Econometrics Final Project Report
A Scrutiny of India's GDP Mis-estimation

Akshat Singh 2016128
Jigme Lobsang Lepcha 2016045
Raghav Sood 2016259
Yash Tomar 2016122

November 14, 2019

1 Background/Motivation

Arvind Subramanian (referred to as 'author' when explicitly not mentioned) was the former Chief Economic Advisor to the Government of India from 2014 to 2018. In June 2019, after his tenure with the Economic Advisory Council, he published a paper titled "India's GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms, and Implications" [1] in which he has claimed that India's GDP did not increase by 7% between the years 2011-12 and 2016-17 but rather by around 4.5% (rather on an average it lies in between 3.5% - 5.5% with a 95% confidence level). The author has claimed that "India has changed its data sources and methodology for estimating real gross domestic product (GDP) for the period since 2011-12"[1], this change has lead to gross overestimation of growth of India's economy.

Soon after Arvind Subramanian's paper was released, the then Economic Advisory Council to the Prime Minister (EAC-PM) critically analyzed the paper, and released a rebuttal of the paper where they reject the author's methodology, arguments and conclusions in the said paper[2]. In our work, we pick up one of the arguments raised by the EAC-PM against the paper, and statistically verify whether the claim of the EAC-PM is correct or not.

2 Problem/Analysis Query

Statement from the PM-EAC Note: "The author mentions that the motivation of his paper is not political and is focused on technical aspects."

2.1 Analysis of Problem Statement

The author in his paper has claimed that his results are not politically motivated because it covers data during the regime of both UPA and NDA governments. He mentions that the methodological changes were initiated under the UPA-2 government but were completed by late 2014 when the NDA-2 government had come into power. Since these changes involved GDP estimations from 2011-12, the new statistical methodology spanned the rule of both the governments. The EAM-PC believes that the paper is technically inefficient in the way it has presented its work and thus the paper can have some other political motives behind it.

In our work, we will perform an econometric analysis on the GDP data of India, and verify whether there is a possibility of some political motivations behind Arvind Subramanian's work or not. The author has performed an analysis over 2 time periods, i.e. pre 2011 and post 2011 with a claim that the second time period spans under 2 different ruling parties. We will instead break the timespan into 4 periods ranging from 1998-2004, 2004-2009, 2009-2014, 2014-18. This is a partisan analysis where the time breaks have been put in the election years of India.

In our analysis we need to identify, which are the periods (if any exist) out of the 4 mentioned above where India behaves differently compared to the other countries. If there are some periods where India is behaving differently, it would indicate that the GDP is misestimated in those periods. Also, if India is behaving differently in only the periods involving a particular ruling party, it will indicate the change of government has had an impact in the GDP estimation of India.

2.2 Literature Review

While creating the mathematical model to formulate our problem statement, we observed some specific trends/anomalies which were creating some sort of disturbance in our analysis of the problem. We delved deeper into the problem to understand these anomalies and tried to study their effect in our analysis.

The Indian economy suffered 2 major shock in the year 2016. The Goods and Services Tax (GST) was the first one followed by Demonetization. This

(modelling)

sort of economic shock created a lot of turbulence in our data set. When the duration of the analysis excluded the period after 2016, we got some sort of explainable results through our model whereas when we included the years after 2016 that is the years 2016 - 2018 the obtained results were quite alarmingly different from the previous ones. Hence we concluded that that these events have led to some noisy data-points which are affecting our analysis and therefore we decided to not include the years after 2016 in the proposed mathematical model.

Another trend which we tried to understand was that of the election years. Every 5 years the incumbent government proposes radical policies to garner votes when nearing an election. These type of policies often involve taking loans or financial assistance which need to be repaid after the elections. All the stakeholders i.e the political parties infuse large amounts of money into the economy around the elections. To correlate this observation with our understanding we segregated out the election years and the years following and preceding an election to get a more unbiased view of the economy. The results obtained helped us to conclude that even after removing these years from our analysis, the trend somewhat remains similar to the original trend and hence there is no need to isolate the election years.

2.3 Proposed Analysis

We have broken down our analysis into 3 seperate parts. In the first part, we perform a partisan analysis on how the correlation between selected indicators varies with the GDP growth. Subramanian observed that post-2011, correlations between most of the indicators and GDP growth broke down. He claimed that unless structural changes occured at the time of GDP methodology revisions (which wasn't the case over here), the correlations should not vary so drastically. We will observe the trends of Correlation vs GDP growth for the indicators across the relevant periods in our analysis.

In the second part, we perform a partisan analysis on the variation of annual average growth of the indicators with the GDP growth. The author claimed that since the GDP growth in the 2 periods (pre and post-2011) is very similar, the average growth for most of the indicators should also be similar across the 2 periods. We will observe the trends of Annual Average

corrn b/w GDP gs & diff-erent indicators
corr b/w GDP gs & gs of diff. indicators.

