



UNITED NATIONS  
UNIVERSITY  
**UNU-WIDER**

# Firm data inventory for South Africa

In this document a number of sources of South African firm level data from Statistics South Africa, the South African Revenue Services, the Companies and Intellectual Property Commission, the World Bank, the Human Sciences Research Council and the Department of Trade and Industry are discussed. Information is provided about variables/questions included, sampling frame, sampling methods, response rates (where applicable), methods of enumeration/collection, the availability of the data and the potential to link the different sources of data discussed.

## Statistics South Africa (StatsSA)

StatsSA conducts enterprise-level surveys which are drawn from a sampling frame called the business sampling frame<sup>1</sup>. The business sampling frame is based on company income tax (CIT), value-added tax (VAT) and Pay as you Earn (PAYE) tax databases obtained from the South African Revenue Services (SARS). The surveys undertaken include the Large Sample surveys (LSS), conducted in each 1 digit industrial sector every three to five years, the Annual Financial Statistics surveys (AFS)<sup>2</sup>, the Quarterly Financial Surveys (QFS) and the Quarterly Employment Statistics (QES) surveys. Thus these surveys only cover firms registered for VAT. Informal firms are excluded because by definition they are not VAT registered.

The surveys generally follow the same sampling methods, although there are small differences between the surveys. Samples are drawn using stratified simple random sampling with the enterprises first stratified by predominant activity according to the SIC and then by size of enterprise. The four size groups used for stratification are modifications of the Department of Trade and Industry's 2003 size categories that vary by SIC and in most cases the firms in the largest size group are completely enumerated. The LSS sample size ranges from 100 enterprises for the electricity, gas and water supply industry, to 10 000 enterprises for the manufacturing industry. The AFS, QFS and QES have a sample size of 13 000, 5 000, and 20 000 respectively. The average response rate for these surveys is over 80%<sup>3</sup>. The QES has been conducted since 2004, the QFS since 2003, the AFS since 2006 and the LSSs since 2001.

The AFS, QFS and LSS collect information on selected income and expenditure items and the consolidated balance sheet (capital expenditure on new and existing assets and the book value of fixed and intangible assets). While the AFS and the QFS focus on financial statistics, the LSS aims to detail "products", i.e. products sold and purchased as well as sample on a more detailed industry level. The QES collects information regarding the number of employees and total wage bill in each firm<sup>4</sup>. The questionnaires are sent to enterprises through post, fax or email and personal visits and telephonic follow up are used. The data has not been made publicly available.

StatsSA data sources can be linked together using the enterprise number created by StatsSA. A cross sectional survey such as the LSS can be turned into a panel because of the stratification and the effective census of firms in the largest size category. However this panel will have a disproportionately high number of large firms. When matching firms across different StatsSA surveys large firms are more likely to be sampled in two different surveys and therefore the result will again be a matched dataset with a disproportionate number of large firms. It should be noted that the QES and QFS are designed as panels- a sample is drawn and all firms are surveyed for a number of quarters in the QES (14 between 2006-2010, 5 between 2005 and 2006) and 4 quarters in the QFS before a new sample is drawn.

Agriculture is excluded from the above surveys. However, StatsSA also conducts annual agricultural surveys and the five-yearly agricultural census. Time series data (based on the surveys from 2005-2012 as well as 2002 and 2007 censuses) is available and includes information on income generated in crop and livestock production, employment, current and capital expenditure and farming debt in the agricultural sector<sup>5</sup>. The sample frame includes

---

1 No surveys at a lower level, for example plant level, are undertaken by Statistics South Africa.

2 Previously called the Economic Activity Survey (EAS) which was discontinued and incorporated into the AFS from 2006 onwards.

3 See the relevant LSS, AFS, QFS and QES reports in the appendix.

4 Further details can be found in the Appendix.

5 A 2012 census was proposed but not conducted.

enterprises which are mainly engaged in farming and farm related services and activities and which are also registered for Value Added Tax (VAT). The sample for the surveys consisted of between 6 800 and 15 000 enterprises (depending on the year of the sample) and followed similar sampling methods to the surveys described above. The average survey and census response rates are 75% and 85% respectively. The questionnaires are sent to enterprises through post, fax or email and personal visits and telephonic follow up are used. The data is not publicly available.

