Gender and Technological Change: Measuring Vulnerability to Technological Displacement in the South African Labour Market

---

Leslie Dwolatzky

Supervisors:
Prof V. Schöer
Prof E. Webster

A report submitted in partial fulfillment of the requirements for
the degree of Master of Arts in the field of e-Science
in the
School of Social Sciences
University of the Witwatersrand, Johannesburg
3 May 2022

Abstract

Across sectors and industries, recent developments in machine learning are having increased applications to processes of economic production. As these systems become more capable of performing complex tasks and become more cost effective, firms are incentivised to incorporate new technologies into their businesses, potentially at the expense of labour. Depending on how easily the occupational tasks that workers perform can be simulated by new technologies, labour-substituting capital has the propensity to become a cause of technological displacement. It is in this context and with reference to the case of the South African labour market that this research project investigates which demographic groups of workers are most vulnerable to replacement by machine learning systems and technological change. Utilising the Brynjolfsson, Mitchell and Rock (2018) 'suitability for machine learning' (SML) framework, the project categorises the South African labour force according to the demographic indicators of race, gender and age. Employing relative distribution methods and quantile regression models, the project finds evidence that South Africa's young, African female population is significantly more vulnerable to replacement by technological systems compared to other demographic groups. This is explained by the types of occupations that South African women perform. Specifically, as the African female population becomes more educated and more sought-after employees in the labour market, they are moving towards more clerical and administrative occupations. While these occupations offer a more stable source of employment and income compared to the elementary occupations they are moving away from, these occupations are theorised to involve tasks which can be more easily performed and automated by new technologies.

Declaration

I, Leslie Dwolatzky, declare that this report is my own, unaided work. It is being submitted for the degree of Master of Arts in the field of e-Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.

Leslie Dwolatzky

3 May 2022

Acknowledgements

Working on this project, I have had the great privilege of being supervised by two leading experts and giants of this academic field. I am deeply indebted to Volker Schöer and Edward Webster for their guidance, patience, enthusiasm, teachings and support. Thank you both for taking me on and providing me with this invaluable wealth of knowledge.

This research would not have been possible without the MA e-Science programme at Wits University and the DST-CSIR National e-Science Postgraduate Teaching and Training Platform, which I see as a vital component in addressing the challenges outlined in this paper. Thank you to Rod Alence, Casey Sparkes and Helen Robertson for ensuring the flawless running of the programme and giving me these life-changing skills.

Thank you to Adeola Oyenubi for aiding this research and sharing your valuable insights.

This paper is the culmination of a monumental twelve month period which I would not have navigated without the support of my family- Rina, Barry and Jodie- and friends- Colin, Isaiah and Megan. Thank you.

The support of the DST-CSIR National e-Science Postgraduate Teaching and Training Platform (NEPTTP) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NEPTTP.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With each wave of industrialisation, economists have sought to conceptualise the effects of emerging technologies on the structure and composition of labour markets. Such investigation is intended to identify the types of workers which are most vulnerable to the threat of technological displacement by labour-substituting technology. With advancements in machine learning (ML) and its increased applicability to processes of economic production, a worker's skill-level is no longer an effective measure of such vulnerability. As processing power becomes faster and data storage and collection become more efficient, ML systems are becoming more capable of replacing workers who have conventionally been thought of as 'high-skilled'. Thus, instead of one's skill-level, what is important in identifying vulnerable workers is examination of the occupational tasks that they perform. Generally, machines and software programmes are less capable of performing more abstract and nonroutine tasks. Following this logic, Brynjolfsson, Mitchell and Rock's (2018) 'suitability for machine learning' (SML) framework offers an effective approach to quantify the extent to which occupational tasks and occupations can be performed by ML systems instead of by humans.

Utilising this framework, the following research project investigates various dynamics of the South African labour market in order to identify the demographic characteristics of workers who perform occupations that are theorised to be most vulnerable to technological diplacement. The Bynjolfsson, Mitchell and Rock (2018) SML measure is applied to the South African labour market by matching this variable to workers and their occupations found in the South African Quarterly Labour Force Survey (QLFS). With the SML measure as the primary variable of interest, relative distribution methods (Handcock & Morris 1999) are employed, and quantile regression models are constructed in order to analyse the demographic and educational characteristics of the SML distribution. This analysis includes both a cross-sectional component of differences between the demographic populations within the 2019 labour market; and a time-series component comparing changes to the SML distribution between 2009 and 2019 for these demographic populations.

The study finds evidence that South African women are significantly more vulnerable to the threat of technological displacement compared to their male counterparts; with young, black women being the most vulnerable demographic segment of the South African labour force. While the vulnerability of the female population as a whole has decreased between 2009 and 2019, the vulnerability of young, black women has increased in this ten year period. This finding is explained by the types of occupations this demographic group are performing. As South Africa's young, black women become more educated and more active participants and sought-after employees in the labour market, they are moving towards more clerical and service-orientated occupations. While these occupations offer a more stable form of employment and a higher income than the elementary occupations they are moving out of, the routine nature of the occupational tasks makes them highly vulnerable to replacement by ML software systems.

1.1 Context and Background of Research Problem

Structural unemployment and income inequality are two of the most pressing challenges facing the South African economy. The country's Gini coefficient of 0.66 in 2018 is one of the highest in the world and has been at this level throughout the post-apartheid period

(Francis & Valodia 2021). The broad unemployment rate has also been consistently high at around 35 per cent. The causes of these challenges are multiple, complex, and interrelated: a concoction of factors- international and domestic, structural and policy-driven-that have played off each other to entrench a socioeconomic system characterised by stark inequality. From the legacies of apartheid, these inequalities are largely racialised, but also include gendered and age-based dimensions.

Apartheid's structural inequality has been exacerbated by factors that have come to the fore in the new Millennium. Globalisation, changing economic trends, and the introduction of new technologies in processes of economic production have affected the sectoral composition of the South African economy and, as such, have impacted the demand for labour in several established South African industries. These phenomena have both worsened apartheid's economic legacies and introduced new structural causes of inequality and unemployment. Most crucially, they have significantly changed the solutions needed to address these issues. It is now not simply sufficient to get marginalised South Africans into the labour force. It is now imperative to get marginalised South Africans into jobs that are resilient to the effects of globalisation and technological changes. There must necessarily be a transformation of the South African economy to one that has a space and function in the global economy and is able to harness the potentials of emerging technologies to complement labour.

It is in this context that this research project is situated. It is important to note that it is not a comprehensive discussion on all the factors relating to unemployment and inequality in South Africa; nor does it offer concrete solutions for addressing these challenges. Rather, it proposes a lens through which to view one of the factors responsible for the country's structural unemployment, displacement and income inequality: technological change and machine automation. As detailed in the theoretical framework below, the effects of new technologies on the demand for labour differ according to the occupation in question. These effects are a function of the occupational tasks that are being performed where workers who perform tasks that can be most easily simulated by a machine or software programme are most at risk of technological replacement. Conceptualising the composition of the South African economy in terms of these categories of occupational tasks and identifying segments of the workforce who perform occupations that are most vulnerable to technological displacement is a crucial first step in the objective of transforming the South African economy.

1.2 Research Question

The primary research question under investigation can be stated as follows: Classified according to demographic and educational indicators, which segments of the South African labour force are more likely to perform occupational tasks that are most vulnerable to the threat of technological replacement through machine learning systems?

1.3 Significance of Study

The conceptualisation of the SML measure is a contemporary approach to investigation into the effects of technological change on labour. As discussed below, it is future-orientated in that it includes in its framework the potential for technological replacement of occupational tasks, beyond what developments in new technologies may allow today. With the current rates of growth in computational power, academic investigation into these forces must be able to adapt alongside these developments in order to remain relevant.

Informed by a review of the literature, this study represents the first application of the SML task-based approach to the South African context. While the measure has been applied to the US labour market (Brynjolfsson, Mitchell and Rock 2018) which experiences a higher penetration of the latest technological advancements in its economic activities and a higher degree of routinisation, applying the measure to a middle-income, developing economy such as South Africa has significant benefits. As ML systems become more cost effective and data across industries become more available, the increased use of these systems in the South African economy becomes more feasible. At this juncture, the country is in a position to prepare for what is to come. This study, thus, represents a contribution both to the literature and to this endeavour by identifying the structure of the South African labour force in terms of its suitability for machine learning and its vulnerability to present and future technological displacement.

1.4 Assumptions and Limitations

Due to various specificities of the data and methodology, the study includes the following assumptions and limitations:

1.4.1 Compatibility of US and South African Occupations

The Byrnjolfsson, Mitchell and Rock (2018) SML measure is constructed at the level of occupational tasks using the US Occupational Information Network (O*NET) classification system and then aggregated to the level of O*NET-defined occupations. For this study, the O*NET occupations and their associated SML scores were matched to the occupations defined by the South African Standard Classification of Occupations (SASCO). Thus, it is assumed that the occupations identified in the US labour market are congruent to those in the South African labour market in terms of the occupational tasks involved. This is a limitation in that, ideally, the SML score would be constructed specifically for the South African context, taking into account the specific occupational tasks involved and the specific occupations which may not translate to other labour markets. While I did attempt to use the disaggregated task data as provided for by StatsSA, it was found that the qualify of this data was not appropriate for analysis. Specifically, the task variable was not codified, was incomplete, and was self reported by respondents meaning that they could not be generalised to the occupational level.

1.4.2 Unchanged Occupational Tasks

Because the data for the SML measure was only collected once using the current O*NET definitions of occupations and their associated occupational tasks, the study also assumes that the occupational tasks necessary to perform an occupation have not changed between 2009 and 2019. For example, it is assumed that the occupational tasks of a bookkeeper (and this occupation's associated SML score) have remained the same in this ten year period. In reality, while an occupation may keep the same nominal title, the tasks required of a worker in that occupation may be dynamic and adapt in order to complement new technologies. This feature is not captured in this study.

1.5 Overview of the Study

The organisation of this paper proceeds as follows. The next chapter outlines the theoretical framework of the study and surveys the literature pertaining to technological change and approaches to measuring its effects on labour. Here, I outline the Brynjolfsson, Mitchell and Rock (2018) suitability for machine learning (SML) framework. I also detail the South African context in terms of sources of unemployment and inequality in the labour market. Chapter Three then outlines the research methodology. I describe the datasets used and provide an overview of the relative distribution methods and quantile regression models which are used in the analysis. I also provide descriptive statistics for the SML measure in the South African context. Chapter Four presents the results of the analysis. This chapter is divided into a cross-sectional and time-series component. The cross-sectional component looks at differences in the SML distributions of different demographic populations within the 2019 labour force; while the time-series component compares the 2019 to the 2009 labour force and identifies changes in the vulnerability of demographic populations in this ten year period. I conclude with recommendations for future studies on this topic.

