



Technical session

Charl van Schoor – UNU-WIDER

Amina Ebrahim – UNU-WIDER

Research that Matters – Stakeholder workshop

15 November 2018



UNITED NATIONS
UNIVERSITY
UNU-WIDER

Outline

- Using Tax Administrative Data
 - CIT-IRP5 Panel (SARS-NT Panel)
 - Individual Panel
 - Research
 - completed, ongoing and future
-



Using Tax Administrative Data



UNITED NATIONS
UNIVERSITY
UNU-WIDER

What is tax admin data?

Think of it as a tool that can be used to further researcher's understanding of their field.

It's a tool with a lot of dimensions allowing for more granular research while having a wide variety of other benefits.

With it we have a new window into the mechanisms of the economy without most of the traditional statistical problems. It's part of the mechanism of the economy, and not an external observation of the economy.

Why is tax data important to researchers and analysts?

- Needed for production of national statistics
 - Evaluate public policy
 - Evidence based policy making
-

Advantages of tax data

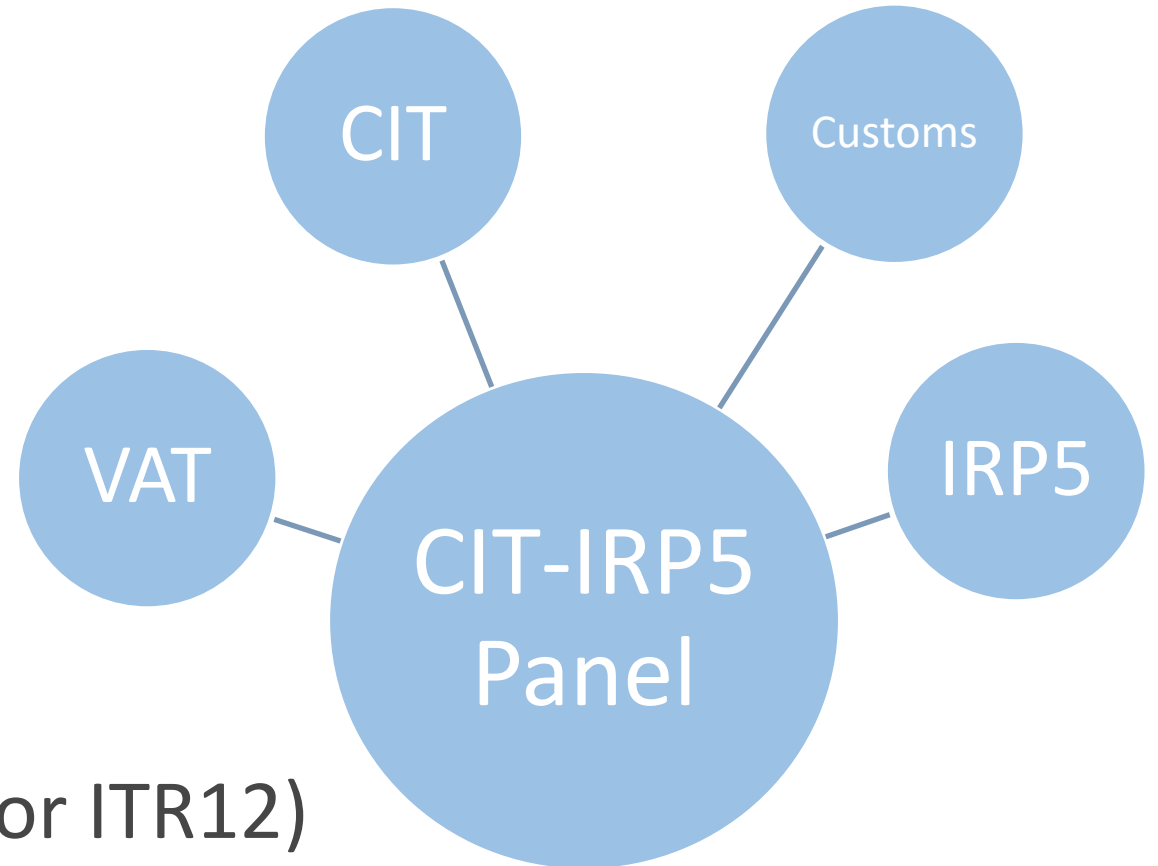
- Much larger sample sizes/full population
 - Higher dimensionality
 - Longitudinal in nature
 - More dependable, no attrition or non-response
 - Lower costs in gathering data
-

