Causality Analysis for SENSEX and CPI in the Indian Economy

Zehaan Naik

July 2024

Contents

1 Motivation

In the SWAP model, as shown in Chow et al. (2015), we describe an intuition of two quantities being co-dependent on each other in a manner that their relationship can be modelled through a doubly stochastic model. This report aims to establish a sense of bi-directional causality between SENSEX and CPI in the Indian economy, which would be an example of such data.

The BSE SENSEX (the S&P Bombay Stock Exchange Sensitive Index or simply SENSEX) is a free-float market-weighted stock market index of 30 well-established and financially sound companies listed on the Bombay Stock Exchange. The 30 constituent companies, some of the largest and most actively traded stocks, represent various industrial sectors of the Indian economy. Recorded since 1 January 1986, the S&P BSE SENSEX is regarded as the pulse of the domestic stock markets in India.

A consumer price index (CPI) is the price index of a weighted average market basket of consumer goods and services purchased by households—changes in measured CPI track changes in prices over time. The CPI is calculated by using a representative basket of goods and services. The basket is updated periodically to reflect changes in consumer spending habits. The prices of the goods and services in the basket are collected monthly from a sample of retail and service establishments. The prices are then adjusted for changes in quality or features. Changes in the CPI can be used to track inflation over time and to compare inflation rates between different countries.

From basic intuition, we can understand that SENSEX and CPI must be co-dependent. Extensive literature on the same claim can be found in many articles such as Kumari (2012), an empirical study of the same. The core fact that we base this intuition on is that the Consumer Price Index (CPI) is a measure of inflation, which can impact the stock market, including the BSE Sensex. Conversely, SENSEX is India's principal stock exchange, which affects the market and liquidity, ultimately affecting CPI.

2 Data

The data set that I use for this analysis is sourced from Kaggle.

We look at monthly data from October 2000 to August 2020. Table 1 contains the summaries for the two data sets.

Along with their summaries, it also helps us to look at their time series plots and perform the Augmented Dickey-Fuller as described in Harris (1992) to test the hypothesis. The null hypothesis for this test claims that the time series under consideration has a unit root, implying that it is not stationary. Hence, we seek to find a small p-value to reject the hypothesis. Figure 1 and Table 2 demonstrate the plots and results for testing the hypothesis mentioned before.

Clearly, neither series is stationary.

Min.	36.73
1^{st} Quantile	45.19
Median	68.47
Mean	73.37
3^{rd} Quantile	101.37
Max	129.30

Min.	36.73
1^{st} Quantile	45.19
Median	68.47
Mean	73.37
3^{rd} Quantile	101.37
Max	129.30

Table 1: Summary for the Data (CPI and SENSEX respectively)

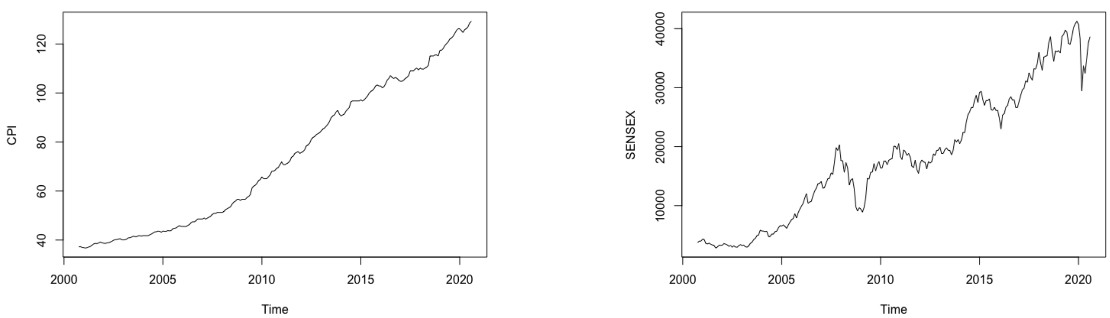

Figure 1: CPI and SENSEX Time Series Plot

CPI	0.5031
SENSEX	0.1628

Table 2: Augmented Dickey-Fuller Test p values

Finally, we conduct a simple Granger test on our two time series to see if they can be used as good predictors for each other. Figure 2 and Figure 3 give us the results.

