UNITED NATIONS
UNIVERSITY
UNU-WIDER

WIDER Working Paper 2020/27

Job spells in an emerging market

Evidence from apartheid and post-apartheid South Africa

Robert Hill, Kezia Lilenstein, and Amy Thornton*

March 2020

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: Few studies exist on job duration in developing labour markets—an important omission both in our understanding of such markets and for the job duration literature, which is mainly based on developed-country case studies, which differ in structural ways. The main reason for this is likely data constraints in developing countries, since job duration analysis has intensive data requirements. Recently, two data sets meeting these requirements became publicly available in South Africa, covering the apartheid and post-apartheid eras. We use these data to provide a broad baseline about job duration in South Africa using survival analysis techniques with three main aims. First, we investigate to what extent stylized facts from the rest of the literature apply to South Africa; second, we analyse trajectories through the labour market; and third, we home in on early-career trajectories. South Africa broadly adheres to stylized facts about job duration: long-term tenure is a common feature of the labour market and the job hazard is non-monotonic and declines with tenure. Trajectories through the South African labour market, though, deviate from the developed-country case and we link this to the importance of labour market segmentation in South Africa. The influence of apartheid-era labour market policy on restricting the freedom of certain groups is clear in the results from this era, and interpretation of results from this time must take this historical context into account or risk reaching seriously misleading conclusions.

Key words: apartheid, dynamics, job duration, job spells, survival analysis, worker vulnerability

JEL classification: C41, J64, J71

Additional information on data used by the authors can be found in the online appendix accessible here (https://www.wider.unu.edu/publication/job-spells-emerging-market).

*Development Policy Research Unit, School of Economics, University of Cape Town, Cape Town, South Africa; corresponding author: kezia.lilenstein@uct.ac.za

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

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-784-2

https://doi.org/10.35188/UNU-WIDER/2020/784-2

Typescript prepared by Gary Smith.

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

1 Introduction

Fundamentally, the two primary features of a job are its wage and how long it lasts (Diebold et al. 1997). Today, there is an extensive literature on wages in the developing world thanks to the expansion of household surveys and other microdata in these countries over the past 50 years. However, far less work has been conducted on job duration,[1] primarily due to data constraints since the requirements for job spell analysis make it highly data intensive. Job duration analysis is useful in two key ways for emerging markets: first, the length of a given job relates directly to popular concerns around declining employment stability and worker vulnerability; and second, at a macro-level, job spells provide signals about dynamism in the labour market in general. In these respects, the developing world represents an important omission in the job duration literature both in terms of filling a gap in our understanding of emerging labour markets and in contributing to a literature that is dominated by developed-country case studies, which differ in structural ways to developing economies. Higher levels of uncertainty, unemployment, inequality, and poverty in emerging markets, for example, can spill over into differences in the distribution of job duration. Recently, two data sets which are ideally set up to study job duration have become publicly available in South Africa, covering the apartheid and post-apartheid periods, respectively. We use these data sets to do three things: (1) characterize job duration in South Africa; (2) investigate worker trajectories through the labour market; and (3) describe trajectories for an early-career sub-sample.

Much of the job duration literature is occupied with ascertaining whether there has been a decline in employment stability (Farber 2009; Hollister and Smith 2014). These concerns arose after the onset of outsourcing in the 1980s, but remain relevant today with the rise of precarious work, the gig economy, and the advance of automation and the fourth industrial revolution (Kalleberg and Vallas 2017; Katz and Krueger 2019). The consensus is that there has been a decline in employment stability and labour market dynamism, indicated by drops in the shares of 'lifetime' jobs (jobs lasting more than 20 years) and, counter-intuitively, a rise in the median length of current tenure in the United States (Farber 2009; Hollister and Smith 2014; Hyatt and Spletzer 2016). In their analysis, Hyatt and Spletzer (2016) link this rightward shift of the current job tenure distribution to a decline in labour market dynamism as hiring rates slackened after recessions in 2001 and 2008: when there are fewer new hires, median tenure rises. Farber (2009) attributes the decline in lifetime jobs to structural changes in private sector-employment, to which the public sector is less susceptible. Specifically, competitive forces have strengthened, driven by four major trends that Kalleberg and Vallas (2017) identify as underpinning an increase in precarious work since the 1980s: globalization, financialization, digitization, and deunionization.

In locating South Africa within these broader trends, it is important to lend a developing-country perspective. For one, the rise in concern about precarious work in the West relates to a resurgence of precarity after roughly a century of setting up the 'standard employment relationship', characterized by stable earnings, long-term job security, social benefits, and statutory protections; whereas for the majority of the developing world, this was never the norm (Kalleberg and Vallas 2017). During most of the twentieth century, South Africa was under the rule of the apartheid racial segregation regime that exploited Black populations for the benefit of a white minority. The standard employment relationship was reserved for whites during this time, while the remainder of the population laboured under extremely vulnerable conditions (Nattrass and Seekings 2011). After the first democratic election in 1994, new legislation was put in place to extend the benefits associated with the standard employment relationship

[1] The literature on unemployment duration is much larger than that on employment duration, and there are a few studies on developing-country markets in this field. Dendir (2006) examines unemployment duration in Ethiopia and Tansel and Taşçı (2010) study the Turkish case.

to the rest of the population.[2] However, population groups structurally exploited during apartheid, such as women and Black people, remain at a disadvantage in the labour market today.

In the post-apartheid era, the South African economy exhibits some of the worst levels of income inequality and unemployment in the world, and about half of the population remains in poverty (Statistics South Africa 2017). In 2014, the Gini coefficient stood at 0.69 and unemployment by the broad definition exceeded 30 per cent in 2010, and remains at this high level today (DPRU 2017). The labour market is by far the dominant driver of aggregate inequality and poverty in South Africa (Leibbrandt et al. 2012), and a formal-sector job is a strong predictor of a household escaping poverty (Schotte et al. 2018). Worker vulnerability is therefore a central political and social concern and a critical dimension of this is uncertainty about current and future employment. Such uncertainty undermines the ability to plan and save for the future and ultimately to maximize welfare over the lifetime. In a recent analysis of attitudinal data, job stability was the characteristic of a job that South Africans valued the most (Mncwango 2016). Many dimensions of worker vulnerability can be evaluated with the annual labour market data collected by the national statistics bureau, Statistics South Africa (StatsSA) (e.g. leave entitlement, average work hours; see Bhorat et al. (2016) for a comprehensive discussion), but until now, job instability has been difficult to capture due to the survey design of much of the South African labour market data.

This study uses a panel of administrative tax data to study job spells in the post-apartheid period, and uses a retrospective survey collected at the end of apartheid which includes episodic employment histories to study the apartheid era. There are three main parts to the analysis in this paper. In the first instance, we want to describe job spells in South Africa. These results are interpreted in light of stylized facts in the existing literature, trends in job duration in the developed world, and the specific South African context. The second and third parts of the analysis exploit the special structure of our data sources to learn about how workers move through different job spells, in general, and at the beginning of their careers, in particular.

Overall, we find that there has been a decline in employment stability as measured by a decline in the share of lifetime jobs, an increase in median current tenure, and a likely decline in the length of the average completed job spell since the end of apartheid. Slightly over half of workers mainly work jobs lasting five years or more; about 40 per cent mainly work jobs shorter than this; and the remainder of just less than 10 per cent work mainly unstable jobs, with women over-represented in the last case. Job duration in South Africa largely adheres to the same stylized facts as job duration in the developed world, and most departures from these tenets can be explained by the influence of apartheid-era labour market policies on the labour dynamics of certain groups, which must be accounted for when interpreting results for this period. Trajectories through the labour market, however, appear to operate differently in South Africa than what has been found in the literature. We link this result to structural differences between developed and developing labour markets, in particular the higher level of segmentation in the South African labour market in comparison with developed labour markets.

2 Theory and background to job duration

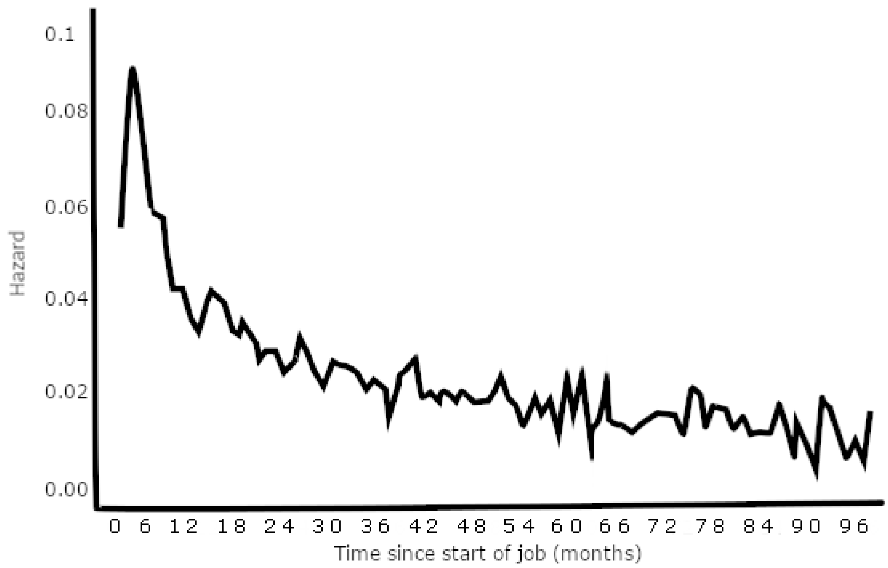
Farber (1999) summarizes three main stylized facts about job duration in his analysis of the United States: (1) long-term employment is common; (2) most new jobs end early; and (3) the hazard of a job

[2] The two main examples here are the Basic Conditions of Employment Act of 1997, which lays out worker entitlements to various types of leave, maximum hours of work, and working conditions, and the country's first sectoral minimum wage in 1999. It is not a requirement for firms to make mandatory medical scheme or pension contributions, but firms are required to contribute to the Unemployment Insurance Fund (UIF), which supports workers for a period of less than a year after termination of employment.

ending decreases with tenure. Farber (1994) notes an additional two stylized facts relating to worker trajectories: (4) mobility is persistent—some people have stable careers and others have more volatile ones; and (5) a worker's most recent labour market mobility is most relevant for their current job duration. These findings have by and large been confirmed by other developed-country studies such as those from Germany (Boockmann and Hagen 2008), Canada (Christofides and McKenna 1993), Portugal (Horny et al. 2012), the UK (Booth et al. 1999), and a comparative study with a developed-country sample (Burgess 1999). In terms of establishing a baseline for South Africa, a useful first exercise is assessing to what extent South Africa conforms with these stylized facts.

The typical shape of the job duration hazard found throughout the literature is presented in a reproduction of a result from Farber (1999) in Figure 1. The shape is at first non-monotonic and then decreasing with time, meaning that most new jobs end early and that the longer a particular employment relationship has lasted, the more likely it is to continue to last. One theory explaining the initial non-monotonicity is that there is a lag in the matching process caused by information asymmetry before the onset of the job (Jovanovic 1979). Match quality is an 'experience good' and can only be ascertained once the job has begun. All information about the match quality becomes available to both employers and employees as time on the job passes, allowing them to evaluate whether the match should continue or be terminated. As such, the peak of the hazard occurs some time after the job has started, at about six months in the United States (Farber 1999) and the United Kingdom (Booth et al. 1999), and better outside options for employees can steepen this peak. With higher wages offered by other employers or a high job-finding rate, on-the-job search increases, leading to shorter jobs.

Figure 1: The job duration hazard in the United States



Source: authors' construction based on figure 5 in Farber (1999).

Thereafter, the hazard declines as the continuation of the employment relationship is reinforced over time by the ongoing accumulation of job-specific capital and investment by both employers and employees (Becker 1964). Job-specific capital includes knowledge that the worker accumulates on the job that is of no value in any other employment relationship. It is costly to accrue this capital for both the employer and employee in terms of the search process and on-the-job training (Becker 1964; Jovanovic 1979); therefore both parties prefer good-quality, long-lasting matches.

Job-specific capital (which includes match quality) stands in contrast to worker heterogeneity as a theory able to explain variation in job duration. This theory models this variation as a consequence of two main 'mobility types'. High-mobility types change their job more often than low-mobility types, meaning there are many short and many long jobs, although this theory cannot explain the initial non-monotonicity of the hazard (Booth et al. 1999; Farber 1999). This is also complicated by the evolution of mobility types, with people usually becoming more stable as they age and enter family arrangements. Farber (1994) attempts to distinguish between these two theories by controlling both for tenure and the number of prior mobilities in a hazard model. There is evidence in favour of both, and we consider them when setting up our models.

The shape of the hazard is therefore informative about two important labour dynamics. The initial peak is informative about how dynamic the job-matching process is. The level and the rate of decline after the initial peak is informative about job security and stability. For example, a steep peak is generally interpreted positively in that matching is free to take place, there are few impediments to search and there is a variety of work opportunities from which to choose. In other words, there is a healthy and dynamic matching process. A lower or more diffused peak suggests this process is more stagnant, possibly because of restrictions on the ability of employers to terminate employment relationships or because workers do not have a diverse set of outside options should they find themselves in an ill-fitting employment relationship. The reason a job ends also affects when the peak of the hazard occurs. In the United Kingdom, hazards peaked sooner (within 12 months) for workers who quit their jobs, but peaked within the first 18 months for workers who were laid off or quit for other reasons (Booth et al. 1999).

A hazard that declines relatively steeply after the peak indicates that there are 'job security returns' to tenure and the mechanism that reinforces these returns—the worker's accumulation of job-specific capital and experience—is functioning well. By contrast, a flatter hazard suggests this mechanism is weaker because job security is reinforced at a lower rate over time. A higher or lower level of the hazard at later job durations implies less or more job security and stability in the longer term, respectively. The institutional environment—ease with which workers can be hired or fired and whether labour policy supports employment stability—has also been found to affect job duration. For example, France, where labour policy is very protective of workers, had much higher shares of long-term tenure than the United States, where there are few obstacles to hiring and firing (Burgess 1999).

In sum, then, an 'ideal' hazard could be one that peaks quickly, sharply, and at a relatively high level, but then steeply declines thereafter and settles at a low level. This would indicate a market in which the matching process is free and uninhibited, but once a suitably matched employment relationship is set up, the job is stable and secure. Such a shape is usually associated with labour market advantage. For example, the more highly educated are more mobile because their skills are more generally applied and they can easily accrue job-specific capital (Burdett 1978). Higher mobility in the job duration literature signals the freedom to leave poorly matched jobs afforded by a high job-finding rate and good alternative job matches in terms of quality and pay.

2.1 South African labour market dynamics

Although there is a perception in the business community and among some scholars (Fedderke 2012; Go et al. 2009) that the South African labour market is very rigid, the weight of the evidence based on panel data suggests that there is actually a surprising amount of mobility (Banerjee et al. 2008; Cichello et al. 2014; Essers 2016; Ranchhod and Dinkelman 2007). Kerr (2018) uses a panel of tax data to analyse worker and job flows between 2012 and 2014. While international comparisons are difficult due to data, time period, and methodological differences, Kerr (2018) finds that the level of worker flows in the formal South African labour market is in the middle-to-upper range for countries for which there are data. Kerr (2018) finds a rate of flow (the sum of hires and separations) of 52–54 per cent of average employment between 2012 and 2014, meaning that more than one in two job matches either forms or breaks up every year.

Zizzamia and Ranchhod (2019) use five waves of the National Income Dynamics Study panel to characterize transitions between being employed and not employed. They detect a high level of volatility between 2008 and 2017. They find that 27 per cent of the sample fell into volatile employment (employed for only two or three of the five waves) and another 27 per cent were persistently unemployed (employed in only one or no waves). Women were much more likely to find themselves in persistent unemployment, with 17.54 per cent never being employed in five waves compared to only 5.74 per cent of men. Men were also much more likely to be in stable employment: 63.7 per cent of men were employed in four or five waves compared to only 34.7 per cent of women. On the other hand, roughly equivalent shares of men and women in the sample experienced volatile employment.

Banerjee et al. (2008) emphasize the importance of dynamic analysis using the rotating Labour Force Survey panel data. In their study, high levels of individual mobility occurred over the same period that aggregate unemployment remained virtually unchanged. For example, in a six-month period between 2002 and 2003, only half of the informally employed remained so by the end; and only half of the workers searching for work were still doing so six months later; but, aggregate unemployment was stable. Overall, consensus is that the South African labour market is surprisingly volatile and that the working lives of women are especially so.

2.2 What is the trend in current tenure in post-apartheid South Africa?

Using South African national labour market survey data, we investigate the trends in current tenure in order to provide context for the forthcoming analysis on job spells. We use the Labour Force Surveys (LFS) (2000–07) and the Quarterly Labour Force Surveys (QLFS) (2008–present) for this analysis.[3] The (Q)LFS question asks respondents how long they have been working for their current employer or running their current business. In other words, this is a worker-level estimate of tenure at a given point in time, which we can expect to yield quite different results to a sample of completed job spells. Specifically, we would expect current tenure to be longer in duration than tenure in general since tenure is accumulative and increases with age (Farber 1999). In cross-sectional data, we are therefore more likely to be seeing people in the job they 'settle' in for a few years, as opposed to the first few short job matches that have already ended—this is especially true for older workers.[4]

Panel A of Figure 2 plots the trend in median current tenure across the post-apartheid period. There is a clear structural break in 2008: the formal-sector trend in median tenure decreases until 2008 and then starts increasing. Median tenure in the formal sector at both the beginning and end of the period is about five years, but is about 3.4 years at its lowest in 2008. This is both the year of the global financial crisis but also the year the survey instrument changed from the LFS to the QLFS. One might be tempted to think increasing tenure is indicative of more stability in the labour market, but longer median job duration may actually belie a less dynamic labour market.

