R. NAGARAJ *
Indira Gandhi Institute of Development Research

AMEY SAPRE †
National Institute of Public Finance and Policy

RAJESWARI SENGUPTA ‡
Indira Gandhi Institute of Development Research

Four Years After the Base-Year Revision: Taking Stock of the Debate Surrounding India's National Accounts Estimates§

ABSTRACT   In 2015, with the release of the 2011–12 base-year GDP series, the Central Statistics Office (CSO) substantially revised the way GDP is calculated in India. According to the new series, India is the fastest growing large economy in the world. However, other trusted measures of the state of the economy convey a discordant picture. This discrepancy has led to an active debate over the last few years. Numerous studies by academic scholars have identified, analyzed, and documented the problems with the kind of data used in the new series as well as with the specific methodologies applied. The criticisms have cast persistent doubts on the new GDP series and have dented the credibility of India's National Accounts Statistics. The debate seems at an impasse. In this paper, we provide a comprehensive summary of the issues surrounding the new GDP series as highlighted by academic experts and outline recommendations for a possible way forward to resolve India's GDP data crisis.

* *nag@igidr.ac.in*
† *amey.sapre@nipfp.org.in*
‡ *rajeswari@igidr.ac.in*

1. Introduction

Gross domestic product or GDP is one of the most important macroeconomic indicators of the level of economic activity in the country.[1] It drives economic policies, is a crucial input in the fiscal calculations of the government, affects investor confidence, and conveys a comprehensive picture about the health of the economy to the rest of the world. In India, while policy outcomes and the performance of the economy are debated extensively using GDP growth numbers, evaluation of the quality of data and assessment of the soundness of the methodology used for making the GDP estimates do not get as much attention. The last few years have been an exception in this regard. In January 2015, the Central Statistics Office (CSO) released the 2011–12 base-year series of the National Accounts Statistics (NAS) to replace the earlier 2004–05 series.[2] Since then, issues in the measurement of GDP have taken center stage in academic and policy debates.

The 2011–12 series, apart from changing the base year of the NAS, also introduced several methodological changes in GDP computation. These changes were done primarily to align the methods with the most recent international guidelines of the United Nations System of National Accounts (UNSNA), 2008 (SNA 2008). New data sources, particularly for the private corporate sector (PCS, which includes organized manufacturing as well as service sector enterprises), were also introduced. As a result of these changes, the growth rates at the aggregate level, as well as for some sectors, changed significantly under the 2011–12 series as compared to the 2004–2005 series, particularly for the years for which data were available in both the series.

In a paper presented at the 2016 India Policy Forum, Nagaraj and Srinivasan (2017) highlighted some of the core issues in the measurement of the 2011–12 series. They summarized the arguments made in studies published after the release of the NAS in 2015. According to Nagaraj and Srinivasan (2017), while a base-year revision usually leads to a marginal rise in the absolute size

---

1. GDP, or gross value added (GVA), is a measure of goods and services produced in an economy in a year, net of intermediate inputs. Broadly speaking, it is a statistical construct based on innumerable estimations of value addition taking place in an economy. GDP is estimated following the United Nations System of National Accounts (UNSNA)—a global template, revised periodically to account for evolving economic activities.

2. GDP is re-based regularly to account for changing production structure, relative prices, and better recording of economic activities. Crucially, the re-basing also allows for introducing newer methodologies and improved databases. Such changes often expand the absolute GDP size because we are able to more accurately capture output. However, annual growth rates usually do not vary too much with re-basing of GDP, implying that the underlying pace of economic expansion has remained the same.

F I G U R E 1 .   **Nominal and Real Growth Rates of GDP at Factor Cost (GDPfc) and GVA at Basic Prices (GVAbp) under the Old and New Series for Comparable Years**

GDPfc and GVAbp Growth Rates, Old and New Series, Current and Constant Prices
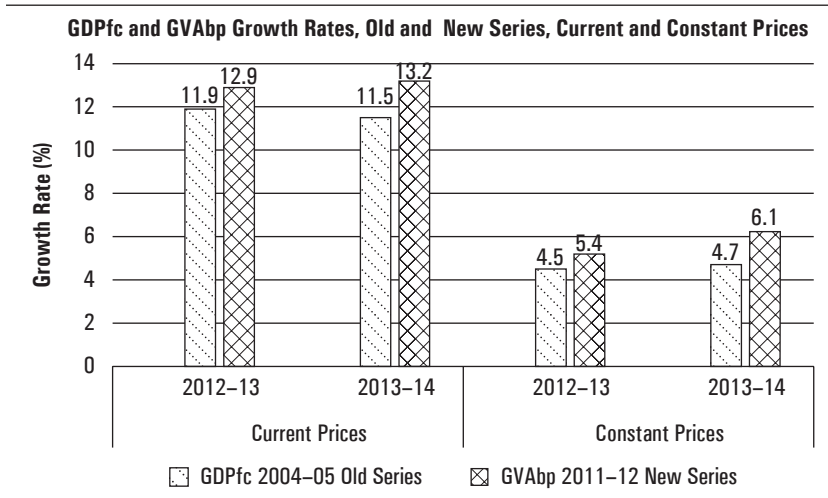
Source: CSO (2014), CSO (2015b).

of the economy owing to better representation, it does not cause a big change in the annual growth rates of GDP estimates. However, the latest base-year revision significantly changed growth rates. It resulted in a 2.3 percentage point *shrinkage* of the absolute size of GDP in the base year (2011–12) and *higher* aggregate GDP growth rates in the subsequent years.

Since the new NAS was released, the biggest doubt has been about the increase in GDP growth rates for the overlapping years for which data on both old and new NAS were available. This led to suspicion about the over-estimation of growth rates for the subsequent years. Changes took place in both real and nominal GDP growth rates, as shown in Figure 1.

Changes in the sectoral real growth rates are presented in Table 1 for the overlapping set of years before the 2004–05 series was discontinued. For instance, the changes in the manufacturing sector led to a revision in growth rates from 1.14 percent to 5.45 percent in 2012–13, and from –0.71 percent to 4.9 percent in 2013–14. Similarly, growth rates for the trade, hotels, and transport sector were significantly revised from 3.02 percent to 6.51 percent for 2013–14 as compared to the 2004–05 series. The revision also altered the institutional composition of India's GDP. In particular, the size of PCS was enlarged, while the unorganized/informal/household (HH) sector got contracted, with the share of the public sector remaining the same.

The methodological changes responsible for these comprehensive revisions have since then been questioned by a number of academic experts and continue to capture the attention of mainstream media, both domestic and

T A B L E 1 . Growth Rates of GDP at Factor Cost (2004–05 Series) and GVA at Basic Prices (2011–12 Series), Constant Prices for Comparable Years

Sector	2004–05 Series Constant Prices			2011–12 Series Constant Prices		
	GDP at Factor Cost			GVA at Basic Prices		
	2011–12	2012–13	2013–14	2011–12	2012–13	2013–14
Agriculture, forestry, and fishing	5.02	1.42	4.71	6.4	1.49	5.57
Mining and quarrying	0.1	−2.16	−1.38	−17.53	0.60	0.19
Manufacturing	7.41	1.14	−0.71	3.13	5.45	4.97
Electricity, gas, and water supply	8.38	2.26	5.92	8.56	2.66	4.16
Construction	10.8	1.11	1.64	13.14	0.35	2.66
Trade, hotels, transport, storage, and communication	4.33	5.07	3.02	6.36	9.77	6.51
Financing, insurance, real estate, and business services	11.35	10.92	12.87	4.49	9.74	11.15
Community, social and, personal services	4.9	5.31	5.55	7.28	4.26	3.85
Total	**6.69**	**4.47**	**4.74**	**5.22**	**5.42**	**6.05**

Source: CSO (2014), CSO (2015b).

international. Over the last four years, a large number of analytical studies have identified and analyzed specific problems in the data and methodology used in the 2011–12 GDP series, over and above those highlighted by Nagaraj and Srinivasan (2017). The common question in these studies has been about the extent to which the revised growth rates paint a true picture of the economy as opposed to being an outcome of problems in the underlying methodology and data used for estimation.

In addition, new controversies related to the 2011–12 series have cropped up in recent times such as: (a) release of two contradictory back series that paint diametrically opposite pictures of the historical performance of the economy, (b) release of first revised estimates for 2016–17, which showed a staggering 8.2 percent growth rate in the year of demonetization when more than 80 percent of the cash in the economy was removed from circulation overnight, dealing a severe blow to the unorganized segment of the population, and (c) release of an NSS service sector survey report (74th Round) in May 2019, which showed several gaps in the sample of firms used for GDP estimation by the CSO.

Despite growing skepticism and the wide range of questions raised by the academic community following the release of the new series, the CSO has defended the series (see, for instance, CSO 2015e; 2018a) citing reasons such as adoption of international best practices, improvements in methods of estimation, and wider coverage of the economy through new datasets.

The findings of the research studies put out in the public domain since June 2016 have raised new questions about the quality of the underlying data sources used in computing the new GDP series, the accuracy of the methods applied, and hence about the credibility of the estimates. It is perhaps time to take stock of all the issues that have been raised in various research studies and explore plausible solutions to the problem. That is what we aim to achieve in this paper.

We approach the issue in a two-step manner. First, we describe the basic changes brought about in the size and composition of various sectors by the new GDP series. Some of these issues were also discussed by Nagaraj and Srinivasan (2017) and we take off from where they had left. We attempt to understand the repercussions of these changes on the sectoral as well as aggregate GDP growth rates. We conclude that a majority of the changes affect the estimates for the private corporate sector, or PCS.[3]

Next, we undertake an examination of the changes in data and methodology used to compute the PCS estimates and discuss the problems therein. Most of these problems seem to stem from the usage of the MCA21 database. In particular, there are three main issues, all of which are related to the way sampling is done by the CSO for estimating the output of the PCS: (a) Which companies are included in the sample? (b) How can we deal with companies that are outside the sample but form a part of the larger universe of companies? (c) How can we deal with companies that cannot be sampled but are included in the PCS? It appears that there are problems in each of these aspects of sampling and we present a detailed discussion of these issues. In addition, we also analyze the problems in GDP growth estimation arising from deflator-related issues, problems in the regional accounts, and issues with the release of two contradictory back series.

We base our analysis largely on the findings of academic experts who have written extensively on these problems. We also take stock of the findings of different committee reports that have dealt with various issues regarding GDP estimation (CSO 2015c; 2015d). Our goal here is to present a

---

3. The PCS includes companies (both financial and non-financial) from the manufacturing and services sectors. MCA21 is an e-Governance initiative of Ministry of Corporate Affairs (MCA), Government of India, that enables secure access to MCA services for corporate entities, professionals and Indian citizens. The MCA21 application is designed to automate processes related to the proactive enforcement and compliance of the legal requirements under the Companies Act, 1956, New Companies Act, 2013 and Limited Liability Partnership Act, 2008.

comprehensive summary of major issues in the new NAS in order to provide deeper insights into the GDP debate, assess the severity of the problem at hand, and discuss a way forward.
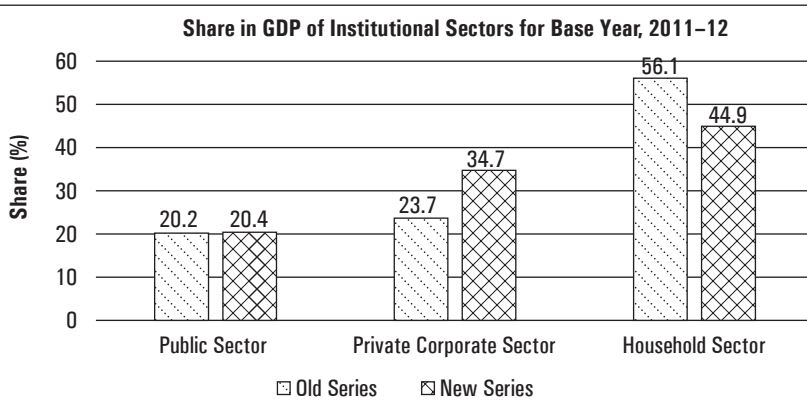
The rest of the paper is organized as follows. In Section 2, we discuss the changes in the shares of various sectors and in the institutional composition of GDP under the new series. In Section 3, we present a detailed analysis of the problems affecting the estimates of the PCS. In Section 4, we discuss issues related to the deflators. In Sections 5 and 6, we talk about the issues with the estimation of the regional accounts and issues with the release of the two back series, respectively. Finally, in Section 7, we summarize the main points and provide recommendations for the way forward.

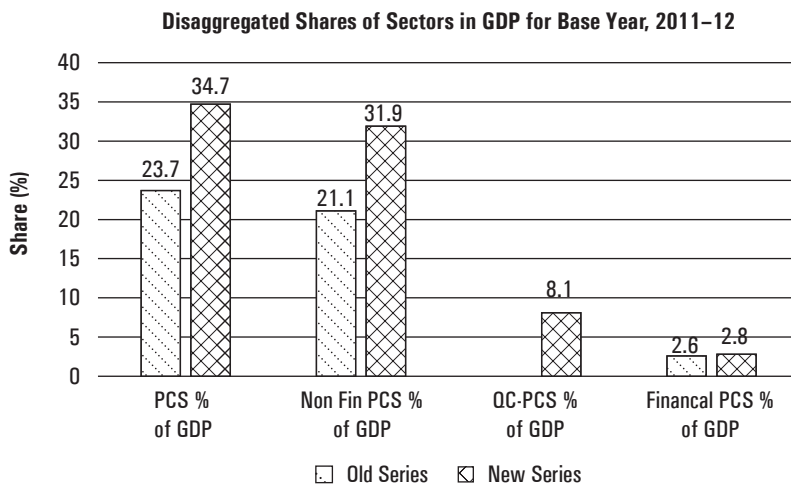## 2. Size, Structure and Evaluation of the Economy as Seen through the NAS

The new NAS has brought about many changes that have altered our image (or understanding) of the structure of the economy. We describe some of the prominent changes with regard to the institutional and sectoral composition of GDP as follows.