Unregistered, informal firms are not included in any of the above surveys. StatsSA has conducted a survey of informal firms using the household based Labour Force Survey as a sampling frame every four years since 2001, called the Survey of Employers and the Self-Employed. The microdata from the first 3 of these surveys is publicly available and can be downloaded from DataFirst<sup>6</sup>. The surveys collect information about costs and revenues, employment and a number of other issues but capital stock is not measured.

## The South African Revenue Service (SARS)

The company income tax (CIT) database contains information on businesses in South Africa. All companies and closed corporations are required to register as provisional tax payers. Since 2011 companies are also required to submit supporting annual financial statements (income statement and balance sheet) to SARS with their tax return. The data is collected from the annual tax returns (ITR14 form submitted online or at a SARS branch). This data is not publicly available but aggregate data on CIT is published annually in the Tax Statistics Bulletin (see Appendix for more details).

Pay As you Earn (PAYE) registered employers must submit tax certificates for all employees regardless of tax liability<sup>7</sup>. These certificates can then be linked to the firm issuing them and thus provide information about the earnings of each employee within each firm. This data is not publicly available.

In addition to PIT and CIT, SARS collects valued-added tax (VAT). VAT is levied at a standard rate of 14 percent by registered vendors on most goods and services. It is compulsory for businesses that earned an income in excess of R1million in any consecutive 12-month period to register as a VAT vendor and smaller businesses may also register. The data is based on the VAT returns submitted by vendors, and includes the value of goods and services supplied to and by each firm and the value added reported by each firm. The data is not publicly available but aggregate data is published annually in the Tax Statistics Bulletin.

SARS also collects trade data. The import data collected includes information on importing company characteristics, HS code, units, destinations, value, origin, tariff rate applied, and exemptions. Export data is also collected and from this data (and the data of 45 other countries) the World Bank has constructed the Exporter Dynamics database covering basic characteristics of exporters (such as their number, size and growth), their concentration and degree of diversification in products and markets, their dynamics in terms of entry, exit and survival, and the average unit prices of the products they trade. This data is publicly available from the World Bank website in aggregated form and for South Africa is available from 2001-2009.

In theory SARS returns data are available from 1995. However electronic submissions only became prevalent in the last five years and this is when SARS believes the data became more reliable- both in terms of completeness and accuracy. There have also been more recent changes in the data that is collected. Thus SARS suggests that VAT return data should be used only from 2010 onwards, the tax certificate data from 2009 onwards and the CIT data from 2010 onwards (bearing in mind that CIT returns changed substantially in March 2013).

---

<sup>6</sup> Further details can be found in the Appendix.

<sup>7</sup> Before tax year 2011 only those individual tax payers earning above the tax threshold were issued certificates. Those with a taxable income below roughly R60 000 per annum – were not required to register with SARS and thus were not issued tax certificates. Since 2011 almost all employees are issued tax certificates.

## The Companies and Intellectual Property Commission (CIPC)<sup>8</sup>

To register as a company in South Africa a company is first required to register with the CIPC. The CIPC Business Registration database contains data for over 3 million enterprises and provides information on the company name, the physical and postal address, the financial year end, date of registration, unique enterprise registration number and company status (active, inactive, deregistered). In addition, some information is available on directors, shareholders and auditors. The Business Returns database covers 730 000 enterprises and provides information on enterprise turnover by unique enterprise registration number. Only around 55 percent of those enterprises have provided at least one positive turnover value to CIPC for the period covered, which is 2003 to 2011. When positive turnover is recorded observations for many companies are only available over intermittent years. Neither database is publicly available.

## World Bank Enterprise Survey

The 2007 World Bank Enterprise Survey targeted establishments based in Johannesburg, Cape Town, Durban and Port Elizabeth. A random sample of 937 establishments was drawn with stratification on firm size, business sector and geographic region for 2007. The data have been made into a panel by linking to data collected from establishments surveyed in the 2003 Investment Climate Survey (ICS) of South Africa<sup>9</sup>. A total potential of 579 firms were identified from the 2003 survey, of which 231 were successfully surveyed in 2007. The ICS data is not separately available but is incorporated into the enterprise survey database. The enterprise surveys were conducted in face-to-face interviews with top managers and business owners. Firm level data and all relevant documentation are publicly available at no charge.

