero mean, implying the presence of a systematic bias or omitted variable in the model, which leads to biasness in the estimated coefficients of the explanatory variables.

c. Interpretation of coefficient estimates:

The intercept ($\beta_0 = 0.416$): is the conditional mean of logarithm of money, reports the percentage change in average demand for money in the US (log of money lm_t) when all other variables (r_t, lp_t, lo_t) are equal to zero.

Coefficient of r_t ($\beta_1 = -0.02601$): suggests a negative relation that, on average, 1 percentage point increase in the interest rate r_t is associated with 0.026 unit decrease in estimated value of

lm_t holding all other variable constant. [The coefficient value of r_t is calculated as t-value = (coefficient/St.error) → coefficient = (t-value*St.error) = -26.01 * 0.001 ≅ -0.02601]

Coefficient of $lp_t(\beta_2 = 0.961)$: represents the elasticity of money demand with respect to prices, implying that 1% increase in the logarithm of prices is associated with 0.961 percentage change estimated increase in logarithm of money on average, ceteris paribus. This depicts that as prices rise, the demand for money increases.

Coefficient of lo_t ($\beta_3 = 0.439$): represents the elasticity of money demand with respect to output, suggesting that, on average, a 1% increase in the logarithm of output is associated with a 0.439% increase in the lm_t, ceteris paribus. This positive relationship implies that higher economic output leads to a higher demand for money.

Substituting the coefficient values from Table 1 to the mathematical model gives us the estimated model:

$$\widehat{lm}_t = 0.416 - (0.026)r_t + (0.961)lp_t + (0.439)lo_t + u_t$$

To **test significance of the parameters** for all variables, we have the following hypotheses to perform two-tail test:

$H_0: \beta_i = 0$ (coefficient is not significant)

$H_1: \beta_i \neq 0$ (coefficient is significant)

Where: i = 1, 2, and 3 indicating slope of r_t, lp_t, lo_t, respectively.

Considering the provided significance level of 5% (α = 0.05), the decision rule is as follows:

Reject the null hypothesis H_0 if:

$$|\hat{t}_i| = \left|\frac{\hat{\beta}_i - \beta_i}{Se(\hat{\beta}_i)}\right| > \left|t_{(\frac{\alpha}{2},df)}\right|$$

Where:

$\hat{t}_i$ is estimated t-value for the individual parameter,

$\hat{\beta}_i$ is expected value of each coefficient,

β_i is true value of coefficient, which is estimated as 0 in null hypothesis,

$Se(\hat{\beta}_i)$ is standard error of coefficient,

$t_{(\frac{\alpha}{2},df)}$ is the critical value, with t-distribution as

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

and degree of freedom (df) = n-k-1 = 480-3-1 = 476; n denotes number of observations, k is number of slopes and 1 is for constant.

Rearranging the equation: put $\beta_i = 0$

$$|\hat{t}_i| = |\frac{\hat{\beta}_i}{Se(\hat{\beta}_i)}| > |t_{(0.025,476)}|$$

The two-tailed 5% critical value $t_{(0.025,476)} \cong \pm 1.9649$ (Source: (T Critical Value Calculator, n.d.))

Therefore, we reject the null hypothesis if:

$$|\hat{t}_i| = |\frac{\hat{\beta}_i}{Se(\hat{\beta}_i)}| > |1.9649|$$

For parameter of Interest rate r_t:

$|\hat{t}_1| = |\frac{\widehat{\beta}_1}{Se(\widehat{\beta}_1)}| \quad \rightarrow \quad |\hat{t}_1| = |\frac{0.02601}{0.001}| \sim 26.01 \quad > \quad 1.9649$

For parameter of logarithm of price lp_t:

$|\hat{t}_2| = |\frac{\widehat{\beta}_2}{Se(\widehat{\beta}_2)}| \quad \rightarrow \quad |\hat{t}_2| = |\frac{0.961}{0.014}| \sim 68.64 \quad > \quad 1.9649$

For parameter of logarithm of output lo_t:

$|\hat{t}_3| = |\frac{\widehat{\beta}_3}{Se(\widehat{\beta}_3)}| \quad \rightarrow \quad |\hat{t}_3| = |\frac{0.439}{0.025}| \sim 17.56 \quad > \quad 1.9649$

Since the estimated t-values of all explanatory variables is greater than the critical value, and p-values are less than 0.05 (given in Table 1), therefore, **reject the null hypothesis**, concluding that the estimated coefficients of r_t, lp_t, lo_t are statistically significant at the 5% level in the estimated regression model. That means, **the interest rate has significant negative impact** on the logarithm of money, while **logarithm of price and logarithm of output have significant positive impact** on the logarithm of money.

Moreover, the intercept also appears to be highly significant, p-value < 0.05, deducing a logical interpretation that there is a constant demand for money even in the absence of explanatory variables.

d. Joint significance of parameters:

F-test is used to test for the joint significance of the parameters of all variables, implying whether explanatory variables (r_t, lp_t, lo_t) jointly have a significant effect on the dependent variable (lm_t).

