ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Ebrahim, Amina; Lilenstein, Kezia

Working Paper

Gender and the South African labour market: Policy relevant research possibilities using South African tax data

WIDER Working Paper, No. 2019/31

Provided in Cooperation with:
United Nations University (UNU), World Institute for Development Economics Research (WIDER)

Suggested Citation: Ebrahim, Amina; Lilenstein, Kezia (2019) : Gender and the South African labour market: Policy relevant research possibilities using South African tax data, WIDER Working Paper, No. 2019/31, ISBN 978-92-9256-665-4, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki, https://doi.org/10.35188/UNU-WIDER/2019/665-4

This Version is available at:
https://hdl.handle.net/10419/211261

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2019/31

Gender and the South African labour market

Policy relevant research possibilities using South African tax data

Amina Ebrahim[1] and Kezia Lilenstein[2]

April 2019

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: This paper has three primary objectives. First, it provides a broad overview of the available South African data used to examine gender in the labour market, with specific reference to the South African Revenue Services (SARS) tax data which has recently been made available. Second, the paper provides a brief overview of the existing South African literature, with a focus on the gender dynamics of the labour market. Last, this paper identifies the areas in which the availability of the SARS data has made new research possible, specifically in the gender dynamics of the South African labour market. We identify six key research areas that can immediately be undertaken using the tax data.

Keywords: administrative data, employment, gender, income inequality, wages
JEL classification: J01, E24

Acknowledgements: The authors gratefully acknowledge support from UNU-WIDER. The recommendations from this paper provided context to the request for research proposals under the Inequality and Public Revenue work streams of the 'Southern Africa—Towards Inclusive Economic Development' project.

[1]UNU-WIDER, Pretoria, South Africa; [2]DPRU, Cape Town, South Africa. Corresponding author: kezia.lilenstein@uct.ac.za.

This study has been prepared within the UNU-WIDER project on 'Southern Africa—Towards Inclusive Economic Development (SA-TIED)'.

Copyright © UNU-WIDER 2019

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-665-4

Typescript prepared by Lesley Ellen.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

South Africa has a rich history of publicly available survey data which has been used extensively to study the labour market and, specifically, its gender dynamics (Mosomi 2019). However, a paucity of firm-level data, and in particular matched employee–employer data, contributes to a number of gaps in the existing South African literature.

The availability of administrative tax microdata (henceforth tax data) has changed the status quo. In 2015, National Treasury (NT), South African Revenue Services (SARS), and UNU-WIDER worked to make tax data available for research in a secure facility in Pretoria. This data has opened new avenues for research across the field of economics. In 2018, SARS derived a gender variable using the South African identity number which is anonymized at SARS before it enters the secure data facility. The addition of the gender variable in the context of the labour market analysis, for the universe of the formally employed, adds a new dimension to the data.

The paucity of matched employer–employee data in South Africa to date has led to labour market research topics such as seasonality of employment, firm-level determinants of wages, worker flows, job churn, tenure, and trade to be underexplored generally. Furthermore, there are few South African studies exploring how gender specifically contributes to our understanding of these matters.

Internationally, administrative data has been used to fill some of these research gaps: in the United States, administrative data has been used to examine hires and separations over time; in the Netherlands, administrative data is used to explain job mobility; and in Spain, administrative data has been used to examine gender wage gaps in the public sector. It is also clear from the international literature that the use of administrative data allows for deeper analysis of the labour market in general and also its gender dynamics. In this paper, we aim to analyse the gaps in the existing literature on gender dynamics in South Africa for areas in which the recently available administrative tax data (Pieterse et al. 2018) can be used to contribute to our knowledge on the experience of women in the labour market.

In Section 2, we provide a brief overview of the strengths and weaknesses of the prominent South African data, as well as an overview of the South African literature relevant to this scoping paper. Section 3 provides an overview of the (SARS) data, the available gender component, and the limitations specific to the tax data. Section 4 highlights areas for future research which have arisen from the availability of the SARS tax data.

2 South African data and literature

The following section provides a brief outline of some of the research limitations and challenges associated with using survey data. Following this, we provide an overview of the South African literature focusing on the gender dynamics of the labour market.

2.1 South African survey data

South Africa has a rich history of publicly available survey data which has been used extensively to study the labour market and, specifically, its gender dynamics. The three key sources of data that have traditionally been used to study women in the labour market are the census, the General

Household Survey (GHS) and the Quarterly Labour Force Survey (QLFS), which feeds into the annual release of the Labour Market Dynamics Study (LMDS).

The South African Census has been conducted in 1996, 2001, and 2011, with plans for additional censuses every ten years (Stats SA 2012). As is typical of a census, it covers a large sample and is the benchmark for all household surveys conducted by the national statistics office, Statistics South Africa (Stats SA).[1] Despite the large sample, the infrequency of data collection as well as the extensive delays in releasing a sample of the data for public use makes this an unsuitable data source for researchers wishing to conduct up-to-date analysis. In addition, Casale and Posel (2002) suggest that the census measurement of employment status may be rudimentary as only a few questions in the census are related to the employment activity. As an example of this, the authors point out that no distinction is made between employment in the formal and informal sectors in the census questionnaire.

The GHS is an annual household survey conducted by Stats SA since 2002, replacing the October Household Survey (OHS) which ran from 1993–99. The GHS is available from 2002 until 2017 at the time of writing and covers six broad areas, namely education, health and social development, housing, households' access to services and facilities, food security, and agriculture (Stats SA 2017a). While some data has been collected on individuals' labour market activities, this survey does not cover a comprehensive range of data related to wages and employment activity.

The Labour Force Survey (LFS) (2000–07) was a bi-annual household survey focusing on labour market dynamics. This became the QLFS in 2008 and is currently on-going. While the QLFS is frequently released and includes data on economic activities, some components of the survey are not released to the public in the quarterly format. Since 2008, the QLFS has been collated into an annual dataset, the LMDS, which includes data from all four quarters of QLFS data for that year. The LMDS also includes a number of variables—including income—which have not been released in the QLFS since 2010 (Stats SA 2017b). However, there are some delays in releasing the annualized LMDS in comparison with the frequency of the QLFS release—the LMDS has currently only been released up to 2017.

Both the GHS and the QLFS (and relatedly the LMDS) are subject to the usual survey data issues such as non-response and under-sampling. Changes in the questionnaire over time and changes of the definitions of employment are some of the particular issues in the OHS and QLFS data described by Casale and Posel (2002). In addition, survey data tends to under-report earnings at the top end of the wage distribution, which may be because of missing information on high wage earners in the survey data, reluctance of high wage earners to report their wages, or wage earners excluding benefits such as pension and bonuses from their gross income (Wittenberg 2017). This may lead, for example, to distortions in measures of income inequality.

