HUL315 Assignment-4

Amaiya Singhal 2021CS50598

1 Estimation

1.1 Task

Investment refers to the purchase or construction of capital goods, including residential and nonresidential buildings, equipment and software used in production, and additions to inventory stocks. It plays a crucial role in determining the long-run productive capacity of the economy.

Our aim is to examine whether investors primarily focus on the real interest rate. We will be performing linear regression on data collected by the Reserve Bank of India (available in the RBI Database on Indian Economy).

1.2 Dataset and Variables

We have been provided with datasets from the RBI Database on Indian Economy which include yearly data on Consumer Price Index, Real GDP, Call Money Rate and Real Investment. Data of the years 1970-11 to 2017-18 are used in the regression. All the data was converted to the base year 2011-12.

1.2.1 Variables

The subscript "t" denotes that the values are for a particular year.

- I_t (**Real Investment**) : Data on Gross Domestic Capital Formation (GDCF) at constant prices was used for this variable.
- i_t (**Nominal Interest Rate**) : Call Money Rate was used as a proxy for this variable.
- CPI_t (**Consumer Price Index**) : The change in logarithm of CPI is used as a measure of inflation.
- Y_t (**real GDP**) : Gross Domestic Product at constant prices was used for this variable.
- t (**Time**) : Since the base year was 2011-12. We have taken t = 0 for the year 2011-12 and accordingly set the values for other years. 2012-13 has t = 1, 2013-14 has t = 2 . . . , similarly 2010-11 has t = -1, 2009-10 has t = -2 and so on.

1.2.2 Regression Equation

$$\ln I_t = \beta_0 + \beta_1 \ln i_t + \beta_2 \Delta \ln CPI_t + \beta_3 \ln Y_t + \beta_4 t + u_t$$

1.2.3 Converting to a Common Base Year

The data collected from the RBI Database on Indian Economy had values with different base years for data of different years. This poses a challenge since we want all the values to be in the same base year to be able to perform a correct regression and assess whether our hypothesis can be rejected or not.

The method we have used to convert the data to a common base year is as follows:

$$Y_t = X_t . \frac{Y_{t+1}}{X_{t+1}}$$

where Y_t represents the data at time t in the base year that we want to convert the data to and X_t is the data at time t at a different base year. Assuming we have data for both base years available for time $t+1$, we can use this equation to calculate the the data at time t in the other base year.

Similarly if we had the data in both base years at time $t-1$ and wanted to calculate the values at time t, we used

$$Y_t = X_t . \frac{Y_{t-1}}{X_{t-1}}$$

Applying this sequentially to all the data points for a particular variable, we converted all the values to the common base year 2011-12. Once all the data was cleaned up and changed to the base year 2011-12, we performed linear regression using the above equation to estimate the coefficients.

- Cleaned up data : Link
- Python Notebook : Link

1.3 Results

1.3.1 Coefficient Estimates

The estimates of the coefficients obtained from the regression are as follows:

Coefficient	Estimate
$\hat{\beta}_0$	-15.9069
$\hat{\beta}_1$	-0.0581
$\hat{\beta}_2$	0.6129
$\hat{\beta}_3$	1.9337
$\hat{\beta}_4$	-0.0325

1.3.2 Variance - Covariance Matrix

The covariance matrix of the estimated coefficients is shown below. These values are used in the next section for performing hypothesis testing.

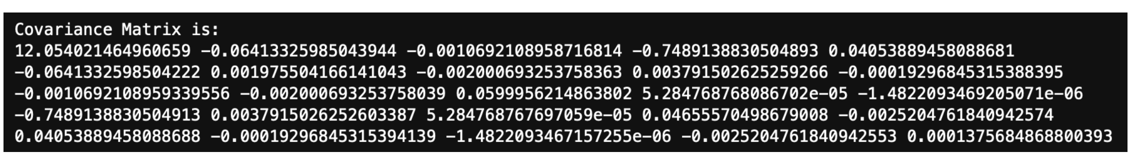
```
Covariance Matrix is:
12.054021464960659 -0.06413325985043944 -0.0010692108958716814 -0.7489138830504893 0.040538894580886881
-0.0641332598504222 0.001975504166141043 -0.002000693253758363 0.003791502625259266 -0.00019296845315388395
-0.0010692108959339556 -0.002000693253758039 0.0599956214863802 5.284768768086702e-05 -1.4822093469205071e-06
-0.7489138830504913 0.0037915026252603387 5.284768767697059e-05 0.04655570498679008 -0.0025204761840942574
0.04053889458088688 -0.00019296845315394139 -1.4822093467157255e-06 -0.0025204761840942553 0.0001375684868800393
```