Growth of indicators vs GDP growth across the relevant periods in our analysis.

In the final part, we perform a partisan analysis on the regression where Subramanian has tried to explain the GDP growth with the help of certain indicators that co-move with GDP growth. In our regressions, we will stick to the same indicators which were chosen by Arvind Subramanian to replicate and build on the results. We will add dummy variables to represent each of the 4 periods separately. We will then conduct statistical tests on the coefficients of the dummy variables and then conclude whether there has been a misestimation in some of these periods or not. We have mathematically formulated the problem statement in the following section.

2.4 Statistical Hypothesis

Arvind Subramanian has estimated GDP growth using the cross-country regression given in equation 1, where i represents the country and T is 0 pre-2011 and 1 post-2011. The author has calculated this for 74 countries including India. Since β_6 represents the coefficient of India in pre 2011 part and $\beta_6 + \beta_5$ in the post 2011 part, this makes the case to check whether the GDP was overestimated or not depending on β_5's estimate. If β_5 is statistically different from 0 then the GDP growth was misestimated in the second period with respect to the first period else the GDP growth was estimated correctly.

$$\begin{aligned} GDP\ Growth_{it} = {} & \beta_0 + \beta_1 Credit\ Growth_{it} + \beta_2 Electricity\ Growth_{it} \\ & + \beta_3 Export\ Growth_{it} + \beta_4 Import\ Growth_{it} + \beta_5 India * T \\ & + \beta_6 India + \beta_7 T + \beta_8 Credit\ Growth_{it} * T \\ & + \beta_9 Electricity\ Growth_{it} * T + \beta_{10} Export\ Growth_{it} * T \\ & + \beta_{11} Import\ Growth_{it} * T + \epsilon_{it} \end{aligned} \tag{1}$$

We have extended the regression model formulated by Arvind Subramanian and formed the linear regression represented by equation 2. This model accounts for the time breaks which occur in the election years. We have added 4 dummy variables, each accounting for one of the time periods. If the value of T_1 is 1 then we conclude that the period is 2001-2004. If T_2 is 1

then we take the period as 2004-2009. If the value of T_3 is 1, it denotes that the period is 2009-14, and if the value of T_4 is 1 then it represents that the period is 2014-2016. At an instant, only one of the 4 dummy variables will take the value 1.

$$\begin{aligned} GDP\ Growth_{it} = {} & \beta_0 T_1 + \beta_1 Credit\ Growth_{it} * T_1 + \beta_2 Electricity\ Growth_{it} * T_1 \\ & + \beta_3 Export\ Growth_{it} * T_1 + \beta_4 Import\ Growth_{it} * T_1 + \beta_5 India * T_1 \\ & + \beta_6 T_2 + \beta_7 Credit\ Growth_{it} * T_2 \\ & + \beta_8 Electricity\ Growth_{it} * T_2 + \beta_9 Export\ Growth_{it} * T_2 \\ & + \beta_{10} Import\ Growth_{it} * T_2 + \beta_{11} India * T_2 + \beta_{12} T_3 \\ & + \beta_{13} Credit\ Growth_{it} * T_3 + \beta_{14} Electricity\ Growth_{it} * T_3 \\ & + \beta_{15} Export\ Growth_{it} * T_3 + + \beta_{16} Import\ Growth_{it} * T_3 \\ & + \beta_{17} India * T_3 + \beta_{18} + \beta_{19} Credit\ Growth_{it} * T_4 \\ & + \beta_{20} Electricity\ Growth_{it} * T_4 + \beta_{21} Export\ Growth_{it} * T_4 \\ & + \beta_{22} Import\ Growth_{it} * T_4 + \beta_{23} India * T_4 + \epsilon_{it} \end{aligned} \tag{2}$$

The coefficients of interest to us are those along with the India dummy variable, i.e. β_5, β_{11}, β_{17}, β_{23}. These coefficients will tell us whether India is performing differently as compared to the other countries of the world in the 4 periods. We have formulated the following tests to check the statistical significance of the coefficients.

- The first null hypothesis in equation 3 checks whether the β_{11} value is statistically different from 0 or not. In another null hypothesis in equation 4 we also check whether the β_5 value is statistically different from 0 or not. In the analysis done by the author, he claims that there were no mis-estimates in the period before 2011. Since β_5 captures how India is different from other countries in 2001-04 and β_{11} captures how India is different from other countries in 2004-09, both values being statistically indifferent from 0 would indicate that there were no mis-estimates in these 2 periods individually which will explain that India's growth pattern was similar to other countries before 2011. If this null hypothesis is rejected, then it indicates there was some sort of mis-estimation before 2011 as well.

