Some problems with India’s GDP data and the way forward

Rajeswari Sengupta (IGIDR)¹

1. Introduction

Gross Domestic Product (GDP) is the most used statistic of macroeconomic performance of a country. It is an indispensable input for effective policymaking and also the main yardstick for assessing the health of an economy in academic research and media commentary alike. Reliable data on GDP is therefore a crucial requirement in any market economy. In India however, the last few years have witnessed widespread controversy about the credibility of the GDP data.

In 2015 the Central Statistics Office (CSO) revised the way GDP is calculated in India. They released a new GDP series with 2011-12 as the base year, replacing the older 2004-05 series. Since then the GDP series has attracted a significant amount of debate and discussion. While the new series showed India to be the fastest growing large economy in the world, various high frequency indicators painted a different picture. Studies have pointed out that the annual GDP growth rates during the 2015-2019 period may have been overestimated on average by 0.36 to 2.5 percentage points. For instance, Subramanian (2019) estimated that instead of the 7 percent growth rate obtained from the new GDP series for the period 2011-12 to 2016-17, the actual growth rate was likely to be around 4.5 percent.

When the new GDP series was introduced in 2015, the CSO published a handbook explaining the methodology (CSO, 2015a). However, despite questions and doubts being raised about the accuracy and veracity of the GDP estimates over the next few years, the CSO did not revise the handbook to reflect any changes that may have been implemented to address the problems. They also did not respond publicly to the criticisms made by the academics. In fact the CSO has continued to defend the new series and have emphasised how it is in line with international best practices, the coverage is better because of the use of new datasets and the methodology used is an improved one (CSO, 2015b, 2018). This

---

¹ Associate Professor at Indira Gandhi Institute of Development Research (IGIDR), Film City Road, Goregaon East, Mumbai 400065. Email: rajeswari@igidr.ac.in. The author is grateful to Josh Felman and Amey Sapre for useful discussions.

² Eventually the CSO was merged with the National Sample Survey Office (NSSO) and has now been redesignated as National Statistics Office (NSO); but for the purpose of this paper we continue to refer to it as the CSO because that is what it was called when the new GDP was introduced.

³ For details of the revisions done in the new series, see Nagaraj, Sapre and Sengupta (2019) for a comprehensive description.
refusal by the CSO to even acknowledge the problems, let alone address them has further entrenched the scepticism in the academic community.

With the inaction of the CSO and with every passing year, the intensity of the controversy seems to have gotten diluted. Nevertheless the problems remain, and it is important to keep the discussion alive, so that at the minimum, all stakeholders can use the data with appropriate caution.

With this objective in mind, in this paper we revisit some of the main issues that have been analysed over the last few years in context of the new GDP series, and offer some recommendations in order to improve the credibility of the data.\textsuperscript{4}

Changing the base year of GDP, which is a common practice all over the world including India, would typically lead to a marginal increase in the absolute size of the economy. This is because the use of improved methodologies and new databases would presumably be able to better capture the output produced in the economy. At the same time the growth rate of GDP would rarely change. In other words, alteration of the base year would have no impact on the pace of expansion or contraction of economic activity, as the case maybe. This however was not the case with the base year revision of 2011-12.

The absolute size of GDP in 2011-12 contracted by 2.3 percent compared to the old series while the annual real GDP growth rate increased drastically from 4.8 percent in the old series to 6.2 percent in the new series, in 2013-14. Most notably the growth rate of the manufacturing sector went up sharply from -0.7 percent in the old series to 5.3 percent in the new series. In a nutshell, the new GDP series showed a booming economy.

On the other hand, many high frequency indicators conveyed an entirely different story, that of a prolonged economic slowdown. A GDP growth rate of 7-7.5 percent seemed inconsistent with slowing bank credit growth, deteriorating balance sheets in both the banking and corporate sectors, weak exports, sluggish private sector investment, poor capacity utilisation in the industrial sector, and declining corporate profits, among other indicators (Subramanian, 2019).

Consequently, various stakeholders began questioning the reliability of the new GDP series. Several economists, independent researchers and analysists wrote extensively about the problems besetting the GDP series.\textsuperscript{5} In the process, a number of issues were unearthed, both with regard to the methodologies and the databases used. The main problems can be categorised into those that pertain to the nominal GDP numbers and those concerning the real growth rates.

\textsuperscript{4} This is not intended to be a comprehensive description of all the problems afflicting the new GDP series but rather a brief summary of some of the main issues that we consider important.

In case of the *nominal* numbers, most of the problems are centred around the use of a new database introduced by the Ministry of Corporate Affairs, called the MCA21 database which contains financial accounts of the private corporate sector. This new database is now used to calculate the GDP of the organised sector. There also exist problems regarding the estimation of output produced by the unorganised sector.

