



South African Revenue Service

The use of tax administrative data in research: a South African experience

Public Economics for Development, Maputo, July 2017

OUTLINE

Introduction –why tax administration data?

Behind the scenes: setting up the first research facility

Lessons learnt from project

Future possibilities

Concluding remarks – benefits of using tax administration data in research

OUTLINE

Introduction –why tax administration data?

Behind the scenes: setting up the first research facility

Lessons learnt from project

Future possibilities

Concluding remarks – benefits of using tax administration data in research

Tax administration data tells economic stories

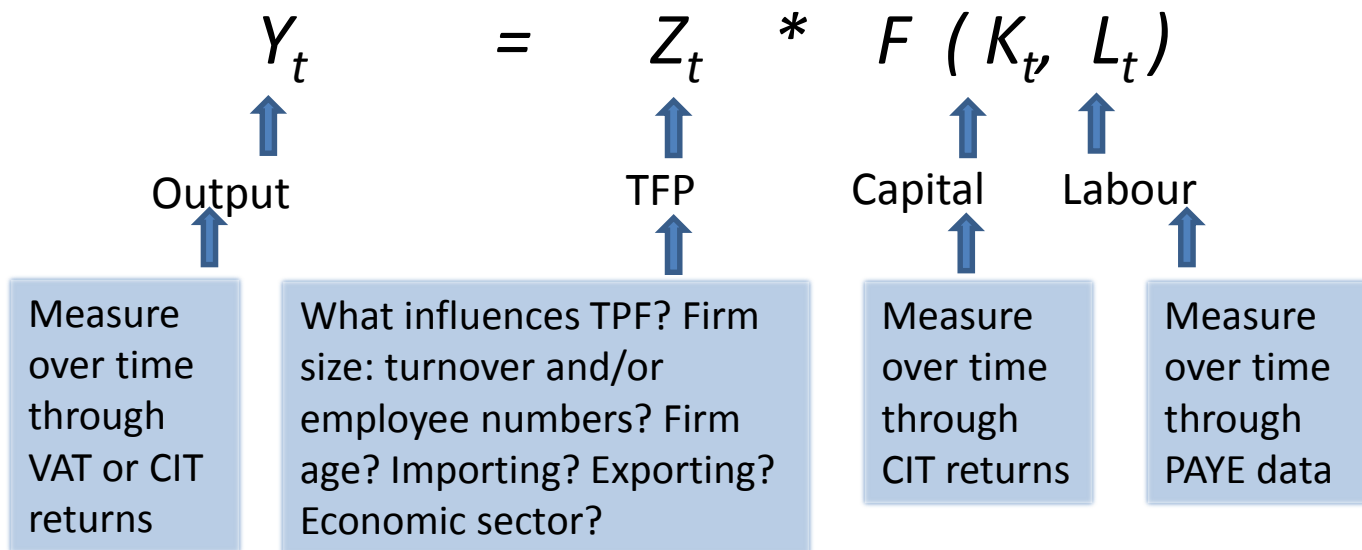
- Taxpayers *register* for taxes, providing information such as contact details
- Taxpayers *provide tax returns*, detailing economic activity – including international trade - relevant for determining taxation for particular periods
- Taxpayers *make payments and receive refunds*
- SARS also receives *other records* for the reconciliation of tax affairs, e.g. tax certificates (IRP5s), pension and retirement annuity contributions
- Information is received from both *businesses and individuals*
- Taxpayers are associated with identifiers (e.g. names, trading names, taxpayer reference numbers, IDs, company registration numbers), that enable their records to be linked *over time* and *across tax types* to yield elements of their ‘economic story’

Tax and customs administrative data available digitally in the South African Revenue Service (SARS)