$$H_0 : \beta_5 = 0 \tag{3}$$

$$H_0 : \beta_{11} = 0 \tag{4}$$

- The third null hypothesis in equation 5 checks whether β_{17} is statistically different from 0 or not. β_{17} captures whether India behaves differently from other countries or not in the period 2011-14. If we reject this hypothesis, it will indicate that India's GDP growth behaved differently from the other countries in the period 2011-14 and thus there was some sort of mis-estimation in this period . Since the ruling party in this period was UPA, it will indicate that there was some sort of misestimation by the UPA party.

$$H_0 : \beta_{17} = 0 \tag{5}$$

- The fourth null hypothesis in equation 6 checks whether β_{23} is statistically different from 0 or not. β_{23} captures whether India behaves differently from other countries or not in the period 2014-16. If we reject this hypothesis, it will indicate that India's GDP growth behaved differently from the other countries in the period 2014-16 and thus there was some sort of mis-estimation in this period . Since the ruling party in this period was NDA, it will indicate that there was some sort of mis-estimation by the NDA party.

$$H_0 : \beta_{23} = 0 \tag{6}$$

- If the author's claim is true that there is some sort of mis-estimation in the post 2011 period , we should be able to reject atleast one of the two hypothesis above. If we fail to reject both of the hypothesis, it means there is no mis-estimation in the post 2011 period. If we only reject one of the hypothesis, it means the particular ruling party in that period caused mis-estimation in GDP growth and the other party did it correctly. Incase both the hypothesis are rejected, it indicates both the parties caused mis-estimation and we need to check the degree of mis-estimation.

Consider revising the language: political parties do not estimate GDP. or "cause" misestimation.

- In case both the hypothesis above are rejected we then compare whether the mis-estimation is similar or not. The fifth null hypothesis in equation 7 checks whether β_{23} is statistically different from β_{17} or not. In case we fail to reject this hypothesis, it means that the degree of misestimation was similar in both the periods. i.e. 2011-14 and 2014-18. If we reject this hypothesis, we then need to check which party has caused more estimation compared than the other.

$$H_0 : \beta_{23} - \beta_{17} = 0 \tag{7}$$

- We know that higher the T value, the greater is the evidence against the null hypothesis and thus there is a higher probability of a significant difference. If the above hypothesis was rejected, we will check whether β_{17} or β_{23} has a higher T value. If there is a significant difference in the two values, then we can claim with a certain level of confidence that there was a higher level of misestimation by one of the two parties. (Higher T value for β_{17} indicates more misestimation by UPA and higher value for β_{23} indicates more misestimation for NDA).

3 Variables and their Descriptions[DATA]

The author accounts for seventeen indicators which are strongly correlated with the growth of the GDP. But out of these seventeen indicators the author only considers four such indicators (mentioned in table 1) for the regression because of easy availability of the data. Also the author claims that these variables are a good enough representation of the entire GDP because of a high R-square value obtained on using these variables, indicating that the model is a good fit. Since we are exactly sticking to the author's model and only replicating it across different time periods, we will also use the same 4 variables for our regression analysis. In the case of correlation vs GDP growth and indicator growth vs GDP growth we have used 13 variables (mentioned in table 1) for which we obtained data from the links mentioned below (in Appendix).

To further extend our analysis we can also add additional indicators to verify the claims made by the author. We can add indicators like agriculture and services to our model to make the model stronger, since they account for a major portion of the Indian GDP.

Agriculture sector contributes to employment, not necessarily GDP.

Services sector contributes to GDP.

4 Summary Statistics

Table 2,3,4,5 indicating description of variables and Figure 1 indicating GDP growth of India has been attached in the Appendix

5 Results

The graph in figure 2 depicts the correlation between selected indicators and the GDP growth. The correlation was computed for 2 periods which are 2011-12 to 2013-14 and 2014-15 to 2017-18 where the horizontal axis is represented by the former and the vertical axis is represented by the latter. Even though 2009 is an election year, we excluded the years 2009-2011 as the change in methodology of GDP calculation happened after the year 2011. We observe that most of the indicators lie around the 45 degree line with the exception of some outliers such as IIP and Petroleum. This shows that even though every variable can have a different structural relationship with the GDP growth but the correlation value remains relatively similar over the two periods for most of the indicators. Therefore barring a few variables whose correlation broke down over the 2 periods , the structural relationship remains same over the duration which involved two different regimes that are UPA-2 and NDA-2. This indicates that there wasn't any drastic change which occured between these 2 periods (UPA-2 and NDA-1) unlike the change Subramanian observed between pre-2011 and post-2011 era.

The graph in figure 3 shows the annual average growth of the selected indicators over the period 2011-12 to 2013-2014 and 2014-15 to 2017-2018. The GDP growth remains almost similar over the two regimes and hence the GDP indicator lies around the 45 degree line. Similarly most of the other selected indicators lie around the 45 degree line which show that the growth rate of the variables remains almost similar over the two periods barring a few outliers. If these values had been lying away from the 45 degree line in the positive or negative direction would have concluded that over the duration of the period even though GDP growth was at certain value the indicator variables were behaving differently between 2011-2018. The results thus observed depict that the behaviour of these variables remained almost similar over the change in government from UPA-2 to

NDA-2 unlike a drastic change in growth observed by Subramanian from pre-2011 to post-2011 era.

