

Singapore's Administrative-Data-First Approach to Business Statistics

Eighth Meeting of the UN Committee of Experts on
Business and Trade Statistics

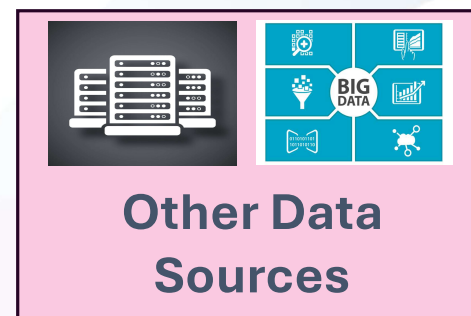
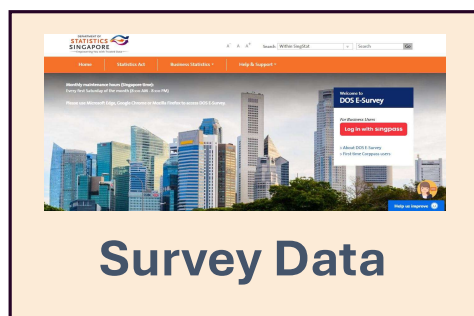
Emerging demand for business statistics

- Requires new data sources (beyond survey) and estimation methods to meet emerging demands:
 - More granular business data
 - New indicators
 - Business data beyond the standard industrial classification, e.g. firm archetypes, ecosystem value chains
 - Implementation of Factoryless Goods Producers (FGPs) classification

DOS's Administrative-Data-First strategy

- Leverage the wealth of information available in administrative data before tapping on survey methods
- Reduce respondent burden and data collection costs
- Support the compilation of disaggregated statistics and better able to meet emerging demands

Increasing use of administrative and other data sources in Business Statistics



- Use of administrative data in DOS's quarterly and annual estimates of the Services sector:
 - Annual: ~50% of the services industries have adopted register-based approach
 - Quarterly: ~80% of sampled firms are estimated using administrative data and not surveyed
- Use of Statistical Business Register (SBR) to support in-depth studies
- Leverage Artificial Intelligence/Machine Learning to access new data sources (e.g. unstructured data, web-based data) to enhance SBR

Challenges with Using Administrative and New Data Sources

- Differing coverage and granularity between different data sources
 - Topline items such as Revenue or Wages usually more complete in terms of coverage from admin data
 - Admin data on more granular income and business cost items may not be available
- Processing and analysis of large volume of unstructured data and Big data require new tools and methods

Estimation Methods/ Tools to Address Challenges

Artificial Intelligence and Machine Learning

- A **random forest machine learning algorithm** is used to predict missing values by training the algorithm for each data item based on firms which reported the data and its relationship with other data items (i.e. **model-based imputation**)
- **Artificial Intelligence tool** to extract and process unstructured data from financial statements

Other Estimation methods

- Use topline items as the control total and estimate the income/expense items using historical firm proportion or industry proportion
- If topline items not available, estimate using historical size (i.e. **historical imputation**) and nearest neighbour (i.e. **donor imputation**)

Establishing International Guidelines and Sharing of Country Practices

Key Learning Points

Important to assess the quality (e.g. definition, coverage) of data fields from administrative and new data sources before using for business statistics

Use of administrative and new data sources for business statistics may require new collection and estimation methods

Suggestions

Data quality assessment framework to guide multiple-source data compilation

Guidelines on data collection and estimation methodology

Thank You

Our Vision

National Statistical Service of Quality, Integrity, and Expertise

Our Mission

We Deliver Insightful Statistics and Trusted Statistical Services that Empower Decision Making



www.singstat.gov.sg



info@singstat.gov.sg



[@SingStat](https://t.me/SingStat)



[@singstat_dos](https://www.instagram.com/singstat_dos)



[@SingStatvideo](https://www.youtube.com/SingStatvideo)



[@sg-department-of-statistics](https://www.linkedin.com/company/sg-department-of-statistics)