On the other hand, misestimation of *real* GDP growth rates under the new series is primarily due to the deflator problem. There are issues both with the deflation procedure traditionally used by the CSO as well as with the specific deflators being applied to convert nominal output (or value added) into real value added. According to some estimates, the deflator problem alone could have resulted in an overestimation of the real GDP growth rate by as much as 2 percent in the first few years after the new series was released.

The problems concerning the deflators have existed for a long time even before the new GDP series was released, so in a way these are legacy problems of GDP measurement in India. However the distortions they create have come to the fore with the new GDP series (Sengupta, 2016b). One reason for the resurgence of the deflator problem is because under the old GDP series, real growth rate was calculated largely using volume based measures, especially for the manufacturing sector for example, whereas in the new series, it is calculated using value based measures. As a result the deflating procedure has become more critical than before for obtaining the real estimates. Yet the deflators used by the CSO or the methodology applied are not adequate for the task.

A combination of the factors mentioned above could lead to either an over or underestimation of the both the level as well as growth rate of GDP depending on which of the factors is more dominating at any given point of time. Irrespective of the exact direction of the mis-estimation, the fact remains that the new GDP series is wraught with problems which have dented its credibility.

This is particularly relevant in the present juncture when the Indian economy is slowly recovering from the large shock of the Covid-19 pandemic. GDP data is being used by academicians as well as policymakers to understand how strong or feeble the recovery is, but if this data itself is fundamentally problematic, then the question remains, can we really say anything meaningful about the economic recovery simply by analysing GDP growth rates?

In the absence of any robust response from the CSO or concerted effort to address the measurement problems, it seems that increasingly alternative indicators must be used for tracking the performance of the Indian economy over time. The advent of a digital age and the growing use of internet and technology in all spheres of the economy have also made this relatively easier today compared to what it would have been a few decades ago. Given the gradual erosion of credibility of official statistics, it might also make sense to start

---

6 One may even argue that over time the CSO has improved the deflation procedure by using a combination of various price indices, still the fundamental problems remain unaddressed.
making greater use of private data sources to complement the official statistics and also to validate the picture emerging from official data as regards the state of the economy and its various facets.

At the same time it is worth noting that private data, however useful, cannot be a substitute for official statistics, it can only be a complement. The use of private data can be helpful as a cross-check on the official GDP numbers. But it is not possible to use indicators to proxy GDP because the Indian economy is highly diverse and because growth rates in various sectors are not well correlated. So we will end up getting very different results depending on the particular indicators that we select. This further underscores the importance of getting the official GDP numbers right, because there is no substitute.

2. The deflator problem

The methodology followed by the CSO is broadly similar to what is done all over the world. It first obtains data on the nominal values of output from financial accounts of firms. It then deflates these figures by price indices to arrive at real GDP estimates. But the CSO deviates from international practice in two ways: the price indices used for deflation and the deflating procedure followed. We consider each deviation in turn.

2.1 Price indices used

In the GDP series, the real figures are derived by deflating the nominal data using price indices. If the deflators used are inappropriate, the estimated real magnitudes will be distorted. In most countries, nominal output is deflated using the producer price index (PPI). India does not have a PPI, and hence the CSO uses the wholesale price index (WPI) instead. The WPI however is beset with shortcomings. To begin with, it does not measure the price of services and services constitute two-thirds of the Indian GDP. Secondly, the WPI assigns a heavy weight to commodities, especially oil. This means that when oil prices fall (or go up) the WPI falls (increases), and this leads to an over (under) estimation of the GDP growth rate, as explained in the next section.

2.2 Deflation procedure

There is a fundamental issue with the deflation procedure used in Indian GDP measurement. Gross value added at current prices is the difference between the value of aggregate output of goods and services and the value of intermediate inputs or raw materials used in the production process. In order to convert this into value added in constant prices i.e. to obtain

\footnote{The deflator related issues in GDP measurement have also been explained in detail in Nagaraj et al (2019).}
real value added, the nominal GVA must be deflated using suitable price deflators. Nearly every major country of the world (with the exception of China and India) follows a methodology called “double deflation”, a standard international practice. Sims (1969) provides a theoretical justification for double deflation. In this method, value of nominal output is deflated using an output price deflator, while value of inputs is deflated using a separate input price deflator. Then, the real inputs are subtracted from real outputs to derive estimates of real GVA.