We performed similar analysis for the period starting from 1998-2009 as well but the results obtained were of very less significance as there was hardly any misestimation and hence were excluded from our analysis.

The first null hypothesis which was introduced in the preceding sections checks whether the β_{11} value is statistically different from 0 or not. In the second null hypothesis in equation 4 we also check whether the β_5 value is statistically different from 0 or not. After performing the mentioned hypothesis testing we fail to reject both the null hypothesis. This indicates that India is not an outlier as compared to other countries in terms of GDP growth in these periods. We can observe this from the table 6 as the p-value of these coefficients is very high (0.92 and 0.55 respectively). This result is similar to what the author had claimed in his paper that there was no mis-estimation in the GDP growth before 2011. As this hypothesis corresponds to the periods 2001-2004 and 2004-2009, this test result helps us to conclude that the period 2001-09 was free from any GDP growth mis-estimation.

The third null hypothesis which checks whether β_{17} value is statistically different from 0 or not helps in determining that mis-estimations were done by UPA-2 or not. After performing the hypothesis testing, we reject the null hypothesis that the value is statistically indifferent from 0. This means that India is an outlier when compared to other countries in this period (2011-14) in terms of GDP growth. We can observe this from the table 6 as the p-value of this coefficient is very low (0.05). This result falls in line with the authors claim that GDP growth mis-estimations were done after 2011. As this β_{17} value corresponds to the duration 2009-2014 , we conclude that mis-estimations were done by the UPA-2 government in their results on GDP growth calculations.

The fourth null hypothesis which checks whether β_{23} value is statistically different from 0 or not helps in determining that mis-estimations were done by NDA-2 or not. After performing the hypothesis testing, we reject the null hypothesis that the value is statistically indifferent from 0.We can observe this from the table 6 as the p-value of this coefficient is very low

(0.02) This means that India is an outlier when compared to other countries in this period (2014-16) in terms of GDP growth. This result falls in line with the authors claim that GDP mis-estimations were done after 2011. As this beta value corresponds to the duration 2014-2016 , we conclude that mis-estimations were done by the NDA-2 government in their results on GDP growth calculations.

We now have concluded that both UPA-2 and NDA-2 performed a mis-estimation in the GDP growth calculation in the periods 2011-14 and 2014-16 respectively. We now need to compare the degree of misestimations.

The fifth null hypothesis which verifies whether the mis-estimation which happened in the years after 2011 were similar for both the regimes or not is tested. On performing the hypothesis testing ,we are able reject the null hypothesis that the difference of the beta values over the duration is statistically similar to zero. Hence we can conclude that in one of the two periods the mis-estimation was higher when compared to the other period.

Further to complete our analysis we compare the T - value of the coefficients β_{17} and $\beta_{23}$and infer with a very low confidence level that the NDA-2 government has mis-estimated the GDP growth more than the UPA-2 government (can be observed by comparing the T-values 1.66 and 2.24 from table 6) and therefore conclude our analysis of the problem that both the governments have mis-estimated the GDP growth, but NDA-2 government has mis-estimated it slightly more than the UPA-2 government.

6 Concluding Remarks

- We compared the correlation coefficient of indicators and their growth vs GDP growth in 2011-14 and 2014-18. We found that there was no drastic change in these values across the periods for most of the indicators
- We conducted a partisan analysis on the regression across the 4 periods after dividing the time period with 4 breaks at the election years that were 2004, 2009 and 2014.
- For periods 2001-04 and 2004-09 there was no misestimation in the GDP growth as inferred with the help of the first two hypothesis.

- We observed that there was misestimation across both periods 2011-14 and 2014-16 with the results obtained from the thirs and furth hypothesis.

- With the final hypothesis of comparing T-values we were able to conclude with a very low significance value that UPA-2 in 2011-14 had lesser mis-estimation of GDP growth than NDA-2 in 2014-16.

References

[1] A. Subramanian, "India's GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms, and Implications," CID Working Papers 354, Center for International Development at Harvard University, June 2019.

[2] B. Debroy, R. Roy, S. Bhalla, C. Singh, and A. Virmani, "GDP Estimation in India- Perspectives and Facts." http://eacpm.gov.in/wp-content/uploads/2019/06/EAC-Paper_GDP-estimation_19-June-2019.pdf, June 2019.