Chapter 2

Theoretical Framework and Literature Review

In 1930, in an essay entitled *Economic Possibilities for our Grandchildren*, John Maynard Keynes defined technological unemployment as "unemployment due to our discovery of means of economising the use of labour outrunning the pace at which we can find new uses for labour" (Keynes 2010:323). This 'discovery of means of economising the use of labour' is indicative of the arrival of a new form of technology that may have a positive impact on the effectiveness or efficiency of production, but a negative effect on the demand for labour. Although he warns that "the very rapidity of these changes is hurting us and bringing difficult problems to solve" (Keynes 2010:323), Keynes' essay is largely optimistic. He argues that technological unemployment is ultimately a short-term phenomenon; a "temporary phase of maladjustment [that] means in the long run *that mankind is solving its economic problem*" (Keynes 2010:323).

More than ninety years after Keynes penned this essay, we are now in a position to investigate the empirical trajectories of technological change, unemployment and displacement in the Twentieth and Twenty-First Centuries. We can examine the forms these changes have taken and locate the groups of workers that have have been most affected in the labour market. Importantly, we must ask whether technological unemployment has indeed been a short-term phenomenon and assess our achievements in finding 'new uses for labour'.

The following chapter traces the theories in the field of labour economics which address these questions. Specifically, I outline two broad approaches to conceptualising the effects of technological change on labour demand: the skill-biased and task-based approaches. Within the task-based approach, I focus on the Brynjolfsson, Mitchell and Rock (2018) 'suitability for machine learning' framework which will be used in the methodology of this study. In the context of this theoretical framework, I then survey the major studies which have utilised these theories and present the important empirical findings relating to technological displacement and unemployment more broadly in South Africa which will inform the analysis in subsequent chapters.

2.1 Approaches to Measuring Technological Change

2.1.1 The Skill-Biased Approach

Labour economists have sought to establish theories that identify the factors that determine what makes a worker more or less vulnerable to technological displacement. The most significant obstacle in this literature has been the unpredictable nature of novel technological changes which are being introduced at an increasingly rapid rate. As such, models have been developed, refuted and revised in this evolving field. For the skill-biased approach, the important factor which determines a worker's vulnerability to technological displacement is their skill level. Specifically, this approach differentiates workers according to the binary groupings of high- and low—skilled. This approach argues that technological change will either complement or substitute workers depending on their skill level. Studies which utilise this approach generally measure skill level in terms of education (Autor, Levy & Murnane 2003).

Empirically, the skill-biased approach is supported by the observed increase in the wage differential between educated (high-skill) and less educated (low-skill) workers in the US and other developed countries in the second half of the Twentieth Century. Studies by Tinbergen (1974) and Katz and Murphy (1991) have shown that technological change is a significant driver in complementing high-skilled labour and substituting low-skilled labour; thus increasing the relative wages of the more educated population (the so-called 'college premium').

The skill-biased approach reflects the type of technological change that may be found on the factory floor. Since the beginning of the Twentieth Century, electrification combined with Fordist methods of mass production have seen innovations in machines that can automate the manual components of the production process. This has seen the replacement of low-skilled, blue collar workers. The result of this, as Frey and Osborne (2017:257) observe, is that "since electrification, the story of the twentieth century has been the race between education and technology".

2.1.2 The Task-Based Approach

Despite its empirical strengths in identifying the observed correlation between education and technological change, the skill-biased approach fails to explain the causality of the relationship (Autor, Levy & Murnane 2003). In addition, as Acemoglu and Autor (2010) argue, the model sees the introduction of new technology as exogenous and technological change as *inherently* skill-biased. Despite a large proportion of observed technological change taking the form of factory automation as described above, there are other forms of technological change that cannot be understood through the skill-biased approach. To account for this, labour economists have conceptualised an alternative model that focuses on the occupational tasks that are being performed instead of a worker's skill level.

An occupational task is defined as an activity that contributes to the production of a good or the fulfilment of a service. An occupation, in turn, can be thought of as a bundle of one or more tasks (Autor, Levy & Murnane 2003). Through innate ability, education, training and experience, a human is endowed with a certain skillset that determines their capacity to perform certain tasks. On the other hand, a machine's capacity to perform a task is a function of the nature of the task; specifically, how easily it can be codified into a set of clearly defined rules and instructions which can be executed statically without deviation.

From this, the task-based approach categorises workers according to the occupational tasks that they perform. Pioneering this approach, Autor, Levy and Murnane (2003) identify four such categories based on two binary classifications: manual and cognitive, and routine and nonroutine. The authors argue that both cognitive and manual routine tasks (for example, record-keeping and factory assembly, respectively) are subject to "substantial substitution" (Autor, Levy & Murnane 2003:1286) by technological change. This is because the routine nature of the tasks makes them easier to automate with machines and software. Nonroutine cognitive tasks performed by, for example, managers and salespeople, offer "strong complementarities" (Autor, Levy & Murnane 2003:1286) to new technologies as the abstract and unpredictable nature of these tasks cannot be easily codified. Finally, occupations involving nonroutine manual tasks such as cleaners and truck drivers offer "limited opportunities for substitution or complementarity" (Autor, Levy & Murnane 2003:1286).

The major strength of the task-based approach is that it includes in its framework the properties that enable technology to substitute labour: the codifiability of the occupational tasks. It incorporates the findings of the skill-biased approach in that it conceptualises routine manual jobs (which are generally low-skilled) as highly vulnerable to technological replacement and nonroutine cognitive tasks (which are generally high-skill) as very secure and even complementary to new technologies. However, it builds on this approach in two important ways. Firstly, it recognises that the effect of technological change on nonroutine manual occupations (which are usually considered low-skill) is more complex than what the skill-biased approach allows. We cannot assume that these occupations are intrinsically more vulnerable simply because they require a relatively lower skill level. Rather, we must determine whether the occupational tasks are sufficiently routine to be executed by a programmed machine (Acemoglu & Autor 2010).

Secondly, the task-based approach recognises the potential vulnerability of routine cognitive occupations. These occupations which include, for example, bookkeepers, secretaries and other clerical jobs generally require mid-level skills and a higher education compared to manual jobs. As such, the potential for these jobs to replacement by software programmes cannot be conceptualised by the skill-biased approach (Acemoglu & Autor 2010).

The task-based approach can be utilised to explain a more recent trend in the wage distributions of developed economies: that of polarisation. Autor, Katz and Kearney (2006) observe that the wage distribution in the US has become more polarised since the 1990s. Employment has increased for both the highest earners and lowest earners relative to mid-level earners. Insofar as this steepening U-shaped distribution is the result of technological change, the task-based model provides an effective explanation for this polarisation. The increase in employment at the bottom of the wage distribution relative to the middle is because nonroutine manual jobs cannot be replaced easily by new technologies, while clerical jobs in the middle of the distribution are being replaced (Autor, Katz & Kearney 2006). This pattern of 'routinisation', which sees the replacement of mid-education and mid-income level jobs, has affected a large number of developed economies coming into the Twenty-First Century (Das & Hilgenstock 2018).

2.1.3 The Revised Task-Based Approach

Despite the strengths of the task-based approach, developments in machine learning (ML) have necessitated a further revision of the model. ML systems are computational programmes which utilise algorithmic models together with a large amount of data. While non-ML programmes are coded with a clearly defined set of rules which are executed statically, ML programmes are coded with algorithmic functions involving X inputs and

Y outputs and then 'trained'. This involves providing the system with data which include the X and Y values. The system is then able to identify patterns in the data between X and Y and revise its execution of the Y outputs dynamically based on these newly learned patterns. The more data an ML system is given, the more it is able to respond to different and nonroutine events (Brynjolfsson & McAfee 2011).

While there are still limits on what these new systems are able to do, ML has necessitated the re-conceptualisation of what the task-based model characterises as 'routine' tasks. Occupational tasks such as translating a document, recognising handwriting and possibly even driving a car are becoming more codifiable through machine learning systems. Such a re-conceptualisation of the task-based model has been attempted in recent studies (Frey & Osborne 2017; Brynjolfsson, Mitchell & Rock 2018). Instead of grouping occupational tasks into the categories or routine/nonroutine and manual/cognitive, the revised task-based model scores tasks on a continuous scale that is designed to capture the likelihood that it can be performed by a machine.

One such example of the revised task-based approach is the 'suitability for machine learning' (SML) framework developed by Brynjolfsson, Mitchell and Rock (2018). The SML framework includes 21 criteria in the form of questions that seek to capture how suitable a task is for a machine learning system to perform on a scale from 1 (not suitable) to 5 (very suitable). An example of the 21 criteria is 'Task is principally concerned with matching information to concepts, labels, predictions, or actions'. For this criterion, the task is given a score between 1 which means 'The task does not have clear, consistent inputs or outputs'; and 5 meaning 'The task has clear inputs and outputs. The purpose of the task is to determine how the inputs affect the outputs (e.g. translating one language to another, matching an image to words describing the image)'[1]. These criteria are applied to a list of direct work actvities (DWAs) as defined by the US Occupational Information Network (O*NET). The scores for the DWAs are then aggregated to the

[1]For the rubric with all 21 criteria and the full construction process, see Brynjolfsson, Mitchell and Rock (2018) replication data at https://www.openicpsr.org/openicpsr/project/114436/version/V1/view

occupational task level and then to the occupation level for a score ranging from 1 to 5. The aggregation process used by the authors will be expanded on in the Methodology chapter of this study.

The SML measure is advantageous in that, in addition to taking into account current developments in machine learning, it is future-orientated. It does not assess occupations based on whether there are technologies currently available that can perform its tasks. Rather, the measure's detailed criteria assess atomised features of the occupational tasks that can potentially be performed by an ML system. In this way, it caters for possible future developments in ML beyond what has been developed today.

In applying the SML framework to the US labour market, Brynjolfsson, Mitchell and Rock (2018) find that most occupations include tasks with both high and low SML scores. Thus, the authors' SML measure ranges from 2.78 to 3.90 with a mean score of 3.47. The occupations with the lowest SML scores include occupations from a range of sectors such as massage therapists, animal scientists and plasterers; while the occupations with the highest suitability for replacement by machine learning systems include concierges, mechanical drafters and brokerage clerks (Bynjolfsson, Mitchell & Rock 2018). By quantifying how suitable occupations are to replacement by technological systems (both now and potentially in the future), the SML measure caters for a deeper level of analysis into the vulnerability of a labour force to the threat of technological displacement. The analysis of this study offers one such example of this by applying the suitability for machine learning framework to the context of the South African labour market.

2.2 Technological displacement and the Broader Labour Market

2.2.1 Empirical Findings of Technological Change and Routinisation

Technological change and the replacement of labour by machines and software systems is one possible driver of unemployment in an economy. The theoretical approaches outlined above offer ways to measure and explain this source of unemployment. However, it is important to note that technological change is not the only source of unemployment and displacement, nor is it an inevitable force in an economy. Empirically, it is observed to differing degrees in different contexts and its prevalence is a function of many other socioeconomic factors including the relative price of labour, the level of public and private investment in a given sector, and state policy that may or may not facilitate technological change. Indeed, in many cases the objective of applying the approaches outlined above is to identify areas of vulnerability in an economy and inform policy that can limit the negative effects of technological change.