The joint hypotheses can be expressed as:

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ (all coefficients are jointly insignificant) → Restricted Model

$H_1: H_0\ is\ false$, at least one $\beta_i \neq 0$ (at least one coefficient is significant) → Unrestricted Model

e. Testing the Economist's hypothesis:

The hypothesis is that the effect of logarithm output is half the effect of logarithm prices:

$H_0: \beta_3 = 0.5 * \beta_2$ or $\beta_3 - 0.5 * \beta_2 = 0$

$H_1: H_0\ is\ false, \beta_3 \neq 0.5 * \beta_2$

Where: $\beta_2\ is\ slope\ of\ lp_t$ and $\beta_3\ is\ slope\ of\ lo_t$

Substituting this restriction in the unrestricted model gives the restricted model as below:

Unrestricted Model → $lm_t = \beta_0 + \beta_1 r_t + \beta_2 lp_t + \beta_3 lo_t + u_t$

Substituting $\beta_3 = 0.5 * \beta_2$ → $lm_t = \beta_0 + \beta_1 r_t + \beta_2 lp_t + 0.5 * \beta_2 lo_t + u_t$

Restricted Model → $lm_t = \beta_0 + \beta_1 r_t + \beta_2 (lp_t + 0.5 * lo_t) + u_t$

Calculating the F-statistic:

$$F = \frac{(RRSS - URSS)/r}{URSS/(n-k-1)}$$

Input the values in equation, as below:

RRSS is Restricted Residual Sum of Squares = 1.760 (as given in Question 1.e)

URSS is Unrestricted Residual Sum of Squares = 1.754 (as given in Table 1)

df_1 → n-k-1 (476 in this case)

df_2 → r is the number of restrictions (1 in this case)

Critical value: $F^{\alpha}_{(df_1, df_2)} \sim F^{5\%}_{(476,\ 1)} \cong 3.86106838$

Condition: Reject the null, H_0, if $F_{(476,\ 1)} > F^{5\%}_{(476,\ 1)}$

Therefore:

$F_{(476,\ 1)} = \frac{(1.760 - 1.754)/1}{1.754/476}$ $\cong$1.62828 < Critical Value:$F^{5\%}_{(476,\ 1)}$ = 3.86106838

Since, the calculated F-statistic (1.628) is less than the critical value of (3.861), we **fail to reject the null hypothesis**.

Hence, based on the F-test, the data supports the economist's hypothesis at 5% significance level, that the effect of logarithm output is half the effect of logarithm prices on the logarithm demand for money.

f. Unit root:

(i) Unit Root Definition and Implications

The unit root is a way to test for non-stationarity of the time-series data. If a series contains a unit root, it is considered non-stationary and exhibits a statistical property known as stochastic trend having persistent effects from random shocks. It infers that the relationship is one-to-one between dependent and independent variable. If a data has unit-root (non-stationarity), it violates one of the key assumptions of standard regression analysis.

The main implications of the presence of unit roots (non-stationarity) in data on the estimated model are:

1. If the data contains a unit root, it implies that the process lm_t may never converge back to the equilibrium on long-run level.
2. If the variables in the regression model contain unit roots (are non-stationary) and are not cointegrated, the standard regression results may be spurious, leading to misleading inferences. Spurious regression occurs when unrelated non-stationary variables appear to be significantly related due to their common trends, rather than a true underlying relationship.
3. The statistical hypothesis tests, such as t-test, F-test, which are based on the assumptions of stationarity (constant mean, variances, covariances) may be invalid in the presence of unit roots. This can result in misspecification of model with misleading conclusions about the significance of estimated coefficients and the overall model fit.

(i) Augmented Dickey Fuller (ADF) test:

The Augmented Dickey-Fuller (ADF) test is a widely used statistical test to check for the presence of a unit root in a time-series. The ADF test adds the lag dependent variable in the equation, the lags are to remove any autocorrelation from the data. The number of lags is selected using one of the information criteria, such as AIC or BIC. Following are the steps to perform ADF test:

1. Specify the model:

As illustrated in Figure 1, the variables are a function of random walk with drift around stochastic trend, therefore:

$$\Delta y_t = \beta_0 + \beta_1 t + \delta y_{t-1} + \sum_{i=1}^{k} \theta_i \Delta y_{t-i} + u_t$$

Where:

$\Delta y_t = y_t - y_{t-1}$ is the first difference of time-series variables to make it stationary,

t is the time trend

δ is the persistence of shock [-1 < δ < 1] or $|\delta| < 1$, captures effect of previous period's value on current change,

$\theta_i \Delta y_{t-i}$ is the lagged value of first difference of variable,

k is the lag order of the autoregressive process, determined by using one of the information criteria, AIC, or BIC.

u_t is the error term.

1. State the hypotheses:

   The null hypothesis is that the time series has a unit root (non-stationarity). The alternative hypothesis is that the time series has no unit root (stationarity).

   $H_0: \delta = 0$ (there is a unit root)

   $H_1: \delta < 0$ (there is no unit root)
2. Estimate the regression equation using OLS method to obtain estimated coefficient ($\hat{\delta}$) and its standard error.
3. Compute the ADF test statistic. The ADF test is calculated in same way as the t-test, as below:

   $$ADF = \frac{\hat{\delta}}{Se(\hat{\delta})}$$

   Where $Se(\hat{\delta})$ is the standard error of estimated coefficient $\hat{\delta}$.
4. Compare the ADF test statistic with critical values from the Dickey-Fuller distribution. If the ADF test value is less than the critical value, we reject the null hypothesis, suggesting the no unit root. If ADF test is greater than the critical value, we fail to reject the null hypothesis, indicating to unit root in the data set.
5. If the null hypothesis is rejected, the time series is considered stationary. If the null hypothesis cannot be rejected, the time-series is non-stationary and further actions may be required to achieve stationary, such as differencing or detrending.

