① Explanation of why quarter 2 is/ can be different.

② summary stats of data need to be added

③ Data References.

Analysis of Quarterly Data For India's GDP - Examining The EAC-PM Rebuttal

Project Report II

- Reasonable work
- Can go up if authors willing to put more effort
- writing/presentation very sloppy.

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

Dr. Subramanian → Subramanian (2019) [SS]

1 Background/Motivation

India changed its data sources and methodology for estimating real Gross Domestic Product (GDP) for the period since 2011-12. The paper by Dr Arvind Subramanian aims to show that due to this change, India's GDP has been significantly overestimated.

Dr Subramanian conducted various statistical tests to support his claim, three of which are within the scope of our analysis. He divided the data into two groups - the first was from 2001 till 2011, and the second from 2012 till 2017.

1. For the first test, he showed that there was a drastic change in the correlation coefficients between the two ~~groups~~ time-periods for the same variable. For example, gs in Petroleum was negatively correlated, with a correlation coefficient of about -0.5, before 2011-12; and was positively correlated, with a coefficient of about +0.6, post-2011-2012. (w GDP gs [SS]) growth [SS]
2. The measured overall real GDP in the two periods is close to the 45-degree line. Thus, we would expect the average growth for all the indicators to also be close to the 45-degree line. However, most indicators lie below the 45-degree line, some of them substantially below. The only indicators that lie above the line are Petroleum and Electricity. Hence Dr Subramanian concludes that the trend of GDP does not follow the trend of the indicators.
3. When a regression equation for GDP growth was run for the two groups of data with a dummy variable for India, it was found that India is an outlier in the group with data after 2011-2012, ~~but not for before~~.

As a rebuttal to the paper published by Dr Subramanian, the Economic Advisory Council to The Prime Minister (EAC-PM) stated that his paper would not stand the scrutiny of policy and research standards as it lacks rigour in terms of the chosen explanatory variables as well as the frequency of data. This means that should the frequency of data, and the number and quality of explanatory variables be increased, one can possibly better explain India's outlier behaviour. We aim to verify this statement released by the EAC-PM by replicating tests 1 and 2 (mentioned above)

→ quotes must be put into quotation marks. [SS]

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

conducted by Dr Subramanian for quarterly data instead of the annual data used by him for India only.

2 Problem/Analysis Query

"However, given the fact that his paper lacks rigor in terms of specific data sources and description; alternative hypothesis; rationale of equation specifications, use of dummies, and robustness-check diagnostics of estimated equations; and choice of countries in the sample and a specific list; it would not stand the scrutiny of academic or policy research standards." (Reference)

As previously mentioned, a part of the EAC-PM's rebuttal is based on the claim that Dr Subramanian's paper is not rigorous enough in terms of specific data sources. Hence according to this claim, should one replicate his analysis with more data points and better quality data, the observed outlier behaviour of India can be better explained. ↳ may not be accurate. [ss]

To test the statement presented by the EAC-PM, we reconduct the tests done by Dr Subramanian using quarterly data. Here, we have increased our dataset in a two-fold manner - first, we have increased the frequency of the data by using quarterly data instead of annual; and second, we have also increased the number of explanatory variables used in our equations.
Our analysis is restricted to India only.

2.1 Test 1

The first test we propose to replicate using quarterly data is the comparison of average growth rates (Test 2 in the previous section). Additionally, we also propose to conduct the same for each individual quarter to check if there is any particular quarter where a particular variable may have a high contribution to outlier behaviour, but not in the other quarters; it might show different behaviour cumulatively when averaged out for all the quarters.

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

For calculating the quarterly average growth rate, we find the percentage change for two consecutive quarters.

$$percentage\ change = \left((final - initial)/initial\right) * 100$$

Then, four subsets are made, one for each quarter. The percentage changes are then averaged out for each quarter.

$$average\ growth\ rate = percentage\ change / n$$

where, n = size of subset

Following this, we plot the average growth rates of GDP along with the explanatory variables for the two time periods. The x-coordinate is the average growth rate before 2011-12, and the y-coordinate is the average growth rate after 2011-12. Thus, points which lie on a 45-degree line would have no change in their average growth rates before and after 2011-12. Points which lie above the 45-degree line have seen an increase in the average growth rate, while points that lie below it have seen a decrease.