In practice, whether firms or industries choose to adopt labour-replacing technology depends on the elasticity and relative price of labour. In analysing cross-national exposures to 'routinisation' (the automation of routine occupational tasks), Das and Hilgenstock (2018:33) find significant differences between developed and developing economies. They find that instances of exposure to routinisation in developing countries is far lower compared to developed economies and argue that this is partially explained by the lower price of labour in developing economies relative to capital. It is also associated with the level of investment in an economy which, for developing countries, has been relatively low. The authors also find that, in developing countries, if the initial exposure to routinisation is relatively low, subsequent onsets of routinisation will be more radical (Das & Hilgenstock 2018).

In looking at exposure to routinisation in South Africa, Davies and van Seventer (2020) find that moderate routinisation has occurred since the turn of the Century; however, there is a significant lag with the adoption of new technologies compared to developed economies. As Das and Hilgenstock (2018) suggest, the moderate exposure to routinisation suggests that subsequent exposures will be more pronounced in the future. Furthermore, the lag that Davies and Seventer (2020) observe suggests that we can look at the adoption of technologies in developed economies to predict which technologies South African firms adopt in the future. Employing a task-based approach to investigate the effects of technological change on wages in the South African labour market, Bhorat, Goga and Stanwix (2013) find that there has been a decrease in wage levels for occupations that involve tasks which are routine and do not require face-to-face interaction. They show that the introduction of new technologies that allow for automation as well as those that allow tasks to become "offshorable" (Bhorat, Goga & Stanwix 2013:21) have indeed had an effect on wages and employment in South Africa.

2.2.2 Unemployment and Inequality in the South African Labour Market

While the analysis of this study focuses exclusively on technological employment in South Africa, it is important to consider the broader socioeconomic factors that affect unemployment and displacement in the context of the South African labour market. Davies and van Seventer (2020) argue that South Africa's unemployment is primarily structural and is caused by a shortage of skilled labour. They identify three factors which determine the demand for skills in an economy. Firstly, growth in total output generally increases the demand for all skills overall. Secondly, structural changes to the sectoral composition of the economy will have an equivalent impact on the demand for skills. Thirdly, and relevant to this study, new technologies are a driver of changes to the demand for skills in an economy (Davies & van Seventer 2020). As South Africa moves towards a move services-orientated economy with a growing finance sector and away from reliance on the mining and agricultural sectors, the demand for skills required to perform more cognitive occupational tasks increases; particularly in administrative and clerical occupations.

The shortage of skilled labour in the South African context is itself a result of various structural inequalities and instances of marginalisation that exist in the country. These inequalities have racialised and gendered dimensions; with the African and female populations, respectively, facing instances of systemic marginalisation in the labour market.

Apartheid explicitly hindered access to employment opportunities, education and skills development for the majority African population. Crankshaw (1997) shows how, during apartheid, state policy facilitated the limited movement of African workers into semi-professional and routine white-collar occupations; but effectively excluded this demographic group from professional- and managerial-type jobs. Within the semi-professional occupational grouping, Crankshaw (1997) shows that there were also substantial differences between the races in terms of educational qualifications and, as a result, salaries. In 1982, 85 percent of African schoolteachers (the most common occupation for African semi-professionals) did not hold a post-Matric qualification; compared to 3 percent of white schoolteachers (Crankshaw 1997:24). While it was already state policy to pay African workers less than their white counterparts, teachers were also paid according to their educational level. This figure is indicative of structural patterns of racial inequality in terms of employment, income and quality of education that are still apparent in the post-apartheid era.

Apartheid's effects on structural inequality also included a gendered component. Policies were patriarchal and catered for the white, male 'breadwinner' at the expense of women who were expected to be homemakers (Mosomi 2019). While African men were required as cheap labour in the growing mining and manufacturing industries, African women were restricted from urban areas except as domestic workers and cleaners. Although upward mobility in employment has occurred in the post-apartheid era, these relative inequalities are still entrenched in the labour market. Women are less likely than men to find employment, and employed women are less likely to be found in occupations associated with higher wages (Espi, Francis & Valodia). These gender-based inequalities are compounded for the female youth population. Young females, and particularly young, black females, are more likely than their male counterparts to be unemployed and more likely to experience labour "churn" (Ingle & Mlatsheni 2017:3). Churn describes the constant movement of a worker between different occupations and different states of employment. This volatility of young, black females in the labour market represents a further instance of structural inequality.

Francis and Valodia (2021) find evidence of a gender pay gap (women are more likely to be paid less than men for the same work) and instances of occupational segregation. Occupational segregation describes the concentration of women in precarious and low-paid jobs (Francis & Valodia 2021) and the under-representation of women, and specifically black women, in high-skilled and managerial occupations (Espi, Francis & Valodia 2019). Related to this, Budlender (2019) explores the gendered nature of different mid-skill level occupations in South Africa. Rather than comparing the gender differences in occupations at the top and bottom of the skills "occupational hierarchy" (Budlender 2019:64) as other studies do, the author is concerned with differences in service-type occupations which fall in the middle of this hierarchy. Despite males and females making up a similar proportion of this group and having similar educational profiles, women are paid 78 percent of the mean male wage in service-type occupations. Budlender (2019) argues that this is because the specific occupations that women perform in this overarching category are predominantly care-service orientated which are under-valued; as opposed to the predominantly male occupations of prison guards and police officers.

However, there have been significant improvements in the situation of female employees in the post-apartheid era. Mosomi (2019) observes a decrease in the gender wage gap from 40 percent in 1993 to 16 percent in 2014. While still relatively slow and lagging behind that of men, women are increasingly gaining more access to education and more stable forms of employment (Posel & Casal 2019). Legislative action introduced by the post-apartheid government to rectify gender- and race-based inequalities in the labour market have been significant in this regard. The introduction of minimum wages in various sectors has been successful in providing a more stable income for women in precarious and low-skilled jobs (Mosomi 2019). The 1998 Employment Equity Act (EEA) has been the most noteworthy piece of legislation that has sought to address labour market disparities. The Act requires firms who employ more than 50 people to report their statistics on the representation and remuneration of their employees by gender and race group, and stipulates various criteria that firms have to meet in order to ensure gender and race-based equity in employment (Espi, Francis & Valodia 2019). Through analysis of the data obtained through EEA reporting, Francis and Valodia (2021) find that the Act has facilitated substantial increases in black workers into skilled and managerial positions, although this group is still proportionally under-represented. The authors also observe the movement of female workers into more professional-type occupations as a result of affirmitive action policy (Francis & Valodia 2021).

2.3 Summary

The reviewed studies show that the South African labour market is highly racialised and gendered. From an employment perspective, there are significant inequalities between the races and genders in terms of access to employment. In terms of occupational segregation, there are disparities with regards to the concentration of each gender in certain types of occupations. Males are found to be over-represented in high-skilled and managerial-type occupations; while females are over-represented in low-sklled and care-based service-type occupations. As a result of these differences but also due to gender discrimination, the gender wage gap has also been found to be significant in the country.

Increased access to education as well as legislative policies, most notably the Employment Equity Act (1998), have been significant in addressing various aspects of the structural inequalities. More women are moving towards managerial and professional positions and various sectoral minimum wages have created more stable employment for workers in more precarious forms of employment. As Mosomi (2019) shows, the gender wage gap has also decreased substantially in the post-apartheid period.

In tracing inequalities in the labour market, the studies reviewed here share common trends in how occupations and occupation types are defined. Occupations are either characterised according to the skill-level required (in line with the skill-biased approach), or defined according to the associated level of expected remuneration, with relationships drawn between these two categorisations. Relatively little has been found that shows how the instances of inequality manifest with regard to vulnerability to technological change and displacement. Similarly, the studies by Bhorat, Goga and Stanwix (2013) and Davies and van Seventer (2020) which have applied the task-based approach to the South African context and found evidence of vulnerability to technological change and routinisation do not show how this vulnerability is distributed among the demographic populations of the South African labour force. Applying the Brynjolfsson, Mitchell and Rock (2018) SML measure, this study seeks to investigate the association between the various structural inequalities found in the South African labour market and vulnerability to technological displacement.

Chapter 3

Research Methodology

3.1 Research design

The central objective of this research project is to conceptualise the South African labour force through the lens of the 'suitability for machine learning' (SML) framework. By matching the Brynjolfsson, Mitchell and Rock (2018) SML measure to occupations found in the South African labour market (and to the workers who perform these occupations), it is possible to identify segments of the workforce that are most vulnerable to technological displacement by machine learning systems in the Twenty-First Century. A higher SML score indicates higher vulnerability; while a lower SML score indicates occupations that cannot easily be performed by an ML system.

In investigating this, this study employs relative distribution methods as developed by Handcock and Morris (1999). Primarily, these methods are concerned with how the distribution of a specified variable (Y) differs between two defined populations, labelled the reference and comparison populations. The study also utilises quantile regression models to infer statistically significant relationships between the SML measure and demographic and educational characteristics.

The research design for this project consists of both a cross-sectional and time-series component. With the SML measure as the variable of interest, Y, the cross-sectional analysis will identify relative differences in the SML distributions of specified demographic segments of the 2019 South African labour force. The time-series component will then define the reference population as the 2009 labour force and the comparison population as the 2019 labour force and analyse significant changes that have occurred within the distributions of different demographic segments in this ten year period.

The research design for this project is largely observational and demonstrative, rather than explanatory. The goal is not to find which groups of workers have been or will be subject to technological displacement; but rather to show how the SML framework can be used to identify segments of the workforce that are *potentially* more or less vulnerable to the threat of technological displacement in the South African labour market. In addition to this, I seek to illustrate how relative distribution methods offer an effective yet underutilised tool of analysis for dissecting differences and trends within the distribution of a variable.

3.2 Data Preparation

The study utilises two separate datasets: the Brynjolfsson, Mitchell and Rock (2018) SML dataset and the South African Quarterly Labour Force Surveys from the second quarters of 2009 and 2019.

3.2.1 Suitability for Machine Learning Dataset

The Brynjolfsson, Mitchell and Rock (2018) SML framework and its associated dataset are made available for public access by the authors. The framework consists of a rubric of 21 criteria that aim to assess how easily a direct work activity (DWA) can be performed by a machine instead of by a human. The authors sourced the DWAs and the associated occupational data from the US-based Occupational Information Network (O*NET). For each criterion, a DWA is given a score ranging from 1 (cannot easily be performed by an ML system) to 5 (can easily be performed by an ML system).

To collect these scores, the authors used the online crowdsourcing tool, *Crowdflower*, to survey 14,783 respondents. Each respondent was given a DWA and asked to score it on a scale of 1 to 5 for each criterion. This respondent-by-criterion matrix as well as the instructions for the aggregation process were made available by the authors. The 2,069 DWAs are matched to 18,396 occupational tasks as defined by O*NET which are then matched to 900 O*NET occupations. Before aggregation to the occupation-level, SML task-level scores are also weighted according to the O*NET 'task importance' index. The weighted occupation-level SML scores are utilised as the central variable of interest for this project.