(ii) ADF test results

Based on the ADF test results provided in Table 2 (Coursework), we determine which variables contain unit root and which do not at 5% significance level.

1. Logarithm of money lm_t:

Test statistic = 0.048

Critical Value at 5% = -1.648

Since the ADF test statistic for lm_t (0.048) is greater than the critical value (-1.648), we fail to reject the null hypotheses of unit root. This indicates that variable lm_t has unit root and is non-stationary.

2. Logarithm of output lo_t:

   Test statistic = -0.909

   Critical Value at 5% = -1.648

Since the ADF test statistic for lo_t (- 0.909) is greater than the critical value (-1.648), we fail to reject the null hypotheses of unit root. This indicates that variable lo_t has unit root and is non-stationary.

3. Logarithm of prices lp_t:

   Test statistic = -0.258

   Critical Value at 5% = -1.648

Since the ADF test statistic for lp_t (- 0.258) is greater than the critical value (-1.648), we fail to reject the null hypotheses of unit root. This indicates that variable lp_t has unit root and is non-stationary.

4. Interest rate r_t:

   Test statistic = -2.612

   Critical Value at 5% = -1.648

Since the ADF test statistic for r_t (-2.612) is less than the critical value (-1.648), we reject the null hypotheses of unit root. This indicates that variable r_t does not contain unit root and is stationary.

Conclusively, based on the ADF test results, the **variables lm_t, lp_t, lo_t contain unit roots** and are non-stationary, while the **variable r_t does not contain unit root** and is stationary.

g. Durbin-Watson Statistic:

The Durbin-Watson (DW) is a test for first order autocorrelation, i.e. it assumes that there is relation between an error and the previous one. The null hypothesis tests that the residuals from an OLS are not autocorrelated, against the alternative hypothesis that states the residuals follow AR1 process (first order autocorrelation).

The DW statistic ranges from 0 to 4, with a value of 2 indicating no serial correlation in the errors (fail to reject the null). Values close to 0 or 4 suggest the presence of positive or negative autocorrelation, respectively.

The decision rule is:

1. If DW < d_L: there is positive serial correlation, reject H_0
2. If DW > $(4 - d_L)$: there is negative serial correlation, reject H_0
3. If d_U < DW < $(4 - d_U)$: there is no evidence of serial correlation, fail to reject H_0
4. If $d_L \leq DW \leq d_U$ or $(4 - d_U) \leq DW \leq (4 - d_L)$, the test is inconclusive.

The hypotheses for DW test are:

H_0: No autocorrelation

H_1: There is autocorrelation.

The DW test statistic can be calculated as:

$$DW = \frac{\sum_{t=2}^{T} u_t u_{t-1}}{\sum_{t=1}^{T} u_t^2}$$

From table 1, we have Durbin-Watson test = 0.122.

To interpret this DW statistic, compare it with the lower and upper critical values from the Durbin-Watson distribution table, which depend on the number of observations (n = 480), the number of explanatory variables (k = 3), and the significance level (α = 5%).

Lower critical value (d_L): 1.832

Upper Critical Value (d_U): 1.859

[Source: values are taken from row 450 and column 3 with Alpha .05 (Durbin-Watson Table, n.d.)]

In this case, the DW statistic value (0.122) is less than the lower critical value (1.832), which indicates to positive serial correlation, therefore, reject the null hypothesis of non-autocorrelation.

The presence of serial correlation violates the assumption of assumption of linear model, that assumes the errors are uncorrelated. The Newey-West test can be performed as next step, to correct for autocorrelation in the data.

Solution 2.

a. Mean, Variance and Covariances with iid process:

Given that $\varepsilon_t \sim iid(0, \sigma^2)$, means that the error term ε_t is independently and identically distributed with a mean of 0 and a constant variance of σ^2.

(i) $\boldsymbol{y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}}$

Where: $|\theta_i| < 1$, i = 1, 2

Mean:

If $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$

→ $E[y_t] = E[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}]$

$E[y_t] = E[\varepsilon_t] + \theta_1 E[\varepsilon_{t-1}] + \theta_2 E[\varepsilon_{t-2}]$

Since ε_t is iid process with mean 0, therefore, E[ε_t] = 0 for all t → ε_t, ε_{t-1}, ε_{t-2} = 0

Hence: $\boldsymbol{E[y_t] = 0}$

Variance:

$$Var[y_t] = Var[(y_t - E[y_t])]$$

Since $E[y_t] = 0$, therefore:

$$Var[y_t] = Var[y_t]$$

Substituting the equation of y_t

$$Var[y_t] = Var[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})]$$

Expanding the equation:

$$Var[y_t] = Var[\varepsilon_t] + {\theta_1}^2 Var[\varepsilon_{t-1}] + {\theta_2}^2 Var[\varepsilon_{t-2}] + 2\theta_1 cov[\varepsilon_{t,}\varepsilon_{t-1}] + 2\theta_2 cov[\varepsilon_t, \varepsilon_{t-2}] + 2\theta_1\theta_2 cov[\varepsilon_{t-1}, \varepsilon_{t-2}]$$

Since ε_t is iid with mean 0 and variance σ^2, therefore:

$Var[\varepsilon_t] = \sigma^2$,

$Var[\varepsilon_{t-1}] = \sigma^2$,

$Var[\varepsilon_{t-2}] = \sigma^2$

$cov[\varepsilon_t, \varepsilon_{t-1}] = cov[\varepsilon_t, \varepsilon_{t-2}] = cov[\varepsilon_{t-1}, \varepsilon_{t-2}] = 0$ (due to iid)

$$Var[y_t] = [\sigma^2 + {\theta_1}^2\sigma^2 + {\theta_2}^2\sigma^2 + 0 + 0 + 0]$$

Hence: $\boldsymbol{Var[y_t] = \sigma^2(1 + {\theta_1}^2 + {\theta_2}^2)}$

Covariance:

where j = 0:

$$cov[y_t, y_{t-j}] = Var[y_t]$$

$$\boxed{\boldsymbol{cov[y_t, y_{t-j}] = \sigma^2(1 + \theta_1{}^2 + \theta_2{}^2)}} \text{ for j = 0}$$

<u>where j = 1</u>:

$$cov[y_t, y_{t-j}] = cov[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2},\ \varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-1}] + cov[\varepsilon_t, \theta_1\varepsilon_{t-2}] + cov[\varepsilon_t, \theta_2\varepsilon_{t-3}] + cov[\theta_1\varepsilon_{t-1}, \varepsilon_{t-1}] + \mathrm{cov}[\theta_1\varepsilon_{t-1}, \theta_1\varepsilon_{t-2}] + cov[\theta_1\varepsilon_{t-1}, \theta_2\varepsilon_{t-3}] + cov[\theta_2\varepsilon_{t-2}, \varepsilon_{t-1}] + cov[\theta_2\varepsilon_{t-2}, \theta_1\varepsilon_{t-2}] + cov[\theta_2\varepsilon_{t-2},\ \theta_2\varepsilon_{t-3}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-1}] + \theta_1 cov[\varepsilon_t, \varepsilon_{t-1}] + \theta_2 cov[\varepsilon_t, \varepsilon_{t-3}] + \theta_1 cov[\varepsilon_{t-1}, \varepsilon_{t-1}] + \theta_1{}^2\, \mathrm{cov}[\varepsilon_{t-1}, \varepsilon_{t-2}] + \theta_1\theta_2 cov[\varepsilon_{t-1}, \varepsilon_{t-3}] + \theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-1}] + \theta_1\theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-2}] + \theta_2{}^2 cov[\varepsilon_{t-2},\ \varepsilon_{t-3}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-1}] + \theta_1 cov[\varepsilon_t, \varepsilon_{t-1}] + \theta_2 cov[\varepsilon_t, \varepsilon_{t-3}] + \theta_1 Var[\varepsilon_{t-1}] + \theta_1{}^2\, \mathrm{cov}[\varepsilon_{t-1}, \varepsilon_{t-2}] + \theta_1\theta_2 cov[\varepsilon_{t-1}, \varepsilon_{t-3}] + \theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-1}] + \theta_1\theta_2 Var[\varepsilon_{t-2}] + \theta_2{}^2 cov[\varepsilon_{t-2},\ \varepsilon_{t-3}]$$

Since $cov\left[\varepsilon_t, \varepsilon_{t-j}\right] = 0$ due to iid

Therefore:

$$= 0 + 0 + 0 + \theta_1 Var[\varepsilon_{t-1}] + 0 + 0 + 0 + \theta_1\theta_2 Var[\varepsilon_{t-2}] + 0$$

$$Var[\varepsilon_t] = Var[\varepsilon_{t-1}] = Var[\varepsilon_{t-2}] = \sigma^2$$

Hence:

$$cov[y_t, y_{t-j}] = \theta_1\sigma^2 + \theta_1\theta_2\sigma^2$$

$$\boxed{\boldsymbol{cov[y_t, y_{t-j}] = \theta_1\sigma^2(1 + \theta_2)}} \qquad \text{for j=1}$$

<u>where j = 2</u>:

$$cov[y_t, y_{t-j}] = cov[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2},\ \varepsilon_{t-2} + \theta_1\varepsilon_{t-3} + \theta_2\varepsilon_{t-4}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-2}] + cov[\varepsilon_t, \theta_1\varepsilon_{t-3}] + cov[\varepsilon_t, \theta_2\varepsilon_{t-4}] + cov[\theta_1\varepsilon_{t-1}, \varepsilon_{t-2}] + \mathrm{cov}[\theta_1\varepsilon_{t-1}, \theta_1\varepsilon_{t-3}] + cov[\theta_1\varepsilon_{t-1}, \theta_2\varepsilon_{t-4}] + cov[\theta_2\varepsilon_{t-2}, \varepsilon_{t-2}] + cov[\theta_2\varepsilon_{t-2}, \theta_1\varepsilon_{t-3}] + cov[\theta_2\varepsilon_{t-2},\ \theta_2\varepsilon_{t-4}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-2}] + \theta_1 cov[\varepsilon_t, \varepsilon_{t-3}] + \theta_2 cov[\varepsilon_t, \varepsilon_{t-4}] + \theta_1 cov[\varepsilon_{t-1}, \varepsilon_{t-2}] + \theta_1{}^2\, \mathrm{cov}[\varepsilon_{t-1}, \varepsilon_{t-3}] + \theta_1\theta_2 cov[\varepsilon_{t-1}, \varepsilon_{t-4}] + \theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-2}] + \theta_1\theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-3}] + \theta_2{}^2 cov[\varepsilon_{t-2},\ \varepsilon_{t-4}]$$