We see that CPI can be a great predictor of SENSEX, but the opposite is not trivially true. This motivates us to consider a lagged causality model on the data set.

3 Causality Analysis

3.1 The Model

Consider the following model:

$$Y \leftarrow \text{CPI}$$

$$X \leftarrow \text{SENSEX}$$

```
Granger causality test

Model 1: SENSEX.dat ~ Lags(SENSEX.dat, 1:1) + Lags(CPI.dat, 1:1)
Model 2: SENSEX.dat ~ Lags(SENSEX.dat, 1:1)
  Res.Df Df      F  Pr(>F)
1    235
2    236 -1 7.3841 0.00707 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2: Granger Test | CPI on SENSEX

```
Granger causality test

Model 1: CPI.dat ~ Lags(CPI.dat, 1:1) + Lags(SENSEX.dat, 1:1)
Model 2: CPI.dat ~ Lags(CPI.dat, 1:1)
  Res.Df Df      F Pr(>F)
1    235
2    236 -1 0.8949 0.3451
```

Figure 3: Granger Test | SENSEX on CPI

Model-1:

$$Y_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \epsilon$$

We observe that the p-value for the co-efficient of X in this model is **0.00203**.

Model-2:

$$X_t = \alpha_0 Y_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + \epsilon$$

We observe that the p-value for the co-efficient of Y in this model is **0.00110**.

Both these values suggest that we have **bi-directionally causal data** that fits well into the SWAP paradigm.

3.2 Residual Analysis

The last thing we wish to check to ensure that we are not getting a spurious fit is the stationarity of residuals. Figure 4. Table 3 gives us the P-values for the ADF test.

