Big data: new insights for economic policy

Gabriel Quirós-Romero,
International Monetary Fund

---

1 This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.
Big Data
New Insights for Economic Policy
Gabriel Quirós-Romero
Deputy Director, STA, IMF

BI-IFC/BIS INTERNATIONAL SEMINAR
BANK INDONESIA
Bali, 26 July 2018

Reproductions of this material, or any parts of it, should refer to the IMF Statistics Department as the source.
Outline

I. Background
II. No straightforward definition; administrative data (?)
III. Potential
IV. Challenges
V. Statistical implications domain by domain
VI. Dos and Don’ts of Big Data for statistics
I. Background

- IMF STA is researching big data as a new data source for statistics, beyond administrative data:
  - potential of big data to benefit macro-economic and financial statistics?
  - organizational, budgetary, and, in particular, methodological challenges that come with incorporating big data?
  - and strategic statistical implications for national and international organizations moving forward
II. No straightforward definition administrative data (?)

- Doug Laney, 2001
- Big data characterized by the “5Vs”
  - High-volume, high-velocity, high-variety
  - Veracity and volatility
- Big data (“found”) as ‘by-products’
  - Social networks
  - Business operations
  - The Internet of Things
III. Where, how is the Potential of Big data for **statistics**?

1. **Big data to answer “new questions” and produce new indicators**

2. **Big data to bridge time-lags of official statistics and support the forecasting of existing indicators**

3. **Big data as an innovative data source in the production of official statistics**

---

**Official Statistics**

Historical Time Series

**TIME LAG**
IV. Challenges

- **Data Quality**
  - quality assessments of indicators will be crucial to minimize governance, political, and reputational risks
  - statistical techniques and methodologies best practices are needed to specifically address *veracity* and *volatility*
  - big data for uncovering meaningful insights, trends, and sentiments may underlie different quality assessments compared to using big data in official statistics
    - continuation of consistent and harmonized historical time series is still needed
  - metadata are key to assess and interpret new data sources
IV. Challenges

Data Access

• Proprietary data held by the private sector
  ➢ Public-Private Partnerships that safeguard independence, privacy and confidentiality
• Data that companies own may evolve from a byproduct to becoming a major asset
• Regular licensing costs come in addition to substantial investments into processing and storage solutions
• Risk of volatility persists
• Best practices for building lasting relationships between data owners and data users are needed (UN Global Working Group)
IV. Challenges

- **New Skill Profiles**
  - Special career stream for data scientists

- **Multi-disciplinary project teams needed to make big data speak**
  - Experts from different professional backgrounds work together

- **IT implications**
  - Sharing of software codes and algorithms; open-source software; cloud-computing
V. Implications by Statistical Domain: **potential** (1/2)

<table>
<thead>
<tr>
<th>Data Origin+</th>
<th>Data Type</th>
<th>Data Source and Techniques</th>
<th>Potential Indicators Derived</th>
<th>Statistical Domains</th>
<th>What May be the Potential?*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Networks</strong></td>
<td></td>
<td>Google trends and search data</td>
<td>now-cast GDP now-cast unemployment consumer sentiment car and property sales</td>
<td>National accounts External sector statistics Financial Statistics Price statistics</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mobile phone system data (electronic money schemes, e.g. M-Pesa) Peer-to-peer transactions</td>
<td>financial inclusion indicators remittances, regional disposable income, consumption patterns poverty reduction SDG “Gender Equality” economic growth</td>
<td>National accounts External sector statistics Financial Statistics Price statistics</td>
<td>1,3</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Twitter tweets</strong></td>
<td>consumer confidence index border mobility, tourism, transitioning of migrants now-cast food prices sentiment and topic trend analysis</td>
<td>Mobility and urban statistics Price statistics Demographic and social statistics</td>
<td>1,2,3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Web-scraping of Facebook posts, Wikipedia articles</td>
<td>geopolitical risk indicators price changes civil protests/labor strikes and national security events consumer sentiment inclusive infrastructure for sustainable development</td>
<td>Price statistics National accounts Demographic and social statistics Labor Statistics</td>
<td>1,2</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Call Detail Record data</strong></td>
<td>SDGs indicators, travel/tourism, transport, migration</td>
<td>Mobility and urban statistics</td>
<td>1,3</td>
<td></td>
</tr>
<tr>
<td><strong>Traditional Business Systems</strong></td>
<td>Data produced by public agencies Administrative data</td>
<td>Taxation registers</td>
<td>consumer spending small business’ income nonresident businesses controlled by resident parent corporations business profiling flight reservation system</td>
<td>National accounts Price Statistics External sector statistics Labor statistics Tourism statistics Transportation statistics</td>
<td>2, 3</td>
</tr>
<tr>
<td></td>
<td>Data produced by businesses 2210. Commercial transactions</td>
<td><strong>SWIFT data</strong> on transaction quantities and financial market prices</td>
<td>multi-sourcing to derive population and housing census population structure global financial flows network concentration cross-border transactions export/import indicators</td>
<td>National accounts Demographic and Social Statistics</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>National accounts Price statistics External sector statistics</td>
<td>2, 3</td>
</tr>
</tbody>
</table>
### V. Implications by Statistical Domain: potential (2/2)