Two other trends also charted in Figure 2 are supportive of this idea, that of the informal sector and that of the unemployment rate. If most new jobs end early, then rising job duration may be indicative of fewer

[3] We use version 3.3 of the Post-Apartheid Labour Market Series (PALMS), which is a harmonized series of South African labour force surveys for the years 1995–2015 (Kerr et al. 2019). The original data for the series come from annual nationally representative cross-sectional labour force surveys collected by StatsSA, the national statistics bureau, since 1995. These were the October Household Surveys (1995–99), LFS (2000–07), and the QLFS (2008–present).

[4] Another reason we may expect the (Q)LFS current tenure estimate to be an overestimate of tenure in general is the phrasing of the question. How long you have been working for an employer versus how long you have been working in a specific job can have slightly different interpretations (Farber 1994). If you work for the same employer for a long period, but on an on-and-off basis, then this type of question could overstate stability of tenure in the sense that we are interested in.

new jobs. This is plausible, considering that the unemployment rate moves closely in tandem with the median job duration. If firms are hiring less after the recession, this may mean that fewer people are able to exit unemployment. Hyatt and Spletzer (2016) note the same trend of increasing median job tenure for the United States from 2000.[5] Their analysis attributes this partly to an ageing working population, but also to a drop in firm births and hiring after the recessions in 2001 and 2008. Counter-intuitively, then, longer median job duration in the post-2008 period could reflect the hardening of the barrier between the employed and unemployed. If this is the case, an increase in median job duration could be associated with increasing inequality.

This is supported by the trend in the informal sector. McKeever (2006) notes that during apartheid, women spent similar amounts of time in formal and informal employment, whereas men spent considerably less time in informal compared to formal employment. McKeever (2006) suggests that this could be because men use the informal sector as a 'springboard' into formal-sector employment, whereas, due to their general disadvantage in the labour market, women find it harder to make this transition and informal employment is a more long-term prospect for them. Therefore the convergence between male and female tenure in the informal sector could indicate that men are finding it harder to make the transition to the formal sector, indicating a less dynamic labour market.

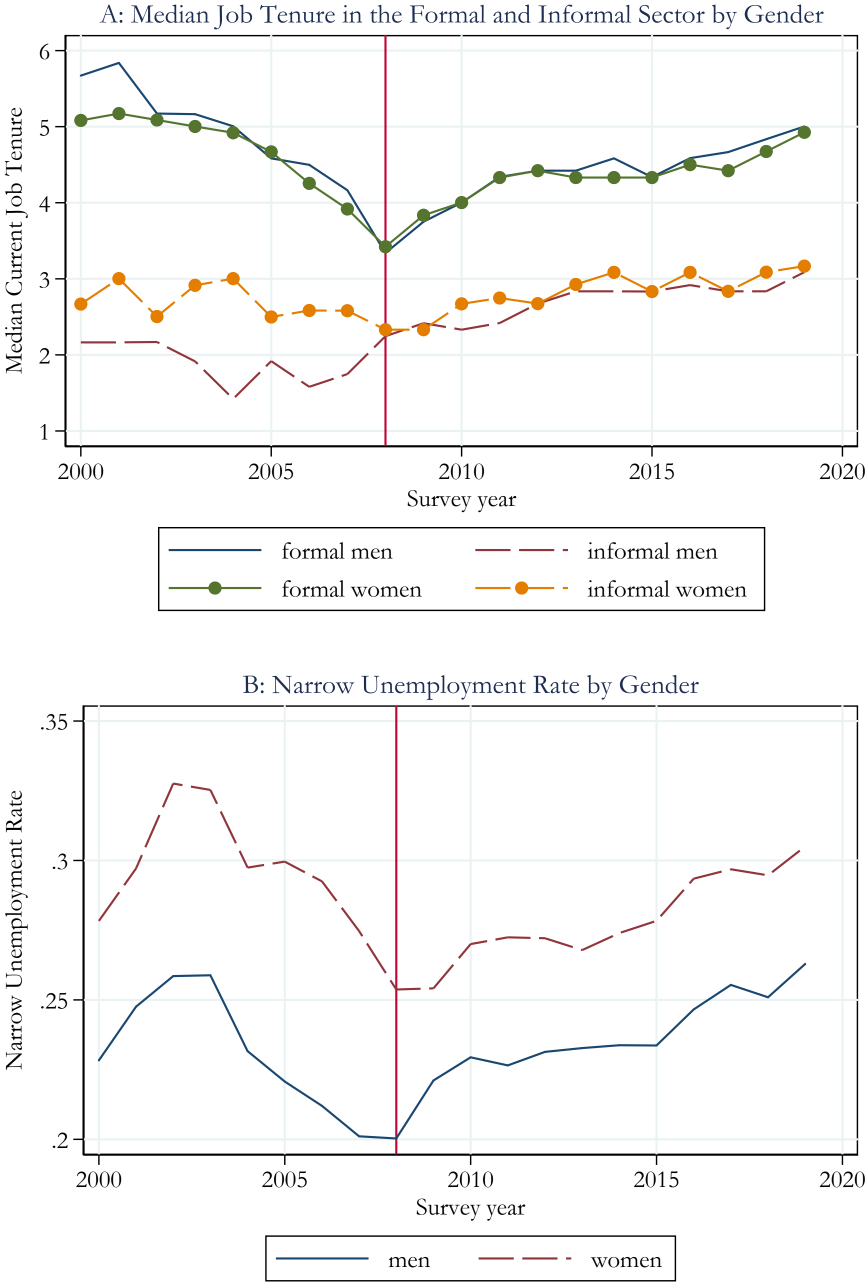
Also of interest is the decline in formal-sector tenure pre-2008, which coincides with a fall in the unemployment rate. This aligns with McKeever's (2006) findings related to job duration during the apartheid era. The apartheid era coincided more generally with the 'golden age' of the standard employment relationship around the world; that is, long-term job security with stable earnings and benefits. This global trend, in tandem with apartheid policies, may have set up the relatively long job durations found by McKeever (2006). During apartheid, Black South Africans required extremely long tenure (e.g. 10–15 years) with a white employer to qualify for Section 10 rights which enabled them to live in urban areas. Jobs were much more numerous in urban areas compared to the Homelands,[6] where labour markets collapsed in the 1970s leading to a surge of unemployment (Nattrass and Seekings 2011). Long job durations during apartheid could therefore be indicative of the line between the employed and unemployed becoming increasingly rigid. In line with these trends, we see both that median job duration and unemployment was high in 2000. The decline of both of these trends in the following seven years may be indicative of increasing dynamism prior to the global recession in 2008.

Lastly, we have previously discussed literature that describes the South African labour market as more dynamic than perceived by the business community. How do these findings square with the idea that the labour market has become less dynamic since 2008? Here, we can differentiate between dynamism as a labour market feature and volatility as an individual worker-level experience. Even as the labour market itself becomes less dynamic, individual workers may be experiencing volatile work trajectories. This is because a smaller and smaller portion of the workforce is enjoying persistent employment, implying a less stable career path for those who are not in this group. As the share of the persistently employed falls over time, the combined share of the inconsistently employed and persistently unemployed must grow. Overall, this is important context for the analysis going forward.

[5] In fact, the United States exhibits the same U-shape except it is more stretched over time. The decline in median tenure started in the 1960s with the onset of increasing world trade. Job tenure reached a low point in the 1990s when much literature on the topic was concerned with declining job tenure as an important policy question for America (Farber 1999). Since 2000 and the two recessions of 2001 and 2008, the US labour market has become less dynamic and median job duration has been rising steadily.

[6] Homelands were rural areas designated for Black South Africans under apartheid

Figure 2: Trends in job tenure and unemployment in South Africa: 2000–19



Notes: the red vertical line marks the year 2008, the year of the global financial crisis, and also the year the survey instrument changed from the LFS to the QLFS.

Source: authors' construction using a cross-entropy weighted stacked series of the PALMS data.

3 Research agenda

Studying job duration is very data intensive, since what is required is job-specific start and end dates which could be collected either in episodic employment history data or panel data that has a suitable number of waves. A major motivating factor for this study is that two data sets with these structures have recently been publicly released, allowing for the study of job duration in post-apartheid South Africa for the first time. The first data set is anonymized administrative tax data collected by SARS and collated into a data set housed in the National Treasury. The data are continually being updated, but the version we use covers the period 2010–17 and theoretically captures all tax-paying individuals in the country. The second data set is the Survey of Socio-Economic Opportunity and Achievement (SSEOA), a retrospective survey collected by a local and international team of researchers in 1991–94 from a nationally representative population of adults, and which includes employment histories. The survey is retrospective and therefore covers the apartheid era.

These two data sets provide nationally representative data on completed employment spells for two very different periods of South Africa's history. Although we can identify the same concept in the data, the two sources are very different. This means one immediate research output is understanding the strengths, weaknesses, and potential complications of using these data sets to study job duration. The SARS data are a panel with close to a complete formal-sector sample, and include some interesting firm-level variables. However, the series only covers seven years, which is relatively short for job spell analysis. Further, the data have limited demographic information since variables are limited to what information is collected on tax forms, meaning we can only identify people as either 'employed' or 'not employed', for example.

The SSEOA data cover a much longer period—we use a period of 1951–91/94, meaning the data can include entire careers in some cases. The data include rich demographic information, including all possible labour market statuses and informal sector information, but no firm-level information. However, the sample size is much smaller, at about 9,000 people, and there is ongoing work on the quality of the data being carried out by DataFirst, the data repository housing the data. The two data sets also suffer from different types of measurement error. The SARS data will be influenced by the idiosyncrasies of filing tax, whereas the SSEOA could be subject to recall bias.

Our aim is to use these two data sets to provide a broad baseline analysis of job duration in South Africa. In this respect, this paper has three main research goals: to describe job duration in South Africa; to learn something about employment trajectories; and to learn something about early-career trajectories, in particular. In the first instance we want to describe job duration in South Africa since few, if any, studies on employment duration in developing countries exist to our knowledge. Higher levels of uncertainty, unemployment, and inequality may lead to different outcomes compared to developed labour markets. As such, our first main research goal is to describe job duration in South Africa compared to stylized facts and trends from the developed world, and locate these within a South African context. The stylized facts mentioned above are: (1) long-term employment is a feature; (2) most new jobs end early; and (3) the hazard declines with tenure (Farber 1999). We investigate the incidence of long-term jobs and use survival analysis techniques to describe the survival and hazard functions of job duration. We run a regression on the determinants of the hazard, controlling for individual- and firm-level variables and worker frailty where feasible. Throughout the paper, results are discussed by gender. This is a common practice in the job duration and tenure literature (Farber 1999; Hollister and Smith 2014), and gender also represents an important dimension of worker vulnerability in South Africa (Zizzamia and Ranchhod 2019).

Our second research goal is to learn something about job trajectories. Few data sets collect information about what South Africans' careers look like over their lifetimes. Our data structures allow us to probe this topic, contributing to our understanding of worker vulnerability: for example, do some people have many short jobs and some people have just a few long jobs? We also investigate how people move through the labour market. Farber (1994) found that a worker's mobility immediately prior to their current mobility was most relevant for the hazard of the current job. Building on this idea, we investigate the impact on the hazard of labour market status immediately prior to a formal job spell using a regression model similar to that used by Boockmann and Steffes (2010) on German data. If you were previously employed in a formal-sector job, does that increase or decrease the risk that your current formal-sector job will end, compared to if you were previously not formally employed? Boockmann and Steffes (2010) find that being previously unemployed or not employed increased the hazard compared to the reference category of job-to-job change, although the size of these effects was greatly reduced by controlling for firm fixed effects. Those who were persistently in the formal sector had more stable formal-sector trajectories.

Our third goal is to investigate employment trajectories, particularly at the beginning of careers. Youth unemployment is excessively high in South Africa; in 2016, the broad unemployment rate for youth aged 15–24 years stood at 60.4 per cent (DPRU 2017). Evidence suggests that unemployment is persistent (Jackman and Layard 1991), undermining the employment prospects for youth who fail to secure a job early on in their careers. It is therefore very important to understand how young people move through the labour market, and one way to do this is to understand heterogeneity across job sequence. Booth et al. (1999), for example, find that the hazards of the first versus the fifth job for Britons in the twentieth century look very different and some of these differences are ascribed to experience in the labour market. Data constraints limit us to comparing first jobs to second jobs in South Africa, but even this comparison is informative. Second jobs may look quite different to first jobs because individuals are able to draw on the labour market experience that they previously did not have.

In the next section, we describe the data in detail. This is followed by a section on method, which explains how survival analysis is applied to job duration analysis. We then describe our empirical analysis, aimed at achieving our three main research goals. Our results follow.

4 Data

4.1 The SARS panel data

The data used in this analysis are primarily from the SARS anonymized individual- and firm-level tax data, made available in collaboration with the National Treasury in the form of a panel data set. The tax data used for this paper are a hybridization of two data sources. The first is the individual-level tax data, which are collected via the employer-issued IRP5 forms and cover information such as total remuneration, period of employment, and some personal characteristics. Currently, IRP5 data are available for the 2008–17 tax years.[7]

The second source of data for the construction of the panel used in this paper is firm-level data, which are collected from corporate income tax (CIT) data. These data are extracted from the income tax return for companies. These firm-level data are also available for the 2008–17 tax years. Using the firm indicator in the IRP5 data, one can merge the firm-level characteristics into the IRP5 data to create an individual-level panel that includes characteristics such as the industry of the firm, or common accounting quantities such as total liabilities of the firm, or debt-to-equity ratios.[8]

An advantage of the data is that they theoretically capture every formal-sector employment spell in South Africa in the period covered. Note the restriction to the formal sector does not constitute a substantial loss in representing the employed population in South Africa, since South Africa has an unusually small informal sector (Banerjee et al. 2008). Further, the data should be free from recall

[7] The South African tax year runs from 1 March of the previous year to 28/29 February of the indicated year. In the case of the 2008 tax year, for example, this would run from 1 March 2007 to 29 February 2008.

[8] It should be noted that the panel is unbalanced due to individuals not appearing in the data every year. This is because entries in the tax data relate to periods of employment, and specifically in the formal sector. Periods of unemployment, or employment in the informal sector, are not captured in the data. As a result, individuals transferring between these states will not be present in every year of the created panel.

bias or measurement error due to fieldworker practice, and can be employed as an individual or a job panel. However, the data are prone to measurement error arising from the idiosyncrasies of tax data, and we return to this issue later when describing how we identify jobs. The main disadvantage of the tax data is the lack of personal demographic information in the panel. The individual-level data include information only on an individual's age and gender at the current time. Other demographic factors thought to influence employment spells, such as race, occupation, or educational attainment, are not available in the data.

The tax data set requires substantial cleaning before it can be used for analysis. To this end, we make certain assumptions regarding what constitutes a valid individual in the tax data. We use the same assumptions as Bhorat et al. (2019) to determine our sample. These assumptions are detailed below:

1. Observations are only kept if they are representative of a natural person. In the IRP5 data, observations can be classified as clubs, estates, partnerships, and welfare organizations. These observations were removed from the data in order to make sure that only valid individuals remained.

2. The sample is restricted to only include those individuals of working age—that is, those between the ages of 15 and 64 years of age.

3. Individuals who do not have any income data are removed from the data set, as it is unclear whether these represent legitimate formal-sector job spells as opposed to administrative measurement error.

After limiting the sample in this way, we discovered that only 3.5 and 3 per cent of the observations in the 2008 and 2009 data were valid observations. As a result, we opted to restrict our analysis to the 2010–17 data sets. It should also be noted that the gender identifier available in the data is taken from the individual's South African ID number. As gender is a key explanatory variable in this analysis, the sample is restricted to individuals with an ID number—that is, citizens and permanent residents of South Africa.

Identifying jobs and job duration in the SARS panel

Defining job spells is complicated by the fact that many individuals have data from multiple IRP5s for the same firm in the same financial year. It is not clear in these cases whether these represent distinct jobs for this individual. On the one hand, an individual may, for example, work in March for a firm and then work for that same firm again in August. This may reasonably be seen as two separate jobs. On the other hand, multiple IRP5s may indicate reissuing of an IRP5 for only one legitimate job spell, or simply an error in the data. In order to address this, we cleaned the data in the following way:

1. We began by cleaning the data of invalid job spells. Job spells are defined as invalid if the recorded start or end dates of the job were not within the tax year in question; where the end date of a period of employment was recorded as occurring before the start date for that employment period; or where at least one of the start or end dates was missing. In the first case—where the recorded job start or end dates lay outside the tax year in question—the date was updated to reflect the first or last day of the relevant tax year. In the second and third cases, the observations were dropped from the data.

2. Where there are two or more IRP5s for an individual in a specific firm in a specific tax year and the start and end dates of those IRP5s are exactly the same, we treat these as duplicate jobs and keep only one observation.

3. Where there are two or more IRP5s for an individual in a specific firm in a specific tax year and the start and end dates of those IRP5s overlap, we combine these into one job spell. We do this by keeping the earliest start date and latest end date of the two or more jobs.

4. In addition, we combine data from IRP5s for the same individual in the same firm into one job spell where the gap between the end date of the first job and the start date of the second job is less than or equal to three months. This is done both within a specific tax year and across tax years. In other words, if an individual worked in a firm until December 2015 in the March 2015 to February 2016 tax year, and then started work again in that same firm in March 2016 in the March 2016 to February 2017 tax year, this was coded as one job. This is done for reasons of measurement error, discussed below.