Since $cov\left[\varepsilon_t, \varepsilon_{t-j}\right] = 0$ due to iid

Therefore:

$$= 0 + 0 + 0 + 0 + 0 + 0 + 0 + \theta_2 Var[\varepsilon_{t-2}] + 0 + 0$$

$$Var[\varepsilon_t] = Var[\varepsilon_{t-1}] = Var[\varepsilon_{t-2}] = \sigma^2$$

Hence:

$$\boxed{\boldsymbol{cov[y_t, y_{t-j}] = \sigma^2}} \text{ for j=2}$$

where j = 3

$$cov[y_t, y_{t-j}] = cov[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}, \quad \varepsilon_{t-3} + \theta_1\varepsilon_{t-4} + \theta_2\varepsilon_{t-5}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-3}] + cov[\varepsilon_t, \theta_1\varepsilon_{t-4}] + cov[\varepsilon_t, \theta_2\varepsilon_{t-5}] + cov[\theta_1\varepsilon_{t-1}, \varepsilon_{t-3}] + \mathrm{cov}[\theta_1\varepsilon_{t-1}, \theta_1\varepsilon_{t-4}] + cov[\theta_1\varepsilon_{t-1}, \theta_2\varepsilon_{t-5}] + cov[\theta_2\varepsilon_{t-2}, \varepsilon_{t-3}] + cov[\theta_2\varepsilon_{t-2}, \theta_1\varepsilon_{t-4}] + cov[\theta_2\varepsilon_{t-2}, \ \theta_2\varepsilon_{t-5}]$$

$$= cov[\varepsilon_t, \varepsilon_{t-3}] + \theta_1 cov[\varepsilon_t, \varepsilon_{t-4}] + \theta_2 cov[\varepsilon_t, \varepsilon_{t-5}] + \theta_1 cov[\varepsilon_{t-1}, \varepsilon_{t-3}] + \theta_1{}^2 \mathrm{cov}[\varepsilon_{t-1}, \varepsilon_{t-4}] + \theta_1\theta_2 cov[\varepsilon_{t-1}, \varepsilon_{t-5}] + \theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-3}] + \theta_1\theta_2 cov[\varepsilon_{t-2}, \varepsilon_{t-4}] + \theta_2{}^2 cov[\varepsilon_{t-2}, \ \varepsilon_{t-5}]$$

Since $cov\left[\varepsilon_t, \varepsilon_{t-j}\right] = 0$ due to iid

Therefore:

$$\boxed{\boldsymbol{cov[y_t, y_{t-j}] = 0}}$$ for j>2

(ii) $\boldsymbol{w_t = w_{t-1} + \varepsilon_t}$

Where: $w_0 = 0$

Mean:

If $w_t = w_{t-1} + \varepsilon_t$

$$E[w_t] = E[w_{t-1}] + E[\varepsilon_t]$$

At, t = 1 $E[w_1] = E[w_0] + E[\varepsilon_1] = 0$ as $w_0 = 0$ and $E[\varepsilon_t] = 0$ (given)

At, t = 2 $E[w_2] = E[w_1] + E[\varepsilon_2] = 0$ as $w_1 = 0$ and ε_t is 0 due to iid

Hence $\boldsymbol{E[w_t] = 0}$ at all values of t

Variance:

$$Var[w_t] = Var[(w_t - E[w_t])^2]$$

Substituting equation of w_t and E[w_t] = 0:

$$Var[w_t] = Var[(w_{t-1} + \varepsilon_t - 0)] = Var[(w_{t-1} + \varepsilon_t)]$$

$$Var[w_t] = Var[w_{t-1}] + 2cov[w_{t-1}, \varepsilon_t] + Var[\varepsilon_t]$$

Since, $cov[w_{t-1}, \varepsilon_t] = 0$ due to iid, and $Var[\varepsilon_t] = \sigma^2$

$$Var[w_t] = Var[w_{t-1}] + \sigma^2$$

Put $Var[w_{t-1}] = Var[(w_{t-2} + \varepsilon_t)] \rightarrow Var[w_{t-1}]$

$$Var[w_t] = Var[w_{t-2}] + 2cov[w_{t-2}, \varepsilon_t] + Var[\varepsilon_t] + \sigma^2$$

By following the same conditions of iid, we get:

$$Var[w_t] = Var[w_{t-2}] + \sigma^2 + \sigma^2$$

Therefore at w_0, we get:

$$Var[w_t] = Var[w_0] + \sigma^2 + \cdots + \sigma^2$$

Since $w_0 = 0$

$$Var[w_t] = t\,\sigma^2$$

Hence: $\boldsymbol{Var[w_t] = t\sigma^2} < \infty$

Covariance:

$$cov[w_t, w_{t-j}] = cov[w_{t-1} + \varepsilon_t, w_{t-j}]$$

At j=0,

$cov[w_t, w_{t-j}] = cov[w_t, w_{t-0}] = cov[w_t, w_t] = Var[w_t]$

Substituting value of $Var[w_t]$ from above equation, therefore:

$\boxed{cov[w_t, w_{t-j}] = t\sigma^2}$ at j = 0

At j=1,

$cov[w_t, w_{t-j}] = cov[w_{t-1} + \varepsilon_t, w_{t-j}] = cov[w_{t-1} + \varepsilon_t, w_{t-1}] = cov[w_{t-1}, w_{t-1}] + cov[\varepsilon_t, w_{t-1}]$

$= Var[w_{t-1}] + 0$

As assumed above $w_t \sim (0, \delta)$, therefore: $\boxed{cov[w_t, w_{t-1}] = t\sigma^2}$

At j > 1,

$cov[w_t, w_{t-j}] = cov[w_{t-1} + \varepsilon_t, w_{t-j}] = cov[w_{t-1} + \varepsilon_t, w_{t-2}] = cov[w_{t-1}, w_{t-2}] + cov[\varepsilon_t, w_{t-2}] = 0$

Therefore: $\boxed{cov[w_t, w_{t-j}] = 0}$ for j > 1

b. Identifying weakly stationary model:

To determine if a time series is weakly stationary (covariance stationary), the following conditions must be satisfied:

1. The mean of the series should be constant and does not depend on time;
2. The variance of the series should be constant and does not depend on time;
3. The covariance between two values in the series should depend only on the time lag between them and not on the specific time periods.