7 Appendix

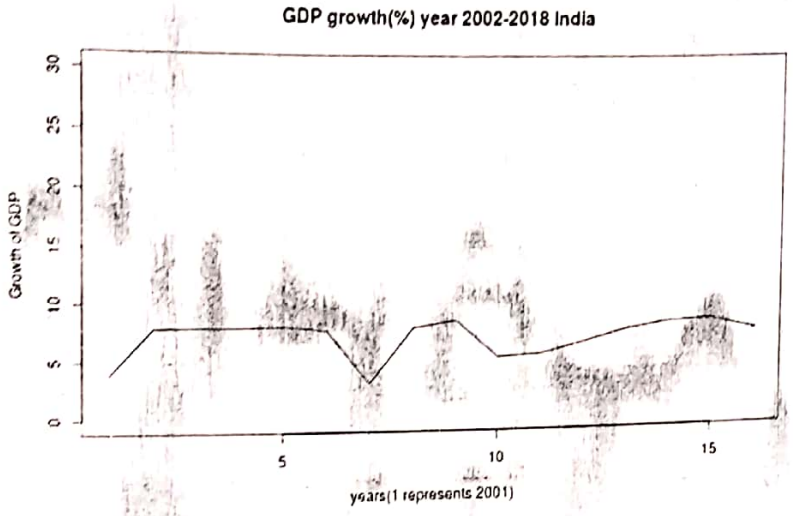


Figure 1: GDP growth of India from 2002-2018

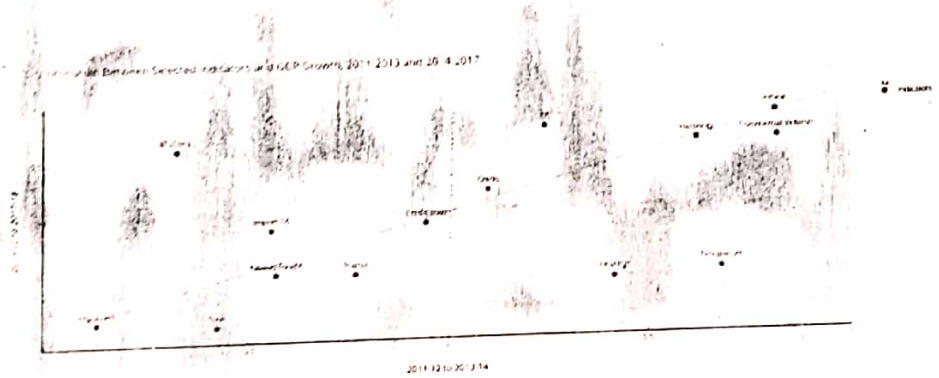
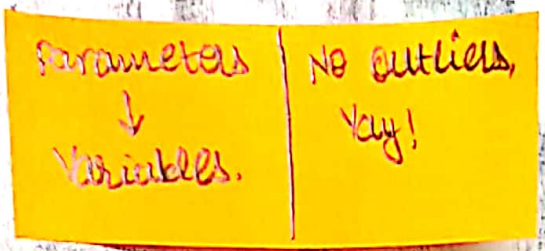


Figure 2: Correlation between selected indicators and GDP growth, 2011-2013 and 2014-2017

S.no.	Variable	Description	Notations
1	Credit	This factor accounts for the credit given to the private sector, it mainly accounts for the financial resources provided to the private sector by financial corporations.	$CreditGrowth_a$
2	Export	This factor accounts for the exports of goods and services which is the value of all goods and rest market services provided to the rest of the world.	$ExportGrowth_a$
3	Import	This factor accounts for the imports of goods and services which is the value of all goods and rest market services received from the rest of the world.	$ImportGrowth_a$
4	Electricity	This factor accounts for the total electricity consumption of a country	$ElectricityGrowth_a$
5	Tractor	This factor accounts for the total production of tractors	NA
6	Steel	This factor accounts for the total production of steel of a country	NA
7	Railway Freight	This factor accounts for the total number of freight train loading	NA
8	Petroleum	This factor accounts for the total , petroleum production of a country	NA
9	Airline	This factor accounts for the airline passenger traffic in a country	NA
10	IIP	This factor accounts for different growth rates in different industry groups	NA
11	IIP (Mfg.)	This factor accounts for the growth rate in different manufacturing industries	NA
12	IIP (Cons.)	This factor accounts for the growth rate for consumer durables	NA
13	Commercial Vehicles	This factor accounts for different growth rates in different industry groups	NA

Table 1: Description of variables. NA denotes that the variable is not used in the regression model, it is only used for calculating correlation coefficient.



