Essays in Spatial and Labour Economics

Matthew John Sharp

London School of Economics & Political Science

A thesis submitted to the London School of Economics & Political Science for the Degree of Doctor of Philosophy

London, August 2020

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 34,000 words (excluding appendices and references).

London, *10 August 2020*

Matthew John Sharp

Acknowledgements

I would like to express my profound gratitude to Vernon Henderson and Henry Overman for their excellent supervision and thoughtful guidance throughout the PhD. I have a learned a great deal from them about how to approach, conduct and communicate academic research. Vernon's deep intellectual curiosity and academic rigor were inspiring. Henry gave wise strategic advice from the start and his razor-sharp mind for econometrics kept me on my toes.

I am also indebted to Olmo Silva for reviewing my upgrade and to both him and Steve Gibbons for being very patient with my technical questions during office hours. Thank you also to Gabriel Ahlfeldt, Felipe Carozzi, Christian Hilber, Simona Iammarino, Neil Lee and Lindsey Relihan for helpful comments during seminar presentations.

This thesis also benefited from comments received from discussants and participants at the North East Universities Development Consortium (NEUDC) Conference, the 12th World Bank-Agence française de développement (AFD) Migration & Development Conference, the Oxford Centre for the Study of African Economies (CSAE) Conference, the European Urban Economics Association (UEA) Conference, the UEA PhD Summer School, the LSE Spatial Economics Research Centre (SERC) Conference and the Southern Africa Labour and Development Research Unit (SALDRU) Seminar at the University of Cape Town.

Generous financial support from the Commonwealth Scholarship Commission and from LSE is gratefully acknowledged. I would also like to thank Takwanisa Machemedze, Alex Montgomery and especially Lynn Woolfrey at DataFirst at the University of Cape Town School of Economics for granting me access to secure data at the centre, for facilitating my visits and for help with data. Sebastian and Tali were gracious hosts in Cape Town.

Thank you to Elizabeth Stuart at Oxford and Mark Roberts at the World Bank for giving me flexibility to focus on my PhD.

I was lucky to learn from and/or be distracted by many PhD friends. Special thanks to Andreas and Lars, who I would take into battle any day. Cong, Dzahmilya, Eduardo, Lindsey, Louise, Margarida, Mook, Sandro, Tanner, Thomas, Ulrich and Yohan from LSE were excellent company. I am also grateful to PhD friends from other universities whom I met at conferences and workshops, especially Tomás at NYU.

Lastly, I am very grateful to my parents and my twin sister for their encouragement, patience and support.

Abstract

This thesis is comprised of three independent chapters on topics in spatial, labour and development economics. I focus on South Africa for which there is a rich and under-exploited set of micro-data and where the peculiar history of the country - including restricted migration and separate development along race lines - makes for a particularly interesting setting.