1. In terms of institutions, the share of the PCS increased by about 11–12 percentage points of GDP (as of 2011–12), with a corresponding decline in the share of the HH/unorganized sector. This was mostly on account of shifting of the proprietary/partnership enterprises from the HH sector to the PCS under a new category, quasi corporations (or QCs), defined as those maintaining accounts. The share of the public sector—defined as general government, public financial enterprises, and public non-financial enterprises—in GDP remained the same across the old and new NAS. This is shown in Figure 2.
2. Within the PCS, the share of private financial enterprises in GDP remained roughly the same in the new NAS, whereas the share of non-financial PCS went up significantly from 21.1 percent to 31.9 percent (Figure 3).[4]

---

4. The PCS constituted 34–35 percent of GDP in 2015–16. Financial PCS accounts for 2–3 percent of GDP. Non-financial PCS consists of: (a) public limited companies (13.4%), (b) private limited companies (11.9%), and (c) QCs (9.6%; their GDP shares are mentioned in parentheses). Roughly speaking, public limited companies represent larger companies, private limited companies are smaller companies, representing medium-sized enterprises, and QCs are smaller enterprises, mostly partnership and proprietary concerns.

**F I G U R E 2 .**   **GDP Shares by Institutional Sector for Base Year 2011–12 in Old and New Series**

Share in GDP of Institutional Sectors for Base Year, 2011–12

Share (%)

- Public Sector: Old Series 20.2, New Series 20.4
- Private Corporate Sector: Old Series 23.7, New Series 34.7
- Household Sector: Old Series 56.1, New Series 44.9

☐ Old Series   ☒ New Series

Source: CSO (2014), CSO (2015b).

F I G U R E 3 .   **GDP Shares of Sub-sectors of PCS for Base Year 2011–12 in Old and New Series**

Disaggregated Shares of Sectors in GDP for Base Year, 2011–12

Share (%)

- PCS % of GDP: Old Series 23.7, New Series 34.7
- Non Fin PCS % of GDP: Old Series 21.1, New Series 31.9
- QC-PCS % of GDP: New Series 8.1
- Financal PCS % of GDP: Old Series 2.6, New Series 2.8

☐ Old Series   ☒ New Series

Source: CSO (2014), CSO (2015b).
Note: PCS = Private Corporate Sector; Non-Fin = Non-financial; QC = Quasi corporations.

3. In terms of output sectors or industries, in 2011–12, the share in GDP of industry (consisting of mining, manufacturing, electricity, gas and water, and construction) went up, somewhat with a corresponding decline in the share of the services sector. The increase in industry's share was mainly on account of manufacturing (Figure 4).

F I G U R E 4 .   **Sectoral Composition of GDP for Base Year 2011–12 in Old and New Series**

**Sectoral Composition of GDP, 2011–12 Base Year**
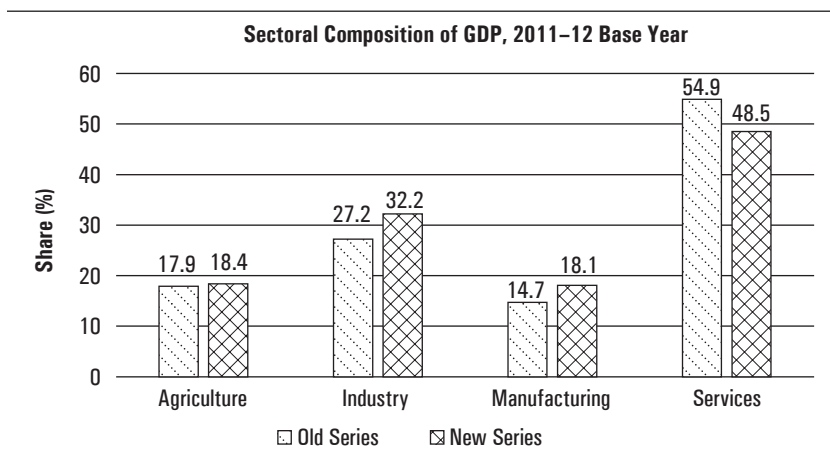
Source: CSO (2014), CSO (2015b).

4. How has the economy evolved over the six years since the new NAS was introduced? In terms of institutions, as shown in Figure 5, the only sector that has gained share is the PCS, within which the share of QCs in GDP has gone up from 8.1 percent in 2011–12 to 9.6 percent in 2015–16, the latest year for which we have the information from the RBI's analysis of the Ministry of Corporate Affairs (MCA) data.

F I G U R E 5 .   **GVA Shares by Institutional Sectors over Time, Current Prices, 2011–12 Series**
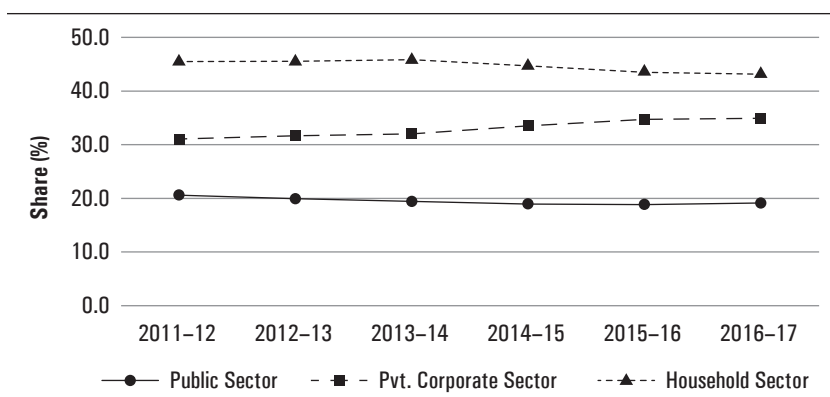
Source: CSO (2018c).

F I G U R E   6 .   **GVA Shares by Major Sectors over Time, Constant Prices, 2011–12 Series**
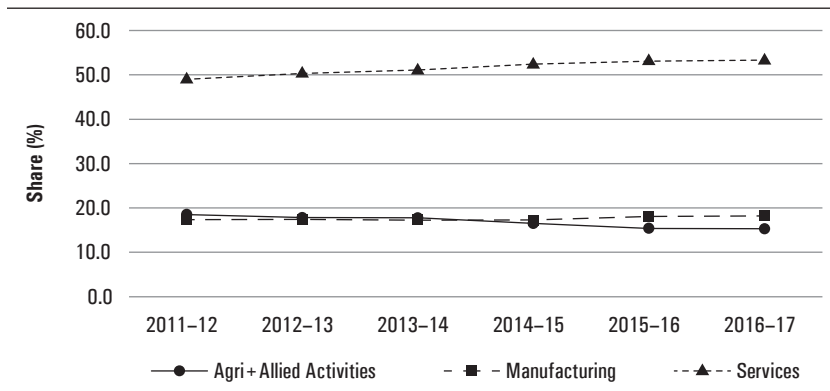
Source: CSO (2018c).

In terms of output, as shown in Figure 6, the shares of agriculture and industry have declined slightly, with a compensating rise in the share of the services sector.

The two big changes introduced in the new NAS are as follows:

1. Shifting the QCs from the HH sector to the PCS; and
2. Use of a new database (MCA21) to compute GVA estimates for the PCS.

If the changes in the shares of sectors and institutions in the aggregate output in the new NAS are an outcome of a mere reshuffling of economic activities, then these should not affect aggregate growth rates. For example, the shifting of QCs to PCS should not increase the aggregate GDP growth rates, given that QCs were already accounted for in the old NAS as part of the HH sector. Yet as we see from Table 1, in the overlapping years for which data on both the old and new NAS are available, the aggregate GDP growth rates were revised upwards in the new series. In terms of coverage, no new sector was captured by the new NAS either, which could have potentially explained the increase in growth rates.

The very fact that the new series reported significantly higher growth rates at the aggregate level for the overlapping years points to the possibility that the changes in methodology and data played a role. Since there has been no substantial change in the methodology used to measure GVA of

the public sector and the HH sector, it may be concluded that the increase in the growth rate of aggregate GDP is mainly due to changes in the PCS, primarily the non-financial PCS since the financial PCS constitutes a small fraction of overall GDP (see Footnote 4).

As shown above, the PCS consists of non-financial companies, financial companies and QCs. Net of QCs and financial companies, the size of the PCS in the new GDP series is higher by 2.9 percent of GDP. This can be attributed to the changes in methodology and introduction of the MCA21 database. The main question here is: Is it a case of more comprehensive capture of the contribution of the PCS, or does it represent an over-estimation?

A number of academic experts have identified and documented multiple problems with the MCA21 database, which under some scenarios might lead to over-estimation of the growth rate of the PCS and of the aggregate GDP growth rate, given the high share of the PCS in overall GDP. Moreover, given the infirmities in the estimation of output of QCs under the new NAS, shifting these entities to the PCS could have potentially contributed to boosting the level and growth rate of PCS GVA and hence aggregate GDP. We discuss these problems in detail in the next section.

3. Issues with Estimates of PCS

The PCS, especially the manufacturing sector, continues to be at the heart of the GDP measurement debate. Since Nagaraj and Srinivasan (2017), a number of new issues concerning the PCS have come up in public debates and these have been chronicled by several academic scholars over the last few years. The bulk of the problem in estimation seems to stem from the shift to the MCA21 database from the Annual Survey of Industries (ASI) database. In what follows we discuss three major issues with regard to the PCS estimates that have surfaced after the introduction of the MCA21 database. These issues, listed below, are primarily related to the manner in which sampling is done by the CSO for the PCS–GVA estimation:

1. What companies are included in the sample?
2. What method is used to account for companies not in the sample but in the larger universe of all companies?
3. What about the companies that cannot be sampled but are included in the PCS?

Below we discuss these issues in detail. In addition, we also analyze the validity of the rationale behind the shift from ASI to the MCA21 database,

issues of misclassification of companies in the PCS, and the problems associated with the shift from an "establishment" to an "enterprise" approach.

3.1. Sample of Companies Used for Estimation

Companies belonging to the PCS, that is, manufacturing as well as services sector companies, file their financial returns in the MCA21 database but not all companies file in every year. The set of companies that file returns at least once in three years is called an "active" set.[5] This is regarded by the CSO as the "universe" of companies for estimating the GVA of the PCS. Within the "active" set, only a fraction of the companies file returns in any given year. For the GVA estimation of any given year, the CSO first considers those companies that have filed their returns in that specific year. This is the "filing" set which constitutes the sample for that year.[6] They then use a blow-up factor to estimate the GVA of the non-filing, active companies.

The first big question with regard to sampling is whether the companies in the sample considered by the CSO are working companies. It would be problematic if the "filing" set consisted of say shell companies that engage in fictitious transactions for the purpose of evading laws and falsely report their returns. The GVA estimates computed on the basis of the returns of these companies are likely to be erroneous. In this context, two key issues are worth looking into. We discuss them sequentially.

3.1.1. Doubts about the Universe and Sample of Companies

In 2016–17, the National Sample Survey Office (NSSO), in its 74th Round, conducted a survey of services sector enterprises, on the way to launching an annual survey of services (on the line of the ASI). With the release of the NSSO's technical report on the services sector survey (hereafter, the

5. We do not know the exact definition of "active" companies in the MCA database. When the MCA passes on the "active" list to the CSO, as per the official documents, the latter considers this "active" set to consist of companies that have filed returns at least once in the last three years. This may not necessarily be the case and there does not seem to be any verification process in place to ensure that this definition indeed correctly identifies the "active" companies given to CSO by the MCA. This itself introduces a layer of uncertainty about the universe of companies that is being considered for the estimation of GVA.

6. The "filing" companies, which constitute the sample set used by the CSO for GVA estimation, vary from year to year because they self-select to file returns. As shown in Table 4B later in this paper, the absolute number of "filing" companies changes every year and so does the ratio of "filing" and "active" companies. This implies that the sample used by the CSO for GVA estimation changes every year. This raises doubts about the comparability of the sectoral GVA estimates over multiple years and the statistical soundness and stability of the estimates obtained.

NSS report) in May 2019, new questions arose regarding the quality and reliability of the MCA21 database, in particular, about the soundness of the sample of companies used by the CSO for its estimation. Official press notes of May 10, 2019, issued by the Ministry of Finance, and May 30, 2019, issued by the Ministry of Statistics and Programme Implementation (MoSPI) have sought to dismiss the doubts, claiming that the MCA database is in fine order for GDP estimation. But if anything, these have raised further doubts about the sample of companies.

One of the three list frames (or universes of enterprises) used for the NSS survey was the list of "active" companies—companies that are said to have filed their statutory returns at least once during the previous three years—obtained from the CSO (called the MCA frame). After due verification of a sample of about 35,000 non-financial companies, the non-response to the survey was found to be as high as 45.5 percent; 21.3 percent of the sampled companies were found to be mis-classified, and 24.2 percent of the companies refused to provide information, or were found to be closed or were non-traceable. Considering the severity of the non-response, NSSO abandoned its project of bringing out two volumes of survey results, and instead settled for a modest technical report. NSSO cautioned data users that "the estimates from the sample are therefore, not likely to be robust over the domains" (NSSO 2019: 16).

Arguing that the non-responding companies could be shell/fake/dubious/ non-existent companies that do not produce goods and services on a regular basis, but perhaps serve as conduits to hide profits or circumvent regulations, critics contended that such companies represent non-working companies. MoSPI defended their GDP estimation procedure (in the May 30 press note) saying that every year MCA has been weeding out an increasingly larger number of companies that are not operating, implying that "active" companies in the MCA's register represent genuinely working companies. Further, the missing/fake/shell companies are outside the set of "active" companies, and hence the database and methodology used by the CSO are correct. MoSPI's May 30 press release also said the following:

…from the 35,456 companies included in the NSS 74th Round, around 34,834 (86.5%) companies had filed their returns in the MCA database and only 622 were untraceable in MCA. In the context of GVA estimation in respect of private corporate sector (PCS), out of the 4,235 units categorised as not traceable at the given address in the 74th Round, around 3,154 units had actually filed returns online on the MCA portal.... For the purposes of National Accounts Estimates, the returns actually filed by the corporates under MCA is duly taken into account and the scaling up factor for the Paid-Up-Capital for the non-response is low.