3.2.2 Quarterly Labour Force Survey

The Quarterly Labour Force Survey (QLFS) has been conducted by StatsSA every quarter since 2008. It contains individual-level data pertaining to respondents' employment histories and current occupations, educational attainment, demographic and geographical indicators, and various job- and industry-related characteristics. This study utilises the surveys from the second quarters of 2009 and 2019. The QLFS surveys were sourced from the Post-Apartheid Labour Market Series, PALMS (Kerr, Lam & Wittenberg 2018). The PALMS dataset offers processed versions of the QLFS variables that are more usable for this study. It also offers a corrected sample weight that is used in this study.

The O*NET occupation codes from the SML dataset are matched to the 4-digit South African Standard Classification of Occupations (SASCO) codes to assign each worker in the QLFS an SML score. In cases where multiple O*NET codes are matched to a single SASCO occupation, the SML scores of the O*NET occupations were averaged to produce a single score. The conversion process was not flawless and was subject to various limitations as detailed in Chapter One of this report. However, manual inspection of the 368 SASCO codes which were successfully assigned SML scores showed satisfactory results.

3.3 Variables of Interest

3.3.1 SML Measure

As stated above, the key variable of interest in this project is the SML score as conceptualised by Brynjolfsson, Mitchell and Rock (2018) and applied to the South African labour market. This is a interval, continuous measure. In theory, it can range from 1 to 5; however, in the context of the South African labour market, it ranges from 2.406 to 3.261[1]. The distribution and statistical features of the SML variable are discussed below.

3.3.2 Independent Indicators and Variables

The demographic variables used in this study include respondents' race, age and gender. The race (population group) variable in the South African QLFS is categorical with four levels: African/Black, Coloured, Indian/Asian, and White. To facilitate analysis through relative distribution methods, this variable is recoded as a binary variable with the African/Black race as one category and all other races pooled into a second category. This distinction is justified as it reflects the structural inequalities and divisions in the South African context. Due to the legacies of apartheid which marginalised the African population as well as subsequent economic policies that have sought to rectify these inequalities, it is necessary to compare this race group to other race groups when analysing instances of inequality and vulnerability in the labour market.

The age variable is also recoded into a binary indicator with South Africa's youth popu-

[1] This range is similar to the range found in the Brynjolfsson, Mitchell and Rock (2018) study of the US labour market and is consistent with their finding that most occupations include both high- and low-scoring DWAs which are then aggregated to the occupational level.

lation, defined as workers aged 15 to 34, compared to older South African workers. This categorisation is in line with the operational definitions of Statistics South Africa (2021) as well as other studies in the literature (Ingle & Mlatsheni 2017; Rankin & Roberts 2011). The gender variable is used as a binary indicator with 'male' and 'female' as the two categories.

The study also utilises the respondents' total years of education as a measure of educational attainment and skill-level. This is an interval-level variable that ranges from 0 to 17 and corresponds to the number of years of education that the respondent has completed.

The 4-digit SASCO occupation codes, which specify individual occupations and are matched to SML scores, can also be aggregated to the 1-digit level to form occupational types. These types include managers, professionals, technicians, clerical workers, services and sales workers, skilled agricultural workers, craft workers, machine operators and workers in elementary occupations. As shown in Chapter 4, there is variation in the SML distributions of these occupation types. Thus, the occupation type variable is also used as an explanatory variable.

3.4 Methods

With the final dataset constructed and the variables specified, this study employs relative distribution methods as conceptualised by Handcock and Morris (1999). As applied to this project, relative distribution methods include the formulation of relative density functions which are constructed using the *reldist* (Handcock & Aldrich 2002) software package in R. Similarly taking into account the differences in the SML distributions in defined populations, quantile regression models are also utilised.

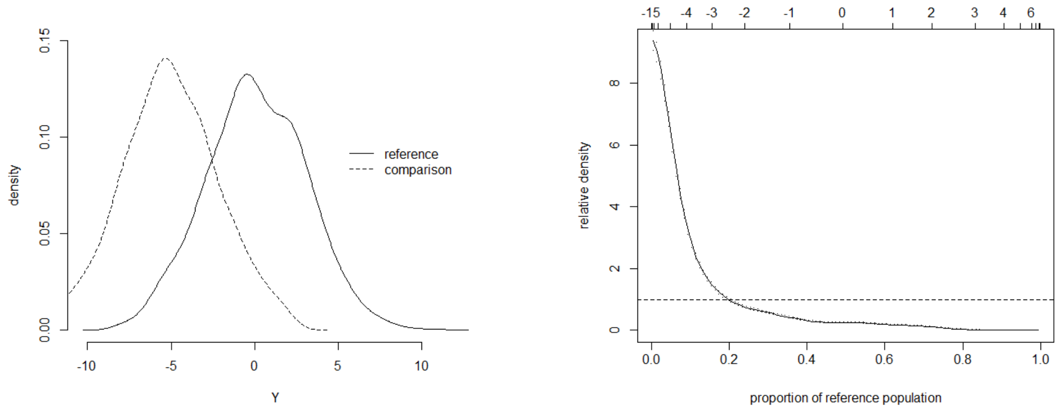
3.4.1 The Relative Density Function

The use of relative distribution methods first requires the definition of two populations, the reference and comparison populations. An instance of the central variable of interest found in the reference population is denoted by Y_0 and in the comparison population as Y. The analysis is, thus, concerned with characteristics of the distribution of Y relative to Y_0. The distribution of Y_0 in the reference population is given by the probability density function, or density function, $f_0(y)$; and in the comparison population as $f(y)$. Following the methodology outlined by Handcock and Morris (1999), the relative density function for the two populations is given by

$$g(r) = \frac{f(y_r)}{f_0(y_r)} y_r \geq 0$$

where y_r is the value of the variable of interest in the r^{th} quantile of the distribution for the reference population. The relative density function is, thus, the density ratio of the distribution of Y in the comparison population to the distribution of Y_0 in terms of the quantiles, r, of the reference population.

The relative density function has the same properties as a standard density function and can be visualised graphically with relative density on the y-axis and quantiles, r, on the x-axis. Where $g(r)$ is greater than 1, the probability of a person being selected in that quantile in the comparison population is higher than the probability of a worker being selected in the reference population. For example, if $g(0.2) = 2$, a worker is twice as likely to be found in the comparison population than in the reference population for values of Y equal to the 20^{th} percentile of the reference population's distribution. For relative density values between 0 and 1, a worker is more likely to be selected from the reference population relative to the comparison population. Figure 3.1 shows a hypothetical example of the density functions of two simulated populations and their associated relative density function.



(a) Overlaid Density Functions (b) Relative Density Function

Figure 3.1: Simulated density functions and relative density function for hypothetical reference and comparison populations

3.4.2 Quantile Regression

With the SML measure as the specified response variable, quantile regression models will be constructed to determine the demographic and educational characteristics of different sections of the SML distribution, defined as quantiles. While OLS regression methods estimate the mean value of the response variable given the covariates, quantile regression estimates the value of the response variable at the n^{th} quantile of the distribution. For example, specifying the quantile $q = 0.5$ would give the median value of the response variable given the covariates. The quantile regression models employed in this study utilise the deciles of the SML distribution ($q \in 0.1, 0.2, ..., 0.9$). The use of quantile regression is advantageous as it allows for the analysis of the demographic and educational characteristics of the SML measure at difference points along its distribution (which indicate increasing levels of vulnerability to technological displacement). The models also account for the presence of outliers in the SML variable.

3.5 SML Overview

Before examining the SML distribution at the worker-level, it is useful to inspect key occupations in the South African labour market and their associated SML scores. Table 3.1 illustrates the departure of the SML framework from more traditional approaches of understanding the effects of technological displacement. The occupations with the lowest SML scores consist of an interesting mix of industries and sectors. Agronomists and civil engineers are considered high-skilled occupations and, thus, conform to the expectations of the skill-biased theory in that they are least vulnerable to replacement. However, the presence of 'low-skilled' sweepers and cleaners on this list show the specificity of the SML framework.

At the other end of the spectrum, medium-skill, cognitive (specifically clerical) occupations scored highly in terms of the SML framework. This is to be expected as these jobs involve tasks which are routine and non-physical, and where data can be provided to teach ML systems how to perform these tasks. The notable exception to this is the 'undertakers and embalmers' occupation. This presents an anomaly as it is not intuitively apparent that this job should be suitable for machine learning. However, this high score is consistent with the Brynjolfsson, Mitchell and Rock (2018) data. Closer inspection of the SML criteria reveals that this occupation was judged to include tasks which are routine and work with a large amount of data which resulted in a higher than average occupational-level score. While this justification is not wholly satisfactory, the observation has been kept in this study.

Applying the SML measure to the 2009 and 2019 QLFS datasets, we can examine the variable at the worker-level. Table 3.2 shows the summary statistics for the SML variable both for the QLFS samples and for the South African population by applying the PALMS survey weights to these statistics. By applying these weights, the 2009 sample of 21,865 observations can be said to represent almost 14 million workers; while the 2019 sample of 16,824 represents nearly 16 million workers. The mean SML score in the South African

Table 3.1: Lowest and Highest SML Occupations in South African Labour Market

	Low SML Occupations	SML	High SML Occupations	SML
1	Agronomists	2.406	Tellers and counter clerks	3.261
2	Librarians	2.413	Data entry operators	3.228
3	Building structure cleaners	2.452	Bookkeepers	3.213
4	Concrete workers	2.452	Undertakers and embalmers	3.201
5	Sweepers	2.462	Travel agency clerks	3.191
6	Civil engineers	2.464	Cashiers	3.190

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

context is approximately equal to 2.77, with a standard deviation of 0.167. Figure 3.2 shows the SML distributions of the 2009 and 2019 labour forces. The abnormal shape has two peaks on either side of the mean as well as a peak on the right-end tail. The irregular distribution of the SML measure suggests that there are distinct characteristics at different levels of the variable which should be investigated in isolation. This makes relative distribution methods and quantile regression (which allow for the dissection of a distribution) an effective approach for analysis.

Table 3.2: Sample and weighted summary statistics for SML measure (2009 and 2019)

Sample	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
2009	21,865	2.767	0.167	2.406	2.647	2.914	3.261
2019	16,824	2.769	0.169	2.406	2.647	2.914	3.261
Weighted	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
2009	13,898,432	2.767	0.167	2.406	2.647	2.905	3.261
2019	15,903,119	2.771	0.167	2.406	2.647	2.914	3.261

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

Table 3.3 shows the most common occupations in the South African labour market and their associated SML scores. They include mainly manual, elementary-type occupations which are below the mean SML score. Protective service workers (security guards) and shop salespersons have the highest SML scores of 2.95 and 2.93, respectively. Domestic workers and famhands constitute the occupations with the highest number of workers for both 2009 and 2019.