However the CSO follows a single deflation practice. It first computes the nominal GVA and then deflates this number using a single deflator to obtain the real GVA. This means that the single deflator has to balance its role as a deflator for the nominal value of output and its role as a deflator for the nominal value of inputs (Nagaraj and Srinivasan, 2016). The CSO (CSO, 2015a) does not explain why it chose to stick with the single deflation procedure even as the GDP estimation methodology underwent drastic changes in the new series.

The implication of using single deflation is that if input prices are in sync with output prices, there is no problem and both methodologies will give similar results. But if the two price series diverge, then single deflation can over, or understate growth by a big margin, depending upon the direction of movement of the two prices. The most commonly used deflator in India as mentioned earlier is the WPI which is heavily weighted towards commodities. Hence, whenever there is a sharp fall in oil prices for example, nominal GDP growth in India will be inflated. This is exactly what had happened for the first few years after the release of the new GDP series. Figure 1 below shows the divergence between WPI quarterly inflation and retail price inflation as captured by the consumer price index (CPI) over the period from June 2012 to December 2021. In the 2014-2017 period, oil prices plunged, causing the WPI to fall sharply relative to the CPI. Between December 2014 and December 2017 WPI inflation averaged -0.16 percent while CPI inflation averaged 4.37 percent. This meant that real GDP growth largely obtained by deflating nominal GDP using WPI, was probably overstated, which in turn drew severe criticisms from various quarters about the reliability of the GDP series.

The reason this happens is because if the price of inputs falls sharply, profits will increase, and nominal GVA will go up. Since real GDP is supposed to be measured at “constant prices”, this increase needs to be deflated away. While double deflation will do this, single deflation will not work in this situation. In fact, if a commodity-weighted deflator like the WPI is used, nominal growth will be inflated, on the grounds that prices are actually falling. In this case, real growth will be seriously overestimated. As the gap between input and output prices starts to close, the problem will diminish. But that in turn could send a misleading signal, because it might seem that growth is slowing, when only the measurement bias is disappearing (Nagaraj et al, 2019).
The opposite problem will happen when oil (or in general commodity) prices start rising, as is the case at present. With the Russian invasion of Ukraine and consequent supply shortages, global crude oil prices have increased sharply. This is getting reflected in a persistent rise in WPI inflation in India. Between April 2021 and February 2022, WPI inflation averaged 12.7 percent, the highest in more than a decade. Whereas consumer prices index (CPI) inflation i.e inflation at the retail output level, has been hovering around 5-6 percent. The rapid increase in the WPI relative to the CPI is imparting an upward bias to the deflator, which increased at the remarkable rate of 8 percent in the second quarter of 2021-22. If this deflator is being overestimated, then real GDP growth rate could be underestimated right now.

![Figure 1: Divergence between WPI and CPI inflation](image)

The numerical example given below further explains the underlying problem. Let us consider a case where actual production is stagnating at 100 units in years 2021-22 and 2022-23 but the output and input prices are changing. Assume that a firm increases its output selling price by 5 percent, in line with CPI inflation, whereas the price of its raw materials goes up by 10 percent because the WPI inflation is 10 percent. Nominal value added, defined as value of sales less value of raw materials, will increase by 3.75 percent from 2021-22 to 2022-23. Since this increase arises entirely from price changes, it needs to be deflated away in order to obtain “real GVA at constant prices”.
Table 1: Single vs Double Deflation illustrated with a numerical example

<table>
<thead>
<tr>
<th></th>
<th>2021-22</th>
<th>2022-23</th>
<th>2022-23</th>
<th>2022-23</th>
<th>2022-23</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal</td>
<td>Nominal</td>
<td>Real</td>
<td>Real</td>
<td>Real</td>
</tr>
<tr>
<td>Sales</td>
<td>100</td>
<td>105</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw materials</td>
<td>20</td>
<td>22</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominal value added</td>
<td>80</td>
<td>83</td>
<td>80</td>
<td>75</td>
<td>82</td>
</tr>
<tr>
<td>CPI inflation (%)</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPI inflation (%)</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth in value added (%)</td>
<td>3.75</td>
<td>0</td>
<td>-6.25</td>
<td>-1.25</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ estimates

As Table 1 shows, using the “double deflation” methodology real GVA growth is zero. This makes sense because actual output has remained unchanged. However, if the CSO’s method of single deflation is applied, we would take the nominal value added and deflate it by the WPI. Since nominal growth is 3.75% and the WPI has gone up 10%, the real GVA growth would be estimated to be -6.25 percent. In other words, growth would be significantly underestimated.\(^8\)

This problem of single vs double deflation is not new in India nor did it start with the release of the 2011-12 base year GDP series. This is an old problem which regained importance once the new series was published. Under the old system of GDP estimation, real growth rate was calculated using volume based measures. Whereas in the new system, it