MoSPI is therefore implying that the above record of filing of returns holds for the PCS as a whole too. This would imply that, say out of about 10.9 lakh "active" companies (as of 2015–16), a majority are filing returns. Non-filing companies form a small fraction of "active" companies whose output is estimated by blowing up the parameters prepared for a majority of the companies. Hence, MoSPI claims that the GDP estimates and its growth rates are valid.

Shortcomings of MoSPI's Contention

The May 30 press note classifies the MCA database into: (a) active companies, and (b) others. An "active" company is taken to mean a working company as it files its financial returns at least once in three years. So, by definition, "others" are non-working companies, whose status, as per the press release, could be "amalgamated," "converted," "unclassified," "under process," "under liquidation," "dissolved," "dormant," etc. There are several problems with MoSPI's claims.

First, contrary to the May 30 press release, the NSS report clearly states that its sample is drawn from the list of *active companies* obtained from the CSO. To quote it, "From the MCA frame *active* private non-financial companies of 2013–14, as available from National Accounts Division was taken into consideration" (p. 3; emphasis added). Hence, all the non-responding/untraceable companies in the sample must also be "active" companies. This means that the active list includes non-working or "others" as well, and hence CSO's list of "active" companies is not watertight, as claimed.

Second, as per MoSPI's note, there are missing companies *within* the set of "active" companies as well as in the set of "filing" companies. Table 2 (abridged from the press note) shows that 2,242 "active" companies belonged to the "casualty" category and 1,845 of these were filing returns. Likewise, 1,357 "active" companies were found to be closed and of these, 990 were filing returns, whereas 3,928 "active" companies were non-traceable and 3,141 of these were filing returns. In other words, both the universe (active) and the sample (filing) of companies used by the CSO for the PCS–GVA estimation appear to be faulty.

Third, as pointed out by MoSPI in the press note, the companies in the "Others" category are also filing returns. Yet by definition "Others" are non-active companies, as explained earlier. This implies that for obtaining the sample GVA estimates (even before blowing up for the universe of "active" companies), the CSO is using financial data of non-active companies, which is theoretically incorrect.

Fourth, MoSPI's (May 30) press release claimed, "In the last few years, nearly 6.3 lakh entities have been de-registered." However, we do not know the distribution of such companies between (a) "filing" companies,

TABLE 2. Status of Companies in the NSS 74th Round Survey

Category	Number of Companies in NSS Survey	In MCA in 2016–17, Active Companies	
		Active Companies	Companies Filing Returns
Surveyed	19,317 (54.5)	18,818 (55.5)	17,612 (57.6)
Casualty (i.e., refused information)	2,428 (6.9)	2,242 (6.6)	1,845 (6.0)
Closed during survey	1,579 (4.5)	1,357 (4.0)	990 (3.2)
Out of coverage, i.e., misclassified	7,573 (21.4)	7,291 (21.5)	6,755 (22.0)
Non-traceable units at the address provided	4,235 (12.0)	3,928 (11.6)	3,141 (10.3)
Total	35,456	33,912	30,583

Source: NSSO (2019).
Note: Figures in brackets refer to percentages of the column total. Percentages do not add to 100 percent, as this is an abridged version. See text for details.

(b) "active" companies, and (c) "others." There is no a priori reason to believe that the act of de-registering companies would remove the unresponsive companies (as highlighted by the NSSO report) from the sample.

Generalizing from the NSS services sector survey results, we now present a fuller picture of the shortcomings of the MCA database for the entire PCS, with the help of Table 3 and a Venn diagram given in Figure 7, using the MCA data for 2015–16. Out of 15.5 lakh companies registered with the MCA, 10.9 lakh or 70 percent were "active" companies (see Table 3) and the rest were "others" (belonging to categories mentioned above). Legally, all "active" companies are said to be working companies. Only 58 percent of the "active" companies filed returns in 2015–16 (i.e., these constitute the "filing" companies). The "filing" companies constitute 85–87 percent of the paid-up capital (PUC) of "active" companies, and hence, MoSPI claims that GDP estimates for the PCS are reliable.

TABLE 3. Details of the MCA 21 Database for 2015–16

Number of registered companies	15.5 lakh
Number of active companies	10.9 lakh
Number of companies filing returns	6.3 lakh
Ratio of active to registered companies	70.1%
Ratio of filing to active companies	58.3%
Ratio of filing to registered companies	40.7%

Source: Ministry of Finance (2019).

However, as described above, the sets of "active" as well as "filing" companies, that is, the universe as well as the sample of companies, appear defective. While the boundaries of various categories of companies in the MCA database are claimed to be watertight, as shown in Figure 7, there seems to be a grey area consisting of shell/fake/dubious/non-existent companies (shaded portion in the figure), whose contours and quantitative dimensions are unknown.

Given that the sample of companies used by the CSO for GVA estimation appears to contain shell companies that engage in fictitious transactions, the sectoral as well as aggregate growth estimates obtained from such a sample are likely to be biased upwards.

3.1.2. FILING VERSUS WORKING COMPANIES

As reported by the CSO (2015b; 2015c), close to 5 lakh "active" companies were a part of the sample used for estimation in the NAS for the initial years of the 2011–12 series. Over the last few years, the number of "active" companies has increased more than 11 lakh (see MCA 2019, for details), while the number of "filing" companies has increased to more than 7 lakh. Tables 4A and 4B show the numbers of registered, active, and filing companies for the years for which data is available. Table 4A tabulates the figures of the total registered and active companies available in the MCA21 database and Table 4B shows the number of "filing" companies in each year. The tables show a steady rise in the numbers of "active" companies and those filing returns. The reality, however, may be different as gleaned from various official documents.

It appears that the CSO uses a set of "common" companies instead of the entire "filing" set for preparing the sample estimates. Common companies are those that have data on returns for the previous year and the current year.

T A B L E   4 A .   **Details of the MCA21 Database: Number of Registered and Active Companies, 2013–19**

Year	Total Registered Companies (in lakhs)	Active Companies (in lakhs)	% Active of Total Registered Companies
2013 (October)	13.45	9.06	67.37
2014 (April)	13.95	9.51	68.20
2015 (April)	14.65	10.26	70.04
2016 (April)	15.47	10.85	70.14
2017 (April)	16.50	13.15	79.73
2018 (April)	17.59	11.76	66.82
2019 (January)	18.50	11.34	61.27

Source: *Monthly Information Bulletin*, Ministry of Corporate Affairs.

T A B L E 4 B . **Number of Active and Filing Companies in the MCA21 Database, from 2012–13 to 2016–17**

Year	Active Companies (in lakhs)	Number of Companies that Filed Returns (in lakhs)	% Filed out of Active
2012–13	8.8	5.6	63.64
2013–14	9.5	6.1	64.21
2014–15	10.1	6.0	59.41
2015–16	10.8	6.3	58.33
2016–17	11.6	7.1	61.21

Source: Ministry of Finance (2019).

This set has remained stable at around 3 lakh companies, a figure just around one-half of the companies touted to be the number of companies filing returns. Ramana Murthy (2018) mentioned,

Accounts of about 5.5 lakh companies (covering both the manufacturing, mining and services sectors) have been analysed and incorporated in the estimation of national accounts series for the above mentioned sectors whereas there are some 11 lakh active companies. The estimates based on the available data were blown up to cover all companies using the active population and ratio of paid-up capital for them. A common company growth based on over 3 lakh companies was used when the data on the whole complement of 5.5 lakh companies were not available.

Therefore, it seems that even though the set of "filing" companies was 5.5 lakh, CSO uses a common set of 3 lakh companies for GVA estimation. It is not clear why this is the case and what happened to the remaining companies. Similarly, the set of companies used by the RBI (as obtained from the MCA) for estimating savings has also remained stable at around 3 lakh companies, as shown in Table 5 (all figures for 2015–16).[7]

T A B L E 5 . **RBI Database of Companies (Obtained from MCA)**

No. of NGNF public limited companies	19,602 with 39.9% of PUC
No. of NGNF private limited companies	2.92 lakh with 32.9% of PUC
Total	3.11 lakh companies (whose PUC would be the weighted average of the PUCs mentioned above)

Source: *Reserve Bank of India Bulletin*, May and June 2017.
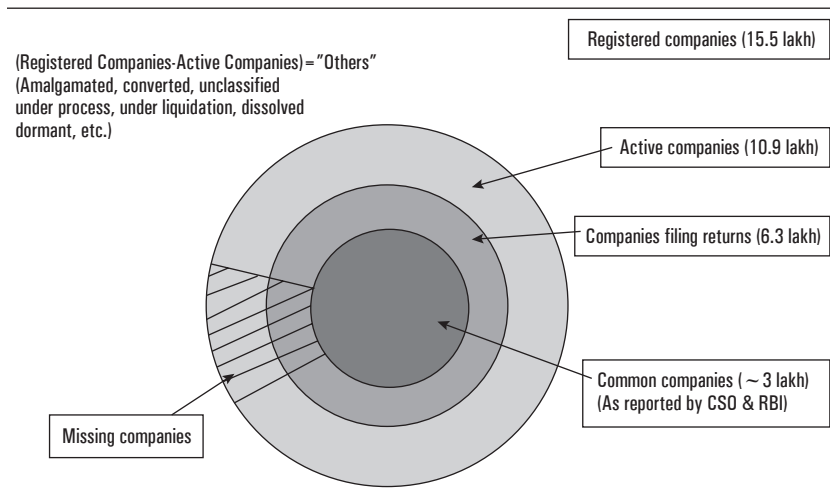Note: NGNF—Non-government, Non-financial.

7. MCA shares the corporate database in its entirety with RBI, as per an agreement in 2015. RBI has been publishing summary results of the MCA data analysis for NGNF public and private limited companies separately.

From the foregoing discussion, we are inclined to infer that an "active" company is merely a legal definition. It does not represent the *economic concept of a working company*, which produces goods and services on a regular basis. Our contention is that the working companies form a subset of: (a) "active" companies and (b) "filing" companies, and perhaps are only 3 lakh in number (the RBI set as well as the set of "common" companies as mentioned by Ramana Murthy [2018]). This is what we show in Figure 7. This anomaly also raises the question as to whether the remaining companies in the "filing" set are shell/fake/dubious/non-existent companies. Moreover, if the actual set used by the CSO for PCS–GVA estimation is only 3 lakh companies, then the reasoning offered by the CSO to defend the use of the MCA21 database based on a comprehensive capture of a larger number of companies is also doubtful.

Until we have a reasonable estimate of the size and composition of working companies, there is no meaningful way of drawing a sample and preparing GVA estimates. If one claims that the difference between the GVA estimates based on the set of working companies and the set of "active" companies is a mere level difference, it would be a leap of faith to say that this does not affect the growth rate.

F I G U R E   7 .   **Venn Diagram for Type of Companies in the "Private Corporate Sector" Category in 2015–16 based on Information from MCA, RBI, and MoSPI**

(Registered Companies-Active Companies)="Others" (Amalgamated, converted, unclassified under process, under liquidation, dissolved dormant, etc.)

Registered companies (15.5 lakh)

Active companies (10.9 lakh)

Companies filing returns (6.3 lakh)

Common companies (~ 3 lakh) (As reported by CSO & RBI)

Missing companies

Source: CSO (2015b); *Monthly Information Bulletin*, March 2016, Ministry of Corporate Affairs; *Reserve Bank of India Bulletin*, May 2016, June 2016, May 2017 and June 2017.
Note: See Table 3 and text for explanations.

3.2. Accounting for Companies Not in the Sample

Under the old NAS, GDP of PCS was not estimated directly. It used to be derived indirectly, as a residual. The saving and investment of the PCS were estimated by the RBI using the balance sheet of selected companies. The RBI sample consisted of about 4,500 large public limited companies and a smaller number of private limited companies. For public limited companies, the PUC of the selected large companies was said to be around 45 percent of the total PUC of public limited companies (as provided to RBI by MCA). Likewise, for private limited companies, the estimates of the selected companies were blown up to cover the entire universe of companies. Separate blow-up factors were used for public and private limited companies.

There was a concern that RBI's blowing up procedure was problematic because the size and composition of the PCS had changed substantially during the last three decades. To overcome the problem, the National Statistical Commission headed by C. Rangarajan recommended conducting a census of working companies. This was not taken up. Instead, MCA's e-filing initiative was seen as a solution to the problem of obtaining the universe of working companies.

Under the new NAS, the CSO does not have data on the returns of the companies that are part of the universe but not part of the sample, that is, the non-filing, active companies. So they use a *blow-up* methodology to calculate the GVA of these companies. The estimates for the non-filing companies are obtained by blowing up the estimates of the filing companies. The blow-up factor used by the CSO (also called the PUC factor) is computed as the reciprocal of the ratio of PUC of "filing" companies to the PUC of all "active" companies (CSO 2015a; 2015c).[8]

This implies that if there are problems in the "non-filing, active" set of companies, then the estimates obtained after blowing up may not convey the true picture of the sectoral growth and hence, of the aggregate growth. Depending on the nature of the problems, there could be overestimation of the growth rates. Several studies have pointed out problems with this blowing up methodology. Here we discuss the two main problems.

---

8. The PUC of a company is the amount for which shares are issued to shareholders. According to the Companies Act, 2013 (Section 64), paid-up share capital is such aggregate amount of money credited as paid-up as is equivalent to the amount received as paid-up in respect of shares issued and includes any amount credited as paid-up in respect of shares of the company. There is a reliance on PUC because in the absence of information on actual production, a physical indicator is required that is closely related to production (or production capacity).

3.2.1. Lack of Correspondence between PUC and GVA

The use of PUC in computing the blow-up factor is based on the assumption that the GVA and PUC have a one-to-one correspondence, and that one can directly infer a company's value addition by analyzing its PUC.