2.2 Test 2

Our data is divided into two groups - one before 2011-2012 and another after. To draw a comparison between the two time periods using quarterly data, we decided to check if there is a quarter, or a group of quarters, where the behaviour is different across the two periods. Hence, the following linear regression was designed:

$$gdp_i = imp_i + exp_i + petr_i + elec_i + cmnt_i + stl_i + iip_i + agr_i + tour_i + imp_i{*}Tdummy + exp_i{*}Tdummy + petr_i{*}Tdummy + elec_i{*}Tdummy + cmnt_i{*}Tdummy + stl_i{*}Tdummy + iip_i{*}Tdummy + agr_i{*}Tdummy + tour_i{*}Tdummy + Q1{*}Tdummy + Q2{*}Tdummy + Q3{*}Tdummy$$

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

Essentially, we have employed the "difference-in-differences" technique to test whether any quarter or group of quarters, was differentially affected in the post-2011 period compared to the others.

On running the regression for equation (1), we obtain a summary table from which we get the p-values, and hence the significance levels, of each of the dummies included in the equation. Using this, we can decide if there is any quarter whose behaviour is different for the two groups of data. If indeed any such quarter exists, the trends in that quarter may be able to explain, or possibly be the reason for, the overestimation of GDP as described in Dr Subramanian's paper.

3 Variables and their Description :

Serial Number	Variable	Description	Notation
1	GDP	Gross Domestic Product	gdp
2	Import Growth	The relative change in the goods and services that are bought from other countries.	imp
3	Export Growth	The relative change in the goods and services that are sold to other countries.	exp
4	Petroleum	Consumption of petroleum products	petr
5	Electricity	Index of electricity production	elec
6	Cement	Index of cement production	cmnt
7	Steel	Index of steel production	stl
8	IIP	Index of Industrial Production (2 digit level)	iip
9	Agriculture and Allied	Agriculture, Forestry and Fishing - GDP at factor cost by economic activity	agr

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

	Activities		
10	Foreign Tourists	Number of foreign tourist arrivals in India	tour

Notes:

Variables 4, 8, 9, 10 were taken from indiastat.com
Variables 2, 3, 5, 6, 7 were taken from epwrfits.in
GDP values were taken from stats.oecd.org

4 Summary Statistics

No Regression results

Excluding variables which don't have available data for all years:

2001 - 2017 Variables	Coefficients
Intercept	-1.323e+04 *** (1.897e+03)
elec	1.305e+02 ** (4.223e+01)
cmnt	-5.560e+01 ** (1.622e+01)
stl	-2.554e+01 (2.161e+01)
petr	7.133e-01 *** (8.529e-02)
tour	1.557e-03 * (6.666e-04)
elec*Tdummy	-2.529e+02 ** (8.365e+01)

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

cmnt*Tdummy	1.701e+02 ** (4.982e+01)
stl*Tdummy	1.189e+02 * (4.898e+01)
petr*Tdummy	-1.240e-01 (1.531e-01)
tour*Tdummy	9.373e-04 (2.071e-03)
q2*Tdummy	3.016e+03 *** (7.305e+02)
q3*Tdummy	-5.467e+02 (1.707e+03)
q4*Tdummy	-2.860e+03 (2.131e+03)
Number of Observations	68
R^2	0.9821

Including all variables for reduced number of years:

2005 - 2015 Variables	Coefficients
Intercept	-1.843e+03 (2.097e+03)
elec	1.092e+02 (3.939e+01)
petr	-8.454e-02 (9.329e-02)
imp	8.998e-03

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

	(3.575e-03)
agr	4.021e-03 (5.020e-03)
elec*Tdummy	-7.835e+01 (2.399e+02)
agr*Tdummy	-2.662e-02 (2.615e-02)
q2*Tdummy	NA
q3*Tdummy	NA
q4*Tdummy	-1.067e+03 (2.097e+03)
Number of Observations	38
R^2	0.9918

