India's GDP mis-estimation: An analysis of the role of Agriculture and Service Sectors

Akash Sharma
Econometrics and Statistics Lab, IIIT Delhi

Background

GDP growth is an important indicator for the economic growth of a country. Accurate estimation of this indicator is crucial for decisions on starting businesses, designing sound policies for future economic growth. Mis-estimation, specifically overestimation of this important measure of economic growth can mislead with regards to the growth of the country due to which the process of calculating the GDP should be carefully checked and validated from multiple sources. Based on their analysis, Subramanian et al. found that the correlations between 17 independent indicators of economic activity and GDP broke down after 2011, leaving India an outlier compared to other countries' economic data concluding the mis-estimation of GDP.

Problem Query

Statement from the EAC to the Prime Minister: *"Further, a cursory look at the indicators suggests a strong link with Industry indicators (a sector that contributes an average of 22% to India's GDP), while the representation of Services (60% of GDP) and agriculture (18% of GDP) is as good as missing. It is difficult to believe that indicators in the Services sector would not correlate with Indian GDP."*

Following is the regression model that was used in Subramanian's paper,

$$GDP = \beta_0 + \beta_1.credit_i + \beta_2.electricity_i + \beta_3.export_i + \beta_4.import_i + \beta_5.India_i + \epsilon_i$$

This regression model considers indicators which are mostly representative of the manufacturing sector which incurs about 22% of the total GDP. However, more important indicators such as agriculture and the service sector which contribute about 18% and 60% of the total GDP respectively have not been considered in the model. This analysis of cherry picking the variables without considering all the factors can possibly introduce a significant bias in the analysis.

Analytical Framework

Research Question: Has the non-inclusion of services and agriculture sectors driven the results of this analysis? Generate statistical tests for or against this claim.

In this study, we consider indicators from the Services and Agriculture sectors in addition to the one's chosen by Subramanian over two time periods, 2002-2011 & 2012-2016, to observe the impact of the result and whether the analysis performed by Subramanian et al. holds true. We propose the following model,

$$GDP = \beta_0 + \beta_1.credit_i + \beta_2.electricity_i + \beta_3.export_i + \beta_4.import_i + \beta_5.India_i + \beta_6.ict_i + \beta_7.travel_i + \beta_8.insurance_i + \beta_9.transport_i + \beta_{10}.fert_i + \beta_{11}.agriland_i + \beta_{12}.cropprod_i + \epsilon_i$$

Hypothesis Testing: We aim to use null hypothesis testing to check whether the parameters of the new regression model are statistically significantly different from zero or not. This would indicate the importance/contribution of the indicator in the model. Formally, we set up the hypothesis test:

$$H_0 : \beta_j = 0 \quad (1)$$

We use the t-statistic for every indicator to test whether it is significantly different from zero to obtain the corresponding results.

Variables and Descriptions

Following are the variables used for Subramanian's analysis:

- Electric power consumption (kWh per capita)
- Imports of goods and services (constant LCU)
- Exports of goods and services (constant LCU)
- Net domestic credit (current LCU)

For our study, we consider the following additional variables(due to their high correlation coefficients with the share of the respective sector to GDP):

- **Agriculture sector indicators:**
  - Fertilizer consumption (kilograms per hectare of arable land)
  - Agricultural land (sq. km)
  - Crop production index (2004-2006 = 100)
- **Service sector indicators:**
  - ICT goods exports (US$)
  - Insurance and Financial services (US$)
  - Travel services (US$)
  - Transport services (US$)

All the variables were obtained from [3]

Results

Regression on Subramanian's model

We performed regression over Subramanian's model of GDP within the same conditions and data sources he used.

	2002-2011		2012-2016	
Variable	Coefficient	p-Values	Coefficient	p-Values
import	4.159e+00 (1.776e+00)	0.0263*	4.567e-01 (2.651e-01)	0.0893
export	-1.270e+00 (3.238e+00)	0.6978	1.368e-00 (1.829e-01)	1.5e***
credit	2.115e-01 (1.409e-01)	0.1442	3.241e-01 (7.174e-03)	2e-16***
India	4.753e+12 (4.120e+13)	0.9090	2.735e+13 (2.063e+12)	2e-16***
electricity	-1.148e+08 (5.568e+08)	0.08381	-2.842e+09 (1.159e+08)	0.0167*