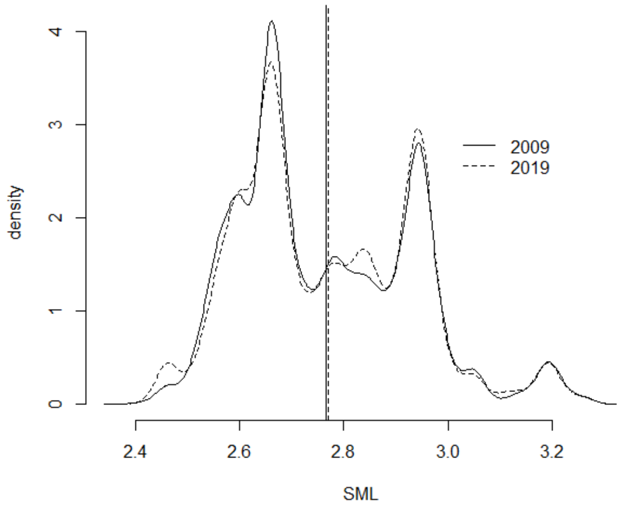


Figure 3.2: SML distribution of 2009 and 2019 labour forces

Table 3.3: Common Occupations in South African Labour Market

2009			2019		
occupation	SML	workers	occupation	SML	workers
Domestic workers	2.658	1,106,885	Farmhands	2.602	1,005,787
Farmhands	2.602	804,632	Domestic workers	2.658	1,004,048
Office clerks	2.669	490,274	Protective services workers	2.948	643,190
Office cleaners	2.560	480,461	Shop salespersons	2.932	577,899
Hand-packers	2.647	443,419	Office clerks	2.669	563,098
Shop salespersons	2.932	430,340	Office cleaners	2.560	551,109

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

Chapter 4

Results

4.1 Introduction

Having applied the SML measure to the South African labour market and investigated the demographic characteristics of this variable, this chapter presents the results of this analysis. Through relative distribution methods and quantile regression models, evidence is found to suggest that South Africa's female workers are significantly more vulnerable to the threat of technological displacement by machine learning systems than their male counterparts. Furthermore, in the ten year period from 2009 to 2019, the vulnerability of young, black females in particular has increased despite the female population as a whole becoming *less* vulnerable.

Although this is in line with our understandings of broader socioeconomic inequalities and marginalisation in South Africa, these findings should not be treated as a given. While young, black female workers have been found to have lower access to employment and the gender wage gap in the country is significant (Casale & Posel 2011; Mosomi 2019), these aspects of inequality are distinct from what the SML variable is measuring. High SML vulnerability indicates that, should advanced forms of labour-substituting technology ever be brought into the South African economy in the future, this demographic group is most at risk of replacement and displacement.

With a focus on clerical occupations, these findings are explained by the types of occupations that female workers are selected into as a result of increases in education and policies that incentivise firms to employ young, African women in mid-skill level jobs.

The analysis of this chapter is divided into a cross-sectional component and a time-series component. In the cross-sectional analysis, I first provide an overview of the 2019 South African labour market in terms of the SML measure. Applying relative distribution methods and quantile regression, I then decompose the SML distributions of the different demographic populations and show evidence that the female population is more vulnerable than the male population. I show the SML distributions for different occupation types and the demographic and educational characteristics of the occupation types. I also find a positive relationship between the SML measure and education, contrary to the expectations of the skill-biased approach.

The time-series component then shows how the SML distributions for different demographic populations have changed between 2009 and 2019, focusing on the female population. I find significant evidence that the female population as a whole has become less vulnerable to technological displacement between 2009 and 2019; but young, black females have become more vulnerable in terms of the SML measure. Again, I explain this finding with reference to the movement of the young, black female population into clerical-type occupations.

4.2 Cross-Sectional Analysis: 2019 Labour Market

4.2.1 Overview

Table 4.1 shows the demographic composition of the 2019 South African labour force and descriptive statistics for key variables of interest that will be utilised in this study. Here, the labour force is divided into eight demographic segments according to gender (male and female), race (black and other races) and age group (youth and older). All statistics shown have been weighted using the PALMS survey weight.

In terms of the SML measure, older males from other race groups have the lowest mean SML score with 2.75, indicating lower average vulnerability to replacement by machine learning systems compared to other demographic groups. Conversely, young, black females are the most vulnerable segment with an average SML score of 2.81. Indeed, viewed through the lens of the SML framework, the inequality between the two genders is immediately apparent. On average, females have a higher SML score compared to males. More than that, the variability of the average SML scores for the four female demographic segments is greater than that of the male segments, suggesting that a female's SML score is more dependent on the other demographic indicators. For all four race and gender groups, the younger population is more vulnerable to replacement by machine learning systems compared to the older population.

Table 4.1: Demographic overview of 2019 labour market (N = 15,903,120)

Gender	Race	Age	SML(mean)	N(%)	educ.(mean years)	perm. contract(%)
Male	other races	> 35	2.75	9.42	12.12	63.52
		15-34	2.76	5.65	11.92	64.91
		total	2.75	15.07	12.04	64.04
	black	> 35	2.76	22.22	10.28	51.86
		15-34	2.76	19.07	10.88	40.59
		total	2.76	41.3	10.56	46.66
	total		2.76	56.37	10.96	51.3
Female	other races	> 35	2.79	6.7	12.29	71.89
		15-34	2.8	4.2	12.36	69.15
		total	2.79	10.9	12.32	70.83
	black	> 35	2.76	19.24	10.32	48.59
		15-34	2.81	13.5	11.6	43.4
		total	2.78	32.73	10.85	46.45
	total		2.79	43.63	11.22	52.54

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

According to the data, males make up 56.37% of the labour force, with older black males being the largest of the eight segments (22.22%). The smallest segment is young females from other races (4.2%). Interestingly, this segment is also the most educated with an average of 12.36 years of education. As will be explored in subsequent sections of this study, there appears to be a negative relationship between the SML measure and education.

Also included in Table 4.1 is the percentage of workers in each segment who are employed with a permanent contract. This potentially captures the stability and security of employment. As shown, a smaller proportion of black workers are employed with a permanent contract compared to the other races with only older black men having a proportion of greater than 50%. Besides for this finding, there is more variability with regards to differences among the other demographic groups. The segment with the highest percentage of workers employed under a permanent contract are females from other races at 70.83%; while only 40.59% of young, black males have permanent contracts.

4.2.2 Demographic Analysis: Quantile Regression and Relative Distributions

Elaborating on these observations, Table 4.2 shows the results of OLS and quantile regression models with the SML measure as the dependent variable. The SML variable is regressed on the demographic and educational indicators discussed above. Column 1 shows the results of the OLS model. The Constant term in the last row, 2.641, serves as the coefficient of the reference group and all other coefficients are relative to this reference group. Consistent with the descriptive statistics shown in Table 4.1, being ‘female’, ‘youth’ or ‘black’ is associated with a relatively higher SML value. Furthermore, an increase in years of education is associated with a higher SML score.

Columns 2 through 10 show the results of the quantile regression model. The utilisation of quantile regression is advantageous as it allows for the disaggregation of the SML

Table 4.2: OLS and quantile regression models for SML on demographic and education indicators

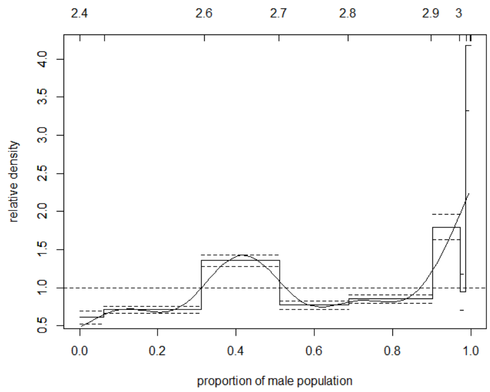
	Dependent variable:									
	SML	SML								
	OLS	*quantile regression*								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
female	0.023***	−0.00003	0.028***	0.008***	−0.008***	0.014***	0.033***	0.056***	0.023***	0.037***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.005)	(0.006)	(0.004)	(0.002)	(0.006)
youth	0.018***	−0.000	−0.006***	−0.007***	−0.005	0.014***	0.033***	0.023***	0.017***	0.017***
	(0.003)	(0.003)	(0.002)	(0.001)	(0.003)	(0.005)	(0.005)	(0.004)	(0.002)	(0.004)
black	0.016***	−0.022***	−0.009***	−0.008***	−0.002	0.031***	0.039***	0.038***	0.036***	0.012***
	(0.003)	(0.005)	(0.003)	(0.002)	(0.003)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)
educ.(years)	0.009***	0.003***	0.006***	0.006***	0.008***	0.013***	0.012***	0.006***	0.002***	0.003***
	(0.0004)	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)
Constant	2.641***	2.565***	2.567***	2.591***	2.604***	2.572***	2.619***	2.747***	2.863***	2.903***
	(0.006)	(0.006)	(0.003)	(0.003)	(0.004)	(0.009)	(0.013)	(0.010)	(0.007)	(0.006)
R^2	0.040									
Adjusted R^2	0.040									
Residual Std. Error	0.165 (df = 16607)									
Note:										*p<0.1; **p<0.05; ***p<0.01

distribution into different levels of vulnerability. Each column corresponds to a decile, q, of the SML distribution and the coefficients for each covariate show the SML score for that demographic group's SML distribution in the q^{th} decile. Across the SML distribution, an increase in years of education is associated with a higher SML score, consistent with the finding of the OLS model.

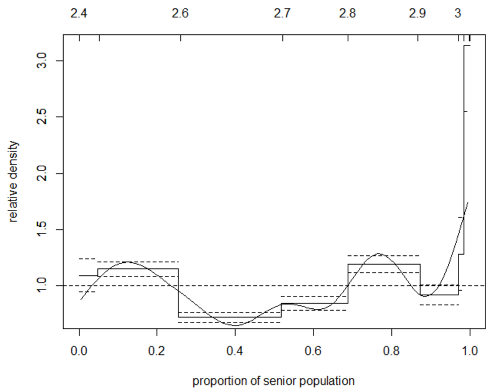
For the 'female' variable, the coefficients are positive and significant for a majority of the quantiles. This means that, at each given level of SML vulnerability, females workers are more vulnerable relative to their male counterparts. The exception to this is in the 4^{th} decile where the female coefficient is negative and significant. As will be explained in subsequent sections, this finding is the result of the fact that the 'domestic worker' occupation is gendered (a large proportion of the workers in this occupation are women) and is located in this quantile.

For both the 'youth' and 'black' variables, the quantile regression model yields interesting results. For the lower deciles, the coefficients are negative and significant, indicating that, for workers at lower levels of SML vulnerability, the youth and black populations have lower SML scores than their demographic counterparts. However, in the upper deciles the coefficients are significant and positive indicating higher vulnerability relative to workers from other races and older workers. This finding suggests that the SML distributions of these two populations are polarised relative to their demographic counterparts. Young and black workers are both the most and least vulnerable in terms of the SML framework, relative to other demographic groups.