S.NO	Parameters	Mean	Min	Max	Standard Deviation
1	Credit	33.37485	31.62626	36.1918	2.463145
2	Imports	16.8445	15.24428	19.64469	2.433281
3	Exports	15.69047	14.26438	17.85912	1.908949
4	GDP Growth(%)	6.5291	3.803975	7.922943	2.360235

Table 2: Variables(2002-2004)

S.NO	Parameters	Mean	Min	Max	Standard Deviation
1	Credit	45.40146	40.06798	49.55937	3.753204
2	Imports	25.37655	22.39642	29.27087	2.518887
3	Exports	21.23415	19.60524	24.09736	1.712902
4	GDP Growth(%)	6.918713	3.086699	8.060729	2.146991

Table 3: Variables(2005-2009)

S.NO	Parameters	Mean	Min	Max	Standard Deviation
1	Credit	51.6002	50.55538	52.38571	0.701431
2	Imports	28.71291	25.95422	31.25929	2.411342
3	Exports	23.97492	22.40093	25.43086	1.249399
4	GDP Growth(%)	6.598325	5.241314	8.497587	1.365368

Table 4: Variables(2010-2014)

S.NO	Parameters	Mean	Min	Max	Standard Deviation
1	Credit	49.94318	48.77986	51.86752	1.368286
2	Imports	22.12252	20.96414	23.43016	1.011691
3	Exports	19.37103	18.78061	19.81319	0.4762714
4	GDP Growth(%)	7.579001	6.982334	8.169527	0.5910007

Table 5: Variables(2015-2018)

Variables	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.49	1.03	3.38	0.00 ***
t1	0.60	1.38	0.43	0.66
I(credit * t1)	0.00	0.01	0.05	0.95
I(exports * t1)	0.05	0.02	1.88	0.06.
I(imports * t1)	-0.05	0.02	-2.03	0.04 *
I(india * t1)	-0.31	3.19	-0.09	0.92
t2	2.07	1.45	1.43	0.15
I(credit * t2)	-0.01	0.01	-0.99	0.31
I(exports * t2)	0.07	0.02	3.07	0.00 **
I(imports * t2)	-0.05	0.02	-2.34	0.01 *
I(india * t2)	1.90	3.18	0.59	0.55
t3	1.48	1.47	1.00	0.31
I(credit * t3)	-0.01	0.01	-0.01	0.95
I(exports * t3)	-0.01	0.02	-0.59	0.55
I(imports * t3)	-0.01	0.02	-0.11	0.90
I(india * t3)	2.12	1.27	1.66	0.05
I(credit * t4)	0.01	0.01	0.75	0.45
I(exports * t4)	-0.01	0.03	-0.19	0.84
I(imports * t4)	-0.01	0.03	-0.27	0.78
I(india * t4)	3.96	1.76	2.24	0.02

Table 6: Esimated Values from model
(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)
Residual standard error: 3.129 on 267 degrees of freedom Multiple R-squared: 0.116, Adjusted R-squared: 0.05307 F-statistic: 1.844 on 19 and 267 DF, p-value: 0.01865

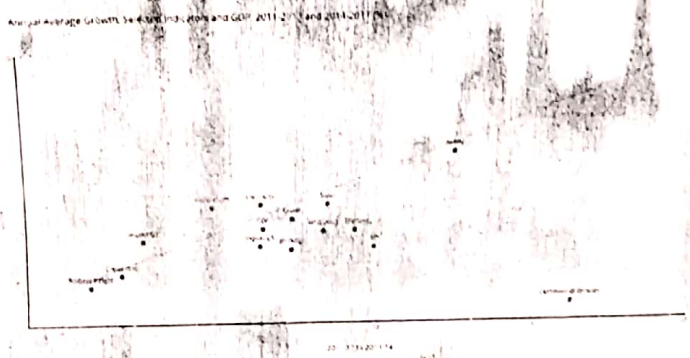


Figure 3: Annual average growth selected indicators and GDP, 2011-2013 and 2014-2017

Econometrics Project Report 1
A Scrutiny of India's GDP Mis-estimation

Akshat Singh
2016128

Jigme Lobsang Lepcha
2016045

Raghav Sood
2016259

Yash Tomar
2016122

September 27, 2019

1 Introduction

Arvind Subramanian (referred to as 'author' when explicitly not mentioned) was the former Chief Economic Advisor to the Government of India from 2014 to 2018. In June 2019, after his tenure with the Economic Advisory Council he published a paper titled "India's GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms, and Implications" [1] in which he has claimed that India's GDP did not increase by 7% on average in the years 2011-12 and 2016-17 but rather by around 4.5% (rather it lies in-between 3.5% - 5.5% with a 95% confidence level). The author has claimed that "India has changed its data sources and methodology for estimating real gross domestic product (GDP) for the period since 2011-12"[1]. this change has lead to gross overestimation of growth of India's economy.

Soon after Arvind Subramanian's paper was released, the then Economic Advisory Council to the Prime Minister (EAC-PM) critically analyzed the paper, and released a rebuttal of the paper where they reject the author's methodology, arguments and conclusions in the said paper[2]. In our work, we pick up one of the arguments raised by the EAC-PM against the paper, and statistically verify whether the claim of the EAC-PM is correct or not.