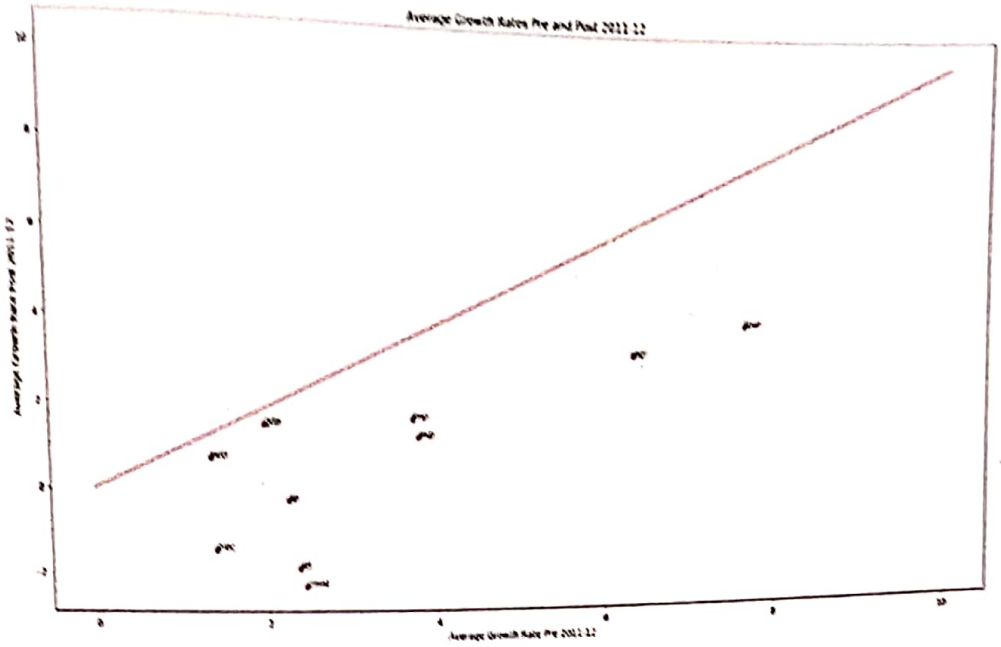
Signif. codes: 0'***' 0.001'**' 0.01'*' 0.05'.'

5 Results

5.1 Test 1

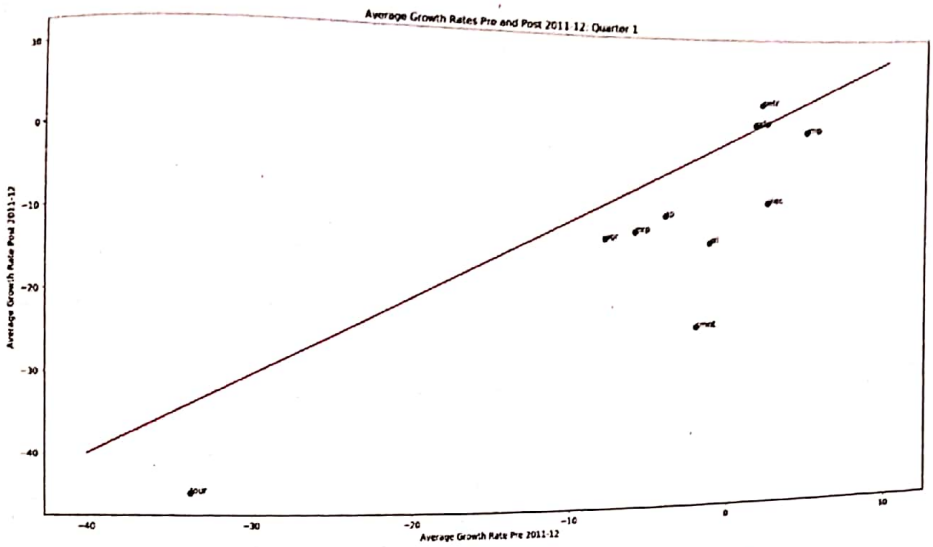
As we see in Figure 1, the average growth rate of GDP is very close to the 45-degree line, i.e., the average GDP growth rate is similar both before and after 2011-12. However, the average growth rate of all the explanatory variables is below it. Since GDP is composed of these explanatory variables, the average growth rate of the explanatory variables should also be similar in both the periods; which is not the case here.