Sapre and Sinha (2016) replicated the process of blow-up of GVA for a comparable sample of firms (from the CMIE Prowess database) that qualify for filing in the XBRL format in the MCA21. They find that GVA and PUC have little or no correspondence, especially in cases where GVA is negative (i.e., a loss-making company). The PUC of a company is by definition always positive. This means that it is possible that by using a PUC-based blow-up factor, estimates are scaled up for companies that are, in reality, loss-making companies with negative GVA. This would potentially lead to an overestimation (see Box 1 for details).[9]

The application of the blow-up methodology requires a detailed analysis of the GVA and PUC of registered companies in the MCA21 database. In response to this problem, NSC (2018) recommends:

Cross-validation study on data on corporate bodies with single manufacturing unit available from the two sources, MCA and the ASI. Additionally, a study of plants covered in ASI data belonging to non-reporting but active companies in the MCA list should be undertaken. In the same vain, the ratio of GVA to PUC should be compared between companies that submit their returns by the specified due date and those that submit their returns after the due date. A related research that may be undertaken using ASI and MCA data is to identify plant covered in ASI data which belong to active but not reporting manufacturing companies in the MCA list. The ratio of GVA to invested capital for such plants should be studied in comparison with plants that belong to companies in the MCA list which are active and reporting. (NSC 2018, III 6.5)

At present the PUC based blow-up factor is determined on the basis of the data of firms that have submitted their data in the required forms by a specific date. Some of the non-reporting firms submit their data later. The ratio of GVA to PUC should be compared between the firms that submit their returns within the specified date and those that submit later. Such research may provide an answer to the question whether the ratio of GVA to PUC is lower for later filers or non-filers as compared to the firms that file their returns in time. (NSC 2018, III 3.3.11)

9. Manna (2017) corroborates this finding by highlighting that a common blow-up factor for all companies would be inappropriate and separate blow-up factors ought to be computed for different size classes of PUC. Both Sapre and Sinha (2016) and Manna (2017) have argued in favor of exploring alternatives other than PUC for blow-up of GVA. Manna (2017) proposed the use of gross fixed assets, and Sapre and Sinha (2016) explored the possibility of using industry-wise growth rates for scaling up GVA of non-filing companies.

3.2.2. ISSUES WITH THE UNAVAILABLE COMPANIES

One key issue in using the MCA21 dataset is in dealing with the problem of non-filing. Given the process of data extraction from the MCA21 database, the non-filling points to a case of potential overestimation. If there are sufficient reasons to consider that non-filing firms are (a) wound up, or de-registered, (b) loss-making, or (c) are fictitious shell companies that exist only on paper and are not undertaking any service or production activities, then scaling up the estimates of the "filing" companies to account for the "non-filing" ones is likely to lead to overestimation of GVA, of the PCS, and possibly of the overall level of GDP as well. As discussed earlier, the NSS report of May 2019 showed that there are indeed serious problems of missing companies in the "active" set and in the set of "non-filing" companies.

The problem with the blowing-up methodology is, therefore, an inevitable consequence of inappropriate sampling where in the set of "non-filing" companies: (a) there could be shell companies with fake accounts, showing growth rates that never happened, (b) there could be dead companies (i.e., companies that have shut down) with zero GVA, whose imputed growth rates will be higher than actual, and (c) there could be loss-making companies, whose value added is overstated, because the PUC is used as a blow-up factor. Since these companies are actually shrinking, the overall growth rates will be overstated because positive growth rates will be imputed to them.

In summary, the main point as discussed in Sections 3.1 and 3.2 is that the extent to which the MCA21 database problems distort the sectoral and aggregate GDP growth rates depends on: (a) the blow-up ratio for the "non-filing" companies, and (b) the nature of the problems (low growth rates, no growth, decline in GVA, negative GVA, etc.), with the "non-filing" companies. Problems would also arise if the "active" set contains shell companies. Unless there is concrete evidence that the "non-filing" set consists of proper companies with positive GVA and growth rates, and that the "active" set does not contain shell companies, it is hard to dismiss the doubts of overestimation, given the sampling and methodological issues outlined above.

3.3. Companies That Cannot Be Sampled

A portion of the PCS under the new NAS consists of entities that cannot be sampled. They do not file returns in the MCA21 database, which means that they are not part of the usual sample of "filing" companies used by the CSO for GVA estimation. The manner in which their growth rate is estimated raises questions about possible overestimation. These entities are QCs. They are perhaps the least understood part of the PCS in the new

NAS, as disaggregated information on the PCS is not available. Here, we piece together the available information on the PCS, and the size and composition of QCs.

In Figure 3, the size and structure of the PCS in the old and the new NAS are discernible (as discussed in Section 2). The size of the PCS relative to GDP in the new series increased substantially, mainly on account of QCs, which in 2011–12, constituted 8.1 percent of GDP.

What Are QCs?

A QC is an enterprise not registered under the Companies Act, yet said to behave like a company. It is a partnership or proprietary enterprise maintaining books of accounts. The underlying idea is that such enterprises are akin to limited liability, profit maximizing firms, as against own-account or HH enterprises engaged in subsistence activities, and often employing family labor. The new NAS, following the SNA 2008 guidelines, introduced the concept of QCs by bifurcating unorganized/HH/informal sector enterprises into QCs, and clubbing them with the non-financial PCS, leaving HH/own-account enterprises in the HH/informal sector. As per the new NAS, QCs consist of:

1. Crop production in plantations, other than those covered in the PCS.
2. Unincorporated enterprises covered in ASI.
3. Unincorporated enterprises of manufacturing that are not covered under the ASI but maintain accounts.
4. Cooperatives providing non-financial services.
5. Unincorporated enterprises providing non-financial services maintaining accounts.
6. Unorganized financial enterprises.[10]

In the earlier NAS, items (1), (2), and (4) were included in the non-financial PCS. The remaining three are the new additions clubbed together under QCs. Table 6 provides the share of institutions in GDP as of 2015–16 (based on the RBI's analysis of MCA data).[11] The QCs' share in GDP was 9.6 percent in 2015–16 and their share in the non-financial PCS GDP was 27.5 percent.

10. It is not clear how unorganized financial enterprises, essentially, informal moneylenders, are included in QCs.
11. It is to be noted that the RBI's analysis of the MCA database provides summary results separately for Non-government Non-financial (NGNF) public and private limited companies. The QCs' share is obtained by subtracting the share of public and private limited companies from the PCS.

T A B L E 6 . Disaggregation of Non-financial PCS Companies and their Percentage Shares in GDP for 2015–16

Institutional Sector	Share in GDP (%)	Share in Non-financial PCS GDP (%)
Public limited companies	13.4	38.5
Private limited companies	11.9	34.0
Quasi-corporations	9.6	27.5
Non-financial PCS	34.9	100

Source: *Reserve Bank of India Bulletin*, May and June 2017.

T A B L E 7 . Percentage GVA shares of QCs by Industry or Sector for 2011–12

Sector	Distribution of QCs (%)
Agriculture and allied	4.2
Mining and quarrying	1.7
Manufacturing	28.3
Electricity, gas, and water	−0.6
Construction	−1.0
Trade, hotel, and restaurants	40.0
Transport and communication	5.7
Community, social, and personal services	9.2

Source: CSO (2015b).
Note: The columns do not add up to 100% as this table excludes QCs from financial services.

Table 7 shows the industry or sectoral distribution of the QCs' output for 2011–12 (more recent data is not available). Evidently, two sectors, namely, manufacturing, and trade, hotels and restaurants accounted for 68.3 percent of the QCs' output in 2011–12. Incidentally, these are also the sectors which had witnessed a significant boost in their growth rates after the GDP revision.

Methodologically, the shift of QCs from the unorganized sector to the PCS is questionable. The SNA (2008) lays down conditions under which such a shift may be done. To quote the SNA, QC is "an unincorporated enterprise that has sufficient information to compile a complete set of accounts as if it were a separate corporation and whose de facto relationship to its owner is that of a corporation to its shareholders" (as mentioned in Subba Rao 2015).

In contrast, what is done in India is the following:

The enterprise survey collects information on whether the enterprise is maintaining books of accounts or not. As recommended by SNA 2008, all these unincorporated enterprises

have been classified as QCs, if they are maintaining accounts, and otherwise as household enterprises. The estimate of GVA from QCs has been added to the GVA of incorporated enterprises in the case of non-financial corporations. (CSO 2015b; 11).

In the GDP revision, proprietary and partnership firms in the ASI, and non-HH enterprises in NSS surveys were *deemed* to maintain accounts, and hence were categorized as QCs. There is no evidence of the NAS revision committees verifying if the QCs, in fact, maintained accounts and whether "…it were a separate corporation and whose de facto relationship to its owner is that of a corporation to its shareholders."

As mentioned in the previous section, the mere shifting of QCs from the unorganized/HH sector to the PCS should not change the aggregate growth estimates. But the concern is that along with the shift, the methods of estimating their output have also undergone some changes, which may have affected the growth rates (Subba Rao 2015). Since the QCs do not form a part of the sample used by the CSO for PCS–GVA estimation, to the best of our knowledge, their growth rate is estimated in two different ways, depending on which sector they belong to. The growth rate for the QCs engaged in manufacturing activities is taken from the ASI data on partnerships and proprietary entities. The growth rate for the QCs engaged in services is taken to be the same as that of the rest of the private corporate sector, which would be an exaggeration given the nature of the QCs. This may have boosted the growth rate of the QCs and, accordingly, of the PCS and aggregate GDP.

The issue really is whether the QCs are growing at the rate of other companies in the PCS. For most periods, we do not have enough information to assess this. They may be more dynamic since they are smaller, or they may be stagnant in which case their growth would be overestimated. It is possible that some are the former, and some the latter. We have no idea what the aggregate situation is for the QCs. It is possible however, that post-demonetization and post-GST, these companies are growing much more slowly than the big companies in the PCS. In that case, the maintained assumption will lead to an overestimation of growth.

3.4. Comparability of MCA21 with ASI-based Estimates

The discussion in Sections 3.1–3.3 shows that there are myriad problems with the MCA21 database used by the CSO to compile the estimates for the PCS under the new NAS. Both the universe and the sample of companies used for the estimation of GVA seem to be riddled with holes and given the manner in which blowing up of estimates is undertaken, there is ample room for overestimation of sectoral and aggregate growth rates.

As mentioned earlier, under the old NAS, for much of the non-government and non-agricultural activities, data used to be collected for the factory sector from the ASI. In addition, various unorganized sector surveys by the NSS were establishment-based surveys. However, there has been a concern that the ASI was increasingly missing out value addition taking place outside of the factory premises, in sites such as service centers, R&D laboratories, and company headquarters. Hence, a view prevailed that the ASI was underestimating the output growth in the manufacturing sector, given its specific approach to data collection.

The replacement of ASI with the MCA21 database for the manufacturing sector was predicated on the foregoing views. The problems with the MCA database that have been uncovered over the past few years by various academic experts raised the following question: Is it really true that the ASI captured value addition taking place only inside the registered premises, ignoring the related or auxiliary activities? Dholakia et al. (2018) examined the question by closely looking at: (a) the ASI schedule, (b) its field investigators' manual (which provides detailed instructions to investigators on how to post the information in the questionnaire), and (c) discussions with the concerned officials responsible for the data collection.

The investigation revealed that the ASI, in fact, gathers information on all activities of a factory, and the data gathered are apportioned to different factories of an enterprise as per standardized procedures. The argument that a shift to the enterprise approach increased the capture of value addition is not entirely correct. In light of this research, the premises of the changeover from an establishment approach to an enterprise approach for GDP estimation and, hence, the use of the MCA database itself appears questionable and unwarranted.

In the face of all the criticisms, the CSO has continued to defend the use of the MCA21 database. In 2015, after the release of the new series, the then Chief Statistician of India, T. C. A. Anant claimed that the use of the new database for the PCS captures the production that was left out in the earlier ASI series. To quote him,

There is a large invisible corporate segment, which we were not adequately describing in the earlier series. We were partially describing it in manufacturing through the ASI. So, there is recognition that there is a need to get better information on this segment as a large part of government policies are aimed at this segment. The 5,000 listed companies are typically not the principal focus of promotional policies. (Sidhartha and Gupta 2015).

While the MCA database may have technically increased the coverage of companies, it is worth noting that more data need not necessarily mean better data, as has been analyzed in the previous sections.

3.5. Other Problems with the MCA21 Database

3.5.1. SHIFT FROM AN ESTABLISHMENT TO AN ENTERPRISE APPROACH

The 2011–12 series makes a conceptual shift by capturing value addition at an "enterprise" level instead of at an "establishment" level. In economic terms, the distinction between the two can be understood as the difference between a factory (or a plant) and a firm (or an enterprise). The former is a technical unit of production, and the latter is an organizational unit of production. Various authors (Nagaraj 2015a; 2015b; 2015c; Nagaraj and Srinivasan 2017; Rajakumar et al. 2015; Sapre and Sinha 2016) have looked into several aspects of the estimation process in detail. In the new series, the GDP of the private corporate sector is estimated using the financial statements of enterprises as a whole as opposed to the earlier method of using the industrial output of factory establishments. This shift leads to a direct comparison with the ASI-based estimates.

Sapre and Sinha (2016) point out that lack of clarity on measures of output and costs at the enterprise level can lead to imprecise estimates of GVA for various sectors. For instance, the activities of enterprises can be much more diverse than those of factories, and not all of these functions would qualify as "manufacturing." Yet under the enterprise approach, all sources of value added of enterprises classified as "manufacturing companies" are included in the calculation of the manufacturing sector's GVA. This approach inflates the level of manufacturing output and possibly also the growth rate of the sector if the ancillary activities are growing faster than the manufacturing ones. To get a sense of the magnitude, we can compare the level of output based on industrial sales, that is, only considering manufacturing output with the total revenue, which includes the overall enterprise activities. Using data from CMIE Prowess, Sapre and Sinha (2017) show the difference in value addition based on two different measures of output for a set of companies.