The author in his paper has claimed that his results are not politically motivated because it covers data during both UPA and NDA governments. He mentions that methodological changes were initiated under the UPA-2 government but were completed by late 2014 when the NDA government had come into power. Since these changes involved GDP estimations from 2011-12, the new statistical methodology spanned the rule of both the

governments. The EAM-PC believes that the paper is technically inefficient in the way it has presented its work and thus the paper can have some other political motives behind it. In our work, we will perform an econometric analysis on the GDP data of India, and verify whether there is a possibility of some political motivations behind Arvind Subramanian's work or not.

2 Problem Statement

In our work, we wish to perform a statistical analysis on Arvind Subramanian's work on the GDP estimation of India. The author has performed an analysis over 2 time periods, i.e. pre 2011 and post 2011 with a claim that the second time period spans under 2 different ruling parties. We will instead break the timespan into 4 periods ranging from 1998-2004, 2004-2009, 2009-2014, 2014-16. This is a partisan analysis where the time breaks have been put in the election years of India. The author in his work compares the GDP movement of India with the other countries. He claims that the GDP of India is misestimated in a period if it behaves differently from the general trend of the GDP relationship of the other countries. The author found that in the post 2011 period, India's GDP behaves differently compared to other countries. In our analysis we need to identify, which are the periods (if any exist) out of the 4 mentioned above where India behaves differently compared to the other countries. If there are some periods where India is behaving differently, it would indicate that the GDP is misestimated in those periods. Also, if India is behaving differently in only some of the periods and not all, it will indicate the change of government has had an impact in the GDP estimation of India. ,

To conduct our analysis, we will stick to the same variables in our model which were chosen by Arvind Subramanian. We will add dummy variables to represent each period seperately. We will then conduct statistical tests on the coefficients of the dummy variables and check which of them are statistically different from each other. If we find some periods where the variables are statistically different, it will indicate that the estimation of GDP has been performed differently/ misestimated in atleast one of the periods.

3 Problem Variables and Dataset

S.no.	Variable	Description	Source
1	Credit	This factor accounts for the credit given to the private sector, it mainly accounts for the financial resources provided to the private sector by financial corporations.	link
2	Export	This factor accounts for the exports of goods and services which is the value of all goods and rest market services provided to the rest of the world.	link
3	Import	This factor accounts for the imports of goods and services which is the value of all goods and rest market services received from the rest of the world.	link
4	Electricity	This factor accounts for the total electricity consumption of a country	link

Table 1: Description of variables

The author accounts for seventeen indicators which are strongly correlated with the growth of the GDP. But out of these seventeen indicators the author only considers four such indicators (mentioned in table 1) because of easy availability of the data. Also the author claims that these variables are a good enough representation of the entire GDP because of a high R-square value obtained on using these variables, indicating that the model is a good fit. Since we are exactly sticking to the author's model and only replicating it across different time periods, we will also use the same 4 variables for our analysis.

The author divides his analysis into two time periods which is pre and post 2011 for which he adds a dummy variable T. Now to consider whether there is any political influence or not, we will extend the author's model with three breaks instead of one and the break points will be for years 2004, 2009 and 2014. So if there are three breaks then we are considering four time periods which is pre 2004, 2004 to 2009, 2009 to 2014 and post 2014. So to account these four time periods we use three dummy variables T_1, T_2 and T_3. (We do not use dummy variable T_4 to avoid dummy variable trap)

To further extend our analysis we can also add additional indicators to verify the claims made by the author. We can add indicators like agriculture and services to our model to make the model stronger, since they account for a major portion of the Indian GDP.

4 Statistical Models and Tests

Arvind Subramanian has estimated GDP growth using the cross-country regression given in equation 1, where i represents the country and T is 0 pre-2011 and 1 post-2011. The author has calculated this for 74 countries including India. Since β_6 represents the coefficient of India in pre 2011 part and $\beta_6 + \beta_5$ in the post 2011 part, this makes the case to check whether the GDP was overestimated or not depending on β_5's estimate. If β_5 is statistically different from 0 then the GDP growth was overestimated in the second period with respect to the first period else the GDP growth was estimated correctly.

$$\begin{aligned} GDP\ Growth_{it} = {} & \beta_0 + \beta_1 Credit\ Growth_{it} + \beta_2 Electricity\ Growth_{it} \\ & + \beta_3 Export\ Growth_{it} + \beta_4 Import\ Growth_{it} + \beta_5 India * T \\ & + \beta_6 India + \beta_7 T + \beta_8 Credit\ Growth_{it} * T \\ & + \beta_9 Electricity\ Growth_{it} * T + \beta_{10} Export\ Growth_{it} * T \\ & + \beta_{11} Import\ Growth_{it} * T + \epsilon_{it} \end{aligned} \tag{1}$$

We have extended the regression model formulated by Arvind Subramaniam and formed the linear regression represented by equation 2. This model accounts for the time breaks which occur in the election years. Now the analysis is spread over multiple periods and the variables added account for each of the specific periods. If the value of T_1 is 1 then we conclude that the period is 2004-2009. If T_2 is 1 then we take the period as 2009-2014 and if the value of T_3 is 1 then it represents that the period is 2014-2018. If the value of all the T's is 0 then the period becomes 1998-2004.