Measurement error is a substantial problem when attempting to analyse job duration in the data, in that it is both widespread and systematic. We have little idea of how widespread the measurement error is, but some idea of how systematic it is. Kerr (2018) argues that there is a severe amount of measurement error in the employed-from and employed-to dates. Kerr (2018) finds a pattern whereby the number of jobs steadily increases over the tax year until reaching a peak in December, followed by a steep drop in the number of jobs in the last two weeks of the tax year. Most industries then 'bump back up' to the right number of jobs at the beginning of the year, with two notable exceptions: agencies (including temporary employment services agencies) and finance, insurance, real estate, and business services, which account for 90 per cent of the difference. This problem, plus additional measurement concerns (like firms issuing multiple IRP5 forms for the same person), motivated us to define a rule to identify a job even if there is some interruption in employment in the data. If a worker experiences a break in employment that is less than three months, and then commences working for the same employer, we count this as one job given the high chance that the interruption could be measurement error. This rule is very similar to that used by Boockmann and Steffes (2010) in their analysis of administrative panel data similar to the SARS panel.[9]

In other words, we face a trade-off between under- and overestimating the number of jobs. The systematic pattern in the number of jobs over the year found by Kerr (2018) suggests that using the data as it is puts us in danger of underestimating the number of jobs at certain times of the year (if too few jobs are reported at the beginning and in the last two weeks of the tax year). This in turn puts us in danger of underestimating true job duration if work interruptions seen in the data are not real interruptions. We therefore combine job spells where the first job occurred close to the end of the financial year and the second job occurred close to the beginning of the subsequent financial year. This is an effort to partly overcome this problem, especially at the beginning and end of the tax year if many workers are experiencing this work interruption at this point in time. There is therefore also a question of to what degree we are introducing error with our rule. If work interruptions of three months or less (for the same employer) are genuine, then we will be overestimating job duration. However, we feel compelled to make an effort to deal with the systematic measurement error that we are aware of and do our best to be as explicit as possible in our data set-up.

Sample selection

We use two samples from the SARS panel in this paper: what we term the full sample and the early-career sub-sample. Because this paper applies survival analysis techniques to the data, it is statistically and theoretically important to define the job start date. We include in our full sample all jobs in 2011 or later

---

[9] These authors use administrative panel data from 1991–2001. They differentiate between what they call a 'job-to-job' change and a 'recall'. A job-to-job change is defined as a worker leaving one job and starting another at a different firm either immediately or within 60 days of the previous one. They define a recall as a worker ending and then returning to the same job within a period of 91 days.

that were not present in 2010. Our full sample can then be defined as including any person who started a job in 2011 or later and includes about 13 million people and 27 million jobs (Table 1).[10] Men make up most of the sample, with about 7.4 million men with 15 million jobs over the period. There are about 5.6 million women with 11 million jobs. This means that on average we are observing two jobs per person over the period 2011–17, with marginally more for the average man (2.1) compared to the average women (1.9). Table 1 also reports the extent of right-censoring; that is, jobs that are ongoing at the time the study period ends. We find that 20 per cent of people in our full sample have jobs that are ongoing at the end of 2017.

We use an early-career sub-sample of this full sample to home in on the dynamics of job sequence; that is, how first jobs look different to second jobs. We therefore need to be able to convincingly identify an individual's first job and, since legal working age in South African starts at 15 and the data are only seven years long, this is impossible for people older than 21 years in these data. When we observe a 30-year-old in our data in 2011, we do not know whether this is their first or fifth job, for example. As such, we are limited to analysing a sample of 15–21-year-olds.[11] Table 1 reports that the sub-sample comprises around 300,000 individuals, working in total about 420,000 jobs over the seven-year period. About 310,000 of these are first jobs and 78,000 are second jobs. Again, there are more men than women, although we observe the same average number of jobs per man and woman (1.4). The sample includes 160,000 men with about 220,000 jobs and 140,000 women with about 200,000 jobs.

Another important consideration for survival analysis is the issue of concurrent jobs, or when an individual is working two or more jobs in different firms simultaneously. Jobs cannot overlap when carrying out regressions on survival data because this violates the assumption that events are independent. However, concurrent jobs occur at a very high rate in the SARS data, as reported in Table 1. About 27 per cent of jobs occur at the same time as another job. We cannot tell if jobs are part-time or full-time in the SARS data so these could be people working multiple part-time jobs, for example. Our sample size is sufficiently large that dropping concurrent jobs should not be a problem from a power point of view. However, we are worried that dropping concurrent jobs will bias our results if concurrent jobs or the types of people who work concurrent jobs are different to their non-overlapping counterparts, which is plausible. Booth et al. (1999), for example, use a rule to select one of the simultaneously worked jobs which is a compromise between conditionality and sample size. The problem with this approach, of course, is that the selected job is still a conditional job even if it doesn't appear so in a cleaned data set. Although we run the risk of violating the independence assumption, we are more interested in having a more representative sample of jobs in South Africa, and so we choose to run our regressions on the full sample of jobs. We run a robustness regression in which we tag only the 'main' job where the main job is defined as the highest-paid of the overlapping set.[12] The effect of only using main jobs reduces the sample by about 16 per cent, as reported in Table 1.

---

[10] This is similar to a rule used by Boockmann and Steffes (2010) in their analysis of similar administrative panel data from Germany. Their data span 1991–2001, and to avoid left-censoring they only include spells that started between 1996 and 2001.

[11] We specially investigated the data in different ways to ensure that 15-year-olds do not appear in the data before their fifteenth year, meaning we can say with confidence that when we observe a 15-year-old in our data it is in fact their first formal-sector job. In our final sample we include only 15-year-olds in 2011; then 15- and 16-year-olds in 2012; then 15-, 16-, and 17-year-olds in 2013; and so on until we also include 21-year-olds in 2017.

[12] Booth et al.'s (1999) rule is to choose the longest job. Although it occurs in a published paper, we do not think this is a good criterion since the authors are selecting observations based on the outcome.

Table 1: Sample description in the SSEOA and the SARS panel

	SSEOA 1951–91/94			SARS panel 2011–17					
	People *N*	Jobs *N*	Censored jobs (share)	People *N*	Jobs PAYE *N*	Jobs tax ref. # *N*	All concurrent jobs (share)	Concurrent Non-main only (share)	Censored jobs (share)
				Full sample					
All	6,162	12,621	0.35	12,957,753	27,129,462	27,059,394	0.27	0.16	0.20
Men	3,739	8,080	0.35	7,358,628	15,324,511	15,165,831	0.27	0.16	0.19
Women	2,423	4,541	0.33	5,599,125	10,895,092	10,818,951	0.27	0.17	0.22
				Early-career sub-sample: job 1 and 2					
All	6,162	9,389	0.33	310,610	422,344	422,714	0.15	0.08	0.30
Men	3,739	5,847	0.33	164,476	223,384	223,626	0.15	0.08	0.30
Women	2,423	3,542	0.32	146,134	198,960	199,088	0.15	0.08	0.30
				Early-career sub-sample: job 1					
All	6,162	6,162	0.30	310,610	310,610				0.29
Men	3,739	3,739	0.30	164,476	164,476				0.30
Women	2,423	2,423	0.30	146,134	146,134				0.29
				Early-career sub-sample: job 2					
All	3,227	3,227	0.38	78,274	78,274				0.32
Men	2,108	2,108	0.39	41,370	41,370				0.31
Women	1,119	1,119	0.36	36,904	36,904				0.32

Source: authors' compilation based on SARS and SSEOA data.

4.2 Retrospective employment history data

The SSEOA was collected by a team led by Donald Treiman from the University of California, Los Angeles (UCLA); Sylvia Moeno, who was part of the Strategic Planning Group at Eskom at the time; and Lawrence Schlemmer, who at the time was the director of the Centre for Policy Studies at the University of the Witwatersrand.[13] The survey was conducted in 1991–94 and collected retrospective life and employment histories from a nationally representative sample of 9,086 respondents over the age of 20. Since the data were retrospective and collected in the early 1990s, many of the employment history data will be representative of the apartheid era, which will form an interesting comparator to the post-apartheid case.

Due to difficulties with sample collection, it was necessary for a second round of sampling of male respondents to improve representivity (Treiman et al. 2017). As such, there are a number of weights available in the data, some of which have been configured by DataFirst, the repository housing the data. To weight our analysis, we use the individual-level cross-entropy weight calibrated on demographic information from both the original and additionally sampled males, called *ceweight2m_ind* (Treiman et al. 2017).

We restrict the sample to formal-sector jobs to improve comparability with the SARS panel and limit the time period to data from 1951 to mitigate against recall bias, following a similar sample restriction by McKeever (2006). This leaves us with a sample size of about 6,000 people working 12,600 jobs in Table 1. Men occur more frequently in our sample and about one-third of jobs are right-censored. Analogously to the SARS panel, we also make use of an early-career sub-sample comprising the first and second jobs people report in their employment histories. The episodic structure of the SSEOA data makes it easy to identify first and second jobs, meaning we do not need the additional age restriction we applied to the SARS panel. Since every person in the sample at least reports a first job, we have about 6,000 first jobs and then about 3,000 second jobs. As usual, men occur more frequently in samples for first and second jobs.

4.3 South African labour market data

In addition to the SARS-NT panel, we use version 3.3 of the PALMS, which is a harmonized series of South African labour force surveys for the years 1995–2015, mainly to describe long-term employment in the post-apartheid period since the SARS panel is too short to do this (Kerr et al. 2019). The original data for the series come from annual nationally representative cross-sectional labour force surveys collected by StatsSA since 1995. These were the October Household Surveys (1995–99), LFSs (2000–07), and the QFLSs (2008–present). Earnings information for the QLFSs is sourced from the Labour Market Dynamics Surveys for the corresponding years. The original surveys all cover approximately 30,000 dwelling units based on about 3,000 primary sampling units drawn from the master sample of the most recent census at the time. A stratified, two-stage cluster sampling design is employed in each case, stratified at the provincial level. Data are self-reported to the enumerator (or by proxy in the case of an absent respondent) and cover the spectrum from basic demographic and household information to detailed labour market data. An advantage of PALMS is that it harmonizes variable definitions across different surveys and includes a set of weights which we use, *ceweight1*, that enables more consistent comparison of the cross-sections over time (Kerr and Wittenberg 2019). Our sample for these data is all employed people between the ages of 15 and 64 with non-missing data on current employment tenure. Current employment tenure is only available from the year 2000 or the beginning of the LFS.

[13] The Human Sciences Research Council was responsible for data collection and the study was funded by the United States National Science Foundation and a group of South African agencies: the Anglo-American/De Beers Chairman's Fund, the Trust for Educational Advancement in South Africa, the Human Sciences Research Council, Johannesburg Consolidated Investments, and the Union Carbide Corporation.

5 Method

This study has three main goals: to describe job duration in South Africa; to learn something about employment trajectories; and to learn something about early-career trajectories, in particular. To do this, we apply survival analysis to the SARS data and the SSEOA data. We therefore first take the time to provide the necessary detail on survival analysis and why and how it has been used in the job duration literature. This is followed up by three subsections describing the specific empirical work carried out to answer the research questions associated with our three different goals.

5.1 Survival analysis

Farber (1994) notes that the two most important sources of variation in job duration are worker heterogeneity and time dependence. This necessitates the use of survival analysis, which is designed to cope with these two aspects. Another advantage of survival analysis is its ability to cope with right-censored data, which is common in duration data and in both of our data sources. Using the usual techniques to describe the distribution of jobs (e.g. the mean, median, density) could be biased because they will exclude ongoing jobs, thus introducing a form of selection bias (Box-Steffensmeier and Jones 2004) if ongoing jobs or people with ongoing jobs are different.

Instead, we can compute the hazard and survivor functions for job duration, which are statistics that account for censoring and time dependence (Box-Steffensmeier and Jones 2004). We define T as a positive random variable denoting survival times with the actual survival time of a particular unit being denoted as t. The variable T, then, is the probability distribution characterized by the probability density function, $f(t)$, and the cumulative density function, $F(t)$. The survivor function, $S(t)$, is the proportion of jobs surviving beyond time t, and is defined as the reverse of the cumulative density function of failure times:

$$S(t) = 1 - F(t) = \Pr(T \geq t) \tag{1}$$

$S(t)$ captures the probability that a survival time, T, is equal to or exceeds time t. $S(0) = 1$ since all jobs are surviving when they begin. This value decreases over time as jobs start to fail, until we reach the maximum time horizon in our study when $S(t)$ is at its lowest. We define start time as the beginning of a given job spell, meaning time resets to zero when an individual begins their second job. Start time in our study is job-related and not related to the year of observation when the survey or panel began.

The second important survival statistic is the hazard rate, $h(t)$. The hazard rate reports the risk that a job will end and accounts for time dependence, or heterogeneity of risk at different job durations. Another way to think about the hazard is as a conditional failure rate: the hazard is the probability of an event failing conditional on an observation surviving until a particular point. In our case, this is the chance that a job will end after t days, given that it has lasted t days. This conditionality 'corrects' the calculation of the chance of a job ending by assessing the chance over the correct risk set. Failing to properly identify the risk set will lead to an underestimation of the risk (Box-Steffensmeier and Jones 2004). Jobs that last one year are different to jobs that last five years, and when assessing the risk that a job of five years will end, we should not be informed by jobs that only lasted one year.

This conditionality is what distinguishes the hazard from the survivor function, which makes the hazard of particular interest. The hazard can also be easily conditioned on covariates and incorporated into a regression framework. The hazard relates mathematically to the survivor function and is defined as follows:

$$h(t) = \frac{f(t)}{S(t)} = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \tag{2}$$

Thus, $h(t)$ is the rate of failure per time unit for the instantaneous interval of $[t, t+\Delta t]$ when Δt is very small, conditional on survival at or beyond time t.

5.2 Empirical strategy

Describing job spells in South Africa

To characterize job duration in South Africa, we aim to assess to what extent South Africa conforms with the three main stylized facts identified by Farber (1994, 1999). These stylized facts are (1) that long-term employment is a feature of the labour market; (2) that most new jobs ended early; and (3) that the risk of a job ending declines with tenure.

One limitation of the SARS data is that the panel is relatively short at only seven years for assessing Farber's (1999) first stylized fact: long-term jobs are a feature. When Farber (1999) and Burgess (1999) discuss long-term jobs, or 'lifetime' jobs, they mean jobs that are 10–20 years long in the United States or rest of the developed world. We therefore utilize a formal-sector sample of the QLFS over a similar period to the SARS data (2010–19) to answer this question. Following Farber (1994) and Burgess (1999), we look at the shares of workers over 45 years of age with current tenure of 20 years or more. We are able to do the same for the formal sector in the SSEOA data that represent the apartheid job market.

Next, we turn to assessing Farber's (1994; 1999) next two stylized facts: that most new jobs end early and that the risk of a job ending declines with tenure. First, we use the SARS and SSEOA job-level panel to organize job spells into the following categories to characterize the job spell distribution, using the full sample in the instance of the SARS panel:

1. Stable job: the share of jobs that have lasted at least five years or more.
2. Moderately stable job: the share of jobs that have lasted more than one year and less than five years.
3. Unstable job: the share of jobs that have lasted more than three months and up to one year.
4. Highly unstable job: the share of jobs that have lasted three months or less.

Following this, we use the job panels to plot the Kaplan–Meier hazard and survivor functions for job duration and by various categories of interest, such as gender, race, occupation, and sector.

The analysis is extended by regressing the hazard on covariates to understand how each influences job duration. Practice in the literature is to apply a Cox proportional hazards model with worker frailty, since worker heterogeneity explains most of the variation in the hazard (Booth et al. 1999; Farber 1994). The Cox model models the baseline hazard non-parametrically and is appropriate when the baseline hazard is not of substantive interest and the researcher wants to model time dependency as noise (Box-Steffensmeier and Jones 2004). However, the non-parametric nature of the Cox model also makes it computationally much less efficient than its parametric counterparts, and this proved problematic since both of our data sets have very large sample sizes compared to what is common in other applications of survival analysis. Essentially, a Cox model with worker frailty failed to compile in both the SSEOA data and the SARS panel.[14]

[14] Boockmann and Steffes (2010) successfully run a Cox model on over 100,000 observations, but this is a sample much smaller than the SARS panel, which has a sample size of 27 million. They also ran the model without worker frailty. We are able to run a Cox model with standard errors robust to clustering at the person level in the SSEOA data, but this didn't compile in the SARS panel after running for three days. One solution here could have been to run the regression on many much smaller sub-samples and then report the distribution of the results.

A non-parametric baseline is appropriate when researchers do not have a strong prior on the shape of the baseline function. However, we actually do have such a strong prior based on previous work in advanced economies consistently showing that a job hazard first increases and then decreases over time. This means that parametric models that allow for such a non-monotonic baseline hazard, like the log-logistic and log-normal, could be appropriate for our analysis. The well-known drawback of parametric models is that results can be quite sensitive to how the baseline is specified (Box-Steffensmeier and Jones 2004). To guard against this, we follow the usual practice of running a set of different models and reporting them all.

We set up a schedule of regressions based on what was computationally feasible for each data set. The frailty model is preferred as it best captures worker heterogeneity by explicitly modelling this variation as a random coefficient. A second-best approach is to model worker heterogeneity in the error and run the regression with standard errors robust to clustering at the person level. Wherever possible, we use the frailty model and otherwise use robust standard errors. For both data sets, we run a log-normal and log-logistic model with worker frailty. We additionally run a generalized gamma, but are limited to using robust standard errors since using frailty with the generalized gamma is not yet available in Stata. Additionally, a Cox model with robust standard errors is run on the SSEOA data with the Efron method to break ties.

As previously mentioned, the SSEOA and SARS data contain quite different sets of covariates. In the interest of having results that are comparable between the apartheid and post-apartheid periods, we run one set of 'comparable' covariates found in both data sets which only includes gender. We run the following for job j belonging to person i:

Comparable descriptive model:

$$h_j = \alpha + \beta gender_i + \varepsilon \tag{3}$$

Following this, we run a regression exploiting the variables available in each data set. In the SARS data we include industrial sector, firm size, firm year-on-year employment growth, average age of the worker over the job spell, and the start year of the job spell.[15] The average age and start year of the job spell are not conventional variables used in a job hazard regression, but since the entire panel is only seven years long we believe they are informative and can be sensibly interpreted. The ability to control for firm characteristics is an advantage as the exclusion of these is an important source of bias in most job duration analyses (Boockmann and Steffes 2010). In the SSEOA data we are able to include education level, race, occupation, and a public-sector dummy. Due to the continuous structure of the data combined with the time period spanning a maximum of 54 years, we exclude the age and year variables as they have no sensible interpretation over this much longer time span.[16] The models we run on the SARS and SSEOA data are respectively:

Multiple descriptive model:

$$\begin{aligned} h_j = \alpha + \beta_1 gender_i + \beta_2 sector_j + \beta_3 firmsize_j + \beta_4 growth_j + \beta_5 avgage_{ij} + \\ \beta_6 startyear_j + \varepsilon \end{aligned} \tag{4}$$

$$h_j = \alpha + \beta_1 gender_i + \beta_2 occupation_j + \beta_3 education_i + \beta_4 race_i + \beta_5 public_j + \varepsilon \tag{5}$$

[15] The data have a continuous structure, which is why we only have average worker age instead of age and start year instead of year. It is possible to transform the data into a discrete structure in future analyses.