As can be seen from Table 8, in comparison to industrial sales, total revenue shows a considerable increase in the level estimates of GVA. Total

TABLE 8. Comparison of GVA Using Industrial Sales and Total Revenue as Measures of Output, from 2011–12 to 2013–14

Year	Based on Sales (₹ crores)	Growth Rate (%)	Based on Dis. Revenue (₹ crores)	Growth Rate (%)	Difference (₹ crores)
2011–12	701,896.6		767,311.7		65,415.1
2012–13	742,237.2	5.74	819,228.5	6.76	76,991.3
2013–14	780,371.1	5.13	872,178.0	6.46	91,806.9

Source: Sapre and Sinha (2016).
Note: "Dis. Revenue" represents fields of disaggregated revenue adding up to total revenue of the company.

revenue fields include revenues from ancillary and related manufacturing activities and other non-operating revenues like treasury operations. The change in the measure of output possibly explains the large upward revisions in levels and on average, it corresponds to a 1 percent increase in growth rate for the manufacturing sector.

In comparison to the establishment-based estimation, the enterprise approach has also complicated the process of GVA estimation to some extent. Conventionally, subtracting the cost of production (of manufactured items) and taxes from the value of output gives an estimate of value addition. However, with diversified activities under one roof in case of an enterprise, identifying the costs of manufacturing activities from financial statements poses serious challenges (see Sapre and Sinha 2016, for details on the process of GVA estimation). Lack of proper identification of cost components can lead to imprecise GVA estimates.

3.5.2. Identification of Firms in the MCA21 Database

In the MCA21 database, the CSO relied on using company identification (CIN) code to identify manufacturing companies. The decision to use the CIN was made as the ITC-HS codes of products were either unreported or unavailable in the XBRL forms (see CSO 2015c, for details).[12] However, in absence of the ITC-HS codes, using the CIN code can potentially lead to a mis-classification of companies in identifying their business activity. Sapre and Sinha (2016) find that within the manufacturing sector, several companies operate as wholesale traders or service providers. These companies may have changed their line of business after they were originally registered (this was reported in the NSS survey of services as discussed earlier). These changes do not get reflected in the CIN code assigned to the companies. Such misclassification of companies will distort the manufacturing estimates, though not the overall GVA. The Sapre and Sinha paper and Table 9 illustrate how firms registered in manufacturing can also be in other activities.

Companies may change their primary activities over time as part of their usual business strategy and even repeatedly. Hence, lack of a proper identification system poses serious challenges for classification and estimation of value addition at the sectoral level. Sapre and Sinha (2016), and Pandey et al.

---

12. The Indian Trade Classification Harmonized System (ITC-HS) is an eight-digit code system used for product identification for import, export operations. XBRL is an Extensive Business Reporting Language e-platform used for filing annual financial statements with the MCA. Companies that have: (a) turnover greater than ₹100 crores, or (b) PUC greater than ₹5 crores, or (c) are listed, file in the XBRL format (see http://www.mca.gov.in/XBRL/pdf/ITC_HS_codes.pdf for details, accessed on May 7, 2021).

T A B L E   9 .   **Sample of Firms with CI Registered in Manufacturing Activities but Operating as Services Companies (2011–12)**

Industry Activity (2-digit NIC 2008)	Number	Industry Activity (2-digit NIC 2008)	Number
Trade in other manufacturing goods	362	Financial services including leasing	328
Other asset financing services	279	Securities investment services	275
Renting services	163	Services	128
Software services	81	Commission agents services	76
Trade in electrical machinery	76	Trade in manufactured products	63
Trade in chemicals	59	Trade in minerals and energy sources	57
Real estate infrastructural services	54	Trade in transport equipment	49
Trade in drugs and medicines	48	Business services	43
Trading in food products	43	Trade in agricultural crops	40
Tech. consultancy and engineering services	31	Info. tech-enabled service/BPO	21
Hotel and restaurant service	22	Other consultancy	17
Fund-based financial services	19	Trade in non-electrical machinery	15
Finance-related allied activities	15	Shipping services	13
Printing and related services	13	Research and development	10
Storage and warehousing services	11		

Source: CMIE Prowess, see Sapre and Sinha (2017) for details.

(2019) show the extent of misclassification that can arise in the absence of a system of identification and classification, and present an illustrative exercise on the frequency of changes in economic activity. They contend that it is of crucial importance to build and use the history of economic activity of companies so as to correctly classify companies into respective sectors based on their primary economic activities. As an illustrative case, Table 9 shows a sample of companies with economic activity different from their CIN-based activity.

In principle, the misclassification is a year-on-year problem and requires a detailed scrutiny of their product schedules. While the problem in using the CIN code was briefly raised in CSO (2015c), no systematic recourse was mentioned to solve this problem. NSC (2018, III 3.3.9) had taken a critical view of the problem by stating:

Moreover, the MCA21 dataset has serious quality issues. The economic activity or activities (NIC codes) perused by a company is extracted out of the CIN (Corporate Identification Number), assigned to the company at the time of registration. The NIC

code reported at time of registration is likely to undergo change in due course of time. The MCA21 dataset is not designed to include all the economic activities pursued by a company. However, it may be possible to tackle this difficulty by using the MGT-7 forms, which contain information regarding activity-mix of the companies.

The extent of distortion in GVA estimates due to misclassification cannot be assumed to be negligible. There are two main concerns: (a) misclassification introduces spurious volatility in levels and growth rates, and such volatility does not represent actual movements, and (b) it distorts the GVA-to-Output (GVA/GVO) ratio, which is significantly different for manufacturing and services. Identification of economic activity remains among the finer aspects of measurement and accuracy of macroeconomic aggregates. The case of the manufacturing or services sector is no different and deciphering information from a large dataset like MCA21 is a challenging task.

4. Deflator-Related Issues

The issues discussed in the previous section pertain to nominal GDP estimation. When it comes to real GDP growth rate estimation under the new NAS, a major issue is related to the kind of deflators that are being used to convert the nominal values to real estimates. There are two main issues in this regard, and we discuss them below.

4.1. Single versus Double Deflation

To get to the heart of the problem, one needs to understand how GDP figures or, almost equivalently, GVA figures are calculated. In the broadest terms, the procedure followed by the CSO is the same as that followed all over the world. It obtains data on the nominal values of output produced in various sectors of the economy from the financial accounts of firms. Then, it deflates these figures by price indices to arrive at estimates of real GDP. The CSO's methodology differs from what is followed in other countries in two specific areas: the deflating procedure it follows and the price indices it uses.

In terms of the deflating procedure, the standard international practice, followed by nearly every major country with the exception of China and India, is to use a methodology called "double deflation." Under this procedure, output prices are deflated by an output deflator, while raw material prices are deflated by a raw material deflator. Then the real input value is subtracted from the real output value to obtain real GVA estimates. The CSO's methodology is different in that it first computes the nominal

GVA, and then deflates this number using a single deflator to obtain the real GVA. If input prices move in tandem with output prices, there is no problem and both methodologies will give similar results. But if the two price series diverge—as they did in India for the first few years after the release of the new GDP series—single deflation can overstate growth by a big margin.[13]

The reason is not difficult to see. If input prices fall sharply, profits will increase, and nominal value added will go up. Since real GDP is supposed to be measured at "constant prices," this increase needs to be *deflated away*. Double deflation will do this easily. But single deflation will not work. In fact, if a commodity-weighted deflator like the Wholesale Price Index (WPI) is used as the single deflator, as is the case under the current methodology, nominal growth will be inflated whenever commodity prices are falling. In this case, real growth will be seriously overestimated.

As the gap between input and output inflation starts to close, the problem will diminish. But that could also send a misleading signal because it might seem that growth is slowing, when only the measurement bias is disappearing. This can be best explained using a numerical example as given in Box 1.

Globally, major developed countries have moved to a double deflator method, particularly for the manufacturing sector. In India, the issue of the deflator regained importance in the 2011–12 series for two reasons. First, while under the old NAS, the real growth rate was calculated largely using volume-based measures, under the new NAS, it is calculated using value-based measures. As a result, the deflating procedure has become more critical than before. Second, starting 2012–13, the WPI and the CPI series diverged substantially from each other, owing to a dramatic fall in global oil prices, which pushed the WPI significantly lower than the CPI. For example, in 2015–16, while WPI inflation fell to –3.7 percent, CPI inflation was 4.9 percent. This divergence continued till 2017–18. Considering that the WPI is the main deflator used by the CSO, this gap between input prices and output prices would have led to an overestimation of the real GDP growth rates under the new series, irrespective of the problems with data and methodology described in the previous section.

The introduction of the MCA21 database has led to new challenges in the construction of a double deflator method in the case of the manufacturing

---

13. For more details about how lack of a double deflation practice may have overstated real GDP growth under the new series, see the article in *LiveMint*: https://www.livemint.com/Opinion/58qihTaOIRd3rPyf1eK09L/Real-GDP-is-growing-at-5-not-71.html (accessed on May 7, 2021).

BOX 1. Single versus Double Deflation: An Illustrative Example

Consider a case where actual production is stagnating at, say 100 units, in years 2014–15 and 2015–16, but output and input prices are changing. Assume that a firm raises its output price by 5 percent, in line with general Consumer Price Index (CPI) inflation, whereas the price of its raw materials falls by 5 percent since WPI is down by 5 percent. In this example, the nominal value added defined as the value of sales less value of raw materials will increase by 8 percent from 2014–15 to 2015–16. Since this increase arises entirely from price changes, it needs to be deflated away in order to obtain "real GVA at constant prices."

	2014–15	2015–16	2015–16	2015–16	2015–16
	Nominal	Nominal	Real	Real	Real
			Estimated Using Double Deflation	Estimated Using Single Deflation and WPI	Estimated Using Single Deflation and CPI
Sales	100	105	100		
Raw materials	20	19	20		
Value added	80	86	80	91	82
CPI inflation (%)		5			
WPI inflation (%)		–5			
Growth (%)		8	0	13	2

As the above table shows, using the double deflation methodology, real GVA growth is zero. In other words, production has remained unchanged, which is correct by construction. If instead the CSO's method of single deflation were applied to this example, one would simply take the nominal value added and deflate it by the WPI. Since nominal growth is 8 percent and the WPI has fallen by 5 percent, real GVA growth is estimated at 13 percent, which is clearly way off the mark. It conveys the impression of a boom when none, in fact, exists.

Source: Authors' estimates.

---

sector. MCA21 is a database of financial statements (such as profit/loss statements and balance sheets) that does not provide information on input or output prices at a commodity level. These are necessary ingredients for constructing a double deflator that can deflate the values of inputs and outputs separately. In most countries, nominal production is deflated by the Producer Price Index (PPI). India lacks a PPI, so the CSO uses the WPI instead. To ensure that the GDP numbers accurately reflect developments in the economy, the CSO needs to develop proper PPIs, and then employ them using the double deflation methodology. The NSC Real Sector Committee recommended that ASI data should be used together with the MCA data to develop a procedure for processing the data, including use of a double deflation procedure.

4.2. WPI versus CPI as Deflator

It will take some time to develop PPIs, and even longer to calibrate double deflation for each sector. There is, however, an interim solution that can be much more easily applied, which would also deal with the other problem with CSO's deflation procedure, in that the WPI suffers from several drawbacks. For one, it does not measure the price of services, and services constitute the bulk—around two-thirds—of India's economy. Instead, the WPI is heavily weighted towards commodities, especially oil. So 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.

One interim solution to this problem, till the time a proper PPI is developed and data on input prices are collected, is to start using the CPI series for the whole of the services sector, instead of the WPI. The change to CPI makes even more sense in the services sector because the CPI has extensive information on price movements in the various services sub-sectors.

Using the CPI would not solve the problems caused by single deflation as that can only be resolved through the introduction of a double deflation methodology, especially for the manufacturing sector, but it would nonetheless help, as shown in the table in Box 1. The main reason is that the CPI at least has the correct sign for the deflator. It is increasing when the deflator needs to increase, rather than falling like the WPI. The result, of course, is not perfect. Using the CPI to deflate the nominal value added leads to an estimated real GVA growth of 2 percent in our example, when the correct answer is zero. But this is much closer to reality than the 13 percent real growth rate obtained by using the WPI as the deflator.

The better "fit" of the CPI is not just an accident of the particular example chosen. It is perfectly general because when commodity prices (such as oil prices) fall, GVA tends to increase, at least in commodity importers such as India (in this case, one should think of GVA as firm profits, which will go up when input prices fall). Since this increase needs to be deflated away to arrive at a real GVA estimate at constant prices, one needs an index that will increase when commodity prices fall, rather than decrease, as the WPI tends to do. The CPI will also tend to decrease, but by much less than the WPI, since commodities constitute a much smaller share of the consumer basket.[14]

---

14. As shown in Sengupta (2016), for the year 2015–16, official statistics showed that nominal growth was 8 percent, and real growth was 9 percent in the manufacturing sector. If the correct deflator was actually around 3 percent, in line with CPI manufacturing inflation, then real growth was only around 5 percent.

5. Issues in Compiling Regional Accounts (Gross State Domestic Product)

The 2011–12 GDP series has led to new challenges for compiling state-level GDP estimates. After the introduction of the MCA21 database, estimates of the organized manufacturing and services sectors are available only at the all-India level. This constraint occurs because the consolidated financial statements of enterprises are not available as per geographical regions, plant locations, and products. As a result, state-level GDP for the organized manufacturing and services sectors is driven largely by allocation rather than by actual estimation done in each state. Relying on an allocation method (e.g., using ASI shares of value added) causes serious measurement issues as such estimates may not entirely reflect ground realities.

For instance, Manna (2018) shows the bias arising out of allocating state-wise GVA based on shares of each compilation category in the total GVA available from ASI. Instead, Manna (2018) argues that a more appropriate allocation method would be to use the shares of the respective compilation category in the total GVA of private companies, as per the ASI. The issue with such an allocation method is that both the MCA21 and ASI frames have different coverage of units and GVA, thus leading to mismatches in growth rates. The problem has also been acknowledged by the Committee on Real Sector Statistics when it stated:

The most important gap in MCA21 data relates to the information at the regional (State) level. For the companies operating in more than one State, there is no way of ascertaining the distribution of GVA of such a company over its States of operation. (NSC 2018, III 3.3.7)

Adding another dimension to the problem, Dholakia and Pandya (2017) in the context of the unorganized services argued that the effective labor input (ELI) method does not take into account variations in productivity at the state level. They argue that labor productivity in sectors such as trade and freight transport services would be necessarily different across states and ignoring such differences can lead to imprecise estimates. In the old labor input (LI) method, although category-wise labor productivity was not explicitly considered, interstate variation was taken into account as the output per worker varied across states. Thus, on theoretical grounds, the new ELI method cannot be assumed to be superior as compared to the simple LI method.