For Model (i): $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$

- Mean: $E[y_t] = 0$ is constant
- Variance: $Var[y_t] = \sigma^2(1 + {\theta_1}^2 + {\theta_2}^2)$ is constant
- Covariances: $cov[y_t, y_{t-j}] = \begin{cases} \sigma^2(1 + {\theta_1}^2 + {\theta_2}^2) & at\ j = 0 \\ \theta_1\sigma^2(1+\theta_2) & at\ j = 1 \\ \sigma^2 & at\ j = 2 \\ 0 & at\ j > 2 \end{cases}$

Since, the mean and variance are constant, and the covariances depend only on the time lag j (and not on t), this series satisfies the conditions of weak stationarity.

For Model (ii): $w_t = w_{t-1} + \varepsilon_t$

- Mean: $E[w_t] = 0$ is constant
- Variance: $Var[w_t]$ is not constant, depend upon t (time variable of w_t)
- Covariances: $cov[w_t, w_{t-j}] = \begin{cases} t\sigma^2 & at\ j = 0 \\ t\sigma^2 & at\ j = 1 \\ 0 & at\ j > 1 \end{cases}$ depending on time lag

While the mean is constant and the covariances depend only on the time lag j, the variance is not constant and depend on t. Since, the condition of constant variance is violated, this series is not weakly stationary.

c. Autocorrelation:

(i) First order autocorrelation:

To determine if the regression suffers from firs-order autocorrelation, the Durbin-Watson statistic can be used.

The DW statistic is given as 1.20.

According to decision rule for the DW test for first order autocorrelation, as mentioned in Solution 1(g) above:

- If DW is approximately equal to 2, there is no first order autocorrelation.
- If DW is close to 0, there is positive first-order autocorrelation.
- If DW is close to 4, there is negative first-order autocorrelation.

- If the test statistic value lies between dL and dU, the test is inconclusive.

Given that, n = 90, k = 2, and alpha = 0.05, the critical values are:

dL = 1.612

dU = 1.703

Since the DW test statistic is less than the lower critical value dL (1.20 < 1.612), therefore, it can be stated that the given regression suffers from positive first order autocorrelation.

(ii) Second order autocorrelation:

The Lagrange Multiplier (LM Test) is the more general test for autocorrelation of any order in the regression residuals – r^{th} order autocorrelation. It is expressed with an auxiliary regression as:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_r \varepsilon_{t-r} + \upsilon_t$$

Where:

ε_t is residual,

ρ_r is the coefficient of autocorrelation,

ε_{t-r} are the lagged residuals up to order, r = 1, 2, 3 ...

$\upsilon_t \sim$ iid(0, σ^2)

The test for second order correlation is carried out as follows:

1. Estimate the original linear regression model using OLS and obtain the residuals $\widehat{\varepsilon_t}$

The original given model is:

$$\widehat{b_t} = 0.49 - 0.27 p_t + 0.22 y_t$$

2. Regress the residuals $\hat{e}_t$ on the original explanatory variables p_t and y_t and lag the residuals upto order 2 [ε_{t-1} and ε_{t-2}], because we require second order correlation.
3. Run the auxiliary regression and test the null hypothesis of no autocorrelation against the alternative. The auxiliary equation and hypotheses are:

$$\widehat{\varepsilon_t} = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \upsilon_t$$

   H_0: $\rho_1 = 0$ and $\rho_2 = 0$ (No autocorrelation)
   H_1: $\rho_1 \neq 0$ or $\rho_2 \neq 0$ (Autocorrelation, at least one $\rho \neq 0$)

4. Obtain the R-squared (R^2) from this auxiliary regression.
5. The test value can be shown as:

$$LM = (T - r)R^2 \sim \chi^2(r)$$

   Where:
   T = number of observations,
   r is the order,
   and $\chi^2(r)$ is the chi-square critical value at r order (df = 2 in this case).
6. If the LM test statistic exceeds the critical value from chi-square table, reject the null hypothesis of no second-order autocorrelation.

In the given output, the LM(2) value is = 7.42

The chi-square distribution with 2 degrees of freedom at 5% significance level: $\chi^2(2)$ = 5.9915

Since, LM(2) = 7.42 > $\chi^2(2)$ = 5.9915 , therefore, **reject the null hypothesis** of no autocorrelation.

Hence, **the given model $\widehat{b_t}$ suffers from second order autocorrelation**.