Disadvantages of tax data

- Not all topics are available in the data. Collected for non-statistical purposes
- Data quality
 - Incomplete
 - Out of date (e.g. address data)
 - Duplicates (tax form revisions)
 - Errors by firms and individuals

Data available at the NT-SDF

- Form Submissions to SARS
 - Company Income Tax
 - Value Added Tax
 - Customs Tax
 - Individual Tax (payroll/IRP5 or ITR12)



Dimensions of the data

- Firm Level
 - 6,426,931 observations by 912 variables (Expected to increase to ~10 mil observations with update)
 - 2008-2014 (2017 with update)

Update

- New extractions from SARS (up to 2017)
- Change in forms or additional fields
- Improved accuracy
- Documentation/Metadata
- Researcher reported issues

Transparency

- Do files and raw datasets are available to researchers
- There is a scale towards constructing a ready to go panel (complexity versus ease of use):
 - Variables are calculated to fit into firm level, but this means a lot of subjective decisions are made
 - To solve this, researchers can construct their own panels, or augment existing ones.

Using the NT-Secure Data Facility (SDF)

- Contains everything you need to start your research
 - Documentation (How to use it)
 - Data (Raw and cleaned)
 - Additional code
 - Papers written on the data



Individual Panel



UNITED NATIONS
UNIVERSITY
UNU-WIDER

Purpose of this dataset

Joins IRP5 and ITR12 information to get a full income distribution.

- The Tax Statistics publication by SARS and National Treasury covers information from the ITR12 returns (e.g. assessed returns through e-filing). Excellent information on top incomes.
- IRP5 information is from employer-provided returns. Covers mostly lower income employees/retirees and all those who don't file a return.

Merging these datasets provides the most complete view of the income distribution from tax returns

Benefits

- Advantages

- More information on each taxpayer, as combining datasets fills in potential gaps
- Source codes detail each type of income, not only employment related
- Simplified structure to aid analysis

