

India's GDP mis-estimation: An analysis of the role of agriculture and service sectors

Akash Sharma

Undergrad Researcher ,IIIT-Delhi, India

Gaurav Arora

Associate Professor,IIIT-Delhi, India

Sumedha Shukla

Phd Scholar ,IIIT-Delhi, India

Arvind Subramanian

Senior Research Staff,Harvard Univ, USA

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."

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.

1. Introduction

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. Investors look at the country's GDP to decide whether or not they should invest in a particular country, This could in turn have important implications to foreign capital.

Mis-estimation, specifically overestimation of this important measure of economic growth can mislead with regards to the growth of the country dues to which the process of calculating the GDP should be carefully checked and validated by multiple sources.

2. Problem Statement

Subraminan et al[1] for the sake of tractability restrict themselves to only four of the indicators and claim that the GDP for India has been mis-estimated. These four indicators include the credit, exports, imports and the electricity consumption. Following is the regression model that was used in the original paper,

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

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.

In this work, we consider the service and agriculture sector to see the impact of the result and whether the analysis performed by Subramanian et al. holds true.

3. Variables

In order to accommodate the data indicators that were not included by Subramanian et al[1], we consider the indicators belonging to two major classes:

- **Service Sector Indicators:** Here, we consider the major services which contribute significantly to the GDP estimation such as Travel, Transport, Communications and High Technology Exports. India being a major technology hub, we considered the contribution to the GDP in terms of the technology based goods being traded.
 - **Logistics and Transport-** Logistics and transport services covers all transport services (sea, air, land, internal waterway, space, and pipeline) performed by residents of one economy for those of another and involving the carriage of passengers, movement of goods (freight), rental of carriers with crew, and related support and auxiliary services.
 - **Information and Communication Technology-** ICT services include services such as international telecommunications, and postal and courier services, computer data, news-related

service transactions between residents and non-residents, construction services, royalties and license fees, miscellaneous business, professional, and technical services.

- **Travel and Tourism services-** Travel and tourism services cover goods and services acquired from an economy by travelers in that economy for their own use during visits of less than one year for business or personal purposes. Travel services include the goods and services consumed by travelers, such as lodging and meals and transport (within the economy visited).
- **Insurance and Financial Services-** Insurance and financial services cover freight insurance on goods exported and other direct insurance such as life insurance; financial inter mediation services such as commissions, foreign exchange transactions, and brokerage services; and auxiliary services such as financial market operational and regulatory services.
- **Agriculture based Indicators:** Agriculture is the most important sector of Indian Economy and includes businesses such as dairy, meat, poultry, fisheries and food grains etc. We consider the following indicators–
 - **Forest area (% of land area)-** Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agro-forestry systems) and trees in urban parks and gardens.
 - **Food production index (2004-2006 = 100)-** Food production index is an indicator of edible food crops that are considered to have nutrients. Coffee and tea are excluded as they have no nutritive value.
 - **Agricultural irrigated land (% of total agricultural land)-** Agricultural irrigated land refers to agricultural areas purposely provided with water, including land irrigated by controlled flooding.
 - **Fertilizer consumption (kilograms per hectare of arable land)-** Fertilizer consumption measures the quantity of plant nutrients used per unit of arable land. Fertilizer products cover nitrogenous, potash, and phosphate fertilizers (including ground rock phosphate).

These indicators shown above are however chosen through a cursory look at their trends . In the paper, we will

be describing our rationale for the choice of indicators specific to the agriculture and service sector. While performing the selection of our indicators for the new model finally, we will be assessing the correlation of various indicators with the value added to their respective sectors (Agriculture or Service). Hence, these indicators listed above are tentative and can be changed in accordance to our correlation results.

Datasets. We used the data and definitions for the indicators from the World Bank dataset ¹ and from the food security portal ².

3.1. Analytic Framework

$$GDP = \beta_0 + \beta_1.credit_i + \beta_2.electricity_i + \beta_3.export_i + \beta_4.import_i + \beta_j.serviceIndicators_i^j \dots + \beta_k.agricultureIndicators_i^k \dots + \epsilon_i$$

Note: Here, *serviceIndicators* represents the different indicators of the service sector and similar is the case for the *agricultureIndicators*.

Hypothesis testing

The validity of the model on incorporating the new variables namely, service and agriculture has to be checked using hypothesis testing. More specifically, we aim to use Null hypothesis testing to check whether the parameters of the new regression model given statistically significantly different from zero or not. This would indicate the importance/contribution of the indicator in the model. Formally, we set up two hypothesis test:

(for all *serviceIndicators*)

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

(for all *agricultureIndicators*)

$$H_1 : \beta_k = 0 \dots \quad (2)$$

This estimation uses the t-statistical test for every new indicator added to check for the corresponding result.

Reduced Model

Instead, of viewing the regression model with the individual indicator variables, we consider the overall contribution (value-added to the GDP) of the agriculture and service sector as a whole. While Subramanian et al. considered only the linear relationships between the different variables

¹<https://data.worldbank.org/>

²<http://www.foodsecurityportal.org/api/countries/agriculture-value-ad>

which have a significant impact on the GDP, it is important to check for any non-linear relationships that might possibly exist and include them as part of the regression model. Hence, here we fit the regression model corresponding to the linear and quadratic relation between the service and agriculture variables.

$$GDP = \beta_0 + \beta_1.credit_i + \beta_2.electricity_i + \beta_3.export_i + \beta_4.import_i + \beta_5.services_i + \beta_6.service_i^2 + \beta_7.agriculture_i + \beta_8.agriculture_i^2 + \epsilon_i$$

Analysis of R^2

After performing the regression analysis, we will be comparing the goodness of fit values of both the models (Subramanian's and the new model) in order to compare the models. However, using R^2 for comparison is only an addition to the argument we will be making about the difference/similarity of the models.

4. Conclusions

GDP estimation is a complex process which includes considering multiple factors varying temporally. In order to verify the estimation of the GDP process, it is important to consider all these aspects without incorporating any bias into the models in the form of assumption. In this work, we aim to address the importance of two major GDP contributors, namely, the service and the agriculture sector for determining the GDP output which were not considered as part of Subramanian's report. We thoroughly study the estimate of the parameters corresponding to the service and agriculture variables using hypothesis testing to check for the validity of the model.

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.