Both residuals are clearly stationary, so the data is fit for a SWAP-type model.

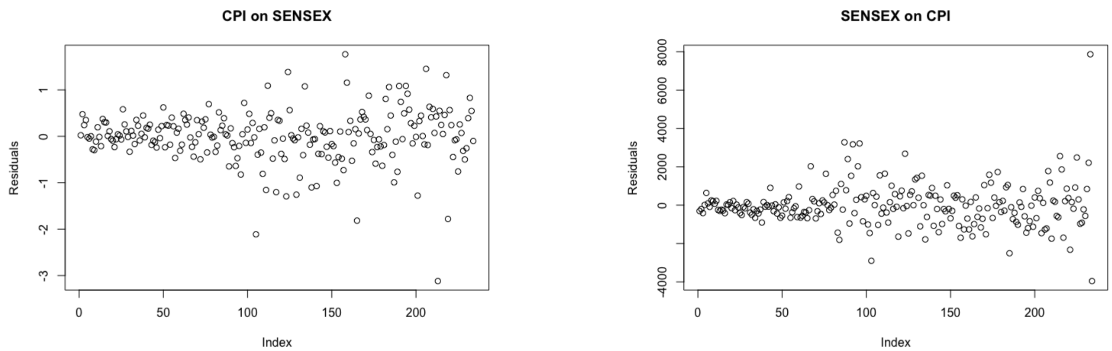

Figure 4: Residual Plots for Causal Model

CPI	0.01
SENSEX	0.01

Table 3: Augmented Dickey-Fuller Test p values

4 The SWAP Model

Finally, we apply the SWAP model to our data and contrast it with a standard linear regression model and a standard quadratic model to see which gives us the best fit.

We show a scatter plot of the data in Figure 5.

4.1 Standard Linear Regression Model

Consider the model:

$$Y \leftarrow \text{SENSEX}$$

$$X \leftarrow \text{CPI}$$

$$\text{Model:} Y = \beta_0 + \beta_1 X + \epsilon$$

The plot for the fitted data is shown in Figure 6.

Finally,

- The R^2 value for the model is **0.9222418**. This shows a decent fit for the model.
- The p value for the Augmented Dickey-Fuller Test is **0.3822**. This means that the residuals are not stationary.

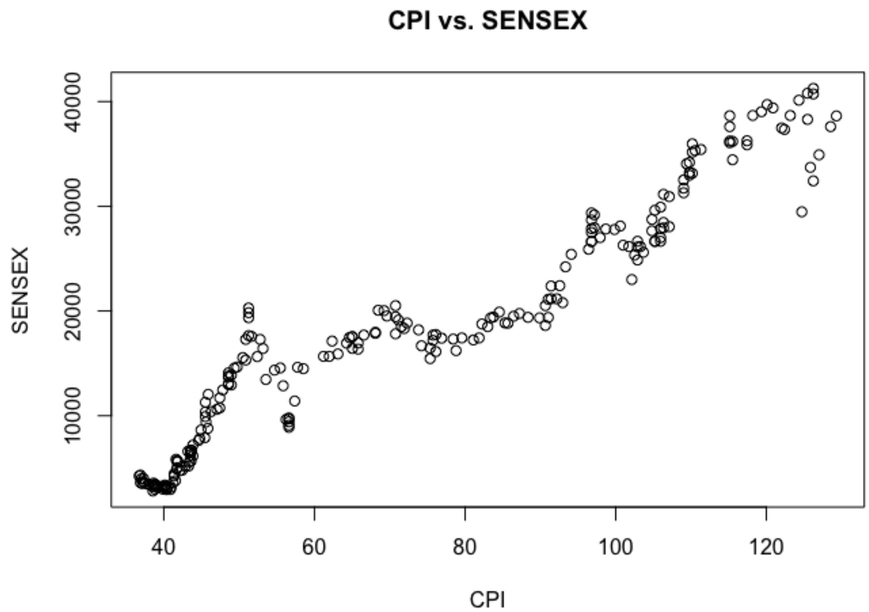

Figure 5: Scatter Plot for the full data

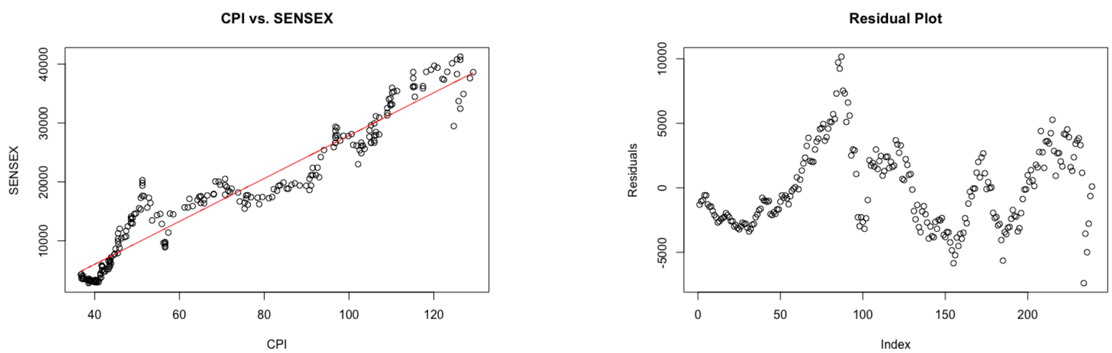

Figure 6: Standard LR model fit

4.2 Standard Linear Regression Model (with X^2)

Consider the model:

$$Y \leftarrow \text{SENSEX}$$

$$X \leftarrow \text{CPI}$$

$$\text{Model:} Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$$

The plot for the fitted data is shown in Figure 7.

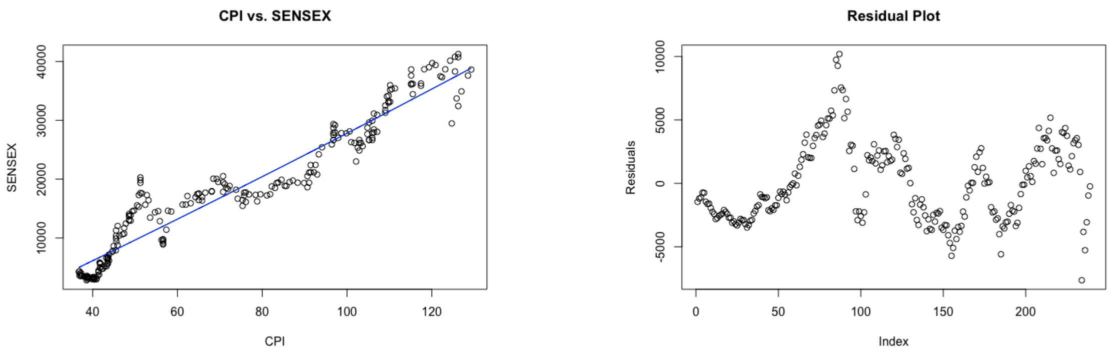

Figure 7: Quadratic LR model fit

Finally,