- Taxpayer details (e.g. addresses) are maintained in tax registers
- Payments (provisional, assessment, penalties, interest) and refunds are recorded digitally
- Most tax returns are provided by taxpayers electronically, e.g.
 - Annual returns by businesses for Corporate Income Tax (CIT)
 - Annual returns by individuals for Personal Income Tax (PIT)
 - Bi-monthly or monthly returns by businesses for Value Added Tax (VAT)
 - Annual reconciliations per employee by employers for Pay-as-you-earn (PAYE) reflecting the earnings of all employees
- Exports and imports by commodity (using HS codes)

Example: Firm-level productivity

- The output of a 'firm' or company is related to its inputs – capital and labour, as well as its 'total factor productivity':
 - the growth in output of firm may change relative to the growth in inputs, and
 - the outputs of two firms may differ , with the same inputs, may differ
- 'Total factor productivity' (TFP) may change over time, and differ between firms



The importance of tax administration data *outside tax administration* is acknowledged in legislation

- The Income Tax Act was amended to enable Statistics South Africa to access Income Tax records: for well over a decade tax Stats SA has drawn samples for economic surveys from a Business Register compiled using tax administration data
- Section 70 of the Tax Administration Act provides for access to taxpayer information by *inter alia* National Treasury and Statistics South Africa for specified purposes
- Section 69 of the Tax Administration Act was amended to make explicit the possibility of using anonymised tax records for research

OUTLINE

Introduction –why tax administration data?

Behind the scenes: setting up the first research facility

Lessons learnt from project

Future possibilities

Concluding remarks – benefits of using tax administration data in research

National Treasury-SARS research project using tax administrative data for firm-level studies

- With support of UNU WIDER, infrastructure was established to enable access to a set of integrated, anonymised tax records (PAYE, CIT, VAT, Customs data)
- Data from 2009 onwards was linked by taxpayer across tax types and time
- Series of public calls for proposals, which were evaluated on the grounds of
 - technical soundness and feasibility,
 - as well as policy relevance
- Researchers provided access to data within a secure data facility/data laboratory

Maintaining confidentiality of taxpayer information was paramount

- Employed best practice in managing a secure data facility
- Multiple layers of protection of the confidentiality of taxpayer data:
 - ‘Obvious’ identifiers removed, such as names and trading names
 - Non-intelligent identifiers replaced ‘recognisable identifiers’, e.g. ID numbers, tax reference numbers
 - Researchers signed confidentiality agreements
 - The facility was physically secure and access controlled
 - Researchers were not able to remove data from the system
 - Results generated from analysis were checked to eliminate the risk of *indirect* identification of taxpayers before being released to researchers

Economic concepts and definitions – how to interpret tax administrative records?

- National Accounts (SNA 2008) definition of ‘firm’:
 - an enterprise is an economic agent having independent economic decision-making power, and whose aim is to produce market goods and services
 - A corporation is a form of enterprise having a legal identity separate from that of its owners

⇒ ***consider a CIT-registered entity to be a firm***
- International Labour Organisation (ILO) definition of employed individual:
 - ‘Persons who ... performed some work for wage or salary in cash or in kind...(or) were temporarily not at work during the reference period...’

⇒ ***Consider income source codes indicative of remuneration to identify employed individuals***

⇒ ***Use the dates on IRP5s to determine months/weeks/days of employment and hence calculate ‘full time equivalents’***

There are complexities to deal with in constructing the research dataset

- The 'reference period' for a CIT return (or set of VAT returns) might not coincide with the period covered by a tax certificate
- Form change over time, e.g. for CIT, ITR14s replaced IT14s
- Some variables may be aggregated/disaggregated from year to year: a 'standardised' database needs to be created

Firm ID	Property, plant and equipment (micro)	Property, plant and equipment (small)	Property (medium – large)	Plant and equipment (medium –large)	Other fixed assets (medium – large)
1	x				
2		y			
3			a	b	c