An excellent resource for researchers wishing to undertake survey data analysis in South Africa is the Post-apartheid Labour Market Series (PALMS). This is a harmonized dataset containing microdata from 61 household surveys, including the OHS, LFS, and QLFS, and was created by DataFirst at the University of Cape Town. PALMS harmonizes variable definitions across the different surveys to reach the most consistent series possible over the time period. Currently, the dataset contains the harmonized series of labour market data for the years 1993 to 2017.

---

[1] The Census 1960–2011, GHS 2002–17, LMDS 2008–17, NIDS 2008–17, SESE 2011–13 data, questionnaires, and metadata are available on the DataFirst Open Data Portal (see DataFirst n.d.).
Specific links for each dataset in this paper's reference list.

In more recent years, the National Income Dynamics Study (NIDS) has been used to examine gender differences in income and gender inequality in the labour market (d'Agostino and Scarlato 2016; Posel et al. 2016). The NIDS is the first national household panel study in South Africa, with the first survey having taken place in 2008. It is conducted every two years, and the most recent wave to have been released was the fifth survey, conducted in 2017 (Brophy et al. 2018). The survey covers a wide range of themes, with a focus on labour market and economic activity. This creates a rich dataset of individual and household level history for researchers to study, enabling them to track individuals and their movements in and out of the labour market over time, thereby uncovering the dynamics of labour force participation in South Africa. However, the NIDS suffers from the same survey data issues as the OHS and QLFS, in addition to challenges associated with survey data. Baigrie and Eyal (2014), who evaluate attrition in the NIDS panel data, write that attrition is the 'Achilles heel' of panel data. Attrition can lead to bias in estimates and can cause complications in weighting the data over time. Baigrie and Eyal (2014) find evidence of differential attrition dynamics for different races, and that there is some evidence of attrition bias in estimated coefficients based on the non-reducing sample when researching the labour market or health specifications. They therefore advise researchers using the NIDS data against generalizing to the overall population without thorough investigation.

There is a general paucity of firm-level surveys in South Africa. Two prominent South African firm surveys are the Quarterly Employment Survey (QES) and the Survey of Employers and the Self-Employed (SESE). The QES is a quarterly enterprise-based survey which collects data on approximately 20,000 VAT-registered private and public enterprises in the formal non-agricultural sector of the South African economy. It is nationally representative and has been conducted since 2004. The survey excludes informal businesses and does not cover the agriculture or domestic services sector. The QES also excludes a description of the unemployed or the employed, their demographic profile, education level, or hours of work. It does, however, contain the number of employees and wage bill in each enterprise. Crucially, the QES is not made publicly available for research. The SESE is a survey of informal firms using the LFS as a sampling frame and has been conducted every four years since 2001 (Stats SA 2013). The most recent survey conducted in 2017 is not yet publicly available. While SESE contains both firm- and individual-level data, it is a small sample limited to non-VAT-registered businesses.

None of the individual-level surveys described above provide matched employee–employer level data. Some progress has been made on matching informal employers and employees using the SESE 2013 and QLFS Quarter 3 2013 surveys (Fourie and Kerr 2017; Lloyd and Leibbrandt 2016), and this dataset would complement any matched employer–employee analysis done using the tax data. Overall, there was limited firm-level data available in South Africa before the tax data became available, and no large-scale dataset containing matched employee–employer data. Therefore, while there is a rich analysis of the individual characteristics associated with labour market outcomes in South Africa, there is a paucity of information on firm-side characteristics or the dynamics between the individual and firm.

2.2 Overview of South African literature

The following section provides a brief overview of the gender-based South African literature, in order to frame the contribution of the tax data.

Gender, employment, and wages

There have been many studies using survey data to analyse the gender dynamics of employment in post-apartheid South Africa. Casale (2004) shows that, post-apartheid, the South African labour market has become increasingly feminized and argues that factors such as lower marriage rates and changes in household structure may have driven this change. However, this has also led to an increase in unemployment for women, who consistently face higher unemployment rates than men in South Africa. Although women are less likely to be employed than men with the same endowments (Oosthuizen 2006; Leibbrandt et al. 2010), Oosthuizen (2006) finds that gender was of decreasing importance for explaining probability of employment between 1995 and 2004. More recently, Kimani (2015) shows that women are less likely than men to transition out of unemployment into employment.

Furthermore, there are a multitude of South African studies that find a gender wage gap between men and women, even after accounting for individual characteristics such as education and age (Burger and Yu 2007; Ntuli 2007; Muller 2009; Casale and Posel 2011), with the wage gap increasing in the post-apartheid years (Burger and Yu 2007; Ntuli 2007). The gender wage gap is largest at the bottom end of the wage distribution (Ntuli 2007; Bhorat and Goga 2013), with pure gender discrimination accounting for most of the gap between the 20th and 50th quantiles (Bhorat and Goga 2013).

While there is a rich body of literature analysing employment and wages and their gender dynamics, the limitations of survey data as well as the scarcity of firm-level data results in gaps in the literature. Relevant to this scoping paper is the fact that survey data—which represents a snapshot in time—does not allow for the analysis of seasonality of employment, or the proportion of the year for which an individual is employed. In addition, limitations to survey data including non-response, attrition, or under-sampling may distort existing analysis of the gender wage gap. Finally, while there is much analysis of the individual-level determinants of wages, there is little analysis of the firm-level determinants of wages due to a lack of appropriate data. The ways in which the tax data can be used to fill these gaps is discussed in more detail in Section 4.

Worker flows and labour demand

There have been a handful of studies estimating worker flow, job flow, and job churn in South Africa; they differ in methodology due to differences in the types of data used. Banerjee et al. (2008) use panel household survey data to estimate gross flows over labour market states between 2002–03. They find that 16 per cent of the formal sector in 2002 had transitioned out of this state by 2003, with males and females equally likely to have retained formal employment over the period. As expected, females are more likely than males to transition out of the labour market when leaving formal employment. However, because this data does not link individuals to firms, they cannot measure worker flow (a measure of hires and separations), which is typically defined from the firm perspective.

Kerr et al. (2014) use the QES firm data[2] to analyse job creation and destruction from the firm perspective, finding that firms create or destroy around 20 per cent of total jobs in a 12-month period. However, analogous to the above, this study does not include individual information and therefore does not analyse these measures from the worker perspective, whether male or female.