OK, but his analysis has other components. Need to explain why you skipped those?

Need to revisit the model

$$
\begin{aligned}
GDP\ Growth_{it} = {} & \beta_0 + \beta_1 Credit\ Growth_{it} + \beta_2 Electricity\ Growth_{it} \\
& + \beta_3 Export\ Growth_{it} + \beta_4 Import\ Growth_{it} + \beta_5 India * T_1 \\
& + \beta_6 India + \beta_7 T_1 + \beta_8 Credit\ Growth_{it} * T_1 \\
& + \beta_9 Electricity\ Growth_{it} * T_1 + \beta_{10} Export\ Growth_{it} * T_1 \\
& + \beta_{11} Import\ Growth_{it} * T_1 + \beta_{12} India * T_2 + \beta_{13} T_2 \\
& + \beta_{14} Credit\ Growth_{it} * T_2 + \beta_{15} Electricity\ Growth_{it} * T_2 \\
& + \beta_{16} Export\ Growth_{it} * T_2 + + \beta_{17} Import\ Growth_{it} * T_2 \\
& + \beta_{18} India * T_3 + \beta_{19} T_3 + \beta_{20} Credit\ Growth_{it} * T_3 \\
& + \beta_{21} Electricity\ Growth_{it} * T_3 + \beta_{22} Export\ Growth_{it} * T_3 \\
& + \beta_{23} Import\ Growth_{it} * T_3 + \epsilon_{it}
\end{aligned} \tag{2}
$$

Resolved in the final report.

- The first null Hypothesis in 3 checks whether β_5, β_{12} and β_{18} values are simultaneously statistically different from 0 or not. We can conclude that if this hypothesis is accepted then the claim made by the author is completely rejected that there has been GDP mis-estimates. For further analysis this null hypothesis should be rejected as Arvind Subramaniam claims that there have been mis-estimates and GDP values have been wrongly calculated. We will proceed further if this null hypothesis is rejected.

Why do you need simultaneity here?

But if India were to be an outlier? TA wouldn't your model miss that detail? Think.

$$H_0 : \beta_5 = 0 \quad and \quad \beta_{12} = 0 \quad and \quad \beta_{18} = 0 \tag{3}$$

- The second null hypothesis in 4 checks whether β_5 value is statistically different from 0 or not. In the analysis done by the author, he claims that there were no mis-estimates in the period before 2011 so in the ideal case ~~scenario~~ we expect the value of β_5 ~~should~~ to be statistically indifferent from zero as he explains that in the period before 2011, India's growth pattern was similar to other countries. If this null hypothesis is rejected then Arvind Subramaniam's claims turns out to be false.

good

$$H_0 : \beta_5 = 0 \tag{4}$$

- The third null hypothesis in 5 checks whether β_{18} - β_{12} is statistically different from 0 or not. If this hypothesis is accepted then it means that the estimates done over 2009-2014 and 2014-2018 are similar and if rejected then it means that there have been some changes in the estimates done by the UPA-2 and NDA-2 government. If the two governements are estimating differently, now we need to compare which of the government is mis-estimating and hence we compare with pre-2009 data.

$$H_0 : \beta_{18} - \beta_{12} = 0 \tag{5}$$

- The fourth null hypothesis in 6 checks whether β_{12} - β_5 is statistically different from 0 or not. If this hypothesis is accepted then it means that the estimates done by the UPA-2 government and UPA-1 were relatively similar and the estimates from 2004-2009 and 2009-2014 were not mis-estimated. This can be inferred as the author claims that there were no mis-estimates in the period before 2011.

$$H_0 : \beta_{12} - \beta_5 = 0 \tag{6}$$

- The fifth null hypothesis in 7 checks whether β_{18} - β_5 is statistically different from 0 or not. If this hypothesis is accepted then it signifies that the relative changes over period 2014-2018 and 2004-2009 were same. Therefore the estimates done by the NDA-2 government and UPA-2 government were similar and were not mis-estimated equal.

$$H_0 : \beta_{18} - \beta_5 = 0 \tag{7}$$

If any of the two hypotheses above turns out to be ~~true~~ FALSE, we can then infer that a change in government has resulted in a change in the GDP estimations of India, which then concludes that the GDP estimation has a political component to it.

This is a well-thought out plan.
I look forward to your analysis.

References

[1] A. Subramanian, "India's GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms, and Implications," CID Working Papers 354, Center for International Development at Harvard University, June 2019.

[2] B. Debroy, R. Roy, S. Bhalla, C. Singh, and A. Virmani, "GDP Estimation in India- Perspectives and Facts." http://eacpm.gov.in/wp-content/uploads/2019/06/EAC-Paper_GDP-estimation_19-June-2019.pdf, June 2019.