[16] Again, these data can be transformed to a discrete structure to overcome this in future analyses.

Employment trajectories in South Africa

This section is concerned with describing how people move through the labour market. Our first exercise is to report job histories at the person level. The SARS sample is of course only seven-year snapshots of people's work trajectories as opposed to the SSEOA data, which are more comprehensive. We split people into four profiles according to the mode of their personal job distribution:

1. Stable trajectory: the share of people who mainly have jobs lasting five years or more.
2. Moderately stable trajectory: the share of people for whom most of their jobs have lasted more than one year and less than five years.
3. Unstable trajectory: the share of people for whom most of their jobs have lasted between more than three months and one year.
4. Highly unstable trajectory: the share of people for whom most of their jobs have lasted three months or less.

To further analyse trajectories, we investigate how previous labour market status impacts the hazard of the current formal-sector job. We use a specification similar to Boockmann and Steffes (2010) and create a dummy variable, $priorstatus_j$, indicating whether the individual was employed or not employed immediately prior to a given employment spell, and include this in a regression on the hazard.[17] The SSEOA data allows us to differentiate between labour market status and formal- or informal-sector work. The variable we create for the SSEOA regression therefore has three parts indicating whether the labour market activity in the spell prior to the current formal-sector job was another formal-sector job, an informal-sector job, or whether the respondent was not employed. We follow the same format for the regression models as above:

Comparable trajectory model:

$$h_j = \alpha + \beta gender_i + \gamma priorstatus_j + \varepsilon \tag{6}$$

Multiple trajectory model for SARS:

$$\begin{aligned} h_j = \alpha + \beta_1 gender_i + \beta_2 sector_j + \beta_3 firmsize_j + \beta_4 growth_j + \beta_5 public_j + \\ \beta_6 avgage_{ij} + \beta_7 startyear_j + \gamma priorstatus_j + \varepsilon \end{aligned} \tag{7}$$

Multiple trajectory model for SSEOA:

$$h_j = \alpha + \beta_1 gender_i + \beta_2 occupation_j + \beta_3 education_i + \beta_4 race_i + \beta_5 public_j + \gamma priorstatus_j + \varepsilon \tag{8}$$

Early-career employment trajectories in South Africa

We take advantage of the panel or episodic structure of the data sources to investigate job sequence, or how first jobs differ from second jobs. To study early careers, we use the early-career sub-sample of the SARS data defined above and also limit the SSEOA data to people's first through second jobs.

[17] We use an indicator variable instead of the hazard of the previous spell (regardless of type), which would be even more informative, as a compromise with the short time period of the SARS data. This is the same as the Boockmann and Steffes (2010) model. Like Boockmann and Steffes (2010), we also impose a rule on not-employed spells in the SARS data to avoid very short spells which are likely to constitute noise. If a not-employed spell lasts less than three months, we ignore the gap as noise and label this worker as previously employed.

As previously, we report descriptive statistics about job duration and plot hazard functions all by job sequence. We run the same series of descriptive regression models as above, but separately by k, where k represents job sequence. In other words, we run the models below for all the first jobs in our sample, then for the second jobs. Our results are informative about how the baseline hazard and other effects vary across job order.

Comparable descriptive model:

$$h_k = \alpha + \beta gender_i + \varepsilon \tag{9}$$

Multiple descriptive model for SARS:

$$\begin{aligned} h_k = \alpha + \beta_1 gender_i + \beta_2 sector_k + \beta_3 firmsize_k + \beta_4 growth_k + \beta_5 public_k + \\ \beta_6 avgage_{ik} + \beta_7 startyear_k + \varepsilon \end{aligned} \tag{10}$$

Multiple descriptive model for SSEOA:

$$h_k = \alpha + \beta_1 gender_i + \beta_2 occupation_k + \beta_3 education_i + \beta_4 race_i + \beta_5 public_k + \gamma priorstatus_k + \varepsilon \tag{11}$$

6 Characterizing job spells in South Africa

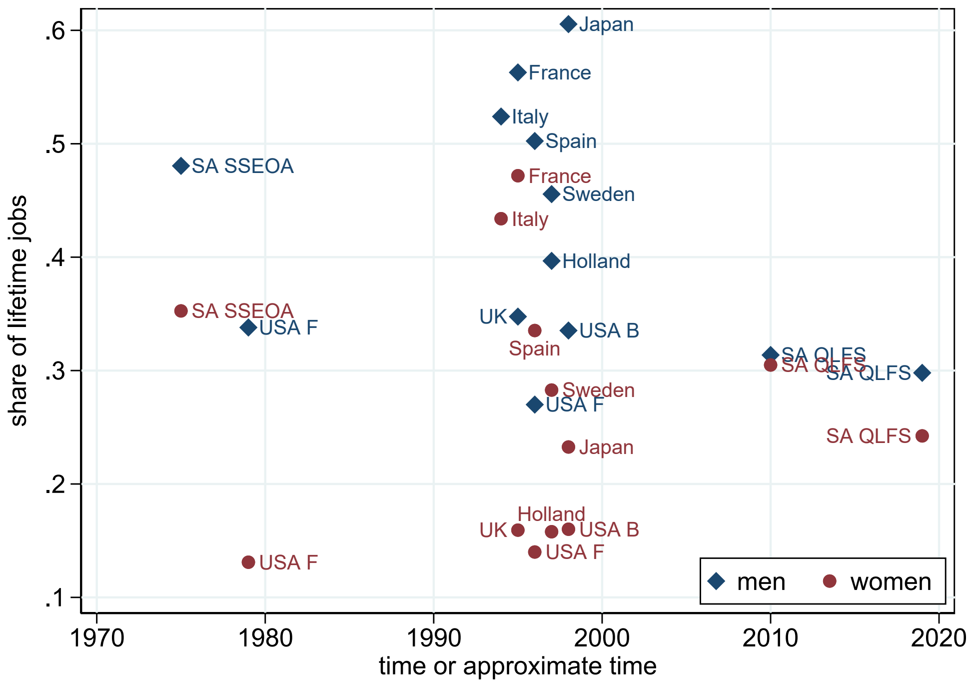
6.1 Is long-term employment common?

The shares of lifetime jobs in South Africa are plotted against an international sample of estimates from Farber (1999) and Burgess (1999) in Figure 3. In an effort to contextualize these estimates with time, we approximate where they fall in cases in which estimates cover a span of time or the calendar date is unclear from the source material. Most of the developed-country data is from the mid-1990s, and a higher share of men than women enjoy lifetime jobs consistently across country and time. The South African estimates from the SSEOA and QLFS fall on either side of the bulk of these estimates, but drawing a trend line between the South African shares makes it appear that South Africa has comparable shares of lifetime jobs to the rest of the sample. These would be roughly similar to Dutch men and Spanish women in the mid-1990s.

Over time, the share of lifetime jobs has declined. In South Africa this happened for women as well, whereas in the United States there was a modest increase in the share for females. However, recent research has concluded that these slight increases in the United States can be attributed to increased female attachment to the labour market over time, and that the overall trend for both men and women has been one of decline in long-term employment until about 2000, following which the shares have been largely stable (Farber 2009; Hollister and Smith 2014; Hyatt and Spletzer 2016). South Africa, though, exhibits a different trend. Between 2010 and 2019, the male lifetime job share was relatively stable, while there was a much steeper drop in the share of women in lifetime formal-sector jobs.

In general, though, it is interesting that South African women prior to the 1990s experienced lifetime jobs at a much higher incidence than American women. Even compared to the mid-1990s estimates, most other countries had much wider discrepancies between their shares of male and female lifetime job incidence. The lifetime tenure gender gap was comparatively smaller in South Africa in the twentieth century, and this is likely related to apartheid labour market policies. To delve deeper, we break down the share of lifetime jobs by race in Table 2. Immediately we can see that the discrepancy between white male and white female lifetime tenure during the apartheid era is highly comparable to the American numbers for 1979 in Figure 3.

Figure 3: Lifetime jobs around the world: the share of employed aged 45 and older reporting current tenure of 20 years or more



Notes: timescale is exact only for the QLFS and the F estimates. The SSEOA estimates span 1950–91/94 and are placed in the middle of this period. All other countries plus USA B have been placed somewhere in the mid-1990s for context based on the author's data-gathering note about calendar time in his paper (Burgess, 1999: 5): 'The calendar date ... was left up to the country expert, with the simple guidance of some time in the mid-1990s when the labour market was roughly in equilibrium'. F = Farber (1999) estimates; B = Burgess (1999).

Source: authors' construction based on SSEOA and QLFS data.

Table 2 reports that during the apartheid era, Black Africans were more likely than white people to have a job for more than 20 years, but the reverse was true in the post-apartheid period. Generally speaking, a higher share of lifetime tenure is interpreted as an advantage. More lifetime tenure among whites in the post-apartheid period is interpreted as greater access to employment stability, which South Africans prioritize (Mncwango 2016). However, our interpretation of the apartheid-era results needs to consider how apartheid labour market policies would have influenced labour market dynamics for different groups. More lifetime tenure among Black Africans during this time is likely related to the long-term tenure requirement of Section 10 rights, mentioned above. Job duration may be shortened if workers have good outside options in terms of how quickly they can find a new or better-paid job (Farber 1999). Jobs, and better-paid jobs in particular, were in shorter supply for Black Africans, especially in the 1980s and 1990s, compared to whites during apartheid. In general, this means that it would have been more rational to remain attached to a job already secured. In other words, a high incidence of lifetime tenure among Black populations during apartheid could be indicative of restricted bargaining power and limited job opportunities, rather than labour market advantage.

The same rationale can explain why lifetime tenure was so common among Coloured and Black African women. Over 60 per cent of Black African women in the SSEOA sample in Table 2 were in 'Elementary' occupations, which includes domestic work. It is plausible that many of these women were domestic workers working for the same white family for decades, which would also secure them the right to live in better-connected and -resourced urban areas. After apartheid ended, these discriminatory policies were abolished, relaxing the importance of long-term tenure for access to a better life and freeing women to expand their job search and take other factors important to them into consideration. By contrast, American women in the twentieth century were not under the same pressure to secure long-term tenure, partly explaining the narrower gender gap in tenure in South Africa over the same period.

Table 2: Formal sector lifetime jobs in South Africa: shares of employed aged 45 and older reporting current tenure of 20 years or more

	SSEOA 1951–91/94			QLFS 2010–19		
	Men	Women	Overall	Men	Women	Overall
Overall	0.48	0.35	0.44	0.30	0.26	0.29
Black African	0.51	0.40	0.47	0.29	0.26	0.28
Coloured	0.39	0.44	0.41	0.29	0.26	0.28
Asian/Indian	0.40	0.24	0.37	0.29	0.26	0.28
White	0.46	0.26	0.40	0.33	0.27	0.30
Public sector	0.60	0.52	0.58	0.53	0.40	0.46
Private sector	0.45	0.32	0.41	0.24	0.19	0.23
N (unweighted)	958	474	1,432	97,888	75,187	173,075
Share (weighted)	0.68	0.32	1.00	0.58	0.42	1.00

Notes: estimates adjusted using sampling weights on a formal-sector sample and reported with a 95 per cent confidence interval.
Source: authors' compilation based on SSEOA and QLFS data.

A final result from Table 2 is that lifetime tenure is consistently more common in the public sector compared to the private sector. This aligns with Farber (2009), who finds that the recent decline in employment stability in the United States is in part due to structural economic changes, to which the private sector is more sensitive than the public sector. As such, median tenure has dropped more quickly in the private compared to the public sector.

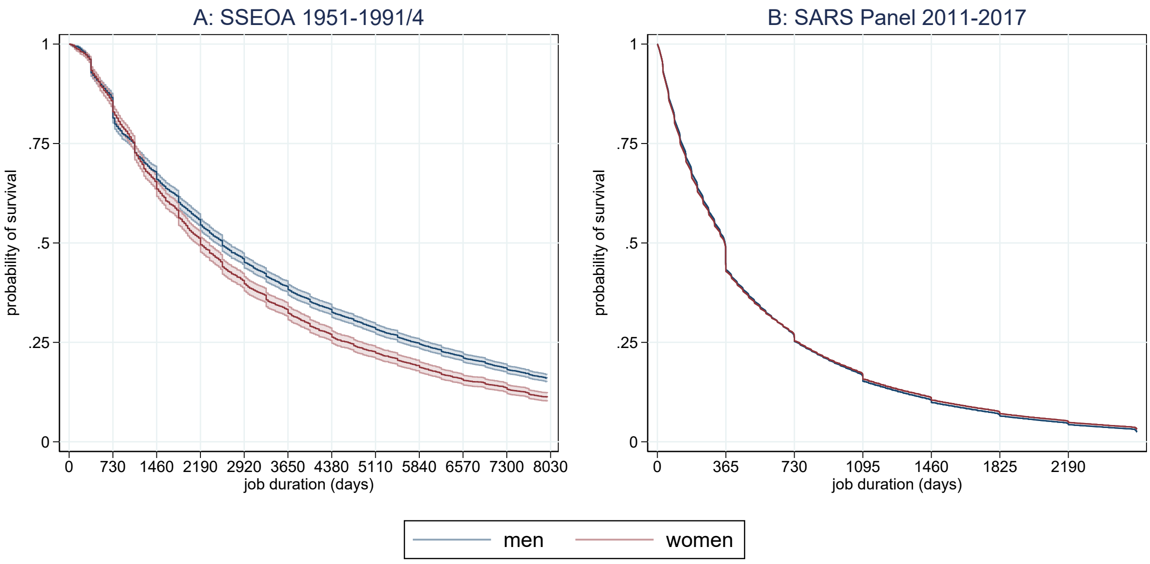
6.2 Job duration during the apartheid and post-apartheid periods

The job survival function is given in panel B of Figure 4. Estimates are reported with confidence intervals, but the enormous sample size makes them very precise. Panel B reveals that half of the jobs in the SARS panel ended after a year, and one-quarter ended within six months. This aligns with Kerr (2018), who found a job flow rate of 52 per cent for 2012–14, also using the SARS data. It is also very similar to the job survival rates for the United States in the 1990s (Farber 1994), but shorter than those for Germany in the mid-1990s. Survival estimates for Germany using a data set constructed very similarly to the SARS panel find that half of jobs survive until about two years (Boockmann and Steffes 2010). Gender is not an important differential in panel B.

The time horizon is much longer in the SSEOA data, given in panel A of Figure 4, and because the survival estimate is not conditional, this means we cannot directly compare panels A and B. What this figure does indicate is that gender becomes increasingly important at longer job durations in the SSEOA data, but the differences are smaller within the first seven years, which is the same time frame available in the SARS panel. During apartheid, men's jobs survived for longer. Half of women's jobs survived for six years; but half of men's jobs survived for approximately seven years.

Table 3 reports statistics describing the distribution of job duration in the SSEOA and the SARS panel. We caution against a direct comparison of these two data sets since these are conventional summary statistics and do not account for time dependence, like the hazard. The time span of the data and when we begin observing jobs in calendar time both impact our results. These factors mean that, structurally, there are more long-term jobs in the SSEOA data. Our samples are restricted to include jobs from the calendar years from which we can reliably identify job start time, this is from 2011 in the SARS panel and 1951 in the SSEOA data. The result is that when we slice across the earlier years in either data set to calculate an average job duration, for example, there are not as many long jobs as there should be.

Figure 4: Kaplan–Meier estimates of job survival in the South African formal sector



Notes: estimates based on a formal-sector sample of jobs and reported with a 95 per cent confidence interval.

Source: authors' construction based on SSEOA and SARS data.

Although essentially the same procedure is applied to both data sets, the SSEOA spans over 40 years. Calendar years with the correct number of long-term jobs then easily outnumber years without and we can interpret the SSEOA estimates as representative of job duration generally. The same cannot be said for the SARS panel, which is only seven years long, meaning the sample restriction at the beginning of the data is influential on the results. The longest possible completed job duration that we can observe in the SARS panel is seven years. This means that job duration in the SSEOA data can be interpreted more generally, whereas the SARS panel reflects job duration for jobs that begin in or after 2011.[18] We can still use these data sets to learn about job duration in these two periods of history and compare broad conclusions. However, the samples are not directly comparable in the sense that it is not appropriate to compare point estimates, for example.

Table 3 indicates that the average job lasted 6.7 years during apartheid, while the median job lasted 4.4 years—an indication of the right-skewness of the job duration distribution. The mean and median may be overestimated for two reasons. First, enumerators were instructed to record start and end years and months, and not days, meaning that the minimum duration of a job was immediately as large as at least a month (28 days) (Treiman et al. 2017). Second, respondents may be prone to only reporting longer periods of labour market activity in a retrospective survey for reasons of recall, among others. The shortest job in the sample is then just under one month and the longest is just over 40 years. Just over one-third of jobs were ongoing at the time of data collection and these were excluded from the calculation of the descriptive statistics in Table 3.

[18] This is not unusual for panel analyses of survival data. See Boockmann and Steffes (2010) for a similar data set-up. What is more unusual is to report conventional summary statistics for a panel like this since the sample is not representative of job duration in general. The difficulty arises when wanting to compare a panel to an episodic data set, for which it is more natural to report summary statistics, as in Booth et al. (1999). There are three potential ways to make the SARS and SSEOA data more comparable in future analyses. The first is that the observation start time for the SARS panel will slowly become less of an issue as time progresses and the panel extends. The second is to make the SSEOA data more directly comparable with the SARS panel by converting both data sets to a discrete data structure and imposing the same rules used to create the SARS panel on sub-samples of the SSEOA data. The third is to impute job durations for existing jobs for which we have no start date into the SARS panel based on age and gender—for example, all 30-year-old women in 2010 have been in their existing jobs for *x* years. Imputation is a widely accepted technique and would assist in 'correcting' the left-hand side of the distribution for left-censored observation start time.