Regression on our model

	2002-2011		2012-2016	
Variable	Coefficient	p-Values	Coefficient	p-Values
import	6.812e-01 (1.106e+00)	0.54418	1.578e+00 (5.891e-01)	0.00963 **
export	-4.088e-01 (2.011e+00)	0.84077	8.061e-01 (3.419e-01)	0.02185*
credit	3.318e-01 (6.105e-02)	1.85e-05 ***	2.676e-01 (1.351e-02)	2e-16 ***
India	-9.291e+13 (5.778e+13)	0.2211	3.362e+13 (2.490e+12)	< 2e-16 ***
electricity	1.663e+08 (1.565e+09)	0.91634	1.095e+09 (3.846e+08)	0.00610 **
agriland	-6.052e+05 (2.898e+06)	0.83648	-1.642e+06 (6.238e+05)	0.01091*
cropprod	-1.759e+11 (1.187e+11)	0.15266	-3.334e+10 (2.920e+10)	0.25833
fert	3.562e+10 (2.544e+10)	0.17549	8.598e+10 (1.773e+10)	9.97e-06 ***
ict	2.198e+03 (7.871e+02)	0.01062 *	-3.430e+02 (1.308e+02)	0.01118 *
finance	-3.890e+03 (1.346e+03)	0.00852 **	-2.284e+02 (5.826e+02)	0.69644
travel	-7.868e+01 (4.678e+02)	0.86798	-1.739e+01 (1.183e+02)	0.88364
transport	1.910e+03 (7.941e+02)	0.02498*	-1.583e+02 (1.077e+02	0.14725

We then,performed regression on our model with added agriculture and service indicators. Here we made the following observations:

- Indicators which were already derived from the Subramanian's model (credit, electricity, imports and exports) remain to be significant in the new model as with the Subramanian's model (in accordance to the p values).
- The coefficient of India's dummy variable now shows up to be significantly different from zero in the post 2011 period than the pre-2011 in our model similar to the Subramanian's model.
- We found that Agriculture Land and ICT to be significant variables for the analysis of the GDP based on the p-value which was not considered in Subramaninan's model.

Note:

- Our analysis was based on data from 15 countries used for determining if India was an outlier or not.
- We defined the significance threshold as 0.05 for our variables.

Non Linearity

Model proposed by Subramanian et al. assumed a simplistic linear regression model. However, the complexity of agriculture and services in India could be better captured by non-linear variables. Here, we consider the dependence of non-linear variables for service and agriculture sectors. For instance, the concentration of IT industry is high in cities like Bangalore compared to other places and productivity might be higher for agriculture in places like Punjab. Further, factors such as weather, festivities play an important role in influencing the non-linearity of factors determining GDP.

2002-2011		
Variable	Coefficient	p-Values
Agriculture (1)	-7.184e+01	0.000545
Agriculture (2)	-1.495e+13	0.003476
Agriculture (3)	7.080e+12	0.110665
Agriculture (4)	-5.680e+12	0.121408
Service (1)	4.197e-01	0.218345
Service (2)	2.651e+12	0.403964
Service (3)	-3.946e+12	0.222988
Service (4)	-1.931e+12	0.549162
2012-2016		
Agriculture (1)	1.226e+03	0.007841
Agriculture (2)	1.806e+14	0.007865
Agriculture (3)	-7.022e+13	0.064634
Agriculture (4)	-5.102e+13	0.028595
Service (1)	-7.072e-03	0.996020
Service (2)	2.143e+13	0.030631
Service (3)	1.670e+13	0.137744
Service (4)	-2.826e+12	0.765309

We compare the model parameters between model proposed by Subramanian et al. with the modified model that includes the agriculture and service variables. Following are the main observations:

- For generic variables of agriculture and service, we observe that the significance of agriculture and service is higher compared to other proposed variables used in the previous model.

Following are the highlights of the above analysis:

- Variables of the power two are more significant compared to linear variables.
- On increasing the power of both services and agriculture sector, the overall statistical significance of the coefficients also decreases.

Conclusions and Discussions

- India was an outlier in the post 2011 period both before and after considering the additional indicators.
- Coefficients of most of the indicators taken for Service and Agriculture sectors were obtained to have a statistically significant difference from 0 in both periods.
- Model non-linearity is important considering that the growth in agriculture and services varies at different rates and from place to place.

References

References

[1] Arvind Subramanian. India's GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms, and Implications. 2019

[2] GDP estimation in India-Perspectives and Facts. Economic Advisory Council to the Prime Minister. 2019.

[3] World Bank. World Development Indicators, https://data.worldbank.org