The revised GDP methodology has affected Gross State Domestic Product (GSDP) estimation, with a sharp rise in "apportionments" and "projections," and a decline in the share of estimates based on state-level primary data, as

demonstrated by Dholakia and Pandya (2017) for Gujarat. This amounts to a regression in the quality of estimation of the GSDP series. It has happened at a time when a greater share of fiscal resources is being managed by the states. In response to the problems in compiling the regional accounts, NSC (2018) clearly outlined that the major issue with regional accounts apart from existing problems was due to the data gap in MCA21. To quote:

In absence of details of a company's State-wise activities, the national-level GVA estimates are allocated to States in proportion to (i) State-level GVA estimates obtained from ASI for manufacturing activities, (ii) the indicators for allocating services sector estimates have been mentioned in para IV.2.4.3 above. [IV. 3.3.6 NSC (2018).]

Given the complexity of the problem at the regional level, lack of credible GSDP estimates could adversely affect the states' ability for resource planning and budgeting. The recommendations of the NSC (2018) for resolving these issues would require a series of policy and regulatory efforts so as to rely less on voluntary compliance by companies and more on data validation and scrutiny checks.

## 6. Issues with the 2011–12 Back Series of GDP

While data and methodology problems remained unresolved, new controversies related to the 2011–12 series have also cropped up in the last one year. Since its release, the 2011–12 NAS did not have a "back-casted series," that is, estimates at 2011–12 prices beginning from 1950–51. The release of a back series of any new base-year series is a routine exercise. Given the substantial changes in data sources and methods of estimation in the 2011–12 NAS series, which introduced inconsistencies with the sources and methods used in the older series, compiling a back series was a major challenge.

However, in 2018, two separate sets of back series based on two different approaches were released, one official and another unofficial, for varying time lengths, leading to an inconclusive debate on the historic growth performance of the economy. First, the Committee on Real Sector Statistics presented its own estimates from 1994–95 to 2013–14 (henceforth, the NSC back series). Subsequently, the CSO released the official version of the back series for only seven years from 2004–05 to 2011–12 (henceforth, the CSO back series).[15] These two series showed different growth rates for 2005–06 to 2011–12. (CSO 2018a; NSC 2018). This is clearly demonstrated in Figures 8 and 9.

15. See NSC (2018) and CSO (2018b) for details and documentation.

FIGURE 8. Comparison of the NSC and the CSO Back Series Growth Rates: Current Prices, 2011–12 Series

Back Series, 2011–12 Series Current Prices
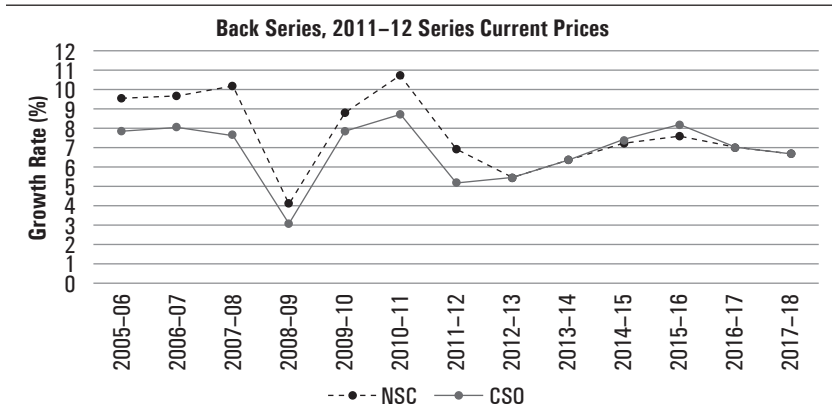
Source: CSO (2018b).

FIGURE 9. Comparison of Average GDP Growth Rates during 2005–12 from the 2004–05 Series, NSC Back Series and CSO Back Series

Average Growth Rate (2005–06 to 2011–12)
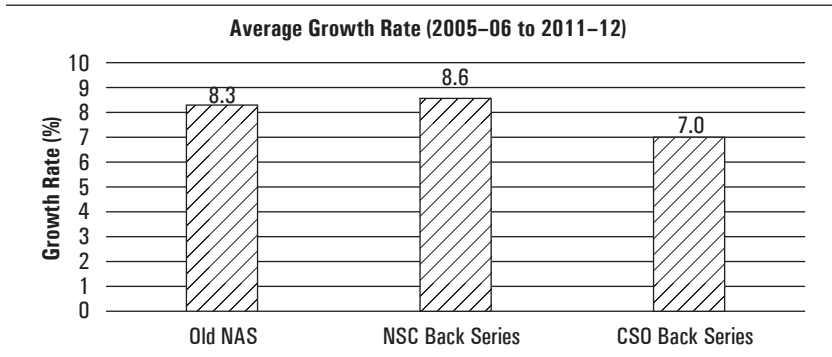
Source: CSO (2018b).

The CSO back series showed lower annual growth rates for all the years from 2005–06 to 2011–12. For the seven-year period, most of which had been so far considered to be an economic boom period for India, the CSO back series reported an average annual growth rate of 7.0 percent, as opposed to the 8.3 percent growth rate reported in the 2004–05 base-year series.

The significant downward revision of growth rates and the diametrically opposite picture painted by the CSO back series compared to the NSC back series raised suspicion about the veracity of the estimates. This was especially because, as mentioned earlier, by most popular accounts, these

seven years recorded unprecedented economic growth, an export boom, a credit boom, and an investment boom when India was hailed as one of the fastest growing economies in the world. The CSO back series changed this piece of Indian economic history.

In addition to changes in the aggregate growth rates, the CSO back series also changed the overall composition of GDP in the following ways:

1. Reduction in the share of the services sector for this seven-year period;
2. A rise in the shares of primary and secondary sectors (corporate manufacturing in particular); and
3. A reduction in the size of the unorganized/informal sector and expansion of the size of the private corporate sector.

The CSO has so far not released the details of all the methods, procedures, and adjustments made in preparing the back series. We can obtain some understanding of how this series was put together from its press release of November 28, 2018. While the NSC back series applied a "production-shift" technique to obtain previous years' growth rates, it seems the CSO back series used a concoction of methods and data sources.

For example, as claimed by the CSO in its press note, till the time the MCA21 database was available, they used this data to calculate the GVA of the PCS. For all the previous years when the MCA data on corporate filings were not available, they resorted to the ASI data (that was used in the older NAS series). They further mention in the same press note:

The methodology for preparing the back-series estimates for the years 2004–05 to 2010–11 is largely the same as the methodology followed in the new base (2011–12). In certain cases, owing to the limitations of the availability of data, either splicing method or ratios observed in the estimates in base year 2011–12 have been applied. …Splicing method has been applied for preparing the estimates in Construction Sector entirely and applied partially in Agriculture and Allied Sectors, Gas Trade, Repair, Hotels and Restaurants, Real Estate, Ownership of Dwelling and Professional Services, Public Administration and Defence and Other Services.

This shows that the CSO back series was estimated using different databases for different periods and different methods for different sectors. This raises serious doubts about the comparability and continuity of the back series with the new 2011–12 GDP series and hence, about the reliability and usability of the back series.

Moreover, as discussed in detail in the previous sections of this paper, many infirmities in the new methodologies and data sources used by the CSO have come to light in the GDP measurement debate and none of these

has been resolved so far. The use of the MCA database, in particular, could have misleadingly enlarged the PCS's share in the Indian economy and its growth rate. Therefore, using the same methods and data sources to backcast the 2011–12 series is likely to result in incorrect estimates as well. In this context, it is worth asking how correct and prudent it is to selectively use some of the contested methods for preparing the back series.

7. Conclusion and Way Forward

In 2015, the CSO introduced a new series of National Accounts Statistics with 2011–12 as the base year, replacing the earlier 2004–05 base-year series—a routine matter for statistical authorities of most countries. The re-basing was carried out to account for changes in the structure of the economy and in relative prices, always following the global template of the UNSNA, the latest one being the 2008 edition. It is also an occasion for statistical authorities to introduce newer databases and better methodologies to improve the data quality.

Typically, as seen in the past, re-basing leads to a slight enlargement of the absolute GDP size, as output that was previously left out or inadequately captured gets recorded after the revision. This does not usually lead to changes in the growth rates, implying that the underlying trend remains the same. The latest NAS revision, however, defied these usual patterns, and reported a slight contraction of the GDP size in the base year, as well as a faster growth rate in the subsequent years. As the growth trends in the new series did not square with related macroeconomic aggregates, widespread skepticism emerged questioning the veracity of the new GDP series. The statistical authorities responded, saying that the newer estimates are sound because they have used the latest UN guidelines, larger databases, and improved methodologies but this failed to carry conviction.

In their IPF paper, Nagaraj and Srinivasan (2017) unpacked the issues and recorded the state of affairs as they were till mid-2016. Since then, fresh research and data releases have uncovered newer problems, thereby strengthening the earlier doubts about the new GDP's veracity and reliability. This has made it imperative to assess the issues, which is the objective of the present paper. Given that much of the newer research and questions are centered on the PCS output estimates, our paper has paid most attention to this aspect of the revision.

A major change in GDP estimation in the new series was the use of regulatory filings of financial returns (in the MCA database) to estimate

output of the PCS, replacing the production accounts obtained under the ASI for manufacturing firms (which account for nearly one-half of the overall corporate sector output). This change was predicated on the view that the production accounts did not capture output outside the factory premises given *its approach* to data collection. The *enterprise approach* used the company balance sheet, which is considered as a solution to this problem. Research undertaken to closely examine the ASI data revealed that the assumption for the shift in approach is factually incorrect, thereby undercutting the very basis of the innovation introduced in the new NAS. Since the manufacturing sector growth rate has been persistently higher in the new series as compared to the picture painted by other macroeconomic indicators, there are apprehensions that the change in approach to data collection may be at the source of the problem.

While the universe of registered companies may have increased substantially, the state of "active" and "filing" companies in the database has serious implications for GVA estimation. The structure of the PCS is such that a small number of large companies contribute a large share to GVA. Limited information from the MCA database suggests that a large number of small companies are unavailable for estimation on an annual basis. Their estimates are obtained through a blowing-up procedure, the details of which have not been released by the CSO. The GVA estimates can be imprecise, especially when the sample size, its fraction, and the universe of working companies are indeterminate.

A recent official data release gave credence to the above suspicion. In 2016–17, NSSO conducted a survey of non-government and non-financial services sector enterprises on its way to launch a full-fledged series of annual survey of services on the lines of the ASI. One of the list frames (i.e., the universe) for drawing the sample, expectedly, was the MCA's list of "active" (i.e., deemed working) companies—a part of the universe of companies CSO uses for estimating the PCS GDP. After due verification, when NSSO launched the survey, it failed to get a response from up to 45 percent of the sampled companies. Admittedly, some of the non-responses could be due to misclassification (which, in principle, could be rectified). But the fact that 24 percent of the sample companies were non-traceable/ failed to respond suggests that the universe of "active" companies used for PCS GDP estimation is unreliable and riddled with holes. This raises doubts about the magnitude and reliability of output estimates (prepared using the same list frame) accounting for over one-third of the economy's GDP.

Another problem with the GDP measurement, which is a legacy issue but became prominent when the new NAS was released, is the manner in which

nominal GDP or GVA values are deflated to obtain the real GDP growth rates. The global best practice is double deflation where different price indices are used for inputs and outputs because if their prices are changing at different rates, then using the same price deflators would yield distorted estimates of value addition. This has not been done for the new GDP series. A single deflator is used in generating the GVA series for a given sector. In the absence of the double deflator method, whenever there is a divergence between output and input prices, the real growth rates would tend to be overestimated. This is exactly what happened in the period 2013–2016 when WPI inflation was significantly lower than the CPI inflation, the net result being a potential overestimation of the real GDP growth rates under the new NAS.

Regional accounts are an integral part of the system of national accounts. The NAS revision process has apparently paid scant attention to the implications of methodological and database changes for the estimation of GSDP estimates. The problem arises because the newer databases used—such as the corporate filings in the MCA database mentioned above—are not geared for producing state-level (let alone at the district level) output estimates. As a result, the state-level estimates are mostly apportionments of the national estimates, grossly distorting the statistical picture of underlying economic reality. With increasing economic decentralization, distorted state income accounts end up affecting the distribution of resources and probably even aggravating inter-regional inequalities.

After claiming for three years that the GDP back series could not be prepared due to the substantial methodological changes in the latest GDP revision, in 2018 the statistical establishment, in quick succession, came out in with two back series with diametrically opposite trends. While the series by a committee of the National Statistical Commission boosted growth rates for the last decade (from 2004–05 to 2011–12), CSO's officially accepted series reversed the trends, drastically lowering growth trends for the previous decade. The conflicting trends and lack of transparency in the methodology used, especially in the official back series confounded data users, and further dented the credibility of the statistical establishment.

Recommendations

If the foregoing analysis is sound and substantial, then it casts a serious doubt on the new GDP series. In response, many private and international financial firms have apparently resorted to their own devices to find proxies for GDP. Some are apparently using World Bank's night lights data as a measure of

economic activity, or high-frequency industry- and sector-specific data, all of which, at best, are second-best solutions.

Going forward, we consider two sets of recommendations, one short run or intermediate remedies, and the second, a longer term and lasting solution. Since the MCA database and the methodologies are the heart of the problem, authorities should immediately release the data in a suitable form for independent verification of the official GDP estimates. As corporate filing is a statutory requirement, the data should, in principle, be easily accessible in the public domain. However, considering its sheer size and complexity, the database needs to be made public in a suitable format via public institutions. MCA could set up data laboratories in leading research institutions and universities, similar to the Census Commission's initiative, to encourage policy-oriented research on the corporate sector. More specifically and immediately, the MCA/CSO can release the following data from 2011–12 onwards:

- Yearly information on the sample size, sample fraction, and the size of universe of "active" companies and their PUCs;
- A break-up of financial and non-financial companies, by various categories; and
- List of companies filing returns with information on selected variables, which could help data users to independently verify data quality.