Solution 3.

a. Models with Cross-sectional dataset

Model 1:

$$\hat{y}_i = 2.3 + 0.2x_{1i} - 6.14x_{2i} - 0.01x_{3i} + 1.5x_{4i} \qquad \text{(Unrestricted Model)}$$

Model 2:

$$\hat{y}_i = 2.6 + 0.25x_{1i} - 6.14x_{2i} \qquad \text{(Restricted Model)}$$

Where:

y_i = rate of return (ROE) for the firm

x_1 = market share

x_2 = measure of firm size

x_3 = industrial growth rate

x_4 = level of world trade in industrial products

(i) Examine the joint hypothesis.

Step 1: State the null and alternative hypotheses:

To examine the joint hypothesis based on given conditions that the coefficients x_3 and x_4 are both equal to zero, the following hypothesis can be stated:

H_0: $\beta_3 = 0 \; and \; \beta_4 = 0$

H_1: $\beta_3 \neq 0 \; or \; \beta_4 \neq 0$

The null hypothesis states that coefficients of x_3 (industrial growth rate) and x_4 (level of world trade) do not have significant impact on ROE, while alternative hypothesis indicates that at least of the variables x_3 and x_4 has significant impact on ROE of the firm.

Step 2: Calculate the F-statistic.

$$F = \frac{(RRSS - URSS)/_r}{URSS/_{(n-k-1)}}$$

Where:

RRSS is the restricted sum of squared residuals = 0.70 (from Model 2)

URSS is the unrestricted sum of squared residuals = 0.60 (from Model 1)

r is the number of restrictions = 2 (in this case)

n is the number of observations = 50 (number of firms, given)

k is the number of slopes in the unrestricted model = 4

By substituting the values in F-test equation:

$$F_{(45,\ 2)} = \frac{(0.70 - 0.60)/_2}{0.60/_{(50-4-1)}} \qquad = 0.375$$

Step 3: Determine the F-critical value at 5% significance:

df_1 = n-k-1 = 50-4-1 = 45, df_2 = r = 2, α = 0.05

$F^{5\%}_{(45,\ 2)} \sim$ 3.2043

Step 4: Interpret the results:

Compare the test value. If the calculated F-test statistic is greater than the critical value of F-distribution, reject the null hypothesis, which means that at least one of the coefficients on x_3 or x_4 is not equal to zero, and these variables should be included in model. However, the F-statistic value (0.375) is less than the F-critical value, therefore, we **fail to reject the null hypothesis**, suggesting that the coefficients x_3 and x_4 are jointly insignificant, and these variables can be excluded from the model. This indicates that the industrial growth rate and level of world trade do not have significant impact on the firm's ROE, therefore the restricted model can be used to estimate the firm's ROE.

(ii) Testing whether financial firms have on average higher ROE:

To assess whether the financial firms have, on average, higher ROE, a dummy intercept variable (D_i) can be incorporated in the model to compare it with the ROE of non-financial firms.

Step 1: Define the two groups:

Treatment Group: Financial firms → D_i = 1

Control Group: Non-financial firms → D_i = 0

Step 2: Regress the model with dummy variable:

Suppose adding the dummy variable to the restricted model:

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} - \beta_2 x_{2i} + \beta_5 D_i + u_i$$

Where:

β_3 and β_4 have been proven as insignificant above,

β_5 indicates the coefficient of dummy variable.

Step 3: Formulate the hypotheses:

H_0: $\beta_5 = 0$ (The mean ROE for financial firms is equal to the mean ROE for non-financial firms)

H_1: $\beta_5 > 0$ (The mean ROE for financial firms is greater than the mean ROE for non-financial firms)

Step 4: Estimate the model's coefficients and calculate the appropriate test statistic, such as t-test, and p-value for the dummy variable coefficient (β_5).

Step 5: Compare the t-statistic with t-critical value at given significance level of 5%.

- If the t-test statistic is greater than t-critical value (p-value < α), then reject the null hypothesis and conclude that there is sufficient evidence to support that financial firms have a higher mean, on average, than non-financial firms.
- If the t-test statistic is less than t-critical value (p-value > α), then fail to reject the null hypothesis and conclude that there is no sufficient evidence to indicate a difference in mean ROE between financial and non-financial firms.

Step 4: If the null hypothesis is rejected, then the coefficient β_5 will measure the average difference in ROE between financial and non-financial firms, ceteris paribus.

ROE of financial firm, where $D_i = 1$:

$$\hat{y}_i = (\beta_0 + \beta_5) + \beta_1 x_{1i} - \beta_2 x_{2i} + u_i$$

ROE of non-financial firm, where $D_i = 0$:

$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} - \beta_2 x_{2i} + u_i$$

b. Test for the 4th order ARCH effects:

The presence of Autoregressive conditional Heteroskedasticity (ARCH) effects indicates that the volatility of the stock market index is time-varying and depends on its own past values.

ARCH model describes the variance of the current error term as a function of the actual previous time period error terms, which is defined as the squares of the previous innovations. To test for the presence of 4th order ARCH effects in a stock market index, the LM test can be used as in the following steps.

Step 1: Estimate the following conditional mean equation using OLS regression:

The ARCH test models variances using ARIMA to pick up conditional heteroskedasticity on variances. Consider following models for the 4th order ARCH effect:

Conditional Mean ARCH(4):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + u_t$$

Where:

y_t is the stock market index at time t

ϕ_0 is constant

ϕ_i is the coefficient for the lagged values of y_t; i=1,2,3,4

u_t is the error term

Step 2: Obtain the residuals u_t from the mean equation in step 1.

Step 3: Square the residuals to get $\hat{u}_t^2$.