[2] While this data is not publicly available, the authors had access to this data for this publication.

One study which does use matched firm-individual data is Kerr (2018), who uses the payroll (IRP5) component of the tax data to evaluate worker flows and job churn in South Africa. The author finds that while worker flows are high on average, there is considerable heterogeneity in worker experience by level of earnings and firm type. However, at the time of the study, information on the gender of employees was not available in the data, precluding any analysis on how these measures may differ across gender.

The recent addition of a gender variable in the tax data therefore makes it possible to evaluate job churn and worker flows by gender using matched employer–employee data, something which is yet to be done in South Africa.

Tenure, employment spells, and wages

International studies suggest a positive relationship between tenure and wages but that returns to tenure are higher for men than for women (Munasinghe et al. 2008; Ioakimidis 2012). The use of administrative data is common when evaluating the relationship between tenure and wages. In South Africa, the unavailability of matched firm–employee data has thus far largely precluded studies of the relationship between tenure and wages for men and women. In addition, the South African data has for the most part not allowed for the examination of the relationship between experience and labour market returns. As a proxy, many studies use a constructed measure of 'potential experience', most often constructed using the individual's age minus their years in education minus six—the age at which most South Africans start school (see, for example, Bosch 2006; Burger and Jafta 2006; Von Fintel 2006; Armstrong 2009; Bhorat and Goga 2012; Magadla et al. 2018). However, Keswell and Poswell (2004) argue that 'potential experience' is a poor proxy for tenure, on-the-job-learning, or other measures which it is supposed to shed light on. This is because of the prevalence of grade repetition and job insecurity (Keswell and Poswell 2004)—many South Africans spend extended periods of time unemployed or out of the labour market.

Similarly, survey data does not allow for analysis of job spells without a detailed, retroactive module on employment history. McKeever (2006) uses data including a retrospective employment questionnaire to examine transitions between the formal and informal economy in South Africa. The author finds that, on average, employment spells in the formal sector are longer for men than for women, but he does not conduct an analysis of the relationship between these job spells and wages. In addition, the data covers a small geographical area and is from 1991.

The availability of the SARS firm and individual data therefore allows for the possibility of examining the relationship between tenure, job spells, and wages by gender in the formal sector, using a vast, up-to-date panel of data.

Trade and wages

Several studies analyse the relationship between trade and wages. It is well-documented internationally that exporting firms pay higher wages than non-exporting firms (Bernard et al. 1995; Bernard et al. 2007; Schank et al. 2007).

In South Africa, the availability of customs data has recently made it possible to analyse this same relationship. Both Bhorat et al. (2017) and Matthee et al. (2017), using the tax data, provide evidence for a firm exporter wage premium in South Africa. Similarly, Edwards et al. (2018) find a wage premium associated with trading firms, with the highest premium for firms which both import and export. However, at the time of both these studies, data on the gender of employees was not available.

The depth and breadth of the survey data in South Africa is vast which, despite its shortcomings, has allowed for rigorous analysis of the labour market. Tax data is, in no way, going to take away from the strengths of survey data. However, there have been some limitations to the research using survey data and the hope is that tax data can fill some of these gaps. In Table 1 below, we summarize the papers we have examined in this section by outlining the data used in each of these papers.

Table 1: Literature examined, and data used

Title	Authors	Data used
Why Has Unemployment Risen in the New South Africa?	Banerjee et al. (2008)	OHS of 1995, 1997, 1999, and the September wave of the Labour Force Surveys (LFS) 2000–04 March 2004, March 2005 LFS Census 1970
Exporters, Jobs, and Wages in US Manufacturing: 1976–1987	Bernard et al. (1995)	Plant-level US Census Bureau's Annual Survey of Manufacturers (ASM)
Firms in International Trade	Bernard et al. (2007)	1997 US Census of Manufacturers and the 2000 Linked-Longitudinal Firm Trade Transaction Database
Firm-level Determinants of Earnings in the Formal Sector of the South African Labour Market	Bhorat et al. (2017)	SARS Administrative Tax Data 2016 (no gender available)
The Gender Wage Gap In Post-Apartheid South Africa: A Re-Examination	Bhorat and Goga (2013)	September Round of the 2007 LFS
Wage Trends in Post-Apartheid South Africa: Constructing an Earnings Series from Household Survey Data	Burger and Yu (2007)	OHS 1995–99, LFS 2000–05
What Has the Feminization of the Labour Market 'Bought' Women In South Africa? Trends in Labour Force Participation, Employment and Earnings, 1995–2001	Casale (2004)	OHS 1995, 1997, and 1999 LFS of September 2001
Unions and the Gender Wage Gap in South Africa	Casale and Posel (2011)	October Household Surveys 1995 And 1999 Surveys
Importing and Firm Export Performance: New Evidence from South Africa	Edwards et al. (2018)	SARS Administrative Tax Data 2016 (No Gender Available)
Gender Wage Gap and Earnings: Predicted by Tenure in the Czech Republic	Ioakimidis (2012)	Structure of Earnings Survey Conducted Annually by the Czech Statistical Office (CZSO) 2004–08
Job Creation and Destruction in South Africa	Kerr et al. (2014)	Quarterly Employment Survey (QES)
Education and Labour Market Outcomes in South Africa: Evidence from the National Income Dynamics Study	Kimani (2015)	NIDS 2008–12, waves 1–3
Trends In South African Income Distribution and Poverty since the Fall of Apartheid	Leibbrandt et al. (2010)	1993 PSLSD Survey, 2000 Income And Expenditure Survey (IES), 2008 National Income Dynamics Study (NIDS)
Labour Demand and the Distribution of Wages in South African Manufacturing Exporters	Matthee et al. (2017)	SARS Administrative Tax Data 2016 (No Gender Available)
Gender Gap in Wage Returns to Job Tenure and Experience	Munasinghe et al. (2008)	Panel Data from the National Longitudinal Surveys of Youth NLSY (1979–94)
Fall Back or Spring Forward? Labour Market Transitions and the Informal Economy in South Africa	Mckeever (2006)	1991 Survey of Socioeconomic Opportunity and Achievement (Household Survey)
Exploring Gender Wage 'Discrimination' in South Africa, 1995-2004: A Quantile Regression Approach	Ntuli (2007)	September (2004) Labour Force Survey (Lfs_2) and the (1995, 1999) October Household Surveys (OHS)
The Post-Apartheid Labour Market: 1995–2004	Oosthuizen (2006)	Ohs 1995, Lfs September 2004
Do Exporters Really Pay Higher Wages? First Evidence from German Linked Employer–Employee Data	Schank et al. (2007)	Matched Employer–Employee German Data

Source: Authors' illustration

3 Administrative tax data

Internationally, the use of administrative data for empirical research is becoming more common (Card et al. 2010). Over and above survey data, administrative data is appealing for research as it offers several advantages. Research using administrative tax records to evaluate public policy has become widespread (Card et al. 2010).