Table 3: Describing the distribution of job duration in the South African formal sector

	SSEOA: 1951–91/94		
	Pooled	Men	Women
Distribution statistics			
Mean	6.7 years	7.1 years	6.1 years
Median	4.4 years	4.8 years	4.0 years
Min.	28 days	28 days	28 days
Max.	40.4 years	39.9 years	40.4 years
Share censored	0.35	0.37	0.34
Share of jobs lasting (x)			
Stable (5+ years)	0.44	0.46	0.40
Moderately stable (>1 year to <5 years)	0.44	0.43	0.45
Unstable (>3 months to 1 year)	0.11	0.09	0.14
Highly unstable ($\leq$ 3 months)	0.02	0.01	0.02
	1.00	1.00	1.00
N (jobs)	7,548	4,718	2,830
	SARS panel: 2011–17		
Distribution statistics			
Mean	1.1 years	1.2 years	1.1 years
Median	8.8 months	9 months	8.4 months
Min.	2 days	2 days	2 days
Max.	7 years	7 years	7 years
Share censored	0.20	0.19	0.22
Share of jobs lasting (x)			
Stable (5+ years)	0.05	0.05	0.05
Moderately stable (>1 year to <5 years)	0.42	0.42	0.41
Unstable (>3 months to 1 year)	0.34	0.34	0.34
Highly unstable ($\leq$3 months)	0.19	0.19	0.19
	1.00	1.00	1.00
N (jobs)	27.6 million	16.1 million	11.5 million

Notes: estimates based on uncensored formal-sector jobs. SSEOA estimates are adjusted using sampling weights.
Source: authors' compilation based on SSEOA and SARS data.

Women tended to work shorter job durations than men during apartheid—consistent with the international literature covering a similar time period. Mean job duration for women is a full year less than that for men, at 6.1 and 7.1 years for women and men, respectively. Zizzamia and Ranchhod (2019) found that in the post-apartheid period, similar proportions of men and women experienced volatile employment (their definition of which roughly corresponds to our moderately stable category), whereas gender differences were more noticeable at the tails of the distribution: many more men enjoyed stable employment and many more women experienced what they termed 'persistent unemployment'. Findings from Table 3 align with this in that a higher share of jobs worked by men were stable and a higher share of jobs worked by women were unstable or highly unstable.

Turning to the SARS panel results, we see that there are much higher shares of short- compared to long-term jobs due to the sample selection rules just discussed. However, we see very similar shares of moderately stable jobs in the SARS data compared to the SSEOA data. The average job lasted just over a year and the median job slightly under a year—also indicating a right-skewed distribution. Time is very granular in the SARS data, so we have the minimum job lasting two days[19] and our maximum is limited to seven years. The mean and median are then likely to be underestimates, although it is worth noting that the job duration hazard peaks at about 6–8 months in the developed world in studies from the 1990s (Booth et al. 1999; Farber 1994). About one-fifth of jobs were ongoing when the data set ended in 2017. Gender differences in the SARS data are minimal.

Overall, the distribution of jobs in both the SSEOA and SARS data are right-skewed. The SSEOA results may be an overestimate and the results from the SARS data are certainly an underestimate. Overall, there is a very broad confidence interval for job duration in South Africa, and a very fuzzy indication that it has declined between the apartheid and post-apartheid periods. Interestingly, there are similar shares of jobs lasting between one year and five years in both periods. This means that the decline in job duration is likely driven by a decline in the incidence of long-term jobs. The trend in the percentage of people reporting current tenure of a year or less in the QLFS between 2010 and 2017 is stable, suggesting it is not an increase in short-term jobs that is drawing down average job duration. The percentage stood between 18.3 and 21.7 per cent. The decline in the number of long-term jobs over time discussed above is likely linked to the global trend of shorter-term jobs in the private sector (Farber 2009), but also to changes to labour market policy specific to South Africa's history.

6.3 What does the job duration hazard look like?

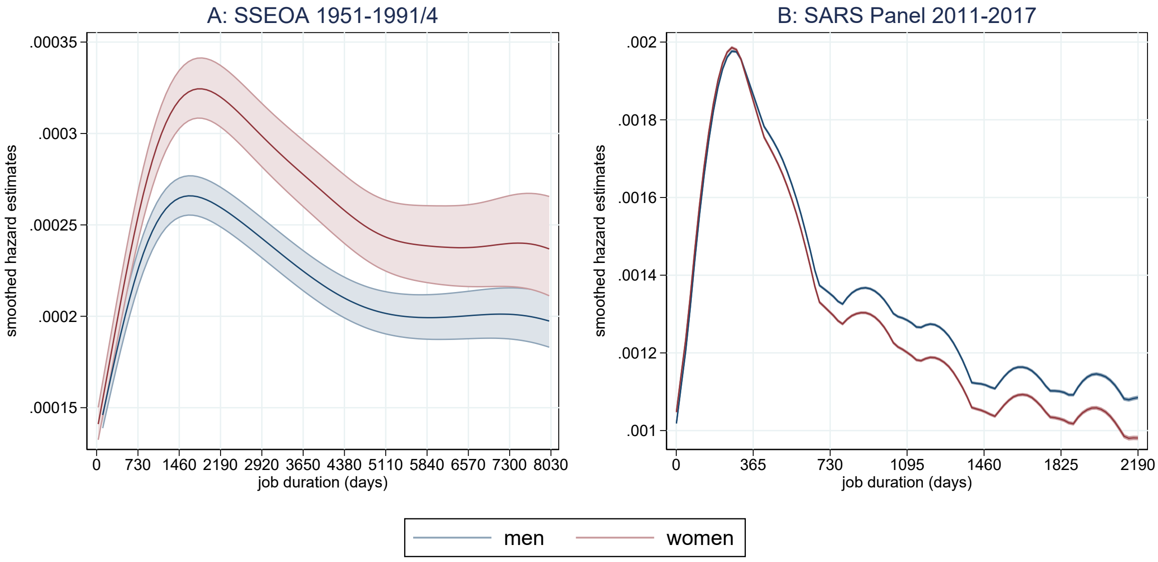
Figure 5 plots the Kaplan–Meier estimate of the job duration hazard in days for two periods in South Africa's history by gender. Note that because the hazard is conditional, it should not be biased by the data structure in the way that the descriptive distribution statistics were in Table 3. The hazard is the probability of failure calculated only over jobs that have lasted a given period long, as opposed to over the full distribution. Because the hazard 'corrects' the risk set, the SSEOA and SARS panel hazards are more comparable. This is an important advantage of survival statistics over describing a distribution more conventionally.

Overall, the shape corresponds with what has been typically found in the developed-country literature: the hazard is non-monotonic, peaking early and decreasing in time thereafter. The post-apartheid hazard in the SARS data peaks at about eight or nine months, which is closely comparable to the results from the United States (Farber 1999) and the United Kingdom (Booth et al. 1999). The hazard in the SSEOA data peaks much later, at about four or five years. We know from Table 2 that long-term jobs were more common during apartheid, so this may be one reason for a later peak in the hazard. A more likely reason, though, is that the SSEOA is retrospective data and people are unlikely to recount very short jobs. Time is also much more granular in the SARS panel, which is able to report information in days. By contrast, enumerators for the SSEOA were instructed to collect start and end dates at a year–month level of detail (Treiman et al. 2017). Compared to the United States and the United Kingdom, the hazard in South Africa seems to drop off at a slower rate after it has peaked, suggesting job security takes longer to take effect.

Figure 5 also shows that the hazard for women is higher than that for men during apartheid, but the opposite is true after apartheid. The idea that women's jobs are at a higher risk of ending is in line with our expectations about female labour market disadvantage and women exiting the labour market for family reasons, especially during apartheid. In the SARS panel, men and women have very similar hazards until about two years, after which men experience a higher risk of a job ending. As the hazard can be raised both by reasons related to labour market advantage and disadvantage, interpretation can be difficult.

[19] We excluded jobs with a duration of one day because they occurred disproportionately and are likely measurement error related to tax filing.

Figure 5: Kaplan–Meier estimates of the job duration hazard in the South African formal sector



Notes: estimates based on a formal-sector sample of jobs and reported with a 95 per cent confidence interval. SSEOA estimates adjusted using sampling weights.

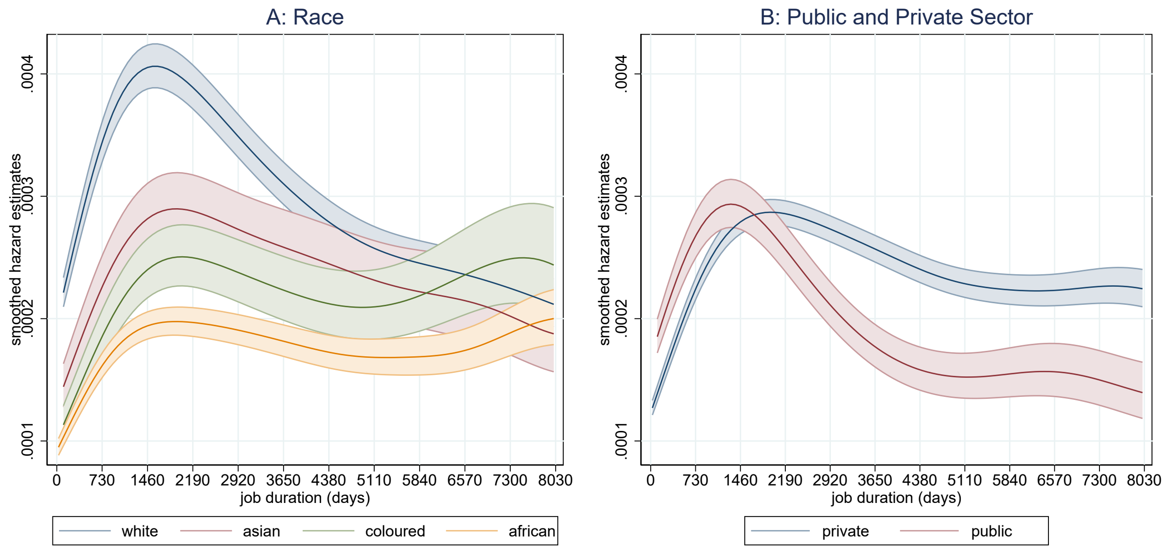
Source: authors' construction based on SSEOA and SARS data.

In Figure 6 we break down the hazard by race and public-sector status in the SSEOA data. Both the public- and private-sector hazards peak at a similar time in panel B, and at a similar level. However, the hazard drops much more steeply after the peak in the public compared to the private sector, indicating that job security is unequivocally higher in the public sector. Panel A gives the breakdown by race. This figure demonstrates how a higher hazard and more mobility can signal labour market advantage. The hazard for white workers is consistently above that of other groups, with Black Africans experiencing the lowest and flattest hazard.

The flat hazard visible for Black Africans indicates that there were almost no 'job security returns' to tenure. A Black African person was as likely to lose a job they had had for five years as one they had had for one year. This is partly because apartheid labour policies restricted Black Africans to performing less skilled jobs—which would undermine the accumulative mechanism that is behind the declining slope in a typical job duration hazard. The idea that the accumulation of job-specific capital in less skilled work has a weaker effect on reinforcing tenure also comes out in our occupation results later (see Figure 7). The peak of the hazard is also quite diffused, suggesting impediments to job matching, perhaps related to the limited outside options and wage ceilings imposed on Black Africans by apartheid labour market policies. By contrast, the hazard for white people takes the shape of the typical hazard, suggesting the job market was more dynamic for white people and tenure was accumulative in the usual sense.

Figure 7 plots hazards by the 10 main occupation groups. Here, we see that senior managers have a flat hazard, similar to that seen for Black Africans previously. However, here it is likely to be related to stability. The typical hazard peaks early as a result of natural volatility in the matching process as employers terminate positions and employees leave jobs to find a better job or a better-paid job. Not only are senior managerial positions good jobs and among the best-paid jobs, but the career trajectory to become a senior manager is relatively structured. People are often promoted into this type of position after working at a company for some time. This means that by the time people are entering these types of jobs, they are less likely to 'chop and change' compared to other occupations, making the matching process less noisy.

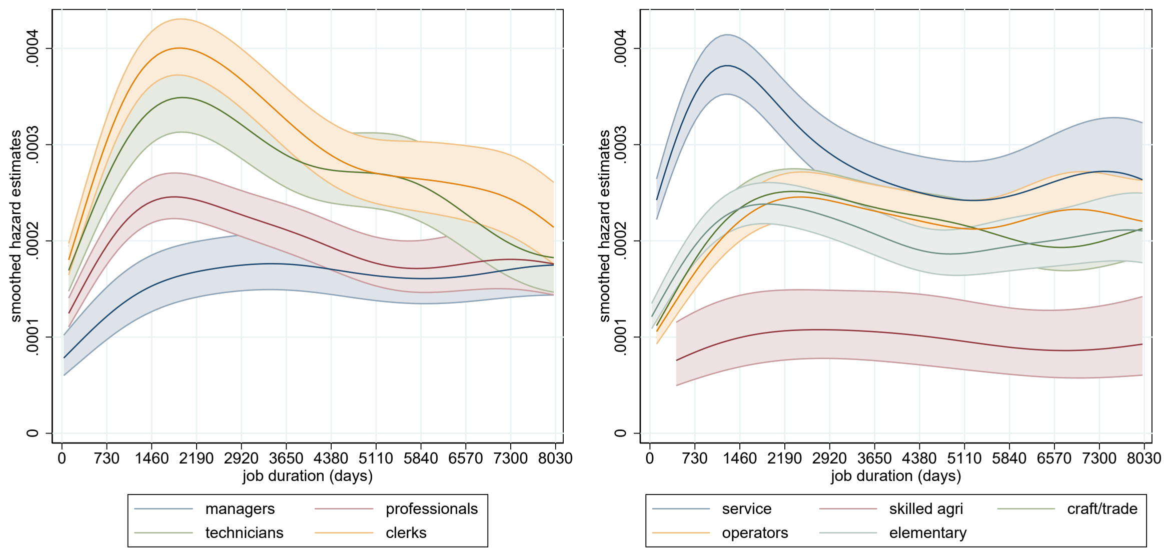
Figure 6: Kaplan–Meier estimates of the job duration hazard in the South African formal sector by race and public sector status, 1951–91/94



Notes: estimates based on a formal-sector sample of jobs; reported with a 9 per cent confidence interval and adjusted using sampling weights.

Source: authors' construction based on SSEOA data.

Figure 7: Kaplan–Meier estimates of the job duration hazard in the South African formal sector by occupation, 1951–91/94



Notes: estimates based on a formal-sector sample of jobs; reported with a 95 per cent confidence interval and adjusted using sampling weights.

Source: authors' construction based on SSEOA data.

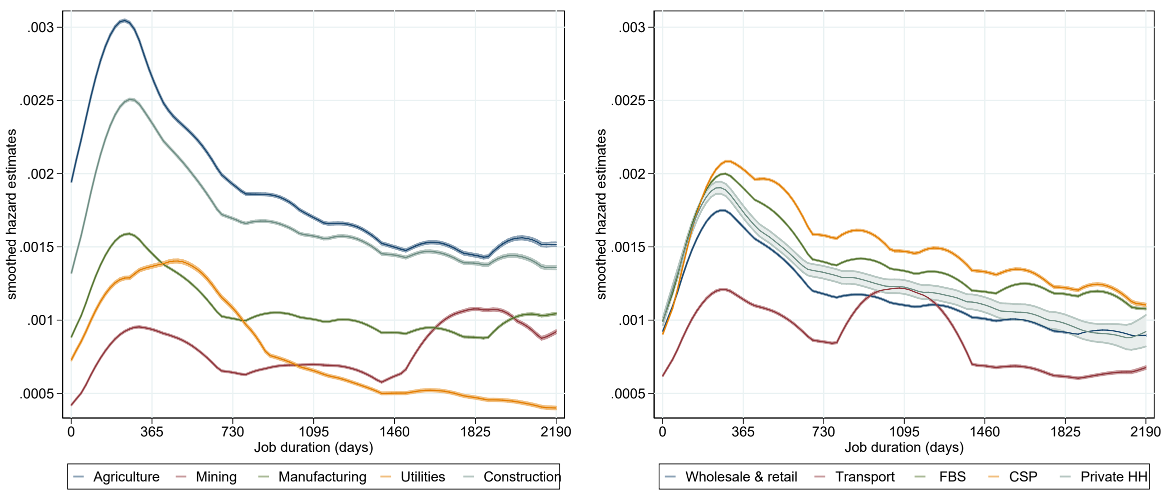
Senior managers are also most likely to experience a job ending after a prolonged job duration, rather than early on, which is also counter to the usual pattern. Senior managerial positions are usually occupied at advanced career stages, meaning these may be many people's last job before they retire. As such, people are more likely to experience a job ending at later job durations since the longer people occupy a job, the older they get and the closer to retirement.

Clerks, service workers, and technicians exhibit a more classical hazard. Job-specific capital and experience are clearly pertinent in these occupations, resulting in a decreasing hazard as job duration lengthens. By contrast, operators, elementary workers, and skilled agriculture experience flatter hazards because their fields are comparatively less skilled, which weakens the importance of accruing job-specific capital and thereby the accumulative effect of the tenure. Blue-collar workers were also found to work longer job durations in Germany, although once firm-level variables were controlled for, this was attributed to these workers selecting into firms covered by work councils and which offered more opportunities for on-the-job training (Boockmann and Steffes 2010).