Step 4: Regress $\hat{u}_t^2$ on a constant and its own past values up to 4th order (4 lags) by estimating an auxiliary regression:

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 \hat{u}_{t-1}^2 + \alpha_2 \hat{u}_{t-2}^2 + \alpha_3 \hat{u}_{t-3}^2 + \alpha_4 \hat{u}_{t-4}^2 + v_t$$

Where:

α_0 is constant

α_i are the coefficient for the lagged squared residuals

v_t is the error term

Obtain the R^2 of the regression.

Step 5: Specify the hypotheses:

H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ (no ARCH effect)

H_1: $\alpha_i \neq 0$ (There is an ARCH effect)

Step 6: Construct the LM statistic:

$$LM = nR^2 \sim \chi^2_{(\alpha,\ k)}$$

Calculate LM statistic and compare with χ^2 distribution at degree of freedom = 4 (due to 4 lags), and significance at 5%.

Step 4: Decision:

Reject the null hypothesis of no ARCH effects if $LM = nR^2$ is > the chi-square critical value $\chi^2_{(5\%,\ 4)}$, indicating presence of 4th order ARCH effect in the stock market index, meaning the index depends on its own past values.

Fail to reject the null hypothesis if $LM = nR^2$ is < the chi-square critical value $\chi^2_{(5\%,\ 4)}$, indicating no 4th order ARCH effect in the stock market index.

c. FTSE100 index monthly data

$$\hat{y}_t = 0.75 \qquad \rightarrow \text{Conditional Mean}$$

$$\hat{\sigma}_t^2 = 1.76 + 0.07u_{t-1}^2 + 0.24\sigma_{t-1}^2 + 0.87u_{t-1}^2 I_{t-1} \qquad \rightarrow \text{Conditional Variance}$$

(i) Interpretation of the final term of the conditional variance:

The final term of the conditional variance ($0.87u_{t-1}^2 I_{t-1}$) captures the asymmetric effect in the market. If the slope is positive and statistically significant then it captures the leverage effect, meaning the bad news have larger or different effect on volatility than good news of the same magnitude.

Where:

u_{t-1}^2 represents the squared residual from the previous period

I_{t-1} is an indicator variable, that takes value 0 when good news (residual $u_{t-1}$from the previous period is positive), and 1 for bad news (residual $u_{t-1}$from the previous period is negative)

The coefficient of 0.87 measures the differential impact indicating to higher impact of negative shocks on the volatility of stock market than the good news.

Good news: $I_{t-1} = 0$

$= \ 0.07u_{t-1}^2 + 0.87u_{t-1}^2 * 0 \quad = \ 0.07u_{t-1}^2$ → The magnitude of shock is 0.07

Bad news: $I_{t-1} = 1$

$= \ 0.07u_{t-1}^2 + 0.87u_{t-1}^2 * 1 \quad = 0.94u_{t-1}^2$ → The magnitude of shock is 0.94

To test the significance of asymmetric effect:

H_0: Asymmetric effect is not significant

H_1: Asymmetric effect is significant

$$\hat{t} = \frac{\hat{\beta}_3}{Se(\hat{\beta}_3)} = \frac{0.87}{0.42} = 2.071$$

t-critical value = $t_{(\frac{\alpha}{2}, df)}$ = $t_{(0.025,152)}$ ~ 1.9757

df = n-k-1 = 156-3-1

n = 13*12 = 156 (monthly data from January 1992 to December 2005)

k = 3

α = 0.05

Since t-statistic (2.071) > t-critical value (1.9757), we reject the null hypothesis, indicating the asymmetric effect is statistically significant, and bad news has 0.87 times higher impact on the index.

(ii) Interpretation of estimated effect of lagged conditional variance (σ_{t-1}^2)

The coefficient of σ_{t-1}^2 is 0.24 that represents the degree of persistence of shock and is captured by the GARCH effect. The GARCH model captures the ARCH infinity.

A value of coefficient of σ_{t-1}^2 close to 1 implies a higher degree of persistence, means the shock lasts for a long time, or in other words, the past volatility has a long-lasting effect on current volatility. A value close to 0 implies a low persistence, where the past volatility has a relatively short-lived effect.

In this case, with an estimated value of 0.24, there is a **moderate degree of persistence** in the volatility process. This indicates that the past volatility has a positive but decaying impact on current volatility over time.

(iii) Measuring positive and negative shock at ± 0.5

$$\hat{\sigma}_t^2 = 1.76 + 0.07u_{t-1}^2 + 0.24\sigma_{t-1}^2 + 0.87u_{t-1}^2 I_{t-1}$$

Estimating value of $\hat{\sigma}_t^2$:

For positive shock:

Substitute the below values in above equation:

$\sigma_{t-1}^2 = 0.74$

$\hat{u}_{t-1} = +0.5$

$I_{t-1} = 0$

$$\hat{\sigma}_t^2 = 1.76 + 0.07(0.5)^2 + 0.24 * 0.74 + 0.87(0.5)^2 * 0$$

$\hat{\sigma}_t^2 = 1.9551$ → volatility when there is positive shock

For negative shock:

Substitute the below values in above equation:

$\sigma_{t-1}^2 = 0.74$

$\hat{u}_{t-1} = -0.5$

$I_{t-1} = 1$

$$\hat{\sigma}_t^2 = 1.76 + 0.07(-0.5)^2 + 0.24 * 0.74 + 0.87(-0.5)^2 * 1$$

$\hat{\sigma}_t^2 = 2.1726$ → risk when there is negative shock