My data vs. yours: Trust in official statistics in a digital age

Rajeswari Sengupta
IGIDR, Mumbai

28 October, 2019
Outline

1. Official data vs. privately collected data: some examples
2. What is going on?
3. Challenges for the NSOs
4. What can be done?
Part I

Divergence of official data from privately collected data: Examples
Example 1: New GDP series released in India in 2015

- The story being conveyed by the official data is that growth has been strong and reasonably steady, since 2013-14.
- Average growth rate between 2013-14 and 2018-19: 7.3%.
Official GDP data shows a certain story
2003-2010: 7.4%; 2011-2018: 6.8%

Annual growth rate of real GDP in India
Unofficial data showed that there had been a sharp slowdown since 2011.

Aggregating microdata is a good way to get an alternative picture.

Look at databases about companies and about investment projects (source: Center for Monitoring Indian Economy).
Privately collected data on corporate profit shows a different story

Real Profit Growth of Corporate Sector
(Annual avg., %)

Before tax
After tax

2002-11
2012-16
Private investment projects under implementation (all industries)

Jan–Mar 2019; 39.48

Rajeswari Sengupta, IGIDs, Mumbai

My data vs. yours: Trust in official statistics in a digital age
Example 2: Demonetisation in India in 2016

- In November, 2016 Indian government made 86% of cash in circulation illegal tender overnight, with new notes gradually introduced over the next several months.
- Official statistics show that in 2016-17 GDP grew at 8.2%.
Research using Night Lights data show that there was significant slowdown in economic activity in the months following demonetisation.

Data on commercial vehicle output, rail freight, service tax receipts and home appliance sales showed slowing growth or contraction.

Extensive survey of small and medium-sized enterprises showed that the cash crunch had severely hampered businesses.

The Chief Statistician of India, defended the official data.
‘Forecasting India’s GDP has become like predicting the English weather.....You never know when it will rain, when it will shine.'

- The message from GDP numbers did not tally with what we saw on the ground.
- Divergence between official statistics and other data created confusion about which data to trust.
- Adversely impacted the credibility of official statistics and of the national statistical agency.
- Potentially misled macro policy making.
In February 2019, Russia’s federal statistical agency Rosstat said that the country’s GDP grew by 2.3% in 2018 vs. 1.6% in 2017— the highest growth rate since 2012.

This came as a surprise. Nobody expected such agility from the Russian economy.

- The agency revised its data on construction growth from 0.5% to 5.3% for 2018.
- Did not tally with other proxies in the construction sector.
- No corresponding pick-up in physical indicators like cement production and transport.

Nigerians do not trust official figures. In 2018 in a tweet, the agriculture ministry quoted a global data aggregation website, Index Mundi, as proof that rice production had gone up by 19%.
Part II

What is going on?
In the pre-IT days, governments had monopoly on data and on dissemination of statistical information. Government provided data was considered the ‘holy grail’. Stakeholders would mostly know about the aggregate picture. It was too expensive for any private person to amass macro data on the economy. Everyone trusted the official statistics.
With computers and internet, it has become easy for private organisations to conduct surveys, collect data from micro units, aggregate and process the data, and make it publicly available.

Other national organisations also publish data on their websites, not just the NSOs.

Result is an information revolution in the field of data and statistics (including core macro indicators such as GDP, Inflation, Unemployment etc).

Official statistics no longer the only source of information about the economy.
Day 44:
Still stranded, with nothing but flat empty water as far as the eye can see.
New players in the data ecosystem

Proliferation of alternative data:
- Centre for Monitoring Indian Economy (CMIE) collects data on firms, and households and is able to measure unemployment, investment, and consumption expenditure.
- Stock exchanges and clearing houses provide data on financial market activity.
- Central banks publish data on banking sector, inflation expectations, consumer and business confidence surveys.
- Other ministries collect and publish sectoral data (railways, aviation, mining, agriculture etc).
- Industry associations (e.g. automobiles) publish data on sectoral performance.
- Satellite data (e.g. night lights data).
- Data obtained by scraping information from websites (e.g. One Billion Prices project at MIT).
Problem 2: Intrinsic issues with macro aggregates produced by NSOs

- GDP is calculated at market prices and assigns zero value to activities that do not cost anything.
- IT revolution has provided us with an abundance of goods and services at effectively zero price.
  - People do not need to buy watches, torches, alarm clocks, cameras, calculators, tv, books.
  - They can do all these things for free on their phones or tablets.
Just because they are priced at zero does not mean that consumers do not find them immensely valuable.

GDP data will capture the reductions in activity in outmoded industries (like the torch or watch industry).

It is unclear that GDP is capturing the consumer value (as opposed to the paid value) obtained from using the new technologies.

Let’s say in the old days people used to spend $2000 on a watch, camera, phone, torch, alarm clock, books, etc.

Now they spend $500 on a cellphone, which they use for very little money.

GDP data would suggest that living standards have fallen, even though they have improved.
People increasingly feel that aggregate macro data do not correlate well with their own life experiences.
Part III

Challenges in the digital age for NSOs
Data divergence and data privacy

Challenge 1: Divergence

- Whenever official data diverges from the data that people privately collect, users begin to wonder whether the official data is correct.
- Whenever aggregates such as GDP paint a picture of the economy that contradicts the narrative conveyed by other data (both private and public sources), users begin to raise questions about the reliability of GDP numbers.

Challenge 2: Data privacy

- Vast increase in volume of digital data (big data) produced by usage of smart phones, social media, web browsing, smart devices etc.
- Lines between private data and public data are getting blurred.
Challenge 3: Inherent complexities in a digital age

- Sitting in Paris, we can order books directly (i.e., online) from London, clothes from the US, and electronics straight from China.
- Our cost of living may have very little to do with the CPI, as measured by going to the neighbouring market.
- Measuring cost of living in this new world is incredibly difficult.
Part IV

What can be done?
Private agencies do not follow universal standards, are not regulated, and hence do not have standardised quality and sanity checks.

Statistical agencies follow an agreed international standard.

That makes them comparable across countries, and easy to understand.

Availability of big data does not obviate the need for statistical agencies.

How can the NSOs re-establish their relevance and credibility in a digital world?
What can the NSOs do?

1. Collaboration with the private sector
   - NSOs can collaborate with private data collecting agencies and devise better ways to measure macro aggregates using technology.
   - NSOs rely on samples, which may not be representative in a world where so many transactions are occurring online rather than in shops.
     - For example, using online data on retail prices to measure inflation on a high frequency basis as opposed to using a basket of goods and services to measure CPI inflation.
   - In collaborations, NSOs can act as a guarantee of data quality.
2. Transparency

- NSOs should make *more* data available, and not less.
  - Make all the underlying micro data (e.g. survey data, administrative data etc) available for research and analysis.
  - Be more transparent about the methodologies used, to demonstrate why their data is reliable and relevant.
  - Engage with the stakeholders to clear suspicion; communicate proactively.

3. Develop alternative measures of consumer welfare, and economic activity to supplement GDP, those that take into account innovations in a digital age.
That's all Folks!