Next we turn to hazards by sector in the SARS panel in Figure 8. Agriculture has the highest hazard of all the sectors, both in terms of the peak being very high but also because the hazard remains high thereafter. This likely means that work in agriculture is volatile and brings little job security and stability. This is intuitive because agricultural work is seasonal for many workers, meaning that jobs are structurally shorter in this sector. Construction also stands out for having a relatively high hazard. Both agriculture and construction are sectors that are made up of large shares of unskilled or low-skilled labour, which can also contribute to less stability. In particular, the high hazard in the construction industry may be related to the Expanded Public Works Programme (EPWP), a labour-intensive public employment scheme that provides temporary employment opportunities to the unemployed. In 2016, the EPWP created over 250,000 work opportunities in the infrastructure sector, with an average duration of four months.

Other sectors that stand out for noticeably lower hazards are utilities, mining, and transport (which includes storage and communication). The utilities hazard is more dispersed early on, but then drops quickly and remains very low. This suggests that although it may take a little longer to secure jobs in this sector, once they are secured they are very stable and workers have good job security. This is unsurprising when considering that many public-sector jobs fall into utilities, such as those belonging to the electricity and transport state-owned enterprises. Mining and transport are also notable because of unusual bulges in the hazard after longer job durations. This happens after about five years in the mining sector and three years in the transport, storage, and communication sectors, and could be related to how contracts are structured in these sectors.

Figure 8: Kaplan–Meier estimates of the job duration hazard in the South African formal sector by industrial sector, 2011–17



Notes: estimates reported with a 95 per cent confidence interval.

Source: authors' construction based on SARS IRP5 panel data.

Job hazards in both periods of South Africa's history conform with expectations. Interpretation of the hazard has proven to be tricky in some cases. We find it hard to explain for example: why women have higher hazards in the SSEOA data but lower hazards in the SARS panel. Further, in some cases, like for Black Africans, flat hazards reflect labour market disadvantages like poor outside options and low skill content. For senior managers, though, flat hazards reflect stability and job matching at advanced career stages. Most findings are plausible and this boosts confidence in using these data for survival analysis of job duration in South Africa.

6.4 Determinants of the hazard

Tables 4 and 5 report results for our *comparable descriptive* and *multiple descriptive* hazard models, respectively. There is agreement across models and specifications that men's jobs lasted longer during apartheid, but evidence is mixed in the post-apartheid panel. Men's jobs lasted 15–17 per cent longer than women's during apartheid according to the first three parametric models in Table 4. The Cox model supports this conclusion. The Cox model is a proportional hazards model and the output is interpreted as meaning that the risk a man's job would end was 83 per cent of the risk that a woman's job would end; or, that being a man reduced the risk of a job ending by 17 percentage points. The Cox model non-parametrically estimates the baseline, hence the absence of a constant term in the output.

The shape parameters for the parametric SSEOA models also generally conform with what the Kaplan–Meier hazards have revealed. The log-normal model yields sigma equal to 1.1, indicating that the baseline hazard is increasing and then decreasing. The log-logistic model reports a gamma of close to 0.64, which means that the baseline hazard is monotonically decreasing. The generalized gamma nests three parametric models—the exponential, the Weibull, and the log-normal—and the level of the shape parameters assists in differentiating which of these models is most appropriate. The best choice is a log-normal if kappa equals zero; a Weibull if kappa equals 1; and an exponential if kappa and the log of sigma equal 1. Table 4 reveals that kappa is significantly different from zero. Our additional test showed that kappa and log sigma also did not equal 1, meaning that while the log-normal was not confirmed, neither were either of the other two. The model fit statistics, however, consistently indicate that the log-normal is the best model.

The shape and model fit parameters for the SARS data arrive at similar conclusions to those for the SSEOA. The main difference is that the sign on the coefficient for male changes between the two log parametric models and the generalized gamma, and this is reinforced by all the coefficients on male being negative when more covariates are added to the specification in Table 5. Table 5 reports than women's jobs last 3–5 per cent longer than men's jobs. Note that we were unable to run a Cox proportional hazards model on the SARS panel because the sample was too large for the model to compile timeously. The positive coefficients on gender only occur in Table 4 for the SARS panel in cases where a frailty model was used. Worker frailty proves to be an important source of variation in all models and data sets. The frailty random coefficient, theta, is significant in all cases in which a frailty model was run.

The expanded specifications in Table 5 allow us to interpret other factors of interest. Public-sector jobs during apartheid lasted 17–25 per cent longer than their private-sector counterparts. Jobs for the degree-educated lasted half as long as those for people who only had primary education or less. The jobs of Black Africans lasted as much as 69 per cent longer than jobs for whites. In the case of both education and race, the most advantaged groups work the shortest job durations on average. This is indicative of the greater bargaining power on the part of these advantaged groups since both have better outside options than their counterparts, should they find themselves in an ill-matched job. The education findings are supported in the developed world (Boockmann and Steffes 2010) and align with theory that highly educated people are more mobile (Burdett 1978) and better able to accumulate job-specific capital. The SARS panel multiple descriptive model also reveals findings in line with expectations. Jobs for older people tend to last slightly longer; jobs in bigger firms seem to last slightly longer as well; and jobs in firms that are growing their employment base also last slightly longer.

Table 4: Regression output for the comparable descriptive hazard model

DEPVAR: job	SSEOA 1951–91/94				SARS panel 2011–17		
duration hazard	Log-normal	Log-logistic	Gen. gamma	Cox PH	Log-normal	Log-logistic	Gen. gamma
Male	0.15***	0.15***	0.17***	–0.17***	0.02***	0.04***	–0.02***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)	(0.00)
_cons	7.63***	7.67***	7.83***		5.492***	5.632***	6.078***
	(0.02)	(0.02)	(0.03)		(0.00)	(0.00)	(0.00)
Shape parameters							
lnsigma	0.10***		0.10***		0.165***		0.216***
	(0.01)		(0.01)		(0.00)		(0.00)
lngamma		–0.45***				–0.408***	
		(0.01)				(0.00)	
kappa			0.31***				0.574***
			(0.04)				(0.00)
Frailty							
Theta	0.27***	0.26***	vce	vce	0.389***	0.345***	vce
	(0.02)	(0.02)			(0.00)	(0.00)	
Model fit							
Log likelihood	–11,618.18	–11,706.79	–11,747.82	–65,888.08	–24,425,543.95	–24,552,852.11	–24,487,502.27
AIC	23,244.35	23,421.59	23,503.63	131,778.20	48,851,095.91	49,105,712.22	49,105,771.61
BIC	23,274.07	23,451.31	23,533.35	131,785.60	48,851,155.29	48,975,012.55	48,975,071.93
Degrees of freedom	4.00	4.00	4.00	1.00	4.00	4.00	4.00
N	12,460	12,460	12,460	12,460	20,709,282	20,709,282	20,709,282

Notes: standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates based on formal-sector jobs. SSEOA estimates are adjusted using sampling weights. Only the Cox model is proportional hazards (PH), all other models are accelerated failure time models. Stars on theta are based on the p-value from the chi-squared LR test that theta equals zero. vce = robust standard errors applied when frailty infeasible for computational or other reasons.

Source: authors' compilation from SSEOA and SARS data.

Table 5: Regression output for the multiple descriptive hazard model

DEPVAR: job	SSEOA 1951–91/94				SARS panel 2011–17		
duration hazard	Log-normal	Log-logistic	Gen. gamma	Cox PH	Log-normal	Log-logistic	Gen. gamma
Male	0.12***	0.13***	0.15***	–0.16***	–0.05***	–0.03***	–0.05***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)	(0.00)
Public sector	0.22***	0.17***	0.25***	–0.23***			
	(0.04)	(0.04)	(0.04)	(0.04)			
Education (base = primary or less)							
Incomplete secondary	–0.19***	–0.22***	–0.18***	0.14**			
	(0.04)	(0.04)	(0.04)	(0.04)			
Secondary	–0.24***	–0.26***	–0.25***	0.22***			
	(0.05)	(0.05)	(0.06)	(0.05)			
Other post-secondary	–0.50***	–0.55***	–0.54***	0.49***			
	(0.07)	(0.07)	(0.07)	(0.06)			
Degree post-secondary	–0.52***	–0.57***	–0.54***	0.48***			
	(0.08)	(0.08)	(0.08)	(0.08)			
Race (base = white)							
Asian	0.37***	0.35***	0.35***	–0.30***			
	(0.06)	(0.05)	(0.05)	(0.05)			
Coloured	0.53***	0.53***	0.52***	–0.44***			
	(0.06)	(0.06)	(0.07)	(0.06)			
Black African	0.69***	0.69***	0.70***	–0.64***			
	(0.04)	(0.04)	(0.04)	(0.04)			
Occupations	YES	YES	YES	YES			
Mean age					0.03***	0.03***	0.03***
					(4.03e–05)	(3.88e–05)	(4.00e–05)
Firm size					1.72e–05***	1.49e–05***	1.15e–05***
					(3.65e–08)	(3.44e–08)	(2.96e–08)
Firm growth					1.78e–06***	1.98e–06***	1.31e–06***
					(2.57e–08)	(3.62e–08)	(1.19e–08)
Sectors					YES	YES	YES
_cons	8.31***	8.45***	8.47***		4.09***	4.20***	5.00***
	(0.09)	(0.08)	(0.09)		(0.00)	(0.00)	(0.00)

Regression output for the multiple descriptive hazard model (continued...)

DEPVAR: job	SSEOA 1951–91/94				SARS panel 2011–17		
duration hazard	Log-normal	Log-logistic	Gen. gamma	Cox PH	Log-normal	Log-logistic	Gen. gamma
Shape parameters							
lnsigma	0.07***		0.07***		0.15***		0.13***
	(0.01)		(0.01)		(0.00)		(0.00)
lngamma		–0.50***				–0.43***	
		(0.01)				(0.00)	
kappa			0.36***				0.77***
			(0.03)				(0.00)
Frailty							
Theta	0.19***	0.18***	vce	vce	0.251***	0.220***	vce
	(0.02)	(0.02)			(0.00)	(0.00)	
Model fit							
Log likelihood	–9,869.81	–9,902.09	–9,884.17	–56,565.15	–16,717,356.27	–16,853,490.57	–16,296,940.18
AIC	19,779.62	19,844.18	19,808.34	113,164.30	33,434,756.53	33,707,025.15	32,593,924.36
BIC	19,925.83	19,990.39	19,954.55	113,288.60	33,435,076.25	33,707,344.86	32,594,244.07
Degrees of freedom	20.00	20.00	20.00	17.00	22.00	22.00	22.00
N	11,055	11,055	11,055	11,055	15,135,340	15,135,340	15,135,340

Notes: standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates based on formal-sector jobs. SSEOA estimates are adjusted using sampling weights. Only the Cox model is proportional hazards (PH), all other models are accelerated failure time models. Stars on theta are based on the *p*-value from the chi-squared LR test that theta equals zero. vce = robust standard errors applied when frailty infeasible for computational or other reasons. Occupation and sector results not shown for reasons of space and are available from the authors upon request.

Source: authors' compilation based on SSEOA and SARS data.

Note the samples here exclude censored jobs and the sample size for the SARS data is very large because we have included all jobs, including those which overlapped (i.e. an individual was working two or more jobs simultaneously). Overlapping jobs, as we have discussed previously, violate the assumption that jobs are independent. To test whether this violation is problematic, we also ran the comparable descriptive model on a sample of non-overlapping jobs, where we selected the highest-paying job as the job to keep in the instance of multiple jobs. The results are reported in Appendix Table A1. Interestingly, this sample restriction serves to exert a downward pressure on all the coefficients, suggesting it is a pertinent restriction. However, we do not know whether the coefficients are different because we have eradicated bias introduced with non-independent jobs, or because second jobs are very different to main primary jobs. If it is the former, then it would be important to restrict the sample in this way.[20] However, if it is the latter, then we argue it is very important to keep these second jobs in the sample since excluding them will lead to biased conclusions. We cannot tell what is behind the different coefficient in Appendix Table A1, but we do know that excluding these jobs has a serious effect on our results.

In sum, the South African case looks much like the developed-country case. We conform with the three main stylized facts: long-term jobs are a feature; most new jobs end early; and the hazard declines with tenure. South Africa also falls broadly within the international trends. There is a decline in long-term tenure and this has occurred even more steeply for women in South Africa than it appears to have occurred for women in the United States. Generally, it appears as if the average job duration has declined between apartheid and the post-apartheid period, although it is difficult to be precise with the wide confidence interval set up by our two data sources and the specific measurement problems attached to each. The median or average job in both periods is one lasting more than one year but less than five, making it moderately stable. During apartheid, men's jobs were more stable, meaning they were more likely than women's jobs to last five years or more.

7 Labour market trajectories in South Africa

In this section we are interested in characterizing how South Africans move through the labour market. An immediate caveat that carries throughout this section is our inability to adjust our results by age because of the continuous, as opposed to discrete, structure of our data.[21] The results presented here are for the samples as a whole at their given distribution of age. Table 6 reveals that, although the SARS panel is much shorter than the SSEOA, people work very similar numbers of jobs. The mean in both data sets is just over one job per person over the period they are observed, and the median is one. The histograms of jobs per person in Figure 9 are also remarkably similar across the data sources. Most people work just one job and women are slightly more likely than men to do so. These results are consistent with the American results from Farber (1994), who found that women held fewer jobs per year than men and that this was due to a lower exit rate from the first job after entry.

[20] Non-independent jobs should serve to attenuate the estimates, but our estimates with the non-independent jobs are larger than those from the sample with only the main job.

[21] The data can be converted to a discrete structure and we leave this for future research.

Table 6: Describing careers in the South African formal sector

	SSEOA: 1951–91/94		
	Pooled	Men	Women
Distribution statistics			
Mean	1.87	1.94	1.76
Median	1.00	1.00	1.00
Min.	1.00	1.00	1.00
Max.	10.00	10.00	10.00
N (people)	5,486	3,275	2 211
Share of people (25 and older) with most jobs lasting (*x*)*			
Stable (5+ years)	0.53	0.57	0.49
Moderately stable (>1 year to <5 years)	0.40	0.38	0.43
Unstable (>3 months to 1 year)	0.06	0.05	0.07
Highly unstable (≤3 months)	0.01	0.01	0.01
	1.00	1.00	1.00
N (people)	3,890	2,319	1 571
	SARS Panel: 2011–17		
Distribution statistics			
Mean	2.02	2.08	1.95
Median	2.00	2.00	1.00
Min.	1.00	1.00	1.00
Max.	139.00	139.00	105.00
Share of people with most jobs lasting (x)*			
Stable (5+ years)	0.10	0.10	0.11
Moderately stable (>1 year to <5 years)	0.36	0.36	0.36
Unstable (>3 months to 1 year)	0.23	0.22	0.24
Highly unstable (≤3 months)	0.11	0.10	0.12
	0.80	0.78	0.83
N (people)	12.9 million	7.4 million	5.6 million

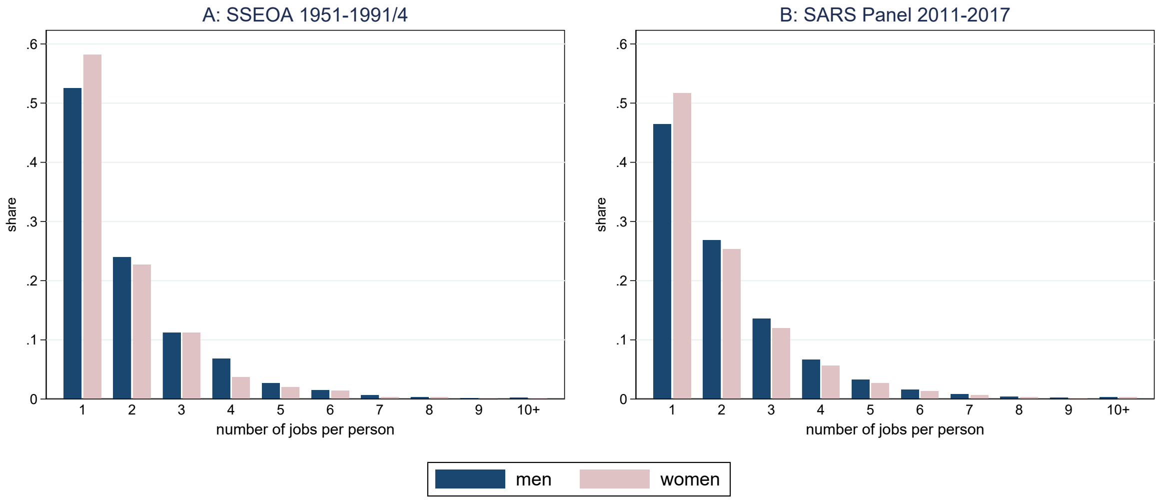
Notes: 'Distribution statistics' estimates include censored jobs but 'Shares of people' categories do not. Estimates based on formal-sector jobs. SSEOA estimates are adjusted using sampling weights. * These categories are calculated slightly differently between the SSEOA and the SARS panel. In the SSEOA, people are categorized according to where the mode falls for their jobs. In the SARS panel, people are placed into a category if 60 per cent of their jobs fall into that category. As a result of this, the shares do not add up to 1 for the SARS panel.

Source: authors' compilation based on SSEOA and SARS data.

Job count and job duration, though, are different dimensions of employment stability. Table 6 uses the same job duration categories as the job-level analysis, but reports the mode at the person level. This gives us some indication of whether most of a person's career in the time we observe has been stable or volatile. In the SSEOA data, about 60 per cent of people mostly experienced stable jobs; about one-third mostly experienced moderately stable jobs; and only about 6 per cent experienced unstable or highly unstable jobs. In the SARS panel, similar proportions of people experienced either moderately stable jobs or jobs that lasted a year or less.[22] However, the SARS results are biased downwards by sample restrictions discussed above, so there are higher proportions of people with mainly unstable jobs and lower proportions with mainly stable jobs.

[22] These shares were calculated slightly differently. They are not a mode, but the share for whom at least 60 per cent of their jobs fell into a particular category.

Figure 9: Number of formal-sector jobs per person in South Africa



Notes: SSEOA estimates adjusted using sampling weights.

Source: authors' construction based on SSEOA and SARS data.