Einav and Levin (2014) showcase the increase in research papers which obtained an exemption for tax data in the American Economic Review. The number rose from 4 per cent in 2006 to 26 per cent in 2014. This is testament to the world's best economics research increasingly using tax data.

Administrative data offers much larger sample sizes than survey data and can even host the full population of interest. The data is often similar to a census, with almost total coverage of the population in question. The large number of observations in administrative data lends well to a large set of research methods and gives us confidence about the statistical power of the results.

The regular collection of administrative records means that administrative data is longitudinal in nature. This enables reliable policy evaluation as it provides data for before, during, and after policy implementation, enabling long-term follow-up.

Slemrod (2016: 1003) describes the use of tax data: 'The expanded availability of tax-return administrative data, especially at the population level, has triggered an explosion of scholarly research addressing tax policy, tax administration, and non-tax questions'. Slemrod (2016) further discusses in more detail the advantages and disadvantages of using tax data versus using survey data.

Administrative records can provide more dependable information than survey data (Einav and Levin 2014). Tax records are often audited, providing an added level of reliability. Technological advances also mean tax forms are sometimes prepopulated with information from employers or information previously provided to the revenue collector. Lastly, the use of administrative data can be less costly than designing surveys and collecting new data.

3.1 Disadvantages of administrative tax data

There are, of course, several challenges when using administrative data. The first challenge is that researchers do not collect the administrative data and have no control over what or how the information is collected. Therefore, the data collected might not be suitable for research, and information not required by administrative data collectors may be incomplete or incorrect. This means that tax data often does not include demographic information of taxpayers. This feeds into the second data challenge, namely a lack of theory and research methods to guide the use of administrative data (Wallgren and Wallgren 2007). This means that information on the how the dataset was constructed is not available through a metadata or data guide, leaving researchers to make their own assumptions about the data, which often cannot be verified.

Thirdly, administrative data gives us information over time through the panel, but the data is drawn at a specific point in time. Any corrections or updates to the administrative data will therefore not be included. Lastly, due to the confidential nature of administrative data, the data is not publicly available. Many research projects using administrative data indicate the examination of the data in a secure facility or where the data has been anonymized for research purposes, or both.

3.2 Overview of the SARS administrative tax data

This section provides more specific information about the tax administrative data available for research at the NT in Pretoria. The tax data from SARS is a rich new source of data allowing for new avenues of research. Tax data has been made available to researchers within the National Treasury Secure Data Facility since 2015. Anonymized Company Income Tax, Value Added Tax, Customs, IRP5 and individual tax assessment data is available in the NT Secure Data Facility. The data originates from SARS, where it is anonymized before it is made available to researchers.

IRP5/IT3(a) certificates are submitted to SARS by employers who are registered for pay-as-you-earn (PAYE). Since 2011, each employee receiving remuneration is issued with an Employee Tax Certificate by their employer irrespective of amount earned. Employers also issue IRP5 certificates for remuneration to individuals for consulting services or from retirement funds. The certificates contain the income, deductions, allowances, benefits, medical scheme contributions, age and employment period information for each job in the formal economy. Gender and other demographic information is not collected by SARS at the time of submission of the IRP5/IT3(a). Gender in the tax data is derived, and the next section explains how this is done. At the time of writing the IRP5 data is available for tax years 2008–18.

The Company Income Tax-IRP5 (CIT-IRP5) panel[3] is an unbalanced panel created by linking company income tax (CIT) data, employee tax data (IRP5), value-added tax (VAT) data, and customs records from trading firms. The CIT data includes firm-level information such as revenue amounts, tax paid, location, and sector of the firm and is derived from the Income Tax Returns for company forms. The CIT-IRP5 panel matches employer–employee variables from the IRP5 and CIT datasets. The panel includes tax information from 2008–15 and is in the process of being updated to include tax information until 2017. Pieterse et al. (2018) provide a more detailed description of the CIT-IRP5 panel. The authors discuss how the panel was constructed and any biases it might contain, and compare the panel with other data sources. We discuss in Section 4 how gender can be examined at the firm level using this panel

The customs data contains the customs records on the import and export activities of trading firms. These records are available at the transaction level and include a firm-level identifier. Transaction entries typically provide information on the value and quantity of imported or exported goods, the destination or origin of the trade, as well as detailed product codes using the Harmonized System (HS) of the World Customs Organization. This includes the HS4, HS6, and HS8 codes.

Entities that are VAT-registered submit forms to SARS on various schedules depending on the size of the firm. The data includes information on the input and output VAT as well as VAT paid on capital expenditure.

The ITR12 data, or Personal Income Tax Return, is completed by the individual. The data includes all remuneration, deductions, and fringe benefits used to calculate the individual's final tax liability for each tax year. Individuals who receive an IRP5/IT3(a) certificate are commonly prepopulated in the ITR12 return. Individuals will manually include additional income from self-employment, investment income or other sources, and any further deductions. Not all taxpayers are required to complete an ITR12. Taxpayers are required to complete an ITR12 return if they have income above the compulsory submission threshold. The compulsory submission threshold was

[3] The panel was created through the joint initiative between SARS, the NT, and UNU-WIDER.

introduced in the 2008 tax year and was set at ZAR120,000. The compulsory threshold was increased each year until the 2015 tax year and has been set at ZAR350,000 since then.

Gender information from IRP5 can be combined with CIT, VAT or customs data to analyse gender at both the firm and the individual levels. For example, combining the customs data and data on formally employed women allows us to examine the difference in wages for women working in trading firms versus non-trading firms, or the differences between the wages of men and women within trading firms. We explore this idea and other ideas in Section 4 of this paper.

3.3 Gender in the tax data

The gender variable only became available in the IRP5 data at the end of 2017. Gender is not a required field on the IRP5/IT3(a) form and is derived from the South African identity number (ID) which includes the individual's gender as the 7th digit. This means that taxpayers without an ID number will automatically be excluded. For example, a foreigner working in South Africa who has no ID number will not have a gender assigned in the tax data. This creates a bias in the gender variable in the tax data. This is something researchers should take into account when conducting any analysis using gender in the tax data.