---

\(^8\) This example assumes that in the CSO’s single deflation procedure, WPI is the only price index that is used. In reality however the CSO uses a combination of WPI and CPI but the WPI plays an overwhelming role as the GDP deflator.
is calculated using value based measures (Nagaraj et al, 2019) which in turn has made the deflator issue more critical. Moreover, in the initial few years after the release of the new series, the WPI and the CPI series diverged significantly with WPI plunging on the back of low oil prices and CPI being higher. In 2015-16, the year when the new GDP series was introduced, WPI inflation was -3.7 percent while CPI rose by 4.9 percent and this divergence was found till 2017-18 implying that purely on this count, the real GDP growth rates in the new series may have been overestimated.

This divergence between the WPI and CPI can keep recurring from time to time because oil and other commodity prices can be quite volatile and are highly susceptible to global economic and political shocks. Whenever there is such a divergence, GDP growth is going to be incorrectly estimated. Hence this is a serious problem that needs urgent attention.

2.3 Service sector deflator

Deflator problems also plague the estimates for the service sector, which accounts for bulk of the Indian economy. Currently, the CSO uses a combination of WPI and CPI to deflate the nominal values for this sector. Using WPI as a deflator is erroneous because the weight of services in the WPI basket is negligible. WPI does not measure the price of services. The fact that it mostly reflects the price of commodities, such as oil implies that when oil prices fall, the WPI falls, and this leads to measured deflation in the services sectors (notably finance and trade) even if service costs could actually be rising. As a result, growth in services could be overstated by a large margin (Sengupta, 2016a).

3. Issues with nominal GDP

While the deflator problems are likely to distort the calculation of real GDP growth rates, there are several problems with the nominal GDP data under the new series. We briefly discuss a couple of them here.

3.1 MCA21 database

The use of the MCA21 database for computing value added in the private corporate sector has perhaps attracted the maximum criticism and has been at the heart of the GDP measurement debate. Among the plethora of problems associated with the use of this database, a major one concerns the “blowing-up” problem. Scaling up or extrapolating from smaller samples to capture the output produced by the universe of all companies while not a new feature of the 2011-12 GDP series has become more complicated with the use of the MCA21 database.

---

9 The MCA 21 database replaced the Annual Survey of Industries (ASI) database in the new GDP series.
In the new methodology, an “active” set of companies consists of those who have filed their annual financial returns at least once in the past three years. This set is what the CSO considers the universe of companies whose GVA must be estimated to get the GVA of the private corporate sector. The sample for any given year on the other hand consists of “available” companies which have filed their returns in that year. Not all companies file returns every year and the CSO does not have data on the returns of the non-filing, active companies i.e. companies that are a part of the universe but not a part of sample so to speak. Hence, the GVA estimates of the non-filing or unavailable companies are obtained by blowing-up the estimates of the available companies.

As discussed in Nagaraj et al (2019), this procedure implies that in case of any problems with the unavailable companies, the blowing-up procedure might lead to erroneous estimates of growth rates. For instance if the unavailable companies have already been wound-up or they are loss-making or they are fictitious, shell companies that merely exist on paper, then scaling up the estimates of available companies might lead to overestimation of growth rates. In fact, it would also be deeply problematic if the “active” set of companies itself contained shell companies that engage in fictitious transactions in order to evade laws and file false returns.

While these problems have been highlighted in several studies (Nagaraj et al, 2019; Sapre and Sinha, 2016, 2017) the CSO till date has not made the year wise number of “available” and “active” companies publicly available. Hence, the exact number of companies for which the GVA is blown up on a yearly basis is unknown. Unless the CSO provides suitable evidence that the set of “active” and “unavailable” companies do not have the problems mentioned above and are indeed properly functioning companies with valid returns, the reliability of the GVA estimates obtained using this database will remain dubious.

### 3.2 Unorganised sector

Estimation of unorganised sector output has always been difficult in India. In case of the unorganised sector, for unincorporated enterprises in manufacturing and services, benchmark GVA estimates are initially prepared at detailed activity level for the base year of (i.e. 2011-12). This uses GVA per worker for every activity which in turn is obtained from the Employment-Unemployment Survey and the Unincorporate Enterprise Survey. Once the estimation is done for the base year, for all subsequent years, GVA is obtained by extrapolating using various indicators that might be relevant to the corresponding economic activity.
For instance, for the unorganised manufacturing sector, the index of industrial production (IIP) is used as an indicator, for “maintenance and repair of motor vehicles and motor cycles”, motor vehicles sales growth is used, for “accommodation and food service activities” or “storage and warehousing” or “real estate activities” corporate sector growth is taken as an indicator and so on.