The MCA and CSO should also create a suitable institutionalized forum for regular and sustained interaction with data users to address the numerous issues that have come up in the course of the GDP measurement debate.

To address the problems arising out of the single deflator issue, till the time producer price indices are generated, which may enable the move towards a double deflator method, the CSO may consider using the WPI as a deflator for the industrial sector and the CPI as a deflator for the services sector. This will help deal with some parts of the problem arising from the deflator issue.

For a lasting overall solution—reiterating Nagaraj and Srinivasan's 2017 recommendation—a statistical audit and a credible expert committee need to be set up to invite the best expertise available globally to review the GDP revision process. Some of the core issues that the expert body may examine are the following:

1. Appropriateness of replacing the establishment approach to data collection with the enterprise approach for the non-farm sector, given

the present level of India's development, and quality and reliability of the available statistical bases;

2. Shifting of QCs from the HH/unorganized sector to the PCS, and its many ramifications for macroeconomic aggregates, and policy; and

3. Critical examination of the incompleteness and unreliability of the MCA database, given the limited state capacity to enforce laws governing private enterprises. There is an urgent need for a thorough investigation to ascertain its suitability for estimating domestic output for an emerging market economy like India.

The objective of the audit/committee would be to investigate in detail the problems in the sources and methods of the new NAS and help come up with the best alternative estimates, preferably before the next base-year revision is conducted, otherwise we may end up perpetuating the defects in the current base-year series and India's GDP will continue to be marred in controversy.

References

CSO. 2014. *National Accounts Statistics - 2014*. New Delhi: Ministry of Statistics and Programme Implementation.

———. 2015a. *Report of the Sundaram Committee on Unorganised Manufacturing and Services Compilation of National Accounts Statistics with Base-Year 2011-12*. New Delhi: Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India.

———. 2015b. *Changes in Methodology and Data Sources in the New Series of National Accounts*, *Base Year 2011–12*. New Delhi: Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India.

———. 2015c. *Final Report of the Sub-Committee on Private Corporate Sector including PPPs*. New Delhi: National Accounts Division, Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India.

———. 2015d. *Report of the Sub-Committee on System of Indian National Accounts*. New Delhi: National Accounts Division, Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India.

———. 2015e, "No Room for Doubts on New GDP Numbers." *Economic & Political Weekly,* 50(16): 86–89.

———. 2018a. *Proceedings of Awareness Workshop on "Challenges and Issues of Official Statistics" for Senior ISS Officers during 18th–19th May, 2018 at Bengaluru* (Part I and II). New Delhi: National Statistical Systems Training Academy, Ministry of Statistics and Programme Implementation, Government of India.

———. 2018b. *Back-series Estimation Base 2011–12 Methodology Document*. New Delhi: Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India.

CSO. 2018c. *National Accounts Statistics - 2018*. New Delhi: Ministry of Statistics and Programme Implementation, Government of India.

Dholakia, Ravindra, and Manish B. Pandya. 2017. "Critique of Recent Revisions with Base Year Change for Estimation of State Income in India." *Journal of Indian School of Political Economy,* 29(1–2), January–June.

Dholakia, Ravindra H., R. Nagaraj, and Manish B. Pandya. 2018, September 1. "Manufacturing Output in New GDP Series: Some Methodological Issues." *Economic & Political Weekly,* 53(35): 10–13.

Manna, G.C. 2017. "An Investigation into Some Contentious Issues of GDP Estimation." *Journal of Indian School of Political Economy,* 29(1–2), January–June.

———. 2018, January. "A Study on the Likely Magnitude of Bias in the Estimates of Gross State Domestic Product for the Private Corporate Segment of Manufacturing Sector as per the New Methodology." *Journal of Indian School of Political Economy,* 30(3–4): 455–468.

Ministry of Finance. 2019. *GDP Estimation: A Clarification*. New Delhi: Press Information Bureau, Government of India, May 10.

Nagaraj, R. 2015a. "Seeds of Doubts on New GDP Numbers: Private Corporate Sector Overestimated?" *Economic & Political Weekly,* 50(13): 14–17.

———. 2015b. "Seeds of Doubt Remain: A Reply to CSO's Rejoinder." *Economic & Political Weekly,* 50(18): 64–66.

———. 2015c. "Growth in GVA in Indian Manufacturing." *Economic & Political Weekly,* 50(24): 117–120.

Nagaraj, R., and T.N. Srinivasan. 2017. "Measuring India's GDP Growth: Unpacking the Analytics and Data Issues behind a Controversy that Refuses to Go Away." In *India Policy Forum*, 13: 73–128. New Delhi: National Council of Applied Economic Research.

NSC. 2018, May. *Report of the Committee on Real Sector Statistics.* New Delhi: National Statistical Commission.

NSSO. 2019. *Technical Report of Services Sector Enterprises in India*. New Delhi: Ministry of Statistics and Programme Implementation, Government of India.

Pandey, Radhika, Amey Sapre, and Pramod Sinha. 2019, January. "What Do We Know about Changing Economic Activity of Firms?" *Working Paper No. 249*. New Delhi: National Institute of Public Finance and Policy.

Rajakumar, Dennis J., Vijayata B. Sawant, and Anita B. Shetty. 2015. "New Estimates of Saving and Capital Formation." *Economic & Political Weekly,* 50(12): 64–66.

Ramana Murthy, S.V. 2018. "Base Change 2011-12 and Implications on GSVA." *Journal of Indian School of Political Economy,* 30(3–4), July–December.

Sapre, Amey, and Pramod Sinha. 2016. "Some Areas of Concern about Indian Manufacturing Sector GDP Estimation", *Working Paper No. 172/2016*. New Delhi: National Institute of Public Finance and Policy.

———. 2017. "Some Unsettled Questions about Indian Manufacturing GDP Estimation." *Journal of Indian School of Political Economy,* 29(1–2), January–June.

Sengupta, Rajeswari. 2016. "Why Inflation Targeting Works." *LiveMint*, August 16. Available at https://www.livemint.com/Opinion/YZFLeMeRm0U94zQzwDVBIK/Why-inflation-targeting-works.html (accessed August 30, 2020).

Sidhartha, and S. Gupta. 2015. "There Is More Acceptance, Credibility of New GDP Data: Chief Statistician." *Times of India*, June 2.

Subba Rao, K.G.K. 2015. "Mystery Private Corporate Saving." *Economic & Political Weekly,* 50(2): 158–162.

To view the entire video of this IPF session and the General Discussion that ended the session, please scan this QR code or use the following URL:
https://www.youtube.com/watch?v=xl0H1a4bEyI

Comments and Discussion[*]

Chair: **Rakesh Mohan**
Yale University

Sudipto Mundle
NCAER

Before getting to the paper itself, let me first give a little bit of the background regarding the report of the committee that the National Statistical Commission (NSC) had set up on real sector statistics (henceforth, the Real Sector Committee). This is partly for reasons of full disclosure, as I happened to be the Chair of that committee, the findings of which are discussed in several parts of the paper and because I will refer to some of the recommendations of the Real Sector Committee in my comments on the present paper.

When I was asked by the NSC to chair the Real Sector Committee, which was supposed to look at different ways of improving real sector statistics in the country, I said I would be prepared to do that only if the terms of reference of that committee included generating the GDP back series because this had become quite urgent by then. There is a convention that whenever a new, rebased GDP series is released, a corresponding GDP back series for earlier years is also released subsequently to enable comparison of what was there before with what came after the rebasing. However, this was not done when the 2011–12 base GDP series was issued and several years had since passed. This left all analysts, whether for policymaking or in the private sector or in academic research, in quite a quandary because tools of time series analysis could not be applied to the new GDP data without the back series.

Fortunately, the terms of reference of the Real Sector Committee were revised as requested and our work proceeded. I should mention that the secretariat for this committee was the CSO, the part of the Statistics Ministry that produces India's national accounts. The member secretary of the committee was the then director general of the CSO, Shri Srivastava, who is now the Chief Statistician and Secretary of the Ministry. When the

---

report was completed, it was reviewed by the committee page by page, line by line, and it was unanimously approved, including by the representative of CSO who also signed off. The report was then submitted to the NSC. In other words, there was strong ownership of the Real Sector Committee report by CSO. In fact, once this report was submitted, it was prominently posted among the new releases of the Ministry on the MoSPI website. This was around July 2018. About a month or so after that, a journalist used the Real Sector Committee's GDP back series to show that the growth rates during the Congress period ending in 2014 were higher than the growth rates after 2014, which is the Bharatiya Janata Party (BJP) period. This was, of course, well known, as the Congress period was one of strong policy stimulus the world over following the 2008 financial crisis. In fact, this growth spurt was evident even in the old 2004–05 base GDP. But once the journalist provocatively politicized the GDP back series issue, politicians got into the act. The Congress spokespersons said that theirs was a much better period. The BJP representatives started saying that the Congress period was one of fiscal profligacy leading to high inflation. With the controversy rising, MoSPI gradually distanced itself from the report, which till then it had fully owned. It was shifted from the prominent "recent reports" section of the Ministry's website to the NSC reports section. Subsequently, the back series that CSO had not been able to produce for the previous four years or so was suddenly completed and released within 2–3 months as the new "official" back series. This "official" back series reversed the comparison, now showing that the growth rates were actually lower during the Congress period than during the BJP period.

I will not go further into this back series issue, but I will use some of the recommendations from other parts of the Real Sector Committee Report while commenting on the present paper.

Let me now get to the paper itself. Most of my comments will focus on issues where I am not on the same page as the authors of this paper, mainly in order to induce discussion. Hence, let me state clearly at the outset that I think this is really an excellent paper, it is very comprehensive and drills down to the details of what has been done in the new national accounts series. That is what is needed to unravel all the puzzles we have been living with for the past few years, ever since the new 2011–12 base GDP series was released.

Following the release of this new series in 2015, doubts were expressed right from the beginning about the reliability of the new series. This was partly because for the three common years for which the old series numbers were given along with the new series, the new series estimates lowered the

growth rate for the first common year, compared to the old series, but made it higher for the subsequent two years. Also, the new growth estimates did not gel with other macroeconomic indicators. Subsequently, a lot of analysis has been undertaken of the possible sources of error, including by the authors of this paper—most famously in the Nagaraj–Srinivasan paper presented at the 2016 IPF (Nagaraj and Srinivasan 2017) and in more work subsequently. All of this is well summarized in the present paper.

There seems to be a general consensus that there is a problem of over-estimation in the new GDP series, especially relating to the MCA data for the industrial sector. Generally, I think it is fair to say that there is a broad consensus that there is growth overestimation, and this paper presents an updated and comprehensive critique of the new GDP series. Certainly, the critique is important and has to be there, but from my perspective, the really important question is about how to fix the problem. The availability of reliable, usable GDP statistics is essential for all economic and other important decision-making in both the public and private sectors, and it is through that lens that I am making my observations.

I think the problems with the new series fall into two broad groups. One has to do with the data itself, especially the MCA data, and how it has been used to generate the GDP numbers as detailed in the paper. The other part is the deflation method that has been used. That is also discussed in the paper. It is useful to compare the new GDP series with the old series or the Real Sector Committee back series with the official back series and see how much of the difference can be explained by differences in the data source and how much by the deflation procedure. I am emphasizing this because while there are differences in nominal growth rates between the two series, these differences are relatively small. The big differences appear when we compare the real GDP growth rates. This suggests that the real problem lies more with the deflation procedures used and not so much with the data. My main reservation with the present paper is that it focuses more on the problem with the MCA data and its unreliability rather than on the problem of the deflation procedure used.

But that being said, let me now first talk about the MCA database, some of the points raised in the paper are regarding the MCA database and the blow-up method that has been used in the new GDP series. As the paper mentions, and illustrates with a Venn diagram, there are at present about 18 lakh companies in the MCA database. Of these, about 12 lakh are the companies which are described as "active," in the sense that they have submitted accounts at least once in the last three years. Of these, about half or 6 lakh companies are described as "filing" companies in the sense

that they filed their accounts with the MCA in the last year. When the NSS tried to use the MCA list of 12 lakh active companies as a sample frame, it found that a large proportion of these companies was either misclassified or non-existent. So the paper concludes that this database is not robust enough to be usable. But the paper also notes that CSO subsequently clarified that a lot of companies reported as missing or non-existent had, in fact, filed their accounts electronically. It was eventually found that just 622 companies were missing, which amounts to about 2 percent of the universe of active companies. Obviously, this error needs to be fixed but it is not such a drastic defect that it calls for the entire MCA database to be junked.

A potentially more serious problem in my view is that the sample of companies filing reports in the current year, the "filing" companies, changes from year to year. It is not the same set of companies which are filing reports every year. So if the set of companies filing accounts this year is different from that which did so last year, *even if the output level has not changed at all*, it could still appear as either an increase or decrease in output. That would obviously be an error. The way CSO has dealt with this is to use the data for what are called "common" companies, which is a subset of 3 lakh companies that have filed accounts with MCA in both years. The paper cites the fact that only 3 lakh companies belong to the set of "common" companies as further weakness of the MCA database. But to me, it is a source of strength because these 3 lakh companies, which form the core of the Venn diagram, comprise a set of companies for which we have robust, comparable data over the years. It is, in fact, still a very large sample, amounting to 25 percent of the universe of active companies.

The blow-up method used by CSO, another weakness cited by the paper, is a genuinely serious problem. The ratio of PUC of the active companies to that of the common companies is used as the blow-up factor to arrive at GVA estimates for the universe of active companies from the data relating to common companies. However, there is no correlation between GVA and paid-up capital, PUC, and this is a genuine weakness in the new GDP series. The Real Sector Committee had suggested that the ASI data should be used together with the MCA data to get more reasonable estimates of GVA.

To sum up on the data side, the MCA data was released and used prematurely in the new GDP series. A lot more homework needed to be done on how to calibrate it better, how to clean it up with the help of ASI data, and so on. I think this can be done by adopting some of the recommendations made by the NSC Real Sector Committee.