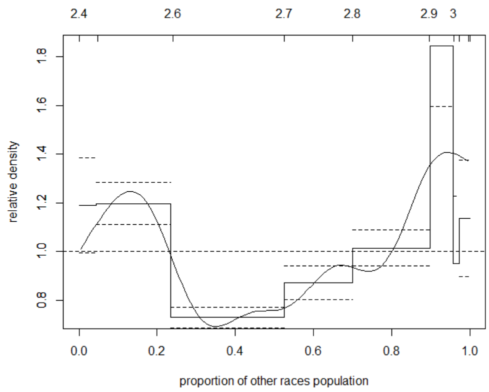
The relative density functions in Figure 4.1 provide further evidence for the key findings of the quantile regression model. The relative density function shows the ratio of the density function of a comparison population over the density function of a reference population in terms of the quantiles of the reference population. Panel A in 4.1 shows the relative density function for the two genders, defining 'male' as the reference population and 'female' as the comparison population. The bottom x-axis shows the quantiles (proportion) of the male distribution while the top x-axis shows the actual SML scores



(a) gender: female as comparison



(b) age: youth as comparison



(c) race: black/African as comparison

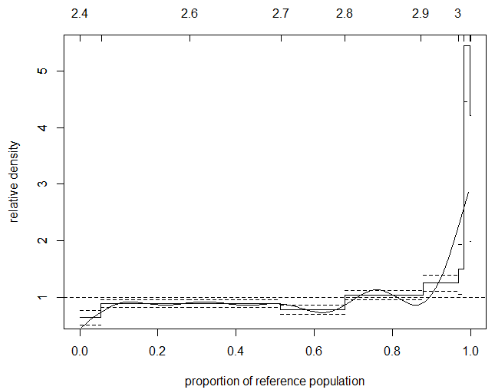
Figure 4.1: Relative density functions for gender, age and race indicators

that correspond to these quantiles. The y-axis shows the relative density which can be interpreted as a likelihood ratio. Where relative density equals 1, there is an equal probability that a worker in that quantile will be drawn from either the reference or comparison populations. Where relative density is greater than 1, there is a higher probability that a worker in that quantile will be from the comparison population; and where it is smaller than 1, the worker is more likely to be from the reference population.

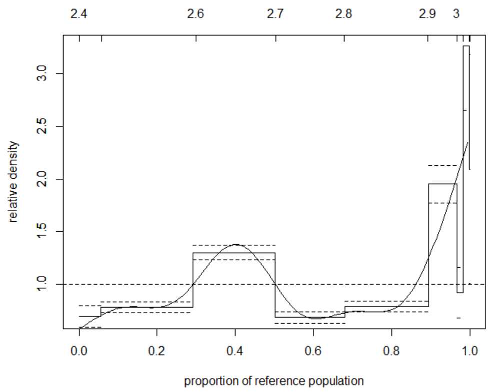
Reflecting the findings in the quantile regression, Panel A shows that at lower levels of SML vulnerability there is a higher probability that a selected worker will be male; while at higher levels, a female worker is more likely to be selected. The exception to this is in the 4th quantile. Again, this is due to the high influence of the domestic worker occupation which is strongly gendered. Panels B and C show the relative density functions for the 'youth' and 'black' indicators, respectively. For Panel B, the reference population is defined as workers over the age of 34 and the comparison population as youth workers. For Panel C, the reference population includes workers from other races and the comparison population is black workers. Both graphs show the polarisation of youth and black workers relative to their demographic counterparts.

However, when these two latter indicators are interacted with the gender variable, the polarisation disappears. The relative density functions in Figure 4.2 show three of these interactions. Here, the comparison populations are defined as: a) young females, b) black females, and c) young, black females; while the reference populations are defined as all other workers. In all three cases, the functions follow the same shape as Panel A of Figure 4.1, with the female population more likely to be found at higher levels of SML vulnerability. This suggests that the most significant determinant of SML vulnerability in the South African labour market is gender, with females being significantly more vulnerable to technological displacement than males.

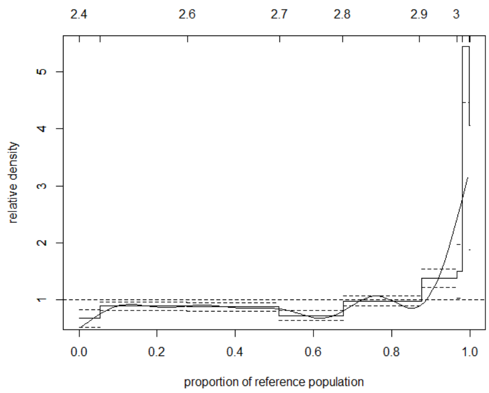
The importance of this finding should not be understated. As discussed in the contextual framework in Chapter Two, there are a myriad examples of gender-based inequality in the South African labour market, with women being the more marginalised population. The gender wage gap has been found to be significant and general access to employment has been found to favour men. Furthermore, these inequalities have been found to be accentuated with young and black workers. While the findings of this analysis are in line with these trends, they represent an additional manifestation of potential inequality: Women who are included in the labour force are significantly more likely to be employed in occupations which are more vulnerable to the threat of technological displacement. In attempting to explain this, the following section looks at the types of occupations women are performing and potential reasons as to why they are selected into high-SML scoring jobs.



(a) comparison: young females



(b) comparison: black females



(c) comparison: young, black females

Figure 4.2: Relative density functions for gender interacted with other demographic indicators

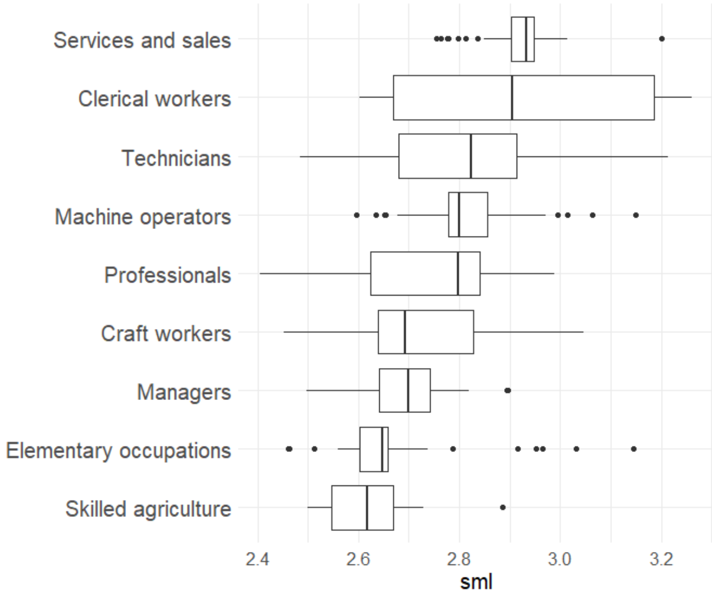


Figure 4.3: SML distributions of occupation types

4.2.3 Analysis of Occupation Type

The PALMS and QLFS datasets specify nine types of occupations into which the occupations are grouped. Ordered by median SML value, the boxplots in Figure 4.3 show the SML distributions for each occupation type. The services and sales occupational grouping has a relatively narrow distribution around the highest median SML score of 2.932. Occupations in the clerical workers grouping, on the other hand, cover a wider range of SML scores but with a similar median score of 2.914. At the other end of the spectrum, the skilled agriculture occupational grouping has a narrow distribution around a median SML score of 2.643. Elementary occupations have a similarly low median score of 2.647 but a wider range with a maximum SML score of 3.144 (corresponding to messengers, porters and deliverers). Managers, meanwhile, have a median SML score of 2.7 and a maximum score of only 2.897. Although the professionals grouping has a mid-level median SML score of 2.798 relative to the other groupings, it has the lowest minimum SML value of 2.406 (corresponding to the agronomists, food scientists and related professionals occupation).

Table 4.3 shows the employment share and average SML score by gender for each occupation type. Females comprise 71.8% of clerical workers which is also the occupation type with the highest average SML score for both males and females. Females also comprise a small majority of workers in elementary occupations (55.4%), technicians (52.8%), and professionals (52.2%). Within the technicians and professionals groups, females have a higher average SML score compared to their male counterparts. Females performing skilled agriculture-type occupations have the lowest average SML score with 2.556, which is 0.091 points lower than their male counterparts. However, females only comprise 14.5% of this occupation group.

Table 4.3: Occupation type employment share and average SML by gender

occ. type	male(%)	female (%)	male SML (avg)	female SML (avg)
Clerical workers	28.2	71.8	2.922	2.921
Services and sales	53.3	46.7	2.920	2.897
Craft workers	89.1	10.9	2.698	2.848
Technicians	47.2	52.8	2.767	2.834
Machine operators	87.6	12.4	2.817	2.832
Professionals	47.8	52.2	2.722	2.759
Managers	70.1	29.9	2.694	2.711
Elementary occupations	44.6	55.4	2.675	2.669
Skilled agriculture	85.5	14.5	2.647	2.556

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

The clerical workers occupation type is of particular interest because such a high proportion of women perform these jobs and it has the highest average SML score. While there are other possible factors, understanding the gender differences within this occupation type will play a part in explaining the gender inequality in SML vulnerability. Table 4.4 shows the most common occupations for males and females and their associated SML scores. For both genders, office clerks (excluding customer service clerks) are the most common clerical occupations. However, compared to the other clerical occupations, the SML score of this occupation is relatively low at 2.669. On the other hand, cashiers and ticket clerks have a very high SML score (3.19) with over 265,000 women performing this job (more than four times the number of male cashiers). Bookkeeping clerks and switchboard operators are also common occupations for female clerical workers and these also have high SML scores of 3.05 and 3.111, respectively.

Table 4.4: Most common clerical occupations by gender

Male			Female		
occupation	N	SML	occupation	N	SML
Office clerks (except customer services clerks)	133,028	2.669	Office clerks (except customer services clerks)	430,070	2.669
Stock clerks	94,898	2.914	Cashiers and ticket clerks	265,028	3.190
Cashiers and ticket clerks	64,398	3.190	Receptionists and information clerks	99,879	2.869
Mail carriers and sorting clerks	22,774	3.007	Accounting and bookkeeping clerks	82,577	3.050
Statistical finance clerks	22,194	2.905	Telephone switchboard operators	66,383	3.111

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

Interestingly, viewed through the lens of the first task-based approaches, these occupations would generally be considered routine-cognitive occupations. Thus, the fact that they are theorised to be highly vulnerable to replacement through the SML framework is in line with the expectations of the broader task-based approach. The routine nature of the tasks of cashiers and the calculations of bookkeepers can be automated fairly easily. In terms of skill-level, these are also occupations which are considered medium-skill. Table 4.5 shows the average years of education by gender for each occupation type. The average clerical worker has over 12 years of education (which is more than general average as shown in Table 4.1); but this is less than the averages for more high-skilled occupations such as managers and professionals. Additionally, a high percentage of clerical workers are employed with permanent contracts (77.3%). This makes clerical occupations an appealing form of stable employment for a demographic population that has faced historical inequality and marginalisation, but is becoming more educated and more active in the labour market.