As the data is anonymized at SARS, the gender variable is derived before the anonymization process. The gender variable is well populated and is described in Table 2.

Table 2: Gender variable in the raw IRP5 data, 2008–18

Tax year	Female	Male	Foreigner	Non-natural entity	Missing	Total
2008	6,046,598	8,435,735	338,102	501,786	48	15,322,269
2009	6,461,279	8,725,662	135,459	304,033	260	15,626,693
2010	6,455,695	8,368,276	237,818	663,792	1,514	15,727,095
2011	6,743,647	8,784,780	363,381	412,567	18	16,304,393
2012	7,195,521	9,173,485	401,829	314,513	4	17,085,352
2013	7,224,912	9,269,647	431,499	303,105	4	17,229,167
2014	7,555,546	9,509,997	465,609	260,400	2	17,791,554
2015	8,674,526	10,331,214	510,771	256,626	1	18,437,708
2016	7,921,309	9,649,796	495,684	249,364	1	17,983,614
2017	8,756,575	9,902,777	514,829	362,974	1	19,537,156
2018	7,629,875	9,034,049	489,956	212,615	1	17,366,496

Source: Authors' own estimates based on IRP5 tax data.

Table 2 includes the raw estimates from all the IRP5 certificates in the tax data. This includes the full population of women working in the formal labour market, including both the private and public sectors. The gender of those individuals without a South African ID cannot be specified. For these individuals, we have only a passport number, and they are identified as 'foreigners' in the table above for this reason. The IRP5 data also includes partnerships, share blocks, and other entities which are represented by the non-natural entity's column. For some individuals, there is a missing ID number and passport number, although the data indicates that they are natural persons. These individuals are represented in the 'missing' column in the table. Again, we cannot identify their gender.

In comparison to other datasets, the QLFS Q2 2018 indicates that 43.8 per cent of total employment in South Africa was made up of women. In the 2018 tax year, we identify in the tax data that 43 per cent of the labour force was made up of women. These figures are very close despite the fact that they exclude all foreign workers included in the QLFS.

3.4 Limitations specific to the tax data

There are limits to the extent to which the tax data, in its current form, can help to fill the gaps in the literature we identify in the next section. Firstly, for the purpose of studying gender in the labour market, the data does not include education level, occupation, race, or hours worked. These are some of the rudimentary variables required for labour market research in South Africa. Secondly, we can only see those employed in the formal sector. The nature of the data means that all informal workers are excluded from the analysis. Individuals who are unemployed are not observed, so we would not know whether they are employed in the informal labour market, unemployed, or not economically active. Thirdly, there is a lag in the availability of the tax data. Companies, for example, have 12 months from their financial year-end in which to complete their tax returns; thus, some company data may be incomplete. While the tax data is panel in nature, any late submissions or revisions submitted after the data is extracted from SARS will not be in the version of the data used for research.

There is very little scope for changes to the tax form to give us more demographic information. The international literature points to merging tax data with other administrative or, more frequently, survey data. Ideally, tax data would include or could be combined with other administrative data to give us all the information we would need to carry out a complete analysis of gender in the labour market.

4 Areas for further research using the tax data

The fact that the tax data takes the form of an employee–employer panel makes it possible to examine the relationship between firm characteristics such as firm size and tenure and the gender wage gap. As discussed, there are some limitations regarding the scope gender-based analysis of the labour market using the tax data. This said, there are a number of areas where the tax data can contribute to the existing body of literature.

4.1 Worker flows and job churn

There are many examples of administrative data being used to analyse worker flows in countries where this type of data is already available. Administrative data has been used to measure the effect of legislation such as dismissal protection on worker flows (Bauer et al. 2007; Kugler and Pica 2007). A handful of studies have analysed firm-level job churn and worker flow in South Africa, with only one of these using matched employer–employee data. The recent addition of a gender variable into the tax data therefore makes it possible to evaluate job churn and worker flows by gender using matched employer–employee data, something which is yet to be done in South Africa. A number of inter-related concepts can be analysed using this data, each of which may differ across gender:

a) job creation and destruction by firms
b) worker flows or the sum of hires and separations[4]
c) job churn or worker flows in excess of those created by job reallocation.

[4] Kerr (2018) provides a useful discussion of these methodological differences in the calculation of worker flows.

The tax data enables an analysis of these concepts, taking into account the available firm characteristics as well as whether these aggregate measures differ by gender.

4.2 Tenure, employment spells, and wages

In the United States, administrative data has been used to map the pattern of hires and separations over time, with the job tenure distribution shifting towards longer tenure jobs since 2000 (Hyatt and Spletzer 2016). Deelan (2012) uses administrative data from the Netherlands to investigate the role of wage-tenure profiles in explaining patterns of job mobility. Administrative data has also been used to examine the effect of job tenure on earnings instability in Italy (Cappellari and Leonardi 2016). The ten-year SARS panel provides the necessary data to track individuals moving in and out of the formal labour market over time. This will allow for evaluation of tenure, employment spells, and subsequent wages earned, by gender. In particular, the data allows for evaluation of the effect of spells outside of the formal labour market on subsequent wages earned, and whether these effects are similar across men and women. While there is limited data on individual characteristics, the panel nature of the data allows for the use of a fixed-effects model in evaluating the effect of movements in and out of the formal labour market on wages earned. This model controls for individual time-invariant characteristics, such as race. There is also room to examine the relationship between firm tenure and wages for men and women, although the lack of substantial individual-level data may preclude econometric analysis.

It is important to note that any study using the tax data to examine formal sector employment spells and wage suppression will not be able to identify the states a worker flows into or out of when leaving or joining the formal sector, as the tax data does not contain information on what individuals are doing when they do not appear in the formal labour market data.

4.3 Seasonality of employment

The tax data makes it possible to measure the seasonality of employment, or the ebbs and flows of employment throughout the year. While survey data will only reveal if an individual was employed within the time period asked about in the questionnaire (usually a week or a month), the tax data allows for an evaluation of the proportion of the year during which men and women are employed in the formal sector. In other words, one can evaluate both the make-up of the formal labour market by gender and the proportion of the year that both men and women spend in formal sector employment. To our knowledge, this is yet to be done on South African data.