The extrapolation assumes that the GVA of the unorganised and organised sectors are growing at the same rate for all years after the base year. This by itself, can be quite a problematic assumption, especially in instances when economic shocks might disproportionately impact the unorganised sector.

From 2016 onwards the unorganised sector has indeed been disproportionately affected by a series of shocks. First came the Demonetisation episode of November 2016 when overnight, more than 85 percent of the currency in India was declared illegal by the government. This was a huge monetary shock which hurt the unorganised sector significantly more than the organised sector. This is because the functioning of majority of the unorganised sector in the Indian economy depends on the use of cash (Bhattacharya and Mitra, 2017). In the aftermath of this shock, the unorganised sector was once again adversely impacted by the shoddy implementation of the Goods and Services Tax in 2017 (Mankar and Shekhar, 2017), which necessitated a particularly difficult and costly adjustment for the enterprises in this sector. In 2018, the non-banking finance (NBFC) sector experienced a mini crisis which in turn impacted unorganised sector firms since they depend heavily on NBFCs for funds. From 2020 onwards, the Covid-19 pandemic has arguably dealt a more severe blow to the unorganised sector compared to the organised sector enterprises.

Given these circumstances and in absence of any adjustment made by the CSO to account for these shocks and their impact on the unorganised sector, the assumption that unorganised sector has been growing at the same rate as the organised sector since 2011-12 is likely to lead to a big overestimation of nominal GDP growth rate.

Furthermore, the GVA per worker estimates used in unorganised sector output estimation are by now more than a decade old. This is because the surveys required for re-estimation of GVA PW need the census for their frame and the census is yet to happen. This implies that the unorganised sector GVA estimates might infact be far from reality.

4. Way forward

The above issues highlight the urgent need to reform India’s statistical system going forward, and more specifically, thoroughly re-examine the changes brought about in the new GDP series.

There appear to be recurrent problems concerning the GDP deflator in India. Steps must be taken to revise the method of deflation and to come up with new, appropriate deflators.

As mentioned earlier, in most countries, nominal values are deflated using the producer price index (PPI). India lacks a PPI, so the CSO uses the WPI instead. This is problematic for reasons already discussed. Efforts must be taken to develop a Producer Price Index (PPI) which would be an ideal deflator for nominal GDP estimates. This is in sync with the practices followed in the developed countries.

Currently there are no good service sector deflators. As discussed, using the WPI as a service sector deflator is inherently problematic. One alternative could be to use the relevant components of the CPI to deflate the service sector estimates. The change to CPI makes sense specifically for the services sector, because the CPI has extensive information on price movements in the various services subsectors (Nagaraj et al 2019). However, eventually an appropriate service sector price index must be developed.

Over time the transition must be made to the double method. This however is problematic now that the MCA21 database is used for GVA estimation. This being a database of financial accounts of companies does not contain information on output or input prices at a commodity level, which are required for doing double deflation.

To some extent using the CPI might also help, especially for the manufacturing sector, though it will not entirely solve the problems of the single deflation method. This is demonstrated in Table 1. Using CPI in the single deflation procedure gives a real GVA growth rate of -1.25 percent which is still better than -6.25 percent (obtained using WPI as the deflator) and is closer to the actual growth of 0 percent in this illustration. Commodities constitute a much smaller share of an average household’s consumption basket and hence CPI does not suffer from the same problem as the WPI.

The estimation of the unorganised sector output needs to be revisited in order to account for periods when the growth rate of this sector might be significantly lower than that of the organised sector. The CSO must make some appropriate adjustments for such periods so that the aggregate GDP growth rate is not overestimated.

Lastly, a detailed statistical audit of the new GDP series needs to be conducted (Sapre and Sengupta, 2019) which would also look into the problems of using the MCA21 database. An independent committee consisting of national and international experts must be constituted, with the mandate and authority to diagnose the problems with the new series and recommend ways to resolve them. The findings of the committee must be made publicly
available for the sake of accountability and transparency. The end objective should be to come up with a statistically sound and credible method of GDP measurement so that the data accurately reflects the conditions in the economy.

A well-defined communication policy is also vital as that would help the CSO to regularly engage with all categories of data users (Sapre and Sengupta, 2019). The release of the new GDP series, and the subsequent controversy has significantly eroded the credibility of the CSO as well. In order to re-establish the credibility of the institution and of the GDP estimates the CSO must keep the data users informed about the changes in data sources and estimation methods in a timely and transparent manner.
References


CSO (2015a), Changes in Methodology and Data Sources in the New Series of National Accounts, Base Year 2011-12, New Delhi Central Statistical Office