4.3 Time Series Analysis: 2009 and 2019

The previous section showed the internal characteristics of the contemporary South African labour market as viewed through the lens of the 'suitability for machine learning' framework and provided evidence that the country's female population is significantly more vulnerable to technological displacement compared to the male population. It

Table 4.5: Educational and contract characteristics of occupation types

occ. type	mean edu (years)	perm. contract (%)
Managers	13.0	51.7
Professionals	15.3	77.6
Technicians	13.2	74.9
Clerical workers	12.3	77.3
Services and sales	11.1	54.1
Skilled agriculture	10.9	16.6
Craft workers	10.3	39.5
Machine operators	10.4	60.2
Elementary occupations	9.1	33.1

Source: StatsSA and author's calculations

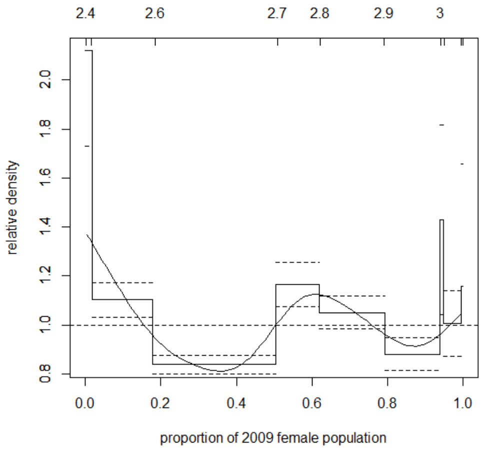
is necessary to place these findings in the context of how the South African economy has changed. It is important to analyse the trajectories of the most vulnerable demographic segments of the labour force, specifically within the female population, and assess whether these groups have moved towards occupations that have increased or decreased their SML scores. The following section compares the 2019 SML distribution to the 2009 SML distribution and analyses what changes have occurred in this ten year period.

4.3.1 Changes to SML Distribution

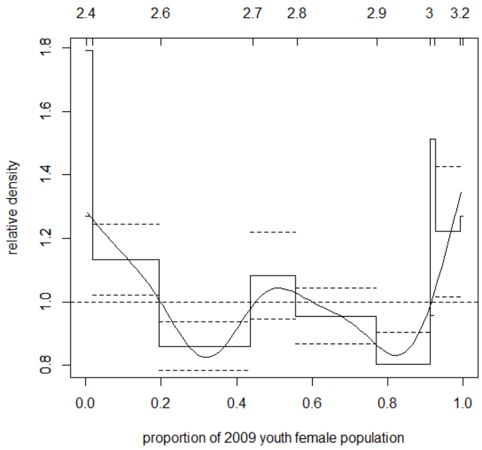
The overall SML distribution of the South African labour market has not experienced significant changes between 2009 and 2019. However, there have been notable changes to the distributions of individual demographic groups within the female population, specifically. Defining the reference population as the 2009 labour force and the comparison population as the 2019 labour force, Figure 4.4 shows the relative density functions for a) the entire female population, b) the female youth population and c) the black, female youth population. Notably, Panel A shows that the entire female population has become *less* vulnerable to technological displacement when comparing 2019 to 2009. At $q = 0.1$ where SML vulnerability is the lowest, there is a significantly higher likelihood that a 2019 female worker will be selected than a 2009 female worker. At the other tail of the distribution ($q = 0.9$), there is a lower likelihood that a 2019 female worker will be selected indicating that female workers in general have moved away from this level of SML vulnerability.

However, when interacted with the other demographic indicators, the changes to the female SML distribution tell a different story. Panel B shows the relative density function for the young female population (15 to 34 years of age). Here, at $q = 0.1$, there has been in increase in the proportion of young female workers in 2019 relative to 2009, indicating that this group has moved into this level of low vulnerability. However, they have also moved into the level of highest vulnerability in this ten year period. This suggests that the SML distribution of the 2019, young female population is polarised relative to the 2009 population. Panel C shows that, for young, black females, there has not been a significant change in the ten year period at the lowest level of SML vulnerability, but there has been a significant change at the highest level. Young, black females are 1.5 times more likely to be found at the highest level of SML vulnerability in 2019 than they were in 2009. This suggests that the increase in female workers at the lowest level of SML vulnerability mainly comprised female workers from other races and the improvements enjoyed by this group is counteracted by a worsened situation for young, black females.

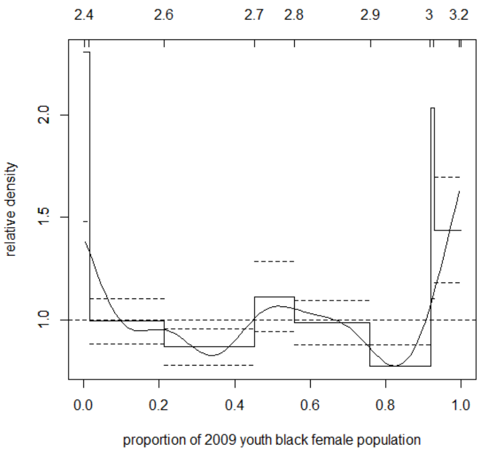
These findings are corroborated by the quantile regression output shown in Table 4.6. Here, the SML measure is regressed on the demographic indicators, a 2019 dummy variable and years of education. The three columns show the first, fifth and ninth deciles of the SML distribution. The Constant terms show the SML scores for older male workers from other races in 2009 in each of these deciles, with the other coefficients being deviations from this group. Confirming the results above, the coefficient for older females from other races in 2019 (female:2019) in the first decile is -0.061 with high significance. This is compared to a significant, positive coefficient of 0.095 for older females from other races in 2009. This means that, at the lowest level of SML vulnerability, older females from other races in 2019 have a SML score that is 0.156 points lower than in 2009. At the highest level of SML vulnerability, young, black female workers in 2019 have a mean SML score of 3.049 $(2.903 + 0.146)$. This is significantly higher than the demographic



(a) female populations



(b) female youth populations



(c) young, black female populations

Figure 4.4: Relative density functions for change in female populations (2009 to 2019)

group's mean score in 2009 of 2.924.

Table 4.6: Quantile regression with demographic and year indicators

	Dependent variable:		
	SML		
	0.1	0.5	0.9
	(1)	(2)	(3)
youth	0.001 (0.014)	0.025*** (0.009)	0.016** (0.007)
black	0.020** (0.009)	0.080*** (0.004)	0.025*** (0.002)
female	0.095*** (0.012)	0.095*** (0.008)	0.065*** (0.023)
2019	0.022** (0.010)	0.017*** (0.005)	0.003 (0.003)
education years	0.002*** (0.0002)	0.011*** (0.0003)	0.002*** (0.0002)
youth:black	−0.014 (0.015)	−0.039*** (0.011)	−0.014* (0.007)
youth:female	−0.016 (0.022)	0.0002 (0.016)	0.039 (0.039)
black:female	−0.089*** (0.013)	−0.135*** (0.010)	−0.043* (0.024)
youth:2019	0.002 (0.017)	−0.025* (0.014)	0.0002 (0.008)
black:2019	−0.021** (0.010)	−0.023*** (0.007)	−0.005 (0.004)
female:2019	−0.061*** (0.017)	−0.040*** (0.013)	−0.018 (0.029)
youth:black:female	0.042* (0.024)	0.071*** (0.020)	0.021 (0.043)
youth:black:2019	0.004 (0.018)	0.031* (0.017)	−0.004 (0.008)
youth:female:2019	0.008 (0.027)	0.010 (0.024)	0.003 (0.059)
black:female:2019	0.043** (0.018)	0.048*** (0.016)	0.014 (0.030)
youth:black:female:2019	−0.022 (0.030)	−0.014 (0.030)	0.146** (0.067)
Constant	2.531*** (0.009)	2.562*** (0.002)	2.903*** (0.002)
Observations	38,256	38,256	38,256
Note:			*p<0.1; **p<0.05; ***p<0.01

4.3.2 Trends in Employment Share by Occupation-Type

Section 4.2.3 illustrated the differences between the genders in terms of the types of occupations that they performed. It was shown that women comprise over 70% of clerical workers in South Africa- a grouping which includes routine-cognitive occupations with high SML scores. Table 4.7 expands on this by showing the changes in employment share for each occupation type. There have been increases in the proportions of the labour force performing services and sales, managerial and professional-type occupations; with services and sales workers now forming the second largest occupation grouping after elementary occupations. Technicians and craft workers have seen the biggest decreases in employment share. There has been a decrease in the proportion of the labour force performing clerical occupations between 2009 and 2019; however, the average SML score in this occupation type has increased from 2.906 to 2.922 in this ten year period.

Table 4.7: Change in employment share and average SML by occupation type

occ. type	2009(%)	2019(%)	perc. change	SML 2009	SML 2019	SML change
Services and sales	12.3	15.5	3.2	2.915	2.910	-0.005
Managers	7.6	9.7	2.1	2.694	2.699	0.005
Professionals	5.1	5.4	0.3	2.728	2.741	0.013
Clerical workers	11.2	11.0	-0.2	2.906	2.922	0.016
Skilled agriculture	0.5	0.3	-0.2	2.622	2.634	0.012
Machine operators	9.4	8.7	-0.7	2.817	2.819	0.002
Elementary occupations	29.9	29.1	-0.8	2.682	2.671	-0.011
Craft workers	13.9	12.4	-1.5	2.714	2.715	0.001
Technicians	10.2	7.9	-2.3	2.798	2.803	0.005

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

Table 4.8 looks at the proportions of females, specifically, in each occupation type. The second column shows the percentage of workers in each occupation type which were female in 2009, and column 3 shows by how much this percentage has increased or decreased between 2009 and 2019. The subsequent columns show how the female population in each occupation type is split among the four internal demographic groups (divided according to the race and age binary indicators). Between 2009 and 2019 in professional occupations, the proportion of females increased from 45.3% to now constituting a majority. Black seniors now make up the largest demographic group of professional female workers, growing from 21.4% in 2009 to 35.4% in 2019.

The decrease in the overall share of clerical workers observed in Table 4.7 has coincided with an increase in the proportion of women who perform these jobs. Within the female population who perform clerical occupations, black youths comprised the largest proportion in 2009 with 29.2% and this figure has grown to 36.6% in 2019. Of particular interest, the proportion of young, black, female bookkeepers has grown from 1% in 2009 to 24% in 2019. These observations appear to provide evidence for the hypothesis that, as they become more educated and more active in the labour market, young, black women are moving towards clerical occupations with high SML vulnerability.

Table 4.8: Proportion of females and female demographic groups by occupation type

			other races				black/African			
occ. type	2009	change	senior '09	change	youth '09	change	senior '09	change	youth '09	change
Professionals	45.3	6.9	35.3	-10.1	23.2	-2.9	21.1	14.3	20.3	-1.3
Clerical workers	69.9	1.9	26.9	-4.1	26.9	-13.2	16.9	10.0	29.2	7.4
Technicians	53.0	-0.2	29.4	-7.6	18.0	-5.5	34.5	4.1	18.1	9.1
Services and sales	47.3	-0.6	10.4	1.6	11.0	-2.7	30.9	9.7	47.7	-8.6
Managers	31.1	-1.2	41.2	-0.4	22.7	-3.8	22.4	3.0	13.6	1.3
Skilled agriculture	16.0	-1.5	7.5	21.1	14.5	-2.4	48.6	-16.9	29.5	-1.9
Craft workers	12.7	-1.8	7.4	-1.1	7.3	-1.6	51.6	0.5	33.8	2.1
Elementary occupations	57.2	-1.8	5.6	0.4	3.9	0.7	58.9	0.7	31.7	-2.0
Machine operators	15.0	-2.6	19.5	-5.0	10.4	-5.7	40.0	5.9	30.2	4.6

Source: StatsSA; Brynjolfsson, Mitchell and Rock (2018); and author's calculations

4.4 Discussion

The analysis of this chapter has provided evidence that, in the South African labour market, female workers in general and young, black female workers in particular are significantly more vulnerable to the threat of technological replacement by machine learning systems compared to other demographic groups. Through quantile regression analysis and by dividing the SML distribution according to levels of vulnerability, women are found to be more likely to perform occupations that can be replaced by machine learning systems. On average, female workers have a higher SML score than that of males in the same quantile of SML vulnerability. These mainly consist of clerical jobs.