Raw ITR14 records



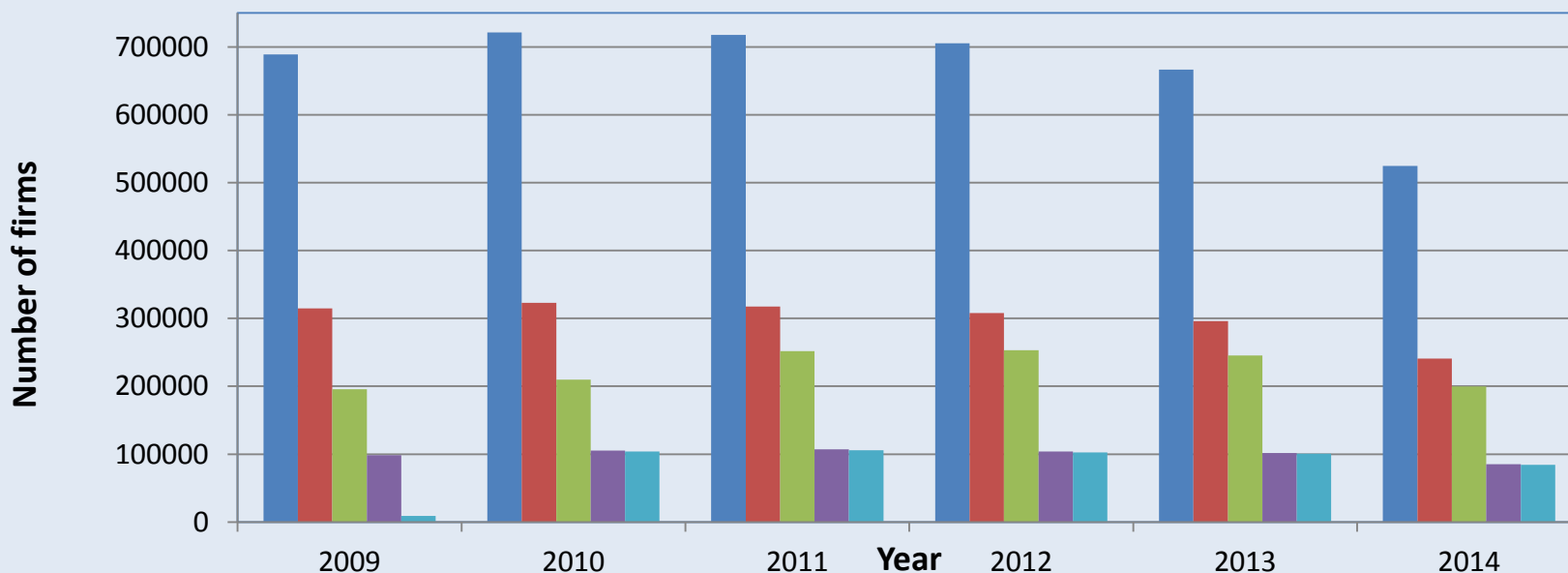
Harmonised ITR14 records

Firm ID	Property, plant and equipment	Other fixed assets
1	x	.
2	y	.
3	a + b	c

What does the CIT-based panel look like?

SARS-NT Firm-level Panel (2015)

- Firms in the SARS-NT Panel
- Firms with non-zero sales
- Firms with non-zero sales and fixed capital stock
- Firms with sales, capital stock and non-zero cost of sales
- Firms with capital stock, cost of sales and linked labour data



How does the panel constructed compare with Stats SA's Quarterly Financial Statistics?



Note: Stats SA's economic sample surveys draw on 'VAT-active' companies

OUTLINE

Introduction –why tax administration data?

