# MSc in Financial Technology

# Quantitative Financial Modelling Module

# Regression & Statistical Analysis


Group Members:

1. Gauri Shingane (20018204)
2. Roshni Dalvi (20012887)
3. Taarandeep Singh (20011757)
4. Sahil Punjabi (20004846)
5. Lakshitha Jayasinghe (20020502)
6. Krishnan Sanjeev Nair (20003928)

Inflation, as an economic indicator plays a pivotal role in shaping the fiscal policies of any nation. Ireland, one of the most rapidly evolving economies within the Eurozone, experiences fluctuations in inflation that are subject to multiple influencing factors. This analysis aims to explore and model the relationship between Inflation (dependent variable) in Ireland and set independent variables that are critical determinants to explain the variation of Inflation.

The independent variables selected for this analysis encompass a range of economic, financial, and external factors of which the most significant variables being Gross Domestic Product (GDP), Trade Balance, Unemployment rates, Industrial Production and VAT.

To make the model accurate and robust we have dropped some other variables that might impact the Inflation of Ireland like Property Price, Consumer Price Index, Consumer Spending etc however there will be a strong Serial Correlation in our Regression Model. The outcome of the model and its variables are explained below.

## ADJUSTED R-SQUARED EXPLAINED:

| *Regression Statistics* | |
|---|---|
| **Multiple R** | 0.778709157 |
| **R Square** | 0.606387951 |
| **Adjusted R Square** | 0.557186445 |
| **Standard Error** | 3.400457486 |
| **Observations** | 46 |

**>> R Square = Residual Sum of Square (RSS) / Total Sum of Square (TSS)**

**>> RSS/ TSS**

**>> 712.55 / 1175.076**

**>> 0.606 or 60%**

In the model, the R Square specifies 60% variability of our dependent variable – Inflation is explained by 5 independent variables/regressors in our model, i.e., GDP, Trade Balance, Unemployment, Industrial Production, and VAT. And remaining 45% of movement is dependent on the error terms/residuals.

The Adjusted R-squared value of 55% specifies goodness of fit and we can infer the model captures a substantial portion of the variability in the dependent variable. The model can predict the dependent variable reasonably well, explaining more than half of the observed variance.

## OLS EQUATION AND BETA COEFFICIENT EXPLANATION:

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| **Intercept** | 6.136330 | 4.309108 | 1.424037 | 0.162190 |
| **GDP** | -0.000225 | 0.000039 | -5.807451 | 0.000001 |
| **Trade Balance** | 0.073434 | 0.023229 | 3.161301 | 0.002992 |
| **Unemployment** | -0.707410 | 0.145028 | -4.877737 | 0.000017 |
| **Industrial Prod** | -0.149683 | 0.066140 | -2.263109 | 0.029127 |
| **VAT** | 0.546961 | 0.170145 | 3.214680 | 0.002584 |

(All values in per cent change)

Y – Inflation

X1 – Gross Domestic Product (GDP)

X2 – Trade Balance

X3 –Unemployment

X4– Industrial Production

X5 – Value Added Tax (VAT)

**Equation:**

**Y = β0 + β1X1 + β2X2 + β3X3 + β4X4 + β5X5**

INFLATION = β0 + β1(GDP) + β2(TRADE BALANCE) + β3(UNEMPLOYMENT) + β6(INDUSTRIAL PROD) + β7(VAT) + ε

Here,

**β0** is the intercept

**β1- β5** values are the regression coefficients or variables for the five independent variables in our regression.

**$\varepsilon$** is the error term that accounts for the variability in Y which is not explained by the regressors.

So, the final equation is,

$\hat{Y}$ = 6.1363– 0.000225 X1 + 0.0734 X2– 0.7074 X3 – 0.1496 X4+ 0.5469 X5

**<u>Beta Coefficients explanation:</u>**

$\beta_0$ – Intercept value - 6.1363

$\beta_1$ – For every per cent change in GDP, the value of inflation will decrease by 0.022%, while holding other variables constant

$\beta_2$ – For every per cent change in Trade Balance, the value of inflation will increase by 7.34%, while holding other variables constant