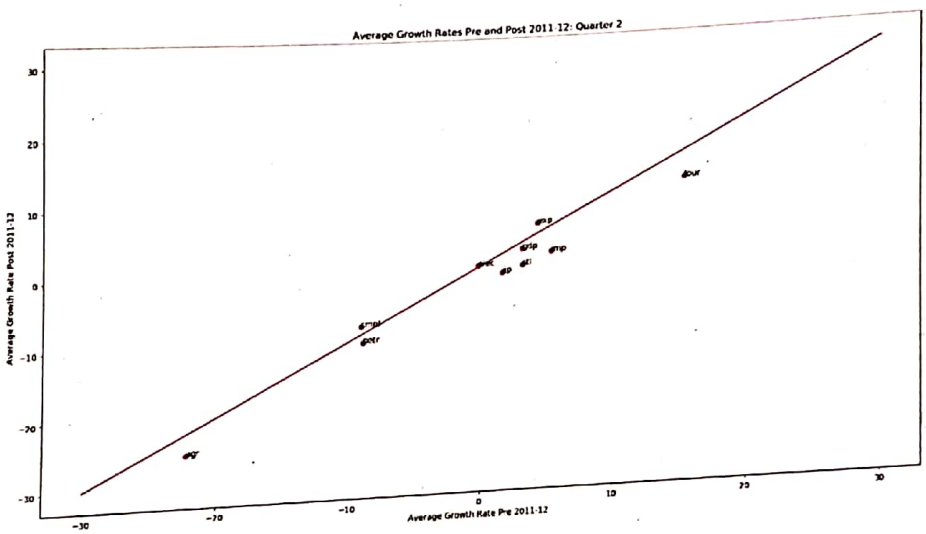
Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248



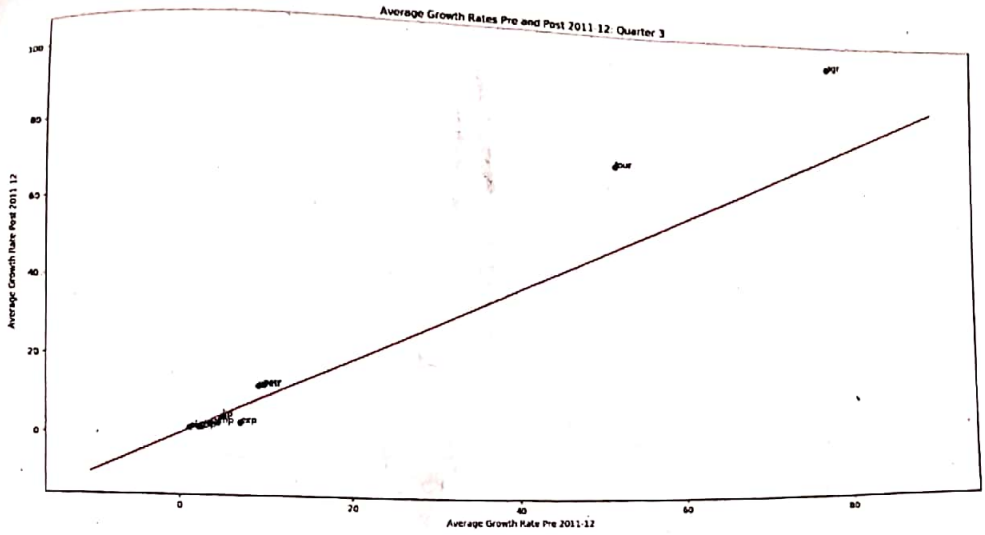
In Figures 2 – 5, we have repeated the same for quarters 1 – 4. In contrast to Figure 1, each of Figures 2 – 5 have some variables which are above the 45-degree line. Since cumulatively the average growth rate of all the variables is below the 45-degree line, the average growth rate for the quarters individually is not similar to that for the entire period. We can also see that the explanatory variables also follow different trends across the quarters for the two periods. For example, in quarter 3, Agriculture is below the 45-degree line, but it is above it in quarter 4. This suggests that there may be a quarter, or a group of quarters, which could be contributing significantly to the GDP. This forms a basis for our next test, which was to determine whether any quarter/group of quarters exhibits different trends before and after 2011-12.