## HSRC Research and Development (R&D) Surveys

The HSRC National Surveys of Research and Experimental Development provide information on research and development (R&D) and were conducted in 1991, 2002, and annually since 2004. The sectors included in the sample frame are business enterprises, government, higher education institutions, not-for-profit organisations, and science councils. The surveys cover in-house R&D expenditure, supply of R&D personnel, as well as the value of outsourced R&D. The sample size is around 240 enterprises for the 2009/10 surveys. The previous surveys had a sample size of around 840 enterprises. Sampling methods differed by sector (see the Appendix for more details). The overall questionnaire response rate for the 2009/10 survey was only about 24%. After imputations were made from secondary sources such as previous R&D surveys and other government databases the survey response rate was reported to be around 75%. The data is not publicly available.

## Department of Trade and Industry

The DTI collects firm data under a number of different projects. Research on “Large, Lead and Dynamic firms” involves data collection on firms in the metals, mining capital and transport equipment sectors. Each chief directorate also collects data for firms relevant to its work. DTI described these data as “uneven,” with some sectors (for example auto manufacturing) having very good data and others have poor data. Neither source of data is based on a representative sample and the research was described as qualitative. These data are not publicly available.

---

<sup>8</sup> The CIPC was established in May 2011 through the amalgamation of the Companies and Property Registration Office (CIPRO) and the Office of Companies and Intellectual Property Enforcement (OCIPE) of the Department of Trade and Industry.

<sup>9</sup> See South Africa 2007 Implementation Notes downloadable from the enterprise survey website. For the ICS survey and report see the appendix.

## Linking data sources from different data producers

Since the Statistics South Africa business sampling frame is based on the SARS tax data any firm surveyed in a Statistics South Africa survey can be linked to its data held by SARS. This has been undertaken by StatsSA in the past but no official documents have been published. Most, but not all, tax paying entities with data held by SARS are included on the StatsSA business sampling frame and thus most of the tax paying entities can be linked to StatsSA data provided that they were sampled. Many firms, particularly small firms, would never be sampled (although represented in the final estimates) and thus information on these firms held by SARS could not be matched to any data collected by StatsSA.

Linking the DTI, HSRC, CIPC or World Bank data to SARS or Stats SA data would be more difficult, although manual matching may be possible. Difficulties in the proliferation of firm identifiers across departments lead to a business registration reform project that was initiated between the DTI, SARS, StatsSA and the National Treasury, with the objective to create a single business registration point with a single firm identifier. The most recent reports indicate that this project has been suspended after progress had been slow and a goal to complete this project in 2013 was not achieved<sup>10</sup>.

---

<sup>10</sup> See the Stats SA 2012 and 2013 work programme documents, page 26 and 17 respectively (the web links are in the appendix).

## Appendix: Reports, surveys, forms and contact details

### StatsSA

#### The Large Sample Surveys (LSS)

The following LSS reports can be downloaded from the StatsSA website at <http://beta2.statssa.gov.za/?s=large+sample+survey&sitem=publications>:

- Manufacturing (P3002) for the years 2001, 2005, 2008 and 2011
- Wholesale Trade (P6101) for the years 2005, 2009 and 2012
- Retail Trade (P6102) for the years 2005, 2009 and 2012
- Motor Trade (P6301) for the years 2002, 2009 and 2012
- Accommodation (P6411) for the years 2004, 2007, 2009 and 2012
- Food and beverages (P6421) for the years 2007, 2009 and 2012
- Construction (P5002) for the years 2007 and 2011
- Electricity, gas and water supply (41-01-02) for the years 2006 and 2010
- Real estate and business services (P8004) for the years 2003, 2006 and 2010
- Transport, post and telecommunication (P7000) for 2006 and 2010

The following LSS questionnaires are available for download:

- Construction - [http://beta2.statssa.gov.za/questionnaires/Large\\_sample\\_survey\\_construction.pdf](http://beta2.statssa.gov.za/questionnaires/Large_sample_survey_construction.pdf)
- Accommodation - [http://beta2.statssa.gov.za/questionnaires/Large\\_sample\\_survey\\_accommodation.pdf](http://beta2.statssa.gov.za/questionnaires/Large_sample_survey_accommodation.pdf)
- Personal services - [http://beta2.statssa.gov.za/questionnaires/Large\\_sample\\_survey\\_personal\\_services.pdf](http://beta2.statssa.gov.za/questionnaires/Large_sample_survey_personal_services.pdf)
- Mining - [http://beta2.statssa.gov.za/questionnaires/Large\\_sample\\_survey\\_mining.pdf](http://beta2.statssa.gov.za/questionnaires/Large_sample_survey_mining.pdf)

#### Annual Financial Statistics Surveys (AFS)

- Reports (P0021) for the years 2006-2012 are available for download from: [http://beta2.statssa.gov.za/?page\\_id=1866&PPN=P0021&SCH=5692](http://beta2.statssa.gov.za/?page_id=1866&PPN=P0021&SCH=5692)
- The AFS questionnaire is available from: [http://beta2.statssa.gov.za/questionnaires/AFS2013\\_P0021-Eng\\_Questionnaire.pdf](http://beta2.statssa.gov.za/questionnaires/AFS2013_P0021-Eng_Questionnaire.pdf)

The following EAS reports (P8001 and P0020) are available for download:

- For 2001-2003: [http://beta2.statssa.gov.za/?page\\_id=1866&PPN=P8001&SCH=3225](http://beta2.statssa.gov.za/?page_id=1866&PPN=P8001&SCH=3225)
- For 2004 and 2005: [http://beta2.statssa.gov.za/?page\\_id=1866&PPN=P0020&SCH=3742](http://beta2.statssa.gov.za/?page_id=1866&PPN=P0020&SCH=3742)

#### Quarterly Financial Statistics Surveys (QFS)

- Reports (P8042 and P0044) for Q12003-Q4 2004 are available for download from: [http://beta2.statssa.gov.za/?page\\_id=1866&PPN=P8042&SCH=2856](http://beta2.statssa.gov.za/?page_id=1866&PPN=P8042&SCH=2856)
- Reports (P8042 and P0044) for Q12005-Q32013 are available for download from: [http://beta2.statssa.gov.za/?page\\_id=1866&PPN=P0044&SCH=5714](http://beta2.statssa.gov.za/?page_id=1866&PPN=P0044&SCH=5714)
- The QFS questionnaire is available from: <http://beta2.statssa.gov.za/questionnaires/QFS2005.pdf>
- A discussion of sampling and the panel nature of the QFS is available from: <http://www.statssa.gov.za/articles/16%20QFS%20article%20Dec%202012.pdf>

## Quarterly Employment Statistics (QES)

- Reports (Po277.1 and Po277) for Q12003 - Q32013 are available for download from the StatsSA website: [http://beta2.statssa.gov.za/?page\\_id=1866&PPN=Po277&SCH=3680&page=1](http://beta2.statssa.gov.za/?page_id=1866&PPN=Po277&SCH=3680&page=1)
- An academic paper discussing the QES data, panel and sampling is available at <http://onlinelibrary.wiley.com/doi/10.1111/saje.12031/abstract>
- An example of the questionnaire (from December 2006) is available from: [http://beta2.statssa.gov.za/questionnaires/QES\\_MAR\\_06.pdf](http://beta2.statssa.gov.za/questionnaires/QES_MAR_06.pdf)

The Agricultural Surveys and Census reports are available from <http://www.statssa.gov.za/agriculture/default.asp>

Stats SA 2012 and 2013 Work Programme documents that discuss the process of creating a common firm id across different departments are available at [http://www.statssa.gov.za/about\\_statssa/work\\_programme\\_2012.pdf](http://www.statssa.gov.za/about_statssa/work_programme_2012.pdf) and [http://www.statssa.gov.za/about\\_statssa/work\\_programme\\_2013.pdf](http://www.statssa.gov.za/about_statssa/work_programme_2013.pdf)

The Survey of Employers and Self-Employed data, documentation and publications for 2001, 2005 and 2009 can be downloaded from DataFirst:

- <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/413>
- <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/223>
- <http://www.datafirst.uct.ac.za/dataportal/index.php/catalog/180>