- The R^2 value for the model is **0.9223582**. This shows a decent fit for the model. It is slightly better than the standard LR case; however, it is not significant.
- The p value for the Augmented Dickey-Fuller Test is **0.3797**. This means the residuals are still not stationary, but we are progressing on the right track.

4.3 SWAP Model

Consider the model:

$$Y \leftarrow \text{SENSEX}$$

$$X \leftarrow \text{CPI}$$

$$g(X) = aX^2 + bX + c$$

$$g^{-1}(Y) = \frac{-b \pm \sqrt{b^2 - 4a(c - y)}}{2a}$$

$$\text{Break Point} = (X = 100)$$

The plot for the fitted data is shown in Figure 8.
Finally,

- The R^2 value for the model is **0.8665867**. This shows a decent fit for the model.
- The p value for the Augmented Dickey-Fuller Test is **0.6076**. This means that the residuals are not stationary.

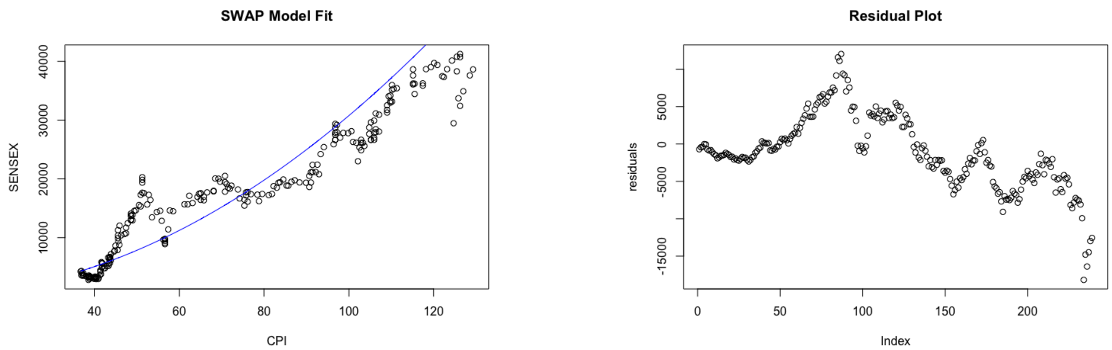

Figure 8: Standard LR model fit

References

Chow, M., Li, B., and Xue, J. Q. (2015). On regression for samples with alternating predictors and its application to psychrometric charts. *Statistica Sinica*, 25(3):1045–1064.

Harris, R. (1992). Testing for unit roots using the augmented dickey-fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics Letters*, 38(4):381–386.

Kumari (2012). Stock returns and inflation in india: An empirical analysis. *The IUP Journal of Monetary Economics*, IX:39–75.

Lee, H. (2008). Using the chow test to analyze regression discontinuities. *Tutorials in Quantitative Methods for Psychology*, 4.