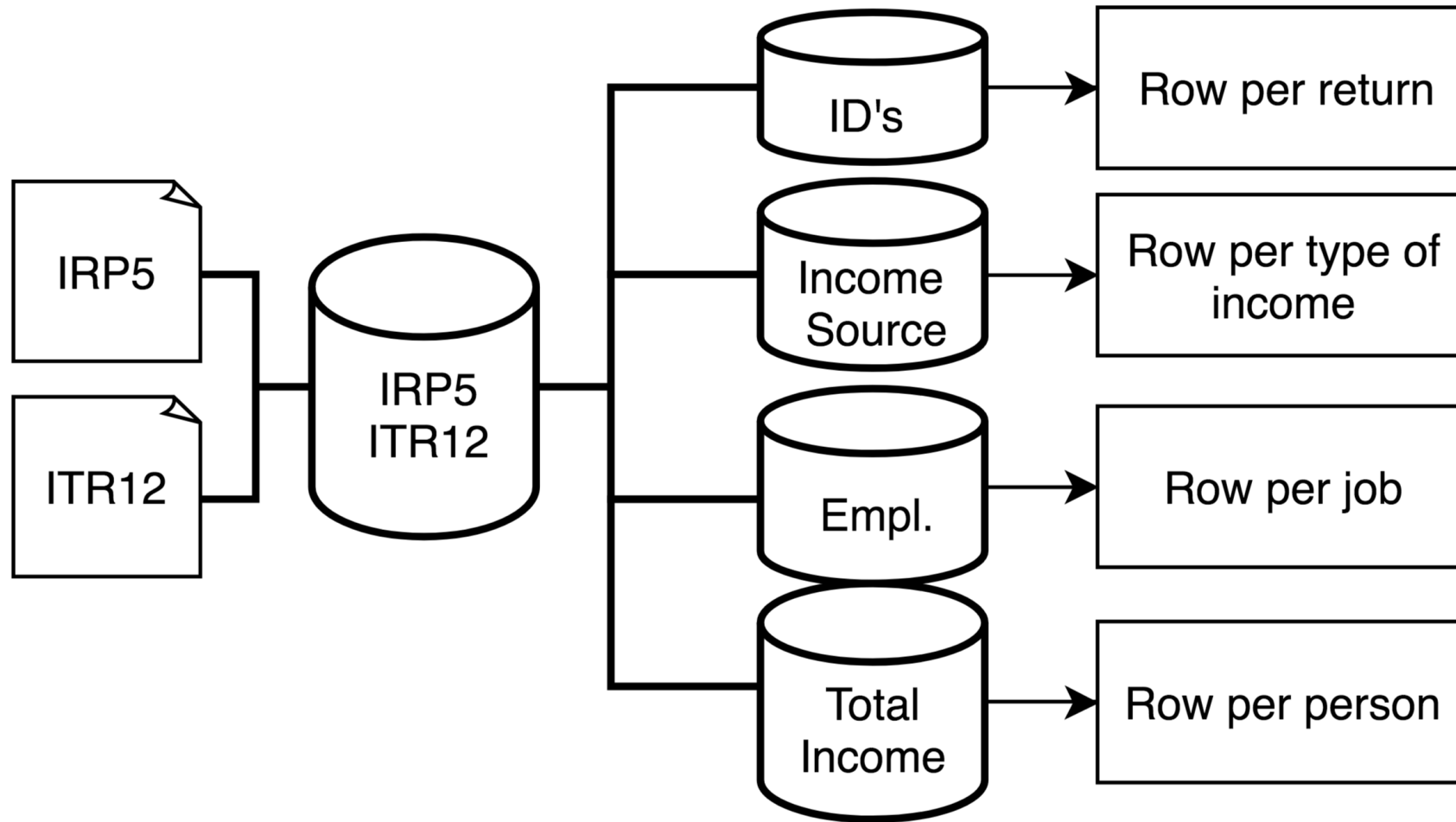
- Disadvantages

- Doesn't cover full income distribution, as informal sector is not covered
- Some types of income are excluded, such as dividends and interest/capital gain income below a threshold (where disclosure voluntary)

Structure of the dataset

- Joined ITR12 and IRP5
- Broken up again into 4 smaller, more usable, datasets with the same relevant information that can be linked by an identifier
- These datasets are:
 - ID's panel (One row per return – creation of a derived ID to link datasets)
 - Income source panel (One row per type of income – 450 million rows)
 - Employment panel (One row per job/income from institution)
 - Total income panel (One row per person)

Structure



Conclusion

- Use of tax data has a lot of potential for research
 - Clean, high-dimensional granular datasets
 - Gender, geospatial, firm and labour broker ID's
- Exciting new dataset that only a few people have ever accessed.



Tax Data research

- completed, ongoing and future



UNITED NATIONS
UNIVERSITY
UNU-WIDER

- Regional growth and development programme (2014 – 2016)
 - 20+ tax data research papers published
- SA-TIED (2017 – 2020)
 - Completed: 5 published papers
 - Ongoing: 15+ expected completion March 2019

Regional growth and development programme

Job flows, worker flows and churning in South Africa (Kerr, 2018)

“The main finding of the paper is that worker flows are substantial, around 53% per year, or 58% when employers classified as engaged in “public administration” are excluded.”

Markups and concentration in South African manufacturing sectors (Fedderke et al., 2018)

“We find both significant markups and significant concentration across most sectors.”

SA-TIED: Completed research - 2018

Innovation activity in South Africa: Measuring the returns to R&D (Steenkamp et al., 2018)

“We find that the return to R&D in South African manufacturing firms is high compared to OECD countries.”

On the persistence of growth for South African firms (Mamburu, 2018)

“It finds a strong, negative serial correlation of growth among South African firms, particularly among smaller firms and those at the tails of the growth distribution

Ongoing research

- Linking to trade: The role of the domestic suppliers network of exporters in South Africa
- Offshoring (and inshoring) within South African manufacturing firms: An analysis of the labour market effects
- The effects of innovator mobility
- Subsidized labour (ETI) and firms

Potential future research

Inequality

- Types of income distributions
- Medical tax credits
- Deductions
- Fringe benefits

Wealth

- Wealth distribution using the investment method

Potential future research

Public revenue mobilization

- Tax incidence
- Behavior of firms given changes in corporate taxation
- Impact of tax incentives on SME's
- Corrective taxation (environmental taxes)

Gender

- Worker flow and job churn
- Tenure and wage intensity of employment
- Gender wage gap
- Trade and wages
- Demand side determinants of female employment and wages

Conclusion

- Lots of great potential for use of administrative data
 - Documentation, cleaning, regular updates important
 - Geospatial, inequality, enterprise level research
- 60+ WIDER research projects planned over the next three years
- Tax data cannot answer all of our research questions. Proposals using other data sources are welcome



UNITED NATIONS
UNIVERSITY
UNU-WIDER

www.wider.unu.edu

Helsinki, Finland

Contact:

charlvanschoor1@gmail.com

amina@wider.unu.edu