In the time period between 2009 and 2019, we find evidence that, although females as a group have decreased their vulnerability to technological replacement; young, black females specifically have become significantly more vulnerable. This is despite an increase in education for this group. The average number of years of education has increased from 11 in 2009 to 11.6 in 2019, the proportion of young black females with a Matric has increased from 50.9% to 51.6%, and the proportion with a higher degree has increased from 3.62% to 7.35%. Indeed, the relationship between education and SML vulnerability is found to be complex. Contrary to the expectations of the skill-biased approach, a significant, positive relationship is found between years of education and SML vulnerability. This is because a large number of manual jobs which would typically be thought of as low-skilled (particularly in the Construction an Agricultural sectors) are associated with low SML scores.

Of course, we cannot conclude that an increase in education is the *cause* of high SML vulnerability or that this relationship is intrinsic and inevitable. But we can suggest that, firstly, an increase in years of education does not by itself guarantee resilience to technological displacement and, secondly, there are possibly other more important factors relating to the types of occupational tasks performed. This is to say that there are potentially other factors at play besides education that have seen the increase in SML vulnerability for young, black females.

In answering this, we must determine the push and pull factors that have led this demographic group to enter occupations with high SML vulnerability. Clerical occupations have potentially provided an attractive source of stable employment and income. They require higher education levels compared to elementary-type occupations and are more likely to offer permanent contracts. From the firm's perspective, there are also important incentives in hiring young, black women. The 1998 Employment Equity Act has required firms to explicitly address the structural inequalities in the labour market. As Espi, Francis and Valodia (2019) have shown, this has seen the increased representation of black males and white women in high-skilled managerial positions. It has also seen the increase of female representation of services occupations (Budlender 2019). As a result of this policy-driven upward mobility, it is likely that black females, specifically, are moving into more clerical and administrative mid-skill level occupations.

These trends of upward mobility are largely positive. There is still much to be done in increasing the proportion of black women in managerial and other high-skilled jobs; but the increases in stable forms of employment represents significant progress. Furthermore, it is important to note that the SML vulnerability identified in this project is still largely theoretical. Whether or not firms do in actuality choose to adopt more forms of labour-substituting capital is a function of several factors which are beyond the scope of this project. However, the finding that there is significant inequality between the two genders in terms of this theoretical vulnerability is noteworthy. It means that, should the adoption of these new technologies be observed, South Africa's women will be affected significantly more than men. Thus, this finding should be used to inform further research on the effects of technological displacement in South Africa.

Chapter 5

Conclusion and Future Work

In addition to the significant findings presented in the previous chapter, this study should also be seen as an attempt to lay the foundation for further investigation into technological displacement and labour force vulnerability in South Africa and other developing economies. The importance of this endeavour has been stated in this study and in the broader literature. The penetration of new technologies into developing economies such as South Africa has not been to the same extent that we have observed in more developed economies. This is partly due to the lower relative price of labour in developing economies compared to developed economies. However, there is evidence to suggest that there is a lag to this penetration and that more forms of labour-substituting capital will be introduced into the South African economy in the future. As technology develops at an increasingly rapid rate and as these technologies become more cost effective, firms become more incentivised to replace labour with capital. As such, we are now in a position to anticipate what the effects of this will be. The burgeoning literature on this topic in the context of developed economies provides us with the tools needed for this undertaking.

One such tool is the Brynjolfsson, Mitchell and Rock (2018) 'suitability for machine learning' measure. As utilised in this study, the SML measure is notable for its capacity to capture how easily *features* of occupational tasks can be performed by machine learning systems by scoring these features on a range of criteria that correspond to the mechanics behind software programming. Thus, as a measure it is future-orientated and does not rely on whether current developments in machine learning and technological change have already relaced specific occupations and tasks. As a scale variable, it allows us to quantify vulnerability to technological displacement and use empirical methods to identify associations and relationships between it and other indicators.

Despite this, it has various shortcomings in its application to contexts outside the US. The SML measure was constructed using the definitions and classifications of occupations as supplied by the Occupational Information Network (O*NET) in the US. While the translatability of this classification system to other contexts is sufficient, it would be a useful exercise for future work to replicate the Brynjolfsson, Mitchell and Rock (2018) methodology to construct a more specified measure for the context of developing economies using more appropriate classifications of context-specific occupations. The SML measure also does not take into account the fact that the tasks associated with a given occupation may have changed over time. The assumption that occupational tasks remain the same and do not evolve to complement new forms of technology is a limitation to any time-series analysis.

This study has focused on variations in the SML distributions of demographic groups in South Africa. While an important starting point, this can be built upon by considering variations in vulnerability to technological displacement along other indicators. Geographical variation, or a focus specifically on urban areas is important in order to determine the ease at which replaced labour can be absorbed by other industries in the geographical vicinity. Related to this, variation between or within specific sectors or industries will aid in contextualising how severe the threat of technological displacement will be in actuality. Some sectors may allow more for the movement of labour between occupations than others. Thus, further study should consider the general equilibrium that the South African economy would reach as the result of technological change.

It is also important to consider the role of the firm in processes of technological change.

This study has argued that, through state policy to promote employment equity, firms in South Africa are incentivised to hire more black and female workers. To expand on this, it is useful to investigate this incentive in relation to the incentives to adopt more forms of labour-substituting capital. As a variable, wages were missing from this study due to the unavailability of data. However, it is important to determine the price of labour for firms and sectors relative to new forms of technology. The role of labour brokers and other external actors has also been found to be significant in this regard (Kenny 2018). Related to this, investigation into the effects of the COVID-19 pandemic and how this has changed the allocation of labour and capital from the firm's perspective has become a vital area of study.

Machine learning and other new technologies have fundamentally changed many aspects of economic production in the Twenty-First Century. They have the potential to create new jobs and complement others. Advancements in processing power and data storage have been important instruments to high-skilled data analysts, and the increased functionality of the internet has reduced many of the limitations to employment imposed by geography. However, technologies also pose a threat to other groups of workers. The tasks performed by clerical, administrative and certain service-type workers can be performed equally as well by software systems and, in cases where these systems are sufficiently cost effective, can replace these workers. In developed countries, this has resulted in technological displacement and a polarisation of the wage distribution. In the South African context, it manifests as an increase in the vulnerability of female workers, specifically young, black females. Further research is required to elaborate on this vulnerability and the true propensity of new technologies to penetrate the South African economy. However, it is clear that any progress made in transforming the South African economy must necessarily include the increased protection of these marginalised groups.

References

Acemoglu, D. and Autor, D., 2010. *Skills, Tasks and Technologies: Implications for Employment and Earnings*. Working Paper 16082. Cambridge, MA: National Bureau of Economic Research.

Autor, D.H., Katz, L.F. and Kearney, M.S., 2006. *The Polarization of the U.S. Labor Market.* Working Paper 11986. Cambridge, MA: National Bureau of Economic Research.

Autor, D.H., Levy, F. and Murnane, R.J., 2003. The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics,* 118(4): 1279-1333.

Bhorat, H., Goga, S. and Stanwix, B., 2013. *Occupational Shifts and Shortages: Skills Challenges Facing the South African Economy.* Pretoria: Labour Market Intelligence Partnership.

Brynjolfsson, E. and McAfee, A., 2011. *Race Against the Machine.* Lexington, MA: Digital Frontier Press.

Brynjolfsson, E., Mitchell, T. and Rock, D., 2018. What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings,* 18: 43-47.

Budlender, D., 2019. Unresolved issues: Equal pay for work of equal value. *Agenda,* 33(4): 62-66.

Crankshaw, O., 1997. *Race, Class and the Changing Division of Labour Under Apartheid.* London: Routledge.

Das, M. and Hilgenstock, B., 2018. *The Exposure to Routinization: Labor Market Implications for Developed and Developing Economies.* IMF Working Paper WP/18/135.

Davies, R.H. and Seventer, D., 2020. *Labour Market Polarization in South Africa: A Decomposition Analysis.* Working Paper 2020/17. Helsinki: World Institute for Development Economics Research.

Espi, G., Francis, D. and Valodia, I., 2019. Gender inequality in the South African labour market: Insights from the Employment Equity Act data. *Agenda,* 33(4): 44-61.

Francis, D. and Valodia, I., 2021. *Black Economic Empowerment: A Review of the Literature.* Working Paper 21. Johannesburg: Southern Centre for Inequality Studies.

Frey, C.B. and Osborne, M.A., 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting & Social Change.* 114: 254-280.

Handcock, M.S. and Aldrich, E.M., 2002. *Applying Relative Distribution Methods in R.* Working Paper 27. Seattle: Center for Statistics and Social Sciences, University of Washington.

Handcock, M.S. and Morris, M., 1999. *Relative Distribution Methods in the Social Sciences.* New York: Springer.

Ingle, K. and Mlatsheni, C., 2017. *The Extent of Churn in the South African Youth Labour Market: Evidence from NIDS 2008-2015.* Working Paper 201. Cape Town: Southern Africa Labour and Development Research Unit.

Katz, L.F. and Murphy, K.M., 1991. *Changes in Relative Wages, 1963-1987: Supply and Demand Factors.* Working Paper 3927. Cambridge, MA: National Bureau of Economic Research.

Kenny, B., 2018. *Retail Worker Politics, Race and Consumption in South Africa: Shelved in the Service Economy.* New York: Macmillan.

Kerr, A., Lam, D. and Wittenberg, M., 2018. *Post-Apartheid Labour Market Series* [dataset]. Version 3.3. Cape Town: DataFirst.

Keynes, J.M., 2010. Economic possibilities for our grandchildren. In *Essays in Persuation.* London: Palgrave Macmillan, 321-332.

Mosomi, J., 2019. An empirical analysis of trends in female labour force participation and the gender wage gap in South Africa. *Agenda*, 33(4): 29-43.

Posel, D. and Casale, D., 2019. Gender and the economy in post-apartheid South Africa: Changes and challenges. *Agenda,* 33(4): 3-10.

Rankin, N.A. and Roberts, G., 2011. Youth unemployment, firm size and reservation wages in South Africa. *South African Journal of Economics,* 79(2): 128-145.

Statistics South Africa, 2021. Youth still find it difficult to secure jobs in South Africa. *StatsSA*. Available at: http://www.statssa.gov.za/?p=14415 [14 February 2022].

Tinbergen, J., 1974. Substitution of graduate by other labour. *Kyklos: International Review for Social Sciences*.