4.4 The gender wage gap

Administrative data is an ideal source for the estimation of the gender wage gap. In their seminal work, Abowd et al. (1999) use French administrative data to show that there was no evidence of change in the male–female wage gap between 1976 and 1987. Kunze (2005) uses German administrative data to show that the gender wage gap exists from entry into first employment and is a permanent feature throughout the employment lifecycle. Administrative data has also been used to examine wage gaps in the public sector wage premia (Hospido and Moral-Benito 2016). Bradley et al. (2015) find a substantial gender wage gap in some occupations in the public sector. If the data in question is rich, it can also be used to answer more targeted questions, such as whether female representation in higher-up positions is related to the gender wage gap. For example, Hirsch (2013) uses a large linked employer–employee dataset made up of both administrative and survey data to show that increasing the female share of management decreases the gender wage gap in German firms.

Wittenberg (2017) shows that the QLFS typically under-reports wage income in comparison with the tax data, particularly at the top end of the distribution. Therefore, existing studies of the gender wage gap in South Africa may be underestimating wage inequality between men and women. This discrepancy may be because of missing information on high wage earners in the survey data, reluctance of high wage earners to report their wages, or wage earners excluding benefits such as pension and bonuses from their gross income (Wittenberg 2017). Because there are large penalties for non-disclosure of income, the tax data includes substantially more complete information on wages earned in the formal labour market at the top end of the distribution. Therefore, studies using this data will be able to undertake a more accurate description of the formal labour market gender wage gap at the top end.

Diaz-Bazan (2015) suggests a methodology to measure inequality which optimally combines household survey information with tax records to construct a complete income distribution. This methodology suggests choosing an income threshold to minimize reliance on household survey data at the bottom end of the distribution, where income data under the threshold is taken from household surveys and data above the threshold is taken from tax records. Typically, this threshold is one where tax fillings become mandatory. This allows the simultaneous use of survey data, which may be more accurate at the bottom end, and tax data, which is likely to be more accurate at the top end of the wage distribution. Hundenborn et al. (2018) use this methodology to combine Personal Income Tax data and data from the NIDS. Combining the two datasets results in a significant decrease in overall inequality of taxable income in South Africa between 2011 and 2014. However, they do not examine individual earnings or report on the gender wage gap using this methodology.

It should be noted, however, that Wittenberg (2017) also finds that the incomes of the self-employed may be under-reported in the tax data in comparison to the QLFS data, possibly due to non-disclosure to SARS by these individuals of some or all income earned. Therefore, researchers who want to use the tax data to measure the gender wage gap in the formal sector should be aware of these heterogeneous discrepancies in both household survey data and the tax data throughout the wage distribution.

4.5 Trade and gender wage gap

While the literature examining the relationship between trade liberalization and wage inequality is not decisive, some literature in developing countries has found that wage inequality tends to increase with trade liberalization (Goldberg and Pavcnik 2007). In the United States, this has been attributed to a reduction in the relative wages of low-skilled workers (Juhn et al. 2014). However, when looking at inequality in wages between men and women specifically, there is evidence that trading decreases the gender wage gap (Black and Brainerd 2004; Juhn et al. 2014). Bøler et al. (2015) use matched employer–employee data to study the differences in the wages between men and women in exporting and non-exporting firms in the Norwegian manufacturing sector. They hypothesize that exporting firms require greater commitment from their employees. If women are perceived as less committed than men, this will increase the gender wage gap in exporting firms. They find that the gender wage gap in exporting firms narrows when the paternity leave is increased in Norway—a policy which they argue decreased the commitment gap between men and women in Norway.

The existing South African literature finds a wage premium associated with trading firms. Lepelle et al. (2017) suggest that the welfare effects of exports are not homogeneous between men and women. The literature on trade and gender is conducted at a national level and Lepelle et. al. (2017) introduce regional heterogeneity when considering gender in the labour market. The literature also

includes country-level evidence of exports broadening the gender wage (see World Bank 2012 for a detailed summary).

The tax data allows for firm-level employment behaviour to be examined. Combined with the customs data, there is scope for an evaluation of the extent to which trade can create opportunities or improve employment for women.

4.6 Firm-level determinants of female employment and wages

There is a host of international literature that uses administrative data to examine the firm-level determinants of wages (Abowd et al. 1999; Gruetter and Lalive 2004; Gürtzgen 2009; Schank et al. 2007; Cerejeira and Guimaraes 2012).For example, Cerejeira and Guimaraes (2012) use Portuguese administrative data to estimates the employer-size wage effect on returns to unobservable skills and measured human capital. They find that the gender wage ratio does not increase as firm size increases. Del Bono and Vuri (2011) examine the relationship between job mobility and the gender wage gap. They find that moves to larger firms represent by far the main source of gender differences in returns to mobility.

As discussed, there are few studies that analyse the firm-level determinants of employment and wages in South Africa. Furthermore, there is—to our knowledge—no literature on whether these firm-level factors affect men and women differently. The addition of the gender variable to the tax data allows for an analysis of the firm-level determinants of female employment by firms. Furthermore, a fixed-effects regression can be used to evaluate the firm-level characteristics that determine the wages paid to women, holding constant time-invariant individual characteristics.

5 Conclusion

The availability of the SARS worker–firm panel of administrative tax data creates a unique opportunity to contribute to the literature on the gender dynamics of the labour market in South Africa. The use of administrative data has a number of advantages over survey data, including a large sample size, regular collection, and more dependable data. In addition, the SARS tax data is the first large-scale matched employer–employee panel dataset in South Africa. Therefore, while South Africa has a strong repository of survey data-based supply-level analysis, the paucity of firm-level data has resulted in limited analysis of demand-side factors affecting women in the labour market.

This paper has identified six areas with scope for further research on the gender dynamics of the formal South African labour market:

1. Worker flows and job churn
2. Tenure, employment spells, and wages
3. Seasonality of employment
4. The formal sector gender wage gap
5. Trade and wages
6. Firm-level determinants of female employment and wages.

While the use of the SARS tax data limits the analysis which can be performed in these areas, there remain a substantial number of areas where the existing South African literature can be enhanced with its use.

References

Abowd, J.M., F. Kramarz, and D.N. Margolis (1999). 'High Wage Workers and High Wage Firms'. *Econometrica*, 67(2): 251–333.

Armstrong, P. (2009). 'Teacher Pay in South Africa: How Attractive is the Teaching Profession'. Stellenbosch Economic Working Paper 04/09. Stellenbosch: Department of Economics and the Bureau for Economic Research at the University of Stellenbosch.

Baigrie, N., and K. Eyal (2014). 'An Evaluation of the Determinants and Implications of Panel Attrition in the National Income Dynamics Survey (2008–2010)'. *South African Journal of Economics*, 82(1): 39–65.

Banerjee, A., S. Galiani, J. Levinsohn, Z. Mclaren, and I. Woolard (2008). 'Why Has Unemployment Risen in the New South Africa?'. *Economics of Transition*, 16(4): 715–40.