Results for both apartheid and, to a lesser extent, the post-apartheid period also find greater gender differences at the edges of the distribution, comparable to the finding by Zizzamia and Ranchhod (2019). Men are much more likely to experience mainly stable careers, with 65 per cent of men mostly experiencing stable jobs in their careers compared to 55 per cent of women. By contrast, 7.76 per cent of women in the SSEOA mostly experience unstable or highly unstable jobs versus less than 5 per cent of men. Women were also 7 percentage points more likely to experience moderately stable jobs compared to men. These results are not adjusted for age, as mentioned previously. If South Africa experienced an increase in female labour force participation towards the second half of the SSEOA (approx. 1975–94), there could be more younger women in the sample, which could be affecting these results. We see the same gender effect to a lesser extent in the SARS panel in that women are 4 percentage points more likely than men to experience unstable or highly unstable jobs.

7.1 Trajectory model regression results

Next, we turn to our regression results in Table 7. Only the output for the previous spell indicator is reported since the other results are much the same. Boockmann and Steffes (2010) run a very similar specification on a data set constructed similarly to the SARS panel, and find that previously unemployed or not employed people worked shorter-duration jobs than those who were previously employed. The results from the SARS panel are the opposite: those previously not employed worked longer job durations than those previously formally employed. Once we control for more covariates in the SARS panel, though, the size of the effect is reduced.

The SSEOA results are mixed depending on whether a frailty model is applied. The frailty model finds previously not employed people worked short-term (or insignificantly shorter) jobs than the reference group. However, the signs turn positive again in our robustness check in Appendix Table A2, which uses robust standard errors instead of frailty for the log-normal and log-logistic. The Cox model also yields positive coefficients, which is similar to the model Boockmann and Steffes (2010) use. These conflicting results highlight the sensitivity of the parametric results to specification. We think on balance the evidence is in favour of longer job durations for those not in the reference group.

These results are counter to the German results and may be related to a slacker labour market in South Africa than in Germany. Unemployment is much higher in South Africa than it is in Germany, and this modulates outside options for a given worker. Someone who has been out of work for a while may prioritize maintaining access to a wage, even if the job is not a good match, if the job-finding rate is low because of high unemployment. Furthermore, these dynamics may be influenced by the structure of the labour market. Numerous authors have characterized the South African labour market as segmented (Heintz and Posel 2008; Kingdon and Knight 1999). A small portion of well-paid highly skilled people can easily obtain secure jobs with good benefits in a well-regulated part of the labour market. The large remainder are less educated and have weaker bargaining power in the face of high open unemployment. They compete for less stable and less skilled jobs in a more poorly regulated part of the labour market.

Potentially, our prior status indicator is proxying for the segment to which workers in our samples belong if we think those belonging to the well-functioning part of the labour market are most likely to be previously formally employed. Those belonging to the well-functioning part of the labour market may be more mobile since they enjoy a higher job-finding rate. In much the same way, we find that more educated workers and more skilled occupations tend to have shorter job durations, because labour market advantage is associated with being more mobile (Burdett 1978). By contrast, those who are able to secure a formal-sector job after unemployment or not being employed may prioritize remaining attached to this job. Those previously informally employed in the SSEOA also work longer job durations than those previously formally employed, supporting this idea.

This variation in employment trajectories would then be an important point of difference between developed- and developing-country labour markets. These differences can likely be attributed to different job-finding rates as a consequence of higher unemployment in the South African labour market, which modulates worker decisions to remain attached to an already-secured job. For example, Boockmann and Steffes (2010) found that the local area unemployment rate reduced the hazard in Germany but only weakly. The direction of the effect is as expected—job durations are longer in areas with more unemployment because people who have jobs may be slower to leave them when they know they have poorer chances of finding another one. The effect in Germany is likely weak because there is not a lot of unemployment in comparison to South Africa, which has experienced excessive unemployment rates of over 25 per cent in the periods covered by both data sets. It is plausible that this effect is much stronger in South Africa, although we unfortunately do not have suitable geographic variables in our data.

Table 7: Regression output for the comparable and multiple trajectory hazard model

DEPVAR: job	SSEOA 1951–91/94				SARS Panel 2011–17		
duration hazard	Log-normal	Log-logistic	Gen. gamma	Cox PH	Log-normal	Log-logistic	Gen. gamma
COMPARABLE TRAJECTORY MODEL							
Prior spell (base = formal sector)							
Informal sector	0.22**	0.13	0.52***	–0.46***			
	(0.08)	(0.08)	(0.08)	(0.07)			
Not employed	–0.08	–0.17***	0.26***	–0.27***	0.13***	0.01***	0.26***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.00)	(0.00)	(0.00)
_cons	7.69***	7.79***	7.62***		5.38***	5.55***	5.83***
	(0.04)	(0.04)	(0.04)		(0.00)	(0.00)	(0.00)
Shape parameters							
lnsigma	0.08***		0.13***		0.17***		0.23***
	(0.01)		(0.01)		(0.00)		(0.00)
lngamma		–0.47***				–0.40***	
		(0.01)				(0.00)	
kappa			0.31***				0.55***
			(0.04)				(0.00)
Frailty							
Theta	0.29***	0.31***	vce	vce	0.38***	0.34***	vce
	(0.02)	(0.02)			(0.00)	(0.00)	
N	12,439	12,439	12,439	12,439	20,709,282	20,709,282	20,709,282

Regression output for the comparable and multiple trajectory hazard model (continued...)

DEPVAR: job	SSEOA 1951–91/94				SARS Panel 2011–17		
duration hazard	Log-normal	Log-logistic	Gen. gamma	Cox PH	Log-normal	Log-logistic	Gen. gamma
MULTIPLE TRAJECTORY MODEL							
Prior spell (base = formal sector)							
Informal sector	0.14	–0.01	0.31***	–0.33***			
	(0.09)	(0.09)	(0.09)	(0.08)			
Not employed	–0.02	–0.13***	0.17***	–0.17***	0.09***	0.04***	0.03***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.00)	(0.00)	(0.00)
_cons	8.31***	8.50***	8.40***		4.01***	4.17***	4.97***
	(0.09)	(0.08)	(0.09)		(0.00)	(0.00)	(0.00)
Shape parameters							
lnsigma	0.06***		0.09***		0.16***		0.13***
	(0.01)		(0.01)		(0.00)		(0.00)
lngamma		–0.51***				–0.42***	
		(0.01)				(0.00)	
kappa			0.36***				0.77***
			(0.03)				(0.00)
Frailty							
Theta	0.18***	0.21***	vce	vce	0.241***	0.216***	vce
	(0.02)	(0.02)			(0.00)	(0.00)	
N	11,035	11,035	11,035	11,035	15,135,340	15,135,340	15,135,340

Notes: standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates based on formal-sector jobs. Only the Cox model is proportional hazards (PH), all other models are accelerated failure time models. Stars on theta are based on the *p*-value from the chi-squared LR test that theta equals zero. vce = robust standard errors applied when frailty infeasible for computational or other reasons. Output for other covariates not reported for reasons of space and are available upon request. Model fit statistics in Appendix Table A3.

Source: authors' compilation based on SSEOA and SARS data.

In sum, job number and job duration are related but different dimensions of employment stability. Results are more mixed in the post-apartheid data, but in the SSEOA women worked both fewer and shorter jobs. Trajectories through the labour market form an important part of characterizing employment stability and worker vulnerability. Regression results indicate that trajectories into formal-sector employment may operate differently in developing versus developed labour markets. More unemployment possibly interacting with labour market segmentation means people who find themselves in formal-sector work when previously they were without it tend to remain in those jobs for longer.

8 Labour market trajectories at early-career stages

In this section we focus on how careers advance at early stages and use a sample of workers' first and second jobs to do so. The sample restriction is described Section 4, but it is worth repeating that this restriction is easier to apply to the SSEOA data because it is episodic, rather than to the SARS data which is a panel. As such, we have a more representative sample of first and second jobs in the SSEOA data; but the sample restriction for the SARS panel had to include an additional age restriction to ensure job sequence was correctly identified. The SARS early-career sample is then representing first and second jobs for those aged 15–21 years of age.

Figure 10 reports how long first and second jobs lasted. Three-quarters of first jobs and over 80 per cent of second jobs are over by the first year in the SARS panel. This is quicker than the whole sample reported in Figure 4, where only 50 per cent of jobs failed by the end of the first year. Gender differences again did not seem pertinent in the SARS panel, but men's jobs lasted longer than women's in the SSEOA, and increasingly so as durations got longer. Male advantage is also more important in second compared to first jobs in Panel A.

Consistently, second jobs fail faster than first jobs in both data sets. In the SSEOA, there is a negative correlation between job sequence and job duration—later jobs are shorter—a result also found in the US case (Farber 1994).[23] This could be evidence in favour of the worker heterogeneity theory for variation in job duration, described in Section 2. People who have many jobs are high-mobility types. As such, job survival and hazard rates for first jobs include in the sample many low-mobility types who extend the survival and reduce the hazard of the first job compared to later jobs, for which the sample is increasingly made up of high-mobility types.

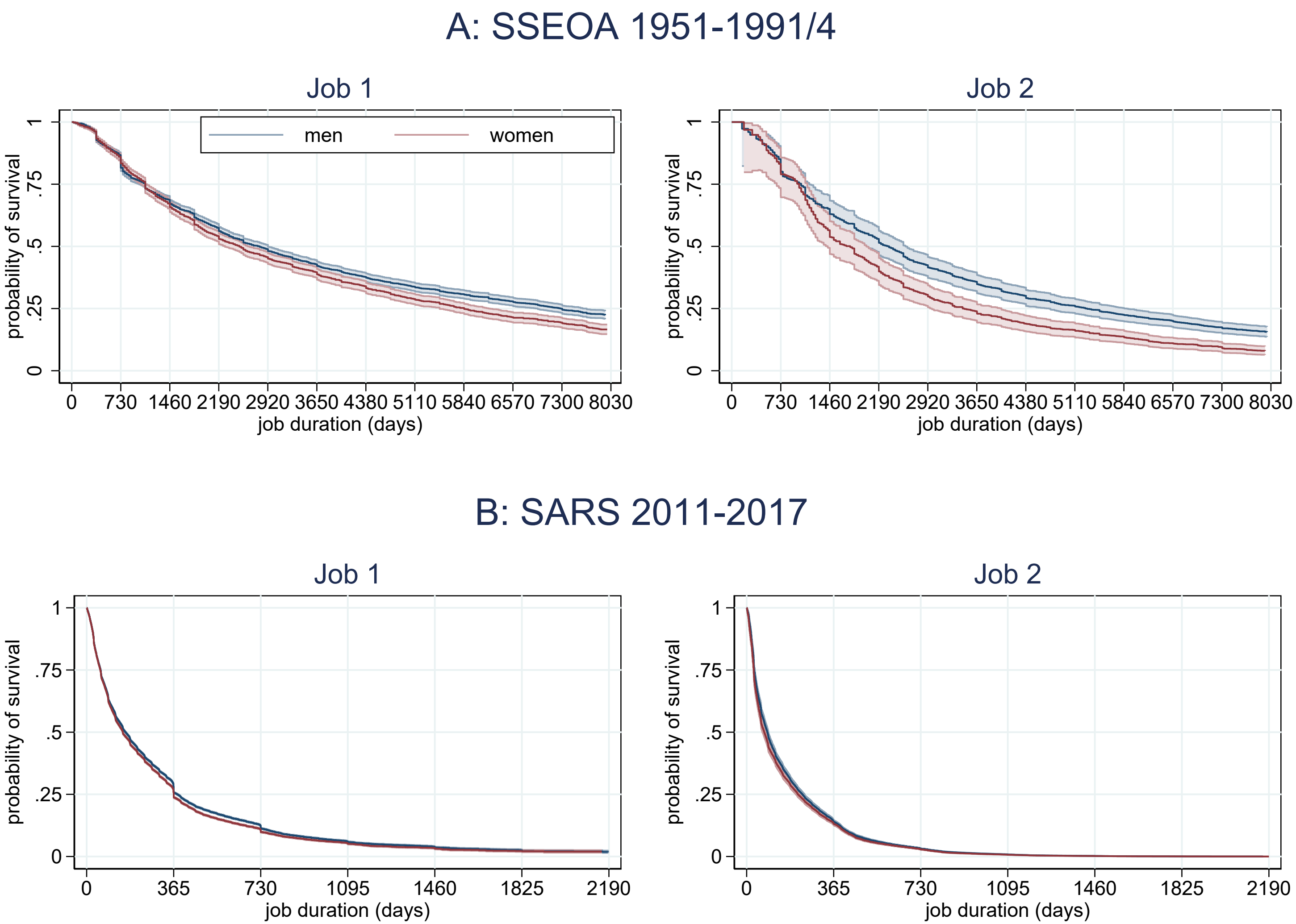
Booth et al. (1999) compare the hazard for first compared to fifth jobs for Britons in the twentieth century. They find that the hazard for the fifth job peaks later than that for the first job.[24] Figure 11 presents hazards by job sequence to understand how first jobs look different to second jobs. The hazards do not peak later as in the British case, but rather second jobs consistently have higher hazards than first jobs. The peak of the first job hazard also looks more dispersed than that of the overall case in Figure 5. A more diffused peak could indicate a slower, more drawn-out matching process, perhaps impeded by a lack of experience about judging job match quality or confidence about leaving an ill-fitting job.

---

[23] A one unit increase in job sequence number raised the job duration hazard by 20 percentage points and was statistically significant at the 1 per cent level in a Cox regression of the hazard on job sequence number with standard errors robust to clustering at the person level.

[24] They interpret this as meaning that learning about match quality is slower by the fifth job; that workers remain longer in jobs as their career advances because they are worried many short jobs send bad signals to prospective employers; and because employers may find it harder to retrench an experienced worker.

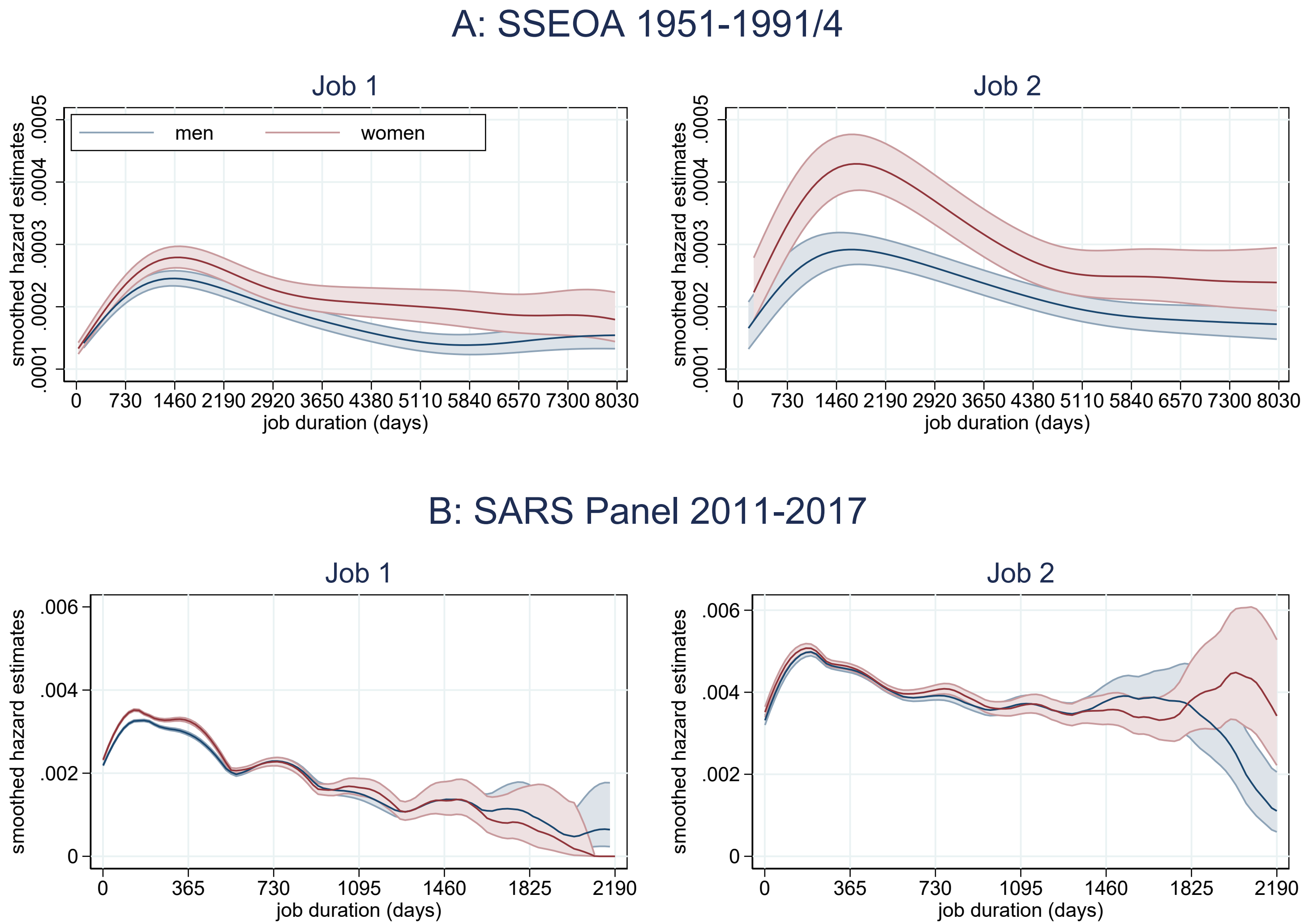
Figure 10: Kaplan–Meier estimates of the job survival functions of first and second jobs in the South African formal sector



Notes: SSEOA estimates adjusted using sampling weights and all estimates reported with a 95 per cent confidence interval.

Source: authors' construction based on SSEOA and SARS data.

Figure 11: Kaplan–Meier estimates of the job duration hazard for first and second jobs in the South African formal sector



Notes: SSEOA estimates adjusted using sampling weights and all estimates reported with a 95 per cent confidence interval.

Source: authors' construction based on SSEOA and SARS data.