Behind the scenes: setting up the first research facility

Lessons learnt from project

Future possibilities

Concluding remarks – benefits of using tax administration data in research

Research learnings for SARS

- Even though the data is not ‘perfect’, administrative records – especially returns - are extremely rich
- Firm-level studies can uncover useful (albeit approximate) relationships that can be used in formulating compliance risk rules
 - Broad relationships between characteristics, e.g. turnover and employees
 - Changes in characteristics over time, e.g. sales, employment
- There is value in looking at the same taxpayer *across tax types* to understand taxpayer behaviour and assess compliance, that could contribute towards tax gap quantification
- Productivity studies may be useful for modelling and forecasting CIT and VAT

Two strategic pillars to optimise research outputs and outcomes

1. Build and utilise a single research database comprising:
 - Well-documented, cleaned, integrated, anonymised tax and customs records, at business- and individual-level
 - Additional administrative and other data ('third party') to complement and/or validate tax and customs administrative data sources
2. Undertake research projects collaboratively with internal and external partners, to:
 - Ensure relevance of research outputs
 - Leverage external skills and expertise
 - Build research and analytical capacity within the public sector

The availability of firm-level data has the ability to influence the direction of economic research

- South Africa has internationally respected poverty and inequality researchers, in part due to the *availability of data*, particularly post 1994
- But, prior to the SARS-National Treasury project, firm-level survey data were rare

“A major gap in the empirical research in South Africa has been in the area of firm-level research. The constraint has been the availability of data. Already we see the benefits arising from the provision of the SARS administered data. Some old insights using industry data have been confirmed. Others refuted. But in each case the foundation of these conclusions is much stronger as it is based on unit-level information.

The provision of the data has also permitted much more detailed and rigorous empirical analysis of tax related issues than previously. This will feed directly into policy design and implementation. It also opens up prospects for serious research, which will make public finance a more attractive field for young economists to engage in. This can only enhance public policy analysis in South Africa.”

- Statement from Economics Professors at five South African universities

OUTLINE

Introduction –why tax administration data?

Behind the scenes: setting up the first research facility

Lessons learnt from project

Future possibilities

Concluding remarks – benefits of using tax administration data in research

There is a demand for the analysis of tax and customs administrative records in research in many areas

Tax policy and administration:
e.g. incidence of tax proposals on taxpayers; likely uptake of incentives; tax gap due to policy and compliance

Economic policy:
e.g. impact of tax incentives on growth, productivity or employment

Wealth and income distributions and dynamics:
e.g. determinants of remuneration; spatial inequalities

Business dynamics:
e.g. firm-level studies as piloted with Treasury; firm survival studies and the impact of firm survival on the tax base

The geography in tax data may enable the provision of lower level estimates of GDP

- Research questions:
 - Can we use tax certificates to disaggregate the Compensation of Employees to actual production locations of multi-location businesses?
 - If yes, what biases will be present if we make the assumption that the Gross Operating Surplus of a business is split proportionally like its Compensation of Employees?
 - If these biases are not ‘show stoppers’, can we use tax administration data to provide credible sub-provincial estimates of value added, i.e. ‘sub-regional GDP’?

The possibilities grow when tax and customs administrative data can be linked to other data sources

- For example:
- Linking qualifications from the South African Qualifications Authority data base, to understand how qualifications influence earnings and income
- Tracking beneficiaries of a youth job creation programme (Harambee) funded through by government
- Understanding intergenerational earning patterns through links with the population register
- Matching house transfer data to capital gains entries in CIT and PIT records

OUTLINE

Introduction –why tax administration data?

Behind the scenes: setting up the first research facility

Lessons learnt from project

Future possibilities

Concluding remarks – benefits of using tax administration data in research

Concluding remarks: there are wide-ranging benefits from research based on administrative data

Some benefits:

- Provides an evidence base for tax, fiscal and economic policy impact analysis and policy formulation, in line with the creation of a capable, developmental State;
- Enables continuous improvement of the quality of administrative data, based on insights gained from deep interrogation of tax and customs records;
- Achieves cost saving to the State through minimising the need for conducting costly surveys, and increases effectiveness through streamlining access to data required by researchers in the public sector;
- Builds research and analytical capacity within the public sector; and
- Builds the policy research capacity of South African researchers in academia and other research institutions.



South African Revenue Service

Thank you