Bauer, T.K., S. Bender, and H. Bonin (2007). 'Dismissal Protection and Worker Flows in Small Establishments'. *Economica*, 74(296): 804–21.

Bernard, A.B., J.B. Jensen, and R.Z. Lawrence (1995). 'Exporters, Jobs, and Wages in US Manufacturing: 1976-1987'. Brookings Papers on Economic Activity. *Microeconomics*, 1995: 67–119.

Bernard, A.B., J.B. Jensen, S.J. Redding, and P.K. Schott (2007). 'Firms in International Trade'. *Journal of Economic Perspectives*, 21(3): 105–30.

Bhorat, H., and S. Goga (2012). 'The Gender Wage Gap in the Post-apartheid South African Labour Market'. University of Cape Town Development Policy Research Unit Working Paper 12/148. Cape Town: University of Cape Town.

Bhorat, H., and S. Goga (2013). 'The Gender Wage Gap in Post-apartheid South Africa: A Re-examination'. *Journal of African Economies*, 22(5): 827–48.

Bhorat, H., M. Oosthuizen, K. Lilenstein, and F. Steenkamp (2017). 'Firm-level Determinants of Earnings in the Formal Sector of the South African Labour Market'. WIDER Working Paper 2017/25. Helsinki: UNU-WIDER.

Black, S.E., and E. Brainerd (2004). 'Importing Equality? The Impact of Globalization on Gender Discrimination'. *ILR Review*, 57(4): 540–59.

Bøler, E.A., B.S. Javorcik, and K.H. Ulltveit-Moe (2015). 'Globalization: A Woman's Best Friend? Exporters and the Gender Wage Gap'. CESifo Working Paper Series 5296. Munich: CESifo Group.

Bosch, A. (2006). 'Determinants of Public and Private-sector Wages in South Africa'. South African Reserve Bank, *Labour Market Frontiers*, 8: 17–24.

Bradley, S., C. Green, and J. Mangan (2015). 'Gender Wage Gaps Within a Public Sector: Evidence from Personnel Data'. *The Manchester School*, 83(4): 379–97.

Brophy, T., N. Branson, R.C. Daniels, M. Leibbrandt, C. Mlatsheni, and I. Woolard (2018). 'National Income Dynamics Study Panel User Manual. Release 2018. Version 1'. Cape Town: Southern Africa Labour and Development Research Unit. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/712 (accessed 26 February 2019).

Burger, R., and R. Jafta (2006). 'Returns to Race: Labour Market Discrimination in Post-apartheid South Africa'. Stellenbosch Economic Working Papers, 04/0. Stellenbosch: Department of Economics and the Bureau for Economic Research at the University of Stellenbosch.

Burger, R., and D. Yu. (2007). 'Wage Trends in Post-apartheid South Africa: Constructing an Earnings Series from Household Survey Data'. DPRU Working Paper 07-117. Cape Town: Development Policy Research Unit.

Cappellari, L., and M. Leonardi (2016). 'Earnings Instability and Tenure'. *The Scandinavian Journal of Economics*, 118(2): 202–34.

Card, D., R. Chetty, M. Feldstein, and E. Saez (2010). 'Expanding Access to Administrative Data for Research in the United States'. National Science Foundation SBE 2020 Planning Activity [Online]. Available at: https://www.nsf.gov/sbe/sbe_2020/2020_pdfs/Card_David_112.pdf (accessed 12 February 2019).

Casale, D. (2004). 'What Has the Feminisation of the Labour Market 'Bought' Women in South Africa? Trends in Labour Force Participation, Employment and Earnings, 1995-2001'. DPRU Working Paper 04/84. Cape Town: Development Policy Research Unit.

Casale, D., and D. Posel (2002). 'The Continued Feminisation of the Labour Force in South Africa: An Analysis of Recent Data and Trends'. *South African Journal of Economics*, 70(1): 156–84.

Casale, D., and D. Posel (2011). 'Unions and the Gender Wage Gap in South Africa'. *Journal of African Economies*, 20(1): 27–59.

Cerejeira, J., and P. Guimaraes (2012). 'The Price of Unobservables and the Employer-size Wage Premium'. *Economics Letters*, 117(3): 878–80.

D'Agostino, G., and M. Scarlato (2016). 'Gender Inequality in the South African Labour Market: The Impact of the Child Support Grant'. MPRA Paper 72523, Munich: Munich Personal RePEc Archive.

DataFirst (n.d.). 'Open Data Portal'. [Online]. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog (accessed 4 December 2019).

Deelen, A.P. (2012). 'Wage-tenure Profiles and Mobility'. *De Economist*, 160(2): 141–55.

Del Bono, E., and D. Vuri (2011). 'Job Mobility and the Gender Wage Gap in Italy'. *Labour Economics*, 18(1): 130–42.

Diaz-Bazan, T. (2015). 'Measuring Inequality from Top to Bottom'. Policy Research Working Paper 7237. Washington, DC: The World Bank.

Edwards, L., M. Sanfilippo, and A. Sundaram (2018). 'Importing and Firm Export Performance: New Evidence from South Africa'. *South African Journal of Economics*, 86 (S1): 79–95.

Einav, L., and J. Levin. (2014). 'Economics in the Age of Big Data'. *Science*, 346.

Fourie, F., and A. Kerr (2017). 'Informal Sector Employment Creation in South Africa: What Can the SESE Enterprise Survey Tell Us?'. REDI3x3 Working Paper 32. Cape Town: SALDRU.

Goldberg, P.K., and N. Pavcnik (2007). 'Distributional Effects of Globalization in Developing Countries'. *Journal of Economic Literature*, 45(1): 39–82.

Gruetter, M., and R. Lalive (2004). 'The Importance of Firms in Wage Determination'. IZA Discussion Paper 1367. Bonn: Institute for the Study of Labor (IZA).

Gürtzgen, N. (2009). 'Rent-Sharing and Collective Bargaining Coverage: Evidence from Linked Employer–Employee Data'. *Scandinavian Journal of Economics*, 111(2): 323–49.

Hospido, L., and E. Moral-Benito (2016). 'The Public Sector Wage Premium in Spain: Evidence from Longitudinal Administrative Data'. *Labour Economics*, 42: 101–22.

Hirsch, B. (2013). 'The Impact of Female Managers on the Gender Pay Gap: Evidence from Linked Employer–Employee Data for Germany'. *Economics Letters*, 119(3): 348–50.