<table>
<thead>
<tr>
<th>Data Origin+</th>
<th>Data Type</th>
<th>Data Source and Techniques</th>
<th>Potential Indicators Derived</th>
<th>Statistical Domains</th>
<th>What May be the Potential?*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet of Things</td>
<td>Data from sensors</td>
<td>Web-scraping to collect price data from online retailers</td>
<td>withdrawal of correspondent banking relationships trade financing</td>
<td>Price statistics</td>
<td>2,3</td>
</tr>
<tr>
<td>(machine-generated data)</td>
<td></td>
<td>Web-scraping of business websites</td>
<td>daily inflation turning points in inflationary trends e-commerce index</td>
<td>Financial statistics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scanner data</td>
<td>enterprise profiling job vacancies</td>
<td>National accounts</td>
<td>2,3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prices and quantities</td>
<td>national and regional consumer prices household income and expenditure</td>
<td>Financial statistics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Credit card data</td>
<td>consumer spending growth trends of the retail sales</td>
<td>National accounts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GPS positioning/tracking data</td>
<td>travel services exports/imports trip duration</td>
<td>National accounts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Traffic/Road sensors</td>
<td>proxy of economic growth/health commuting time traffic intensity</td>
<td>National accounts</td>
<td>1,2,3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Satellite imagery</td>
<td>improved geographical localization of statistical units and assets spatial sampling frame for output measurement land use and geostatistical cartography crop planting area, land use and agricultural output population and asset location as proxy for SDG &quot;Gender Equality&quot;</td>
<td>National accounts Price statistics External sector statistics Demographic and social statistics Transport statistics Agricultural statistics Demographic and urban statistics</td>
<td>1,3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smart meters (energy consumption measures)</td>
<td>non-occupancy rates household consumption electricity supply and consumption price differentials household structure and size</td>
<td>Environmental and Energy statistics National Accounts Price statistics Demographic and Social Statistics Transportation Statistics Geo-Spatial Statistics Agricultural Statistics Rural and Population Statistics</td>
<td>1,2,3</td>
</tr>
</tbody>
</table>

*Indicates potential but without specific data sources.
VI. Dos and Don’ts of Big Data

- In connection to the respective statistical domains, a number of Dos and Don’ts from big data can be identified, which are unevenly distributed across statistical domains.
- The Dos and Don’ts are largely driven by the essence of big data: by-products of private technological and business models that capture behavior of consumers, corporates, banks, individuals or government agencies.
- Big data are particularly promising to enhance directly or indirectly statistics on transactions, less so on stocks.
VI. Dos and don’ts of big data

Dos

Big data, particularly promising at helping measure:

- “soft” information: sentiment, alerts, reactions...
- consumer behavior and patterns (e.g. Amazon, Google searches and ‘clicks’, social networks,...)
- Tourism and private consumption (e.g. roaming information, Google searches, credit cards, click-stream data, scanner ...)
- Financial flows (e.g. SWIFT, mobile phones, ...)
- Prices (scanner data,...)
- Job vacancies and labor skills (e.g. LinkedIn,...)
- Agricultural and construction (satellite images,..)
- big data provides granular, microdata
- .....
VI. Dos and Don’ts of Big Data

Don’ts

• Sample representativeness: bias towards more modern and dynamic economic activities and social behavior
• Big data less suited for stocks, i.e. total financial assets and liabilities of firms, households, government, non residents, both at micro and macro levels
• Revaluation and other volume changes, particularly important in monetary and financial statistics
• As by-product, long time-series based on big data are inexistent and will be fragile because instability from business and technological changes, discontinuity in data provision
• Privacy and confidentiality of personal, firm-level data
IMF Publication on Big Data

*Big Data: Potential, Challenges, and Statistical Implications*

C Hammer, D. Kostroch, G Quirós-Romero, and STA Group, 2017