$\beta_3$ - For every per cent change in Unemployment, the value of inflation will decrease by 70.74%, while holding other variables constant

$\beta_4$ - For every per cent change in Industrial Production, the value of inflation will decrease by 14.97%, while holding other variables constant

$\beta_5$ - For every per cent change in VAT, the value of inflation will increase by 54.69%, while holding other variables constant

## SIGNIFICANCE TEST

| ANOVA | | | | | |
|---|---|---|---|---|---|
| | ***df*** | ***SS*** | ***MS*** | ***F*** | ***Significance F*** |
| **Regression** | 5 | 712.5525008 | 142.5105002 | 12.32458105 | 2.90258E-07 |
| **Residual** | 40 | 462.5244446 | 11.56311112 | | |
| **Total** | 45 | 1175.076945 | | | |

Based on the ANOVA table,

At $\alpha = 0.05$, i.e., 95% confidence interval, we carry out the F-Test for the overall significance of the relationship between variables of our regression model.

$H_0$ - Null Hypothesis

$H_0 : \beta 1 = \beta 2 = \beta 3 = \beta 4 = \beta 5 = \beta 6 = \beta 7 = 0$

The regression coefficients of all seven variables of our regression model are zero, i.e., there is no significant relationship between the dependent variable and any of the independent variables.

$H_1$ – Alternative Hypothesis

$H_1$ : At least one $\beta_i \neq 0$

At least one of the coefficients is not zero, i.e., at least one of the independent variables is linearly related to the dependent variable.

We know that,

F-stat = 12.32458105

Critical value $F_\alpha$ (5, 40) ≈ 2.45

Rejection Region: F-stat > 2.45

Here,

$$9.751351827 > 2.45$$

i.e., our F-stat value is greater than the Critical value and hence we reject the Null Hypothesis using the F-Test.

Also,

$$6.84166E\text{-}07 < 2.45$$

i.e., calculated P-value is less than the Critical value, we again reject the Null Hypothesis by the P-value significance test.

In essence, both methods lead to the same conclusion.

The overall model is significant or valid, i.e., there is a significant relationship between the dependent variable and at least one independent variable.

## SERIAL CORRELATION

Serial correlation explains the relation between the variable's current value and the past values of the variable. The serial correlation happens when the current values of a variable are not independent of the past values in a time series data set.

If the serial correlation calculates to zero then there is no correlation in the time series data and the observations are independent, whereas if the correlation value nears value of one then the observations are serially correlated and the previous observations affect the current observations.

For instance, if you are examining monthly sales data for a product and the sales figures in one month are positively correlated with the sales figures from the previous month, it indicates a positive serial correlation, i.e, if the sales were positive last month they will most probably remain positive this month as well. On the other hand, if there's a negative correlation, it suggests a negative serial correlation, i.e, if the sales were positive last month there is high possibility that the sales will fall this month.

Serial correlation can be measured using various statistics like Durbin Watson test. According to the Durbin Watson test, the value of between zero and less than two indicates no postitive correlation whereas values between two and four indicates negative correlation.

## DURBIN WATSON STATISTICAL ANALYSIS OF MODEL:

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

As per our calculations,

**Durbin Watson statistics** = 1.03950061

**dL=**1.287

**dU=**1.776

**OLS Assumption 4,** emphasize that Error Terms are Uncorrelated to each other i.e. No Serial or Auto Correlation among error terms. To establish the assumption, we have conducted Durbin Watson test of our model. With reference to the above chart, DW statistics is 1.039 and as per the DW Table with a 5% Significance level with n=46 and k=5 (number of regressor in the model) the dL = 1.287 & dU = 1.776. We can observe our DW statistical value is 1.039 which is below the dL value hence we can conclude there is **NO SERIAL CORRELATION or AUTOCORRELATION among the RESIDUALS.**

**REFERENCES FOR DATASETS:**

https://www.cso.ie/en/statistics/

https://data.ecb.europa.eu/data/datasets

https://www.centralbank.ie/statistics/data-and-analysis

https://population.un.org/wpp/

https://fred.stlouisfed.org

https://www.revenue.ie/en/vat/vat-rates/historical-vat-rates/index.aspx