Hundenborn, J., I. Woolard, and J. Jellema (2018). 'The Effect of Top Incomes on Inequality in South Africa'. WIDER Working Paper 2018/90. Helsinki: UNU-WIDER.

Hyatt, H.R., and J.R. Spletzer (2016). 'The Shifting Job Tenure Distribution'. *Labour Economics*, 41: 363–77.

Ioakimidis, M. (2012). 'Gender Wage Gap and Earnings: Predicted by Tenure in the Czech Republic'. *The Journal of Developing Areas*, 46(1): 31–43.

Juhn, C., G. Ujhelyi, and C. Villegas-Sanchez (2014). 'Men, Women, and Machines: How Trade Impacts Gender Inequality'. *Journal of Development Economics*, 106(C): 179–93.

Kerr, A. (2018). 'Job Flows, Worker Flows and Churning in South Africa'. *South African Journal of Economics*, 86: 141–66.

Kerr, A., Wittenberg, M. & Arrow, J. (2014). 'Job Creation and Destruction in South Africa'. *South African Journal of Economics*, 82(SI): 1–18.

Keswell, M., and L. Poswell (2004). 'Returns to Education in South Africa: A Retrospective Sensitivity Analysis of the Available Evidence'. *South African Journal of Economics*, 72(4): 834–60.

Kimani, E.M. (2015). 'Education and Labor Market Outcomes in South Africa: Evidence from the National Income Dynamics Study'. Doctor of Philosophy Thesis. Cape Town: University of Cape Town.

Kugler, A., and G. Pica (2008). 'Effects of Employment Protection on Worker and Job Flows: Evidence from the 1990 Italian Reform'. *Labour Economics*, 15(1): 78–95.

Kunze, A. (2005). 'The Evolution of the Gender Wage Gap'. *Labour Economics*, 12(1): 73–97.

Leibbrandt, M., I. Woolard, A. Finn, and J. Argen. (2010). 'Trends in South African Income Distribution and Poverty Since the Fall of Apartheid'. OECD Social, Employment and Migration Working Paper 101. Paris: OECD Publishing.

Lepelle, R., L. Edwards, and M. Leibbrant (2017). 'Trade Effects on Gendered Labour Market Outcomes in Post-apartheid South Africa'. Paper Presented at the Bi-annual Conference of the Economic Society of South Africa, Rhodes University, South Africa, 30 August to 1 September.

Lloyd, N., and M. Leibbrandt. (2016). 'Entry into and Exit from Informal Enterprise Ownership in South Africa: An Analysis of the 2013 Survey of Employers and Self Employed and 2013 Quarterly Labour Force Survey'. REDI3x3 Working paper 17. Cape Town: SALDRU.

Magadla, S.S., M. Leibbrandt, and C. Mlatsheni (2018). 'Does a Motherhood Penalty Exist in the Post-apartheid South African Labour Market?'. Pretoria: DPME.

Matthee, M., N. Rankin, and C. Bezuidenhour (2017). 'Labour Demand and the Distribution of Wages in South African Manufacturing Exporters'. WIDER Working Paper 2017/11. Helsinki: UNU-WIDER.

McKeever, M. (2006). 'Fall Back or Spring Forward?: Labor Market Transitions and the Informal Economy in South Africa'. *Research in Social Stratification and Mobility*, 24: 73–87.

Mosomi, J. (2019). 'Distributional Changes in the Gender Wage Gap in the Post-apartheid South African Labour Market'. WIDER Working Paper 2019/17. Helsinki: UNU-WIDER.

Muller, C. (2009). 'Trends in the Gender Wage Gap and Gender Discrimination Among Part-time and Full-time Workers in Post-apartheid South Africa'. Cape Town: Development Policy Research Unit, School of Economics, University of Cape Town.

Munasinghe, L., T. Reif, and A. Henriques (2008). 'Gender Gap in Wage Returns to Job Tenure and Experience'. *Labour Economics*, 15: 1296–316.

Ntuli, M. (2007). 'Exploring Gender Wage "Discrimination" in South Africa, 1995-2004: a Quantile Regression Approach'. IPC Working Paper 56. Ann Arbor, MI: Internatonal Policy Center.

Oosthuizen, M. (2006). 'The Post-apartheid Labour Market: 1995-2004'. Development Policy Research Unit Working Paper 06/103. Cape Town: Development Policy Research Unit, University of Cape Town.

Pieterse, D., E. Gavin, and C.F. Kreuser (2018). 'Introduction to the South African Revenue Service and National Treasury Firm-Level Panel'. *South African Journal of Economics*, 86: 6–39.

Posel, D., D. Casale, and E. Grapsa (2016). 'Re-estimating Gender Differences in Income in South Africa: The Implications of Equivalence Scales'. *Development Southern Africa*, 33: 425–41.

Schank, T., C. Schnabel, and J. Wagner (2007). 'Do Exporters Really Pay Higher Wages? First Evidence from German Linked Employer–Employee Data'. *Journal of International Economics*, 72(1): 52–74.

Slemrod, J. (2016). 'Caveats to the Research use of Tax-return Administrative Data'. *National Tax* Journal, 69(4): 1003–20.

Stats SA (2012). 'Census 2011: Metadata'. Pretoria: Statistics South Africa, Cape Town: DataFirst. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/485/download/8285.

Stats SA (2013). 'Survey of Employers and the Self-Employed 2013: Metadata'. Pretoria: Statistics South Africa, Cape Town: DataFirst. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/485/download/8285.

Stats SA (2017a). 'General Household Survey 2017: Metadata'. Pretoria: Statistics South Africa. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/723/download/9842.

Stats SA (2017b). 'Labour Market Dynamics Survey 2017: Metadata'. Pretoria: Statistics South Africa. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/727/download/9904.

Von Fintel, D. (2006). 'Earnings Bracket Obstacles in Household Surveys—How Sharp Are the Tools in the Shed?'. Stellenbosch Economic Working Paper 08/06. Stellenbosch: Department of Economics and the Bureau for Economic Research at the University of Stellenbosch.

Wallgren, A., and B. Wallgren (2007). *Register-based Statistics: Administrative Data for Statistical Purposes*. Vol. 553. New York, NY: John Wiley & Sons.

Wittenberg, M. (2017). 'Measurement of Earnings: Comparing South African Tax and Survey Data'. REDI3x3 Working Paper 212. Cape Town: SALDRU.

World Bank (2012). 'Chapter 6: Globalization's Impact on Gender Equality: What's Happened and What's Needed.' In *World Development Report 2012: Gender Equality and Development*. Washington, DC: World Bank.