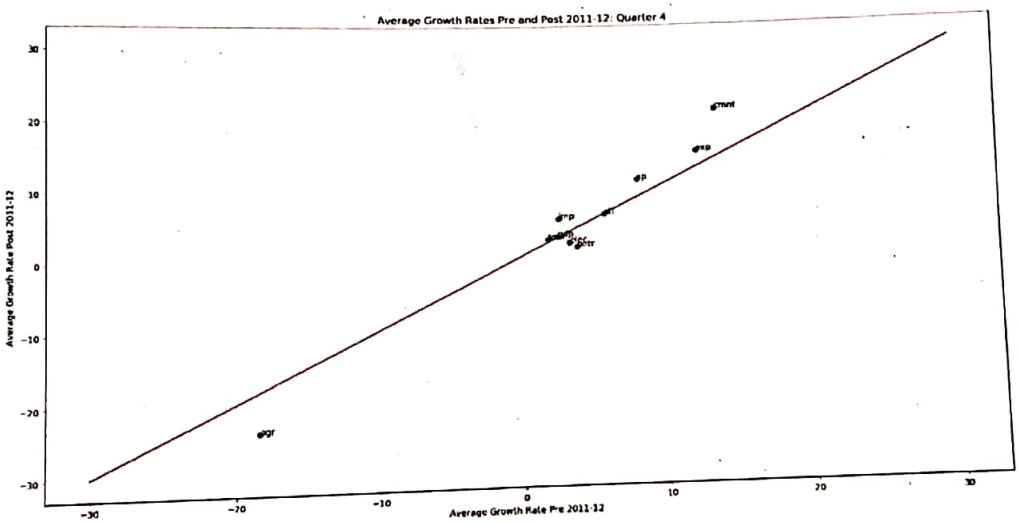
Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248





Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248





5.2 Test 2

As we can see from the summary table, the variable Q2*Tdummy comes out to be statistically significant. So, Quarter 2 behaves differently in the second time period as

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

compared to the first in comparison to how the other Quarters behave in the second period compared to the first.
This might lead to speculation that it is the Quarter 2 data which makes the post-2012 GDP over-estimated.

6 Discussion and Concluding Remarks

- On expanding the frequency of data, we realise that the quarterly trends do not match the annual trends.
- Test-2 leads to the speculation that it is the Quarter 2 data which makes the post-2012 GDP different from pre-2012.

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

16

Analysis of Quarterly Data For India's GDP - Examining The EAC-PM Rebuttal

Project Report I

Team 16
Ritoma Sen, 2018258
Shivangi Dhiman, 2018265
Rahul Kukreja, 2018254
Nikhil Krishnan, 2018248

1 Outline of the Analysis

The EAC-PM's rebuttal to the paper published by Dr Arvind Subramanian claims that his paper would not stand the scrutiny of policy and research standards as it lacks rigour in terms of the chosen explanatory variables as well as the frequency of data. This means that should the frequency of data and the number of explanatory variables be increased, one ~~should ideally be able to~~ can possibly better explain India's outlier behaviour.

also the quality I guess

We know from the paper that the average growth rate before and after 2012 is similar – 7.5% vs 6.9%. Thus, the average growth rate of the explanatory variables should also be similar for the two time periods. Following Dr Subramanian's methodology, we test this for our selected explanatory variables using their quarterly data. Here we are expanding our dataset in a two-fold manner - first, we increase the number of explanatory variables and second, we increase the frequency of the data by using quarterly data instead of annual. Since one of the arguments of the EAC-PM's rebuttal is a lack of sufficient explanatory variables that are inherent for India's GDP as well as a lack of enough data points, these two changes should help towards explaining India's higher GDP compared to other countries.

it's not a box plot, init?

The main idea behind the box plot (Figure 2) in Dr Subramanian's paper is that if we plot the average growth rate before 2012 and after 2012 on the x and y axes respectively, points which are closer to the 45-degree line of equality have the same average growth rate in both the periods. Points below the line of equality show a decline in the average growth rate post-2012, and points above it show an increase. From this, we may say that the variables whose average growth rate increases after 2012 could be the reason for India's high GDP (such as Electricity and Petroleum).

Given the yearly box plot in Dr Subramanian's paper (Figure 2), we first run the same for quarterly data for the selected variables. We would like to see if increasing the frequency of data points gives a different average growth for the variables and GDP, and this may further tell us if there are certain other variables which might be contributing heavily to the outlier behaviour of India.

should check this statistically?

India is a 2+ trillion dollar economy. The difference in 7.5% of 2+ trillion & 6.9% of 2+ trillion is not trivial

Be careful in raising your arguments.

We also conduct the same for each individual quarter to check if there is any particular quarter where a particular variable has a high contribution to outlier behaviour, but not in the other quarters; it might show different behaviour cumulatively when averaged out for all the quarters.