Let me now quickly come to the deflation problem, which I think is the main problem with the new 2011–12 base GDP series. As I indicated earlier,

the differences between the different GDP series are quite modest in nominal terms. The major differences appear in the real estimates after deflation. The global best practice is double deflation where different price indices are used for the inputs and outputs because if their prices are changing at different rates, then using the same price deflators would yield distorted estimates of value addition. This has not been done for the new GDP series. A single deflator is used in generating the GVA series for a given sector. The reason is that the MCA data does not lend itself easily to this kind of double deflation because it is enterprise data and not establishment data, and it cannot be classified by products or location of production. Here again, the NSC Real Sector Committee recommended that ASI data should be used together with the MCA data to develop a procedure for processing the data, including use of the double deflation procedure. However, switching to these new procedures will take time, mainly because India still does not have a producer price series and CSO still makes do with the CPI and WPI price indices. This is a major gap in India's price statistics. Pending the generation of producer price indices, the paper offers an excellent suggestion and makes a compelling case for using WPI indices for the industrial sector and CPI for the services sector in the interim period. I think that is a good way to go.

Reference

Nagaraj, R., and T.N. Srinivasan. 2017. "Measuring India's GDP Growth: Unpacking the Analytics and Data Issues behind a Controversy That Has Refused to Go Away." *India Policy Forum*, 13: 73–128. New Delhi: National Council of Applied Economic Research.

N. R. Bhanumurthy
NIPFP

I will try to limit my comments on the back series work that I have done for the Mundle Committee. I will also highlight and show some of the issues that Dr Mundle has already pointed out in terms of what the difference is between the new official back series and the National Statistical Commission's back series numbers. The main contention of the paper is that the base year revisions should not cause big changes in the growth rates. For this, we need to look at the nominal GDP numbers. When the revision happens, it does so on the nominal series. If you compare the old and new nominal series as

the paper has already pointed out, there was a decline in the nominal GDP numbers in 2011–12, not an increase as many of us believe. In fact, there was also a decline in the years 2012–13 and 2013–14.

The paper talks about sharp revisions in the shares of the private corporate sector and the shifts in the shares of overall manufacturing as well as trade. The authors feel that the revised growth rates do not actually reflect the kind of changes that are happening in the Indian economy. They have not highlighted in their presentation but mention in the paper issues related to two contradicting back series data. The paper also raises concern about high GDP growth during the demonetization period in 2016–17, and it takes the help of the NSS survey to highlight the issue of use of MCA21 data. The paper also questions the quality and accuracy of the methods as well as the databases that have been used in GDP estimation.

The authors conclude that there seems to be overestimation of the growth rates, especially in one particular sector, that is, PCS, and within the non-financial component. The paper raises a lot of issues regarding the reasons for the upward bias in the growth rates. The paper also discusses the relation between the paid-up capital and the GVA, especially that being used for the blow-up. Another issue raised in the paper concerns non-filing companies and the issue related to shifting of the entire methodology from establishment to enterprises as well as the double-deflation problem, which is very useful for further research.

I believe that the paper is very comprehensive. The authors have summarized most of the issues that have been raised since 2015, and in fact, those who have read their IPF July 2016 paper (Nagaraj and Srinivasan 2017) will know that this new paper has updated most of the issues that have been raised since then. In many places, the paper also makes very strong recommendations and attributes those recommendations to a government committee. In a sense, these recommendations are already there in the public domain and I am sure MoSPI is also aware of this.

As regards the use of WPI and CPI in the new series, it has been observed that mostly CPI is used instead of WPI. At one place, the paper says that the NSC back series actually boosted up the growth rates. I believe that if you take the numbers from 1993–94, the difference between the old series and new series, on an average, is higher by only 0.2 percent. The paper also points out that we have done splicing. We have been opposing this method and suggesting that it should not be used mechanically. We have basically used the *production shift approach*, which is very different from splicing. Again, at one place, the paper says that estimation of the NSC back

series is not really transparent. My view is that many people have read our committee report and in fact some have re-estimated and they are getting the same numbers as ours. So, in that sense, the Mundle Committee report recommendations on back series is replicable and is much more transparent than any other method.

The paper also recommends the creation of a *credible* expert committee, thus in a sense appearing to question the integrity of committees and their members. As part of the Advisory Committee on National Accounts Statistics, we do not even get to see the numbers. We only discuss the method and suggest the methodology that should be adopted, and we have not really seen the original databases that have been used. It seems that the paper repeatedly wants to show that GDP is overestimated. Most of the examples in the paper (Box 1) aim to show how the MoSPI figures have been overestimated. However, one can flip back the examples and say that GDP can also be underestimated. For example, the paper assumes a CPI of 5 percent and WPI of –5 percent. But if one assumes that these figures can be taken to be the other way round, one can find that you will be underestimating the GDP numbers. I thought that may be looked at.

The paper is ostensibly saying that the past method was more accurate than the current method. However, let us look at some of the issues. In terms of the nominal GDP growth estimates, the old series and new series are almost the same. I think it is very clear from the two graphs in my presentation that there is no change in the nominal GDP growth. But if you look at the real growth rates, you see that the new series is slightly lower in the pre-2011–12 period but in the post-2011–12 period, you will see a completely different trend. This is exactly what Dr Mundle pointed out in terms of the deflators. I do have some estimates on the deflators that are derived from the new back series.

So what we did as part of the Mundle Committee was basically that we tried to adopt a production shift approach. If we look at both GVA growth rates and GDP growth rates from this approach, we get a smooth series as compared to what we had in the past. In fact, one of the basic objectives of GDP estimation when you look at the back series is that we would like to have something free of artificial breaks and free of deflator issues. In fact, the revision happened only on the production side, not on the prices side. So I don't know why we should see the change in the deflators when we revise the GDP series.

In terms of deflators, the new series of CPI is used instead of CPI-AL agricultural laborer or industrial worker, but for GDP data prior to 2010,

we are not clear what kind of CPI numbers have been used as there was no comparable back series for the new CPI estimates. Ultimately, the problem arises when MoSPI does not provide us details about what exactly is getting into the estimation of their new back series numbers. So as regards the lower growth in the new series as compared to the old series, as seen in the graph, the nominal seems to be fine but there is a problem with real numbers. I think that is where we are seeing some kind of *deflated* growth in the new back series.

The alternative could have been to retain the old deflators, which would have prevented any controversy. If you look at the deflators, we tried to estimate the deflators for the period 2004–11, and it is very clear that actually it is the new deflators that are higher than the old ones, while there should not be any change in the way CPI or WPI or price has been estimated in the pre-2011–12 series. So the entire difference between the new back series and the NSC back series can be explained by the changes in the deflator numbers. This is also true at the sectoral level by looking at the sectoral deflators.

The last question I want to raise is: Has the past method given consistent numbers? Basically, IIP numbers are used to estimate manufacturing sector growth. If you look at the IIP numbers, you will see almost a flat kind of a curve in the recent period. In fact, even in 2016–17, the year of demonetization, the IIP series did not show any dip as compared to the previous year. We use these IIP numbers to estimate the manufacturing sector growth in the past. On the other hand, if you look at ASI as well as MCA21 data, the pattern seems almost similar. But when you use IIP numbers, you do not really get a real picture of the manufacturing sector. Here, I am trying to say that the IIP numbers we had in the past and the way we estimated the manufacturing sector value addition in the past were also flawed. I think that we were perhaps underestimating past growth.

At present, we get the ASI numbers with a lag of two years. Now if you look at the final numbers, after two years, you should be seeing a sharp revision between the ASI numbers and the numbers that have been generated by IIP. We don't see those kinds of sharp revisions when we get the final numbers on the manufacturing sector.

Let me conclude by saying that the paper has raised many relevant issues. I am sure these will be useful for the future. But as Dr Mundle pointed out, we need to really work with NSO, and I hope MCA will release the database for researchers so that more work can be done. In terms of methodology, the past method is very different from the current method. Hence, comparing the past and the present may not really help much when you want to come up with credible GDP numbers.

General Discussion

Rakesh Mohan, the chair, initiated the discussion by stressing that GDP estimation is a serious issue for the country for the interpretation of what is going on in the economy and how it impacts thinking about what to do in the future. He asked that since it is only CSO that knows which companies are included in the MCA21 and which establishments in the ASI, why is it not possible for CSO to do a complete correspondence to clarify what is going on?

Bishwanath Goldar clarified why 2012 and 2013 manufacturing growth rates in the old series were much lower than the growth rates in the new series. He explained that the standard practice is to estimate growth rates initially on the basis of the Index of Industrial Production and then to revise them subsequently. However, the shift to the new series meant that the revisions for 2012 and 2013 never took place, so the growth rates for these two years remained low. On double deflation, he pointed out that using data available on the RBI website, it is possible to compute the double-deflated GVA series for manufacturing or for any other sector and compare it with the single-deflated series from the national accounts.

Goldar talked about how factory-based growth numbers from the ASI would typically be lower than company-based growth numbers simply because companies include more activities beyond production, and these may be growing faster. He noted that factory data from the ASI could now be compared with the corresponding MCA company data since the ASI now has a field for the CIN number of its associated company. He also pointed to the need to resolve the new difficulty of apportioning GVA national estimates (estimated using MCA company data) to the states (using the ASI data). This was not a problem in the past when the ASI was also being used for GVA.

Arvind Subramanian noted that the discussion on GDP estimates was taking place in a very narrow way because of lack of access to expenditure side estimates of GDP; we should be looking in parallel at estimates of investment, consumption, and exports. He asked a question about the size of the sample of companies in the MCA database that was being used to calculate GVA and thought it was much less than the 3 lakhs that the paper spoke about and was more like 3,000–6,000 firms.

Ramana Murthy, the head of CSO's Economics and Statistics Wing, and until recently the head of its Industrial Wing handling the ASI, noted that the report in the paper that NSSO, in preparing for the Annual Survey of Services, had found that 45 percent of the service companies in MCA's list of active companies were untraceable or did not respond was a misrepresentation. He explained that a total of 35,456 companies were sampled out of

1,100,000 supposedly active companies (of which 700,000 were companies filing returns). Hence, the NSSO had taken a small portion of companies belonging to the services sector covered by the NSS 74th Round. Financial enterprises, air transport, and many other service sector enterprises were excluded. Some 21 percent of these were "out-of-survey" units because their NIC codes, the National Industrial Classification codes, fell outside the coverage of the survey, so that constituted a portion of the 45 percent. Further, there were closed units. There were also establishments in the MCA database that were dropped since we were following an enterprise approach, but that did not mean that they were fake or non-existent units. Then there were "non-traceable" units that could not be found at the given addresses, but that again did not mean they were closed, since they could have shifted to a new location and were in operation. There were also about 6 percent "casualty units" that declined to share their data even though they were operational. All of these categories added up to 45 percent, but that did not mean that they did not exist. But that should not be the reason to drop the MCA data, which is much larger than what was used for the NSSO survey.

He also noted that "quasi corporations," mostly proprietary, partnership enterprises maintaining accounts, unincorporated units that are nonetheless counted in the ASI under the Factories Act, are organized sector units and should be treated as part of the PCS in the national accounts according to the best-practice recommendations of the UNSNA. That is exactly what the CSO had done.

Srinivas Murthy, also from the CSO, adding to Professor Goldar's explanation, pointed out that Professor Nagaraj's first slide comparing growth rates for GDP, manufacturing, and trade/hotels had been repeated elsewhere and creates confusion because it does not distinguish between the many stages of finalizing GDP/GVA estimates, ranging from First and Second Advance Estimates, then the second stage Provisional Estimate, and the final First, Second, and Third Revised Estimates. All three have different coverage as more data becomes available. This is all made clear on the MoSPI website. So it is really important to compare like with like.

As regards the use of the MCA21 data, there is a conceptual difference between the First Revised and the Second Revised Estimates of the nominal national account numbers. The common sample of 350,000 companies from the MCA21 cited in the paper relates to the First Revised Estimate, but it is a truncated sample. The Second Revised Estimate uses a larger sample of companies from the MCA21 and uses the provisional ASI data. It is not the number of companies that matters but the PUC of the sample vis-à-vis the universe that matters, and we may be covering some 85 percent of that by the Second Revised Estimate, which is where the blow-up factor matters.

He agreed that PUC and GVA may not be highly correlated, but that is the metric put out by the MCA in public, and even the RBI has used it.

Brijender Singh from MoSPI gave some examples to counter the charge of potential overestimation; he suggested that comparing the ASI manufacturing growth rate and private corporate manufacturing growth rate over time does not always point to overestimation in the case of manufacturing for private corporates. Further, the divergence between First Revised Estimate, which is based on the growth rate of common companies, and the Second Revised Estimate, which is based on the scaling-up factor, does not indicate that the scaling-up factor systematically leads to an increase in the growth rate, and in fact, it sometimes actually leads to a downward revision.

Kaushik Basu pointed out that the share of registered but non-active companies in total companies jumped a lot during individual years. For instance, there was a precipitous drop in 2017–18, but the figure remained more or less constant during the period 2013–16. These fluctuations may provide insights about the identity of these companies, and the impact on growth rates.

Arvind Virmani noted that it is unfair to blame MoSPI for lack of access to the MCA21 data since it belongs to MCA. He highlighted the need for making the data public. He also stressed that while there is a standard methodology for estimating the SNA, there is no standard methodology for the back series. So it is important to distinguish between problems with the back series and those with the estimation.

Dilip Mookherjee said that double deflation versus single deflation seemed to account for most of the difference in the growth rate. So it was not clear why, from an economic standpoint, double deflation was the right way to go. What we are interested in ultimately is national income and not physical production.

Rakesh Mohan concluded the session by urging the government, CSO, and MCA to jointly clarify the doubts being raised about the GDP figures. He noted that since confidence in India's national accounts has fallen, these figures need to be corrected. He suggested that NCAER and CSO could hold a meeting to discuss and clarify these doubts.

The session video, the paper, and all presentations for this IPF session are hyperlinked on the IPF program available on the NCAER website by scanning this QR code or going to the following URL:
https://www.ncaer.org/IPF2019/Agenda/Agenda_IPF_2019.pdf