## SARS

Tax forms and guides to filling out forms are found on the SARS e-filing website: [http://www.sarsefiling.co.za/Forms\\_Guides.aspx](http://www.sarsefiling.co.za/Forms_Guides.aspx)

- For Company Income tax: <http://online.sars.gov.za/AllDocs/OpsDocs/Guides/IT-ELEC-03-G01%20-%20How%20to%20complete%20the%20company%20Income%20Tax%20return%20ITR14%20eFiling%20-%20External%20Guide.pdf>
- PAYE employee guide: <http://online.sars.gov.za/AllDocs/OpsDocs/Guides/PAYE-REG-03-G01%20-%20Registration%20of%20Employees%20for%20Income%20Tax%20Purposes%20-%20External%20Guide.pdf>

The Tax Statistics reports for 2008-2013 are available from the SARS website: <http://www.sars.gov.za/About/SATaxSystem/Pages/Tax-Statistics.aspx>

Information on SARS export data is available from the World Bank Website:

- Reports, questionnaires and methodology can be found at <http://microdata.worldbank.org/index.php/catalog/1031/study-description>

Examples of a VAT submission form and an employee tax certificate form can be found at the end of this document.

## CIPC

- Annual CIPRO reports from 2005 to 2011 are available from: [http://www.cipc.co.za/AnnualReports\\_CIPRO.aspx](http://www.cipc.co.za/AnnualReports_CIPRO.aspx)
- The 2012/2013 annual CIPC report and strategic documents are available from: <http://www.cipc.co.za/AnnualReports.aspx>
- An academic paper which uses and discusses the CIPC business returns data is available at: [www.essa2013.org.za/fullpaper/essa2013\\_2594.pdf](http://www.essa2013.org.za/fullpaper/essa2013_2594.pdf)

## World Bank Enterprise Survey

- Data for South African enterprises is available for the year 2007.
- The current survey instruments and manuals used in the enterprise surveys are available from: <http://www.enterprisesurveys.org/Methodology>
- Raw individual country datasets, aggregated datasets (across countries and years), panel datasets, and all relevant survey documentation are publicly available from the following site: <https://www.enterprisesurveys.org/Portal/Login.aspx?ReturnUrl=%2fportal%2felibrary.aspx%3flibid%3d14&libid=14>
- The ICS survey and report can be downloaded from: <https://openknowledge.worldbank.org/bitstream/handle/10986/6738/399900REPLACEM101OFFICIALoUSEoONLY1.pdf?sequence=1>

## HSRC R&D Surveys

The reports and questionnaires can be downloaded from the following sites:

- The Department of Science and Technology: <http://www.dst.gov.za/index.php/resource-center/rad-reports>, or
- The Human Sciences Research Council website: [http://www.hsrc.ac.za/en/research-areas/Research\\_Areas\\_PPHSI/CeSTii/previous-survey-reports](http://www.hsrc.ac.za/en/research-areas/Research_Areas_PPHSI/CeSTii/previous-survey-reports)

## Contact Details for Data Sources:

StatsSA (AFS, QFS, LSS):

Contact person: Joe de Beer

Email: [joedb@statssa.gov.za](mailto:joedb@statssa.gov.za)

StatsSA (QES):

Contact person: Peter Buwembo

Email: [PeterB@statssa.gov.za](mailto:PeterB@statssa.gov.za)

StatsSA (SESE):

Contact person: Malerato Mosiane

Email: [MaleratoM@statssa.gov.za](mailto:MaleratoM@statssa.gov.za)

SARS:

Contact person: Liz Gavin

Email: [EGavin@sars.gov.za](mailto:EGavin@sars.gov.za)

Tel number:

World Bank Exporter Data:

Contact person: Liz Gavin (SARS)

Email: [EGavin@sars.gov.za](mailto:EGavin@sars.gov.za)

World Bank Enterprise Survey Data:

Data is easily accessible on their website.

HSRC:

Contact person: HSRC front desk

Email: [info@hsrc.ac.za](mailto:info@hsrc.ac.za)

Tel number: +2712 302 2000

OR

Contact person: Anneke Jordaan

Email: [ajordaan@hsrc.ac.za](mailto:ajordaan@hsrc.ac.za)

Tel number: +2712 302 2340