Additionally, we also hypothesise that a comparison between estimated and actual values of GDP should be able to tell us if the overestimation of GDP as stated in Dr. Subramanian's paper can be attributed to the introduction of the new methodology in 2012. We calculate the estimated values of the GDP from the regression equation (equation 1) and plot it against the actual values of GDP obtained from our dataset. If indeed India's GDP is being overestimated, and it is due to the introduction of the new method, the differences between the estimated and actual values should be much less before 2012 as compared to after 2012.

We will also draw this comparison for quarterly data, and then compare our results for quarterly and annual datasets. For quarterly data, we will change our regression equation slightly and add dummy variables for the quarters (equation 2). Hence, as opposed to the 18 data points that the annual data has, we can use 72 data points by using quarterly data and get a richer analysis, which we can then use to see if averaging out the data over the year causes certain things to be hidden.

As a further analysis, if we are able to obtain quarterly data for all these explanatory variables, we would be able to pinpoint exactly which quarters India acts as an outlier for, and perhaps been able to explain why India behaves as an outlier for those quarters.

2 Variables and Data Sources

Variable	Variable Name in Regression Equation	Description	Link	
Import Growth	import	The relative change in the goods and services that are bought from other countries.	http://www.epwrfits.in/NAS_Series.aspx	http://www.epwrfits.in/NASDisplayData.aspx
Export Growth	export	The relative change in the goods and services that are sold to other countries.	http://www.epwrfits.in/NAS_Series.aspx	http://www.epwrfits.in/NASDisplayData.aspx
Electricity	electricity	Index of electricity production	http://www.epwrfits.in/IIP_Series.aspx?SearchNode=149	
Petroleum Consumption	petroleum	Consumption of petroleum products	https://www.indiastat.com/petroleum-data/25/sales-marketing-and-consumption-of-petroleum-products/248/consumption-of-petroleum-products-1950-2019/379578/stats.aspx	
Foreign Tourist	tourist	Number of foreign tourist arrivals in India	https://www.indiastat.com/table/tourism/29/month-wise-foreign-tourist-arrivals-1988-2018/449535/372103/data.aspx	
Cement	cement	Index of cement production	http://www.epwrfits.in/IIP_Series.aspx?SearchNode=171	
Steel	steel	Index of steel production	http://www.epwrfits.in/IIP_Series.aspx?SearchNode=149	
Value of Output from Agriculture and Allied Activities	agriculture	Agricultural yield	http://www.epwrfits.in/Agriculture_All_India_State.aspx	
Railways	railways	Passenger Traffic on	https://www.indiastat.com/table/transpor	

Team 16

		Indian Railways	t-data/30/railways/237/289233/data.aspx
Credit Growth	credit	Relative increase in loans	https://insights.ceicdata.com/Untitled-insight/myseries
IIP	iip	Index of Industrial Production	https://www.indiastat.com/Searchresult.aspx
GDP	gdp	Gross Domestic Product	https://stats.oecd.org/index.aspx?queryid=60702#

We have tried to include each of the 17 variables which according to Dr Subramanian co-move with the GDP growth, and explain a fair amount of its variation (subject to availability of quarterly data).
Some of the datasets are for monthly data and not quarterly, and hence will need to be averaged for every three months.

3 Statistical Model and Tests

This is a time-series regression for only India where i suffixes year from 2001 – 2018.

$$GDP_i = \beta_0 + \beta_1 import_i + \beta_2 export_i + \beta_3 electricity_i + \beta_4 petroleum_i + \beta_5 tourist_i + \beta_6 cement_i + \beta_7 steel_i + \beta_8 agriculture_i + \beta_9 railways_i + \beta_{10} credit_i + \beta_{11} iip_i \quad \text{---- (equation 1)}$$

This is again a time-series regression for only India but for quarterly data where i suffixes quarters from 2001–2018 and Q_i is a dummy variable for $Quarter_i$.

$$GDP_i = \beta_0 + \beta_1 import_i + \beta_2 export_i + \beta_3 electricity_i + \beta_4 petroleum_i + \beta_5 tourist_i + \beta_6 cement_i + \beta_7 steel_i + \beta_8 agriculture_i + \beta_9 railways_i + \beta_{10} credit_i + \beta_{11} iip_i + \beta_{12} Q1 + \beta_{13} Q2 + \beta_{14} Q3 \quad \text{------ (equation 2)}$$

In this equation, the base quarter is Q4 since the intercept corresponds to Quarter4 when all Q_i are equal to 0.

Analytical tests have been described in the analysis section (I) of this report.