Regression results by job sequence support the conclusions drawn from the hazards. Regression output is in Table 8 and model fit statistics are in Appendix Table A4. In some cases, the models did not compile. There are substantial gender differences in the SSEOA: men's jobs last longer in general, but this is especially the case for second jobs. Men's second jobs last 27 per cent longer than women's according to the generalized gamma result for the multiple descriptive model, and their second job hazard is reduced by 20 percentage points according to the Cox model. By contrast, in the SARS panel, the gender gap is more important in first jobs, although the magnitude is small. Men's first jobs last 3 per cent longer than women's using the multiple descriptive model, but second jobs are insignificantly different in duration to women's.

Table 8: Regression output for the comparable descriptive and multiple descriptive hazard model for first and second jobs

DEPVAR: job	SSEOA 1951–91/94				SARS panel 2011–17	
duration hazard	Job 1		Job 2		Job 1	Job 2
	Gen. gamma	Cox PH	Gen. gamma	Cox PH	Log-normal	Log-normal
COMPARABLE DESCRIPTIVE MODEL						
Male	0.09*	–0.12***	0.46***	–0.31***	0.05***	0.04**
	(0.04)	(0.03)	(0.07)	(0.05)	(0.01)	(0.02)
_cons	7.84***		7.44***		4.99***	4.65***
	(0.04)		(0.08)		(0.00)	(0.02)
Shape parameters						
lnsigma	0.23***				0.35***	0.02**
	(0.01)				(0.00)	(0.00)
kappa	0.03					
	(0.05)					
N	6,104	6,104	3,181	3,181	302,570	65,888
MULTIPLE DESCRIPTIVE MODEL						
Male	0.14***	–0.18***	0.27***	–0.20***	0.03***	–0.00
	(0.04)	(0.04)	(0.07)	(0.05)	(0.01)	(0.02)
_cons	8.84***		8.16***		–0.46***	–3.71***
	(0.13)		(0.19)		(0.14)	(0.65)
Shape parameters						
lnsigma	0.15***		0.14***		0.29***	–0.13***
	(0.01)		(0.04)		(0.00)	(0.01)
kappa	0.22***		0.41**			
	(0.05)		(0.13)			
N	5,404	5,404	2,824	2,824	244,997	54,088

Notes: robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates based on formal-sector jobs. Only the Cox model is proportional hazards (PH), all other models are accelerated failure time models. Output for other covariates in the multiple descriptive models not presented for reasons of space and are available from the authors upon request. All models use standard errors robust to clustering at the person level.

Source: authors' compilation based on SSEOA and SARS data.

9 Conclusion

This study has taken a broad view to exploring how two new data sets, the SSEOA and the SARS panel, can be used to study job duration in South Africa for the first time. These data sets provide new opportunities to answer questions about job duration in a developing country and in two very different parts of South African history. However, important differences in the data structure—episodic versus panel structure—influence the sample of jobs observed. These differences are not important when using each data set in isolation, and precedent exists for using each data structure for job duration analysis in the international literature. The difficulty lies in trying to make a comparison of the apartheid and post-apartheid labour markets and must be kept in mind when trying to do so.

We provide a baseline description of job duration in South Africa and find that we adhere to the three main stylized facts about job duration described in the job duration literature on the developed world. First, long-term employment is common and the levels in South Africa are comparable to other developed countries in the post-apartheid period: 31 per cent of men and 27 per cent of women over the age of 46 are currently in jobs that have lasted 20 years or more. Second, most new jobs end early in South Africa. Half of new jobs are over after the first year in the SARS panel. This survival rate is very similar to the United States, but shorter than Germany for the mid-1990s. Third, the job hazard declines with tenure. It declines more slowly in the apartheid case than the post-apartheid case, but these differences could be related to differences in data structures and granularity of time. Overall, the average job lasts more than one year but less than five, making it moderately stable, and job duration has probably declined since apartheid, although data constraints make it hard to be precise.

The unique structure of the SSEOA data and the SARS panel also allow us to investigate how people move through the labour market. In South Africa, people who were previously not employed or informally employed worked longer job durations than those previously formally employed. This is the opposite pattern to what was found using German data; those previously formally employed worked longer jobs. We interpret this as an important difference between trajectories through developed- versus developing-country labour markets related to segmentation in the South African case.

Gender is an important variable in most job duration analyses. We find that gender becomes an important predictor of job duration towards the extremes of distributions. During apartheid, South Africa had a smaller gender gap in long-term tenure than other developed nations. Coloured and Black African women had very high levels of long-term tenure compared to American women or white South African women at the same time, and these higher long-term tenure levels were driven by segregationist labour policy. Interpretation of job duration in the SSEOA must carefully take the effects of discriminatory labour policies into account. In the post-apartheid period, studies on employment dynamics in South Africa have concluded that women lead more volatile work lives than men (Zizzamia and Ranchhod 2019). Similarly, we find more women experiencing our definitions of highly unstable and unstable jobs and fewer women experiencing stable jobs. Initially, men and women have similar job survival rates, but as job duration lengthens, men's jobs start outlasting women's.

This study has provided a broad descriptive baseline for the study of job duration in South Africa. The research agenda is wide open for future work. One immediate direction is to flesh out the baseline by converting the data to a discrete format. This would allow for better adjustment of all the results by age, which is a key dimension of job duration and employment trajectories. Other promising avenues include richer work on trajectories and incorporation of wages. In general, we have found that labour market advantage is associated with working shorter jobs: more-educated and higher-skilled workers work shorter jobs (and during apartheid white people worked shorter jobs). This may sound counter-intuitive when most South Africans prioritize job security and stability. However, according to theory, jobs become shorter when outside options are better. A healthy, well-functioning labour market yields a steep peak on the hazard, but thereafter we want the hazard to decline quickly, representing good job security after matching. Hazard analysis reveals better job security in white-collar occupations, where job-specific capital is more important, and in the public sector.

References

Banerjee, A., S. Galiani, J. Levinsohn, Z. McLaren, and I. Woolard (2008). 'Why Has Unemployment Risen in the New South Africa? 1'. *Economics of Transition*, 16(4): 715–40.

Becker, G.S. (1964). Human Capital: A Theoretical and Empirical Analysis With Special Reference to Education. Chicago, IL: Chicago University Press.

Bhorat, H., R. Hill, S. Khan, K. Lilenstein, and B. Stanwix (2019). 'Quantitative Review of the Employment Tax Incentive'. Report. Pretoria: Development Policy Research Unit, South African National Treasury.

Bhorat, H., K. Lilenstein, M. Oosthuizen, and A. Thornton (2016). 'Vulnerability in Employment: Evidence from South Africa'. Working Paper 201604. Cape Town: Development Policy Research Unit, University of Cape Town.

Boockmann, B., and T. Hagen (2008). 'Fixed-Term Contracts as Sorting Mechanisms: Evidence from Job Durations in West Germany'. *Labour Economics*, 15(5): 984–1005.

Boockmann, B., and S. Steffes (2010). 'Workers, Firms, or Institutions: What Determines Job Duration for Male Employees in Germany?' *ILR Review*, 64(1): 109–27.

Booth, A.L., M. Francesconi, and C. Garcia-Serrano (1999). 'Job Tenure and Job Mobility in Britain'. *ILR Review*, 53(1): 43–70.

Box-Steffensmeier, J.M., and B.S. Jones (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge: Cambridge University Press.

Burdett, K. (1978). 'A Theory of Employee Job Search and Quit Rates'. *American Economic Review*, 68(1): 212–20.

Burgess, S. (1999). 'The Reallocation of Labour: an International Comparison Using Job Tenure'. CEPDP Working Paper 416. London: Centre for Economic Performance, London School of Economics and Political Science.

Christofides, L.N., and C.J. McKenna (1993). 'Employment Flows and Job Tenure in Canada'. *Canadian Public Policy/Analyse de Politiques*, 19(2): 145–61.

Cichello, P., M. Leibbrandt, and I. Woolard (2014). 'Winners and Losers: South African Labour-Market Dynamics Between 2008 and 2010'. *Development Southern Africa*, 31(1): 65–84.

Dendir, S. (2006). 'Unemployment Duration in Poor Developing Economies: Evidence from Urban Ethiopia'. *Journal of Developing Areas*, 40(1): 181–201.

Diebold, F.X., D. Neumark, and D. Polsky (1997). 'Job Stability in the United States'. *Journal of Labor Economics*, 15(2): 206–33.

DPRU (2017). 'An Overview of the South African Labour Market for the Year Ending 2016 Quarter 4'. Factsheet 17. Cape Town: Development Policy Research Unit, University of Cape Town.

Essers, D. (2016). 'South African Labour Market Transitions since the Global Financial and Economic Crisis: Evidence from Two Longitudinal Datasets'. *Journal of African Economies*, 26(2): 192–222.

Farber, H.S. (1994). 'The Analysis of Interfirm Worker Mobility'. *Journal of Labor Economics*, 12(4): 554–93.

Farber, H.S. (1999). ‘Mobility and Stability: The Dynamics of Job Change in Labor Markets’. *Handbook of Labor Economics*, 3: 2439–83.

Farber, H.S. (2009). ‘Job Loss and the Decline in Job Security in the United States’. Working Paper 520. Princeton, NJ: Industrial Relations Section, Princeton University.

Fedderke, J. (2012). ‘The Cost of Rigidity: The Case of the South African Labor Market’. *Comparative Economic Studies*, 54(4): 809–42.

Go, D.S., M. Kearney, V. Korman, S. Robinson, and K. Thierfelder (2009). Wage Subsidy and Labor Market Flexibility in South Africa. Working Paper. Washington, DC: World Bank.

Heintz, J., and D. Posel (2008). ‘Revisiting Informal Employment and Segmentation in the South African Labour Market’. *South African Journal of Economics*, 76(1): 26–44.

Hollister, M.N., and K.E. Smith (2014). ‘Unmasking the Conflicting Trends in Job Tenure by Gender in the United States, 1983–2008’. *American Sociological Review*, 79(1): 159–81.

Horny, G., R. Mendes, and G.J. van den Berg (2012). ‘Job Durations with Worker-and Firm-Specific Effects: MCMC Estimation with Longitudinal Employer–Employee Data’. *Journal of Business & Economic Statistics*, 30(3): 468–80.

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

Jackman, R., and R. Layard (1991). ‘Does Long-Term Unemployment Reduce a Person’s Chance of a Job? A Time-Series Test’. *Economica*, 58(229): 93–106.

Jovanovic, B. (1979). ‘Job Matching and the Theory of Turnover’. *Journal of Political Economy*, 87(5): 972–90.

Kalleberg, A.L., and S.P. Vallas (2017). ‘Probing Precarious Work: Theory, Research, and Politics’. In *Precarious Work*. Bingley: Emerald Publishing.

Katz, L.F., and A.B. Krueger (2019). ‘The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015’. *ILR Review*, 72(2): 382–416.

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

Kerr, A., D. Lam, and M. Wittenberg (2019). ‘Post-Apartheid Labour Market Series’. Dataset. Version 3.3. Cape Town: DataFirst (producer and distributor).

Kerr, A., and M. Wittenberg (2019). ‘A Guide to Version 3.3 of the Post-Apartheid Labour Market Series (PALMS)’. Technical Report. Cape Town: DataFirst.

Kingdon, G., and J.B. Knight (1999). ‘Unemployment and Wages in South Africa: A Spatial Approach’. Working Paper 99-12. Oxford: Centre for the Study of African Economies, Oxford University.

Leibbrandt, M., A. Finn, and I. Woolard (2012). ‘Describing and Decomposing Post-Apartheid Income Inequality in South Africa’. *Development Southern Africa*, 29(1): 19–34.

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(1): 73–87.

Mncwango, B. (2016). ‘Public attitudes to work in South Africa’. Report 16. Pretoria: Labour Market Intelligence Partnership.

Nattrass, N., and J. Seekings (2011). *Cambridge History of South Africa, vol. 2: The Economy and Poverty in the Twentieth Century*. Cambridge: Cambridge University Press.

Ranchhod, V., and T. Dinkelman (2007). 'Labour Market Transitions in South Africa: What Can We Learn from Matched Labour Force Survey Data?' Working Paper 14. Cape Town: SALDRU, University of Cape Town.

Schotte, S., R. Zizzamia, and M. Leibbrandt (2018). 'A Poverty Dynamics Approach to Social Stratification: The South African Case'. *World Development*, 110: 88–103.

Statistics South Africa (2017). 'Poverty Trends in South Africa: An Examination of Absolute Poverty Between 2006 and 2015'. Report 03-10-06. Pretoria: Government of South Africa.

Tansel, A., and H.M. Taşçı, (2010). 'Hazard Analysis of Unemployment Duration by Gender in a Developing Country: The Case of Turkey'. *Labour*, 24(4): 501–30.

Treiman, D.J., A. Lewin, Y. Lu, M. McKeever, G. Espi-Sanchis, M. Wittenberg, and A. Kerr (2017). 'Survey of Socioeconomic Opportunity and Achievement in South Africa: Codebook (Updated)'. Technical Report. Cape Town: DataFirst.

Zizzamia, R., and V. Ranchhod (2019). 'Measuring Employment Volatility in South Africa Using NIDS: 2008–2017'. Discussion Paper 2019/13. Cape Town: SALDRU, University of Cape Town.

Appendix 1: regression robustness checks

Table A1: Regression output for the comparable descriptive hazard model using a sample of 'main' jobs in the SARS panel

DEPVAR: job duration hazard	Log-normal	Log-logistic	Gen. gamma
Male	0.01***	0.03***	–0.03***
	(0.00)	(0.00)	(0.00)
_cons	5.55***	5.67***	6.09***
	(0.00)	(0.00)	(0.00)
Shape parameters			
ln(sigma)	0.22***		0.27***
	(0.00)		(0.00)
ln(gamma)		–0.3***	
		(0.00)	
Kappa			0.51***
			(0.00)
Frailty			
Theta	0.34***	0.31***	vce
	(0.00)	(0.00)	
Model fit			
Log likelihood	–22,968,016	–23,036,249	–22,997,855
AIC	45,936,040	46,072,505	45,995,717
BIC	45,936,099	46,072,564	45,995,776
Degrees of freedom	4	4	4
N	18,397,062	18,397,062	18,397,062

Notes: standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates based on formal-sector jobs. All models are accelerated failure time models with standard errors robust to clustering at the person level. 'Main' jobs defined as jobs that are either not overlapping or are the highest-paid job in the case of job overlap.
Source: authors' compilation based on ther SARS IRP5 panel.

Table A2: Trajectory models in the SSEOA with robust standard errors instead of frailty for the log-normal and log-logistic models

DEPVAR: job duration hazard	Comparable trajectory model		Multiple trajectory model	
	Log-normal	Log-logistic	Log-normal	Log-logistic
Prior spell (base = formal sector)				
Informal sector	0.55***	0.45***	0.33***	0.19*
	(0.08)	(0.09)	(0.09)	(0.09)
Not employed	0.28***	0.17***	0.20***	0.08*
	(0.04)	(0.04)	(0.04)	(0.04)
_cons	7.44***	7.55***	8.19***	8.38***
	(0.04)	(0.04)	(0.10)	(0.10)
Shape parameters				
lnsigma	0.15***		0.11***	
	(0.01)		(0.01)	
lngamma		–0.40***		–0.47***
		(0.01)		(0.01)
Model fit				
Log Likelihood	–11,721.78	–11,825.72	–9,912.314	–9,954.434
AIC	23,453.56	23,661.44	19,866.63	19,950.87
BIC	23,490.71	23,698.59	20,020.11	20,104.35
Degrees of freedom	5	5	21	21
N	12,439	12,439	11,035	11,035

Notes: standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates based on formal-sector jobs. All models are accelerated failure time models with standard errors robust to clustering at the person level. Output for other covariates not reported for reasons of space and are available upon request.

Source: authors' compilation based on SSEOA data.

Appendix 2: extra model fit statistics

Table A3: Model fit statistics for regressions in Table 7

	SSEOA 1951–91/94				SARS panel 2011–17		
	Log-normal	Log-logistic	Gen. gamma	Cox PH	Log-normal	Log-logistic	Gen. gamma
COMPARABLE DESCRIPTIVE MODEL							
Log likelihood	–11,585	–11,667	–11,677	–65,694	–24,420,825	–24,550,201	–24,450,963
AIC	23,183	23,346	23,366	131,394	48,841,661	49,100,412	48,901,936
BIC	23,227	23,390	23,410	131,417	48,841,735	49,100,486	48,902,010
Degrees of freedom	6	6	6	3	5	5	5
MULTIPLE TRAJECTORY MODEL							
Log likelihood	–9,846	–9,873	–9,846	–56,426	–16,715,334	–16,853,100	–16,296,403
AIC	19,736	19,790	19,735	112,891	33,430,716	33,706,248	32,592,853
BIC	19,896	19,950	19,896	113,030	33,431,065	33,706,597	32,593,202
Degrees of freedom	22	22	22	19	24	24	24

Table A4: Model fit statistics for regressions in Table 8

	SSEOA 1951–91/94				SARS panel 2011–17	
	Job 1		Job 2		Job 1	Job 2
	Gen. gamma	Cox PH	Gen. gamma	Cox PH	Log-normal	Log-normal
COMPARABLE DESCRIPTIVE MODEL						
Log likelihood	–8,337.967	–33,576.61	–2,256.398	–13,043.41	–443,570.88	–27,739.49
AIC	16,683.93	67,155.23	4,520.797	26,088.81	887,147.77	55,484.99
BIC	16,710.8	67,161.95	4,545.057	26,094.88	887,179.63	55,512.27
Degrees of freedom	4	1	4	1	3	3
MULTIPLE DESCRIPTIVE MODEL						
Log likelihood	–7,047.577	–28,594	–1,893.253	–11,162.74	–340,491.09	–20,656.28
AIC	14,135.15	57,222	3,826.506	22,359.48	681,026.17	41,356.56
BIC	14,267.05	57,334.11	3,945.425	22,460.56	681,255.17	41,552.32
Degrees of freedom	20	17	20	17	22	22