Figure 1: Covariance Matrix of Estimated Parameters

1.3.3 Interpretation

- $\hat{\beta}_0$ (Constant) : This represents the intercept or baseline of the regression model is -15.9069 i.e. the value when all the independent are set to 0.
- $\hat{\beta}_1$ (Coefficient of lni_t) : Represents that the percentage change in real investment (I_t) is -0.0581 per percentage change in the nominal interest rate (i_t). However upon calculating, the p-value of 0.198 is more than 0.05 (at 5 % significance level), hence we do not have sufficient statistical evidence to claim that $\beta_1 \neq 0$.
- $\hat{\beta}_2$ (Coefficient of ΔlnCPI_t) : Represents that the percentage change in real investment (I_t) is 0.6129 per percentage change in inflation (measured with change in log of CPI). Upon calculation, the p-value of 0.016 is less than 0.05 (at 5% significance level), we reject the hypothesis that $\beta_2 = 0$.
- $\hat{\beta}_3$ (Coefficient of lnY_t) : Represents that the percentage change in real investment (I_t) is 1.9337 per percentage change in log of GDP. Using the p-value, we also reject the hypothesis that $\beta_3 = 0$.
- $\hat{\beta}_4$ (Coefficient of t) : Represents that the percentage change in real investment (I_t) is -0.0325 per unit change in t (i.e. per year). The p-value of 0.008 is less than 0.05 (at 5% significance level), we reject the hypothesis that $\beta_4 = 0$.

Hence we can statistically conclude that inflation, GDP and time play a role in real investment. However we do not have enough statistical evidence to say the same about nominal interest rate.

2 Hypothesis Testing

We want to test the hypothesis that $\beta_1 = -\beta_2$ to examine whether investors primarily focus on the real interest rate. We first define the Null Hypothesis and the Alternate Hypothesis.

- β_1 and β_2 are the regression coefficients of the equation in 1.2.2.
- **Null Hypothesis (H_0)**: $\beta_1 + \beta_2 = 0$
- **Alternate Hypothesis (H_1)**: $\beta_1 + \beta_2 \neq 0$

Calculations

We now calculate the t-test statistics to decide whether or not to reject the null hypothesis.

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2}{s.e.(\hat{\beta}_1 + \hat{\beta}_2)}$$

The value of $s.e.(\hat{\beta}_1 + \hat{\beta}_2)$ is calculated below

$$s.e.(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{s.e.(\hat{\beta}_1)^2 + s.e.(\hat{\beta}_1)^2 + 2.\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$$
$$s.e.(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{0.001976 + 0.059996 + 2 * (-0.002001)}$$
$$s.e.(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{0.05797}$$
$$s.e.(\hat{\beta}_1 + \hat{\beta}_2) = 0.2408$$

Using the above calculated value of standard error, we calculate the t - statistic.

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2}{s.e.(\hat{\beta}_1 + \hat{\beta}_2)}$$
$$t = \frac{(-0.0581) + (0.6129)}{0.2408}$$

$$t = \frac{0.5548}{0.2408} = 2.304$$

The degrees of freedom are

$$n - k - 1 = 46 - 4 - 1 = 41$$

Result

For 5% significance level, at $\alpha = 0.05$, the critical value is 2.02 for degrees of freedom = 41 (2 tailed test).

Since $2.304 > 2.02$, we reject the null hypothesis. Hence we conclude that the alternative hypothesis stands and that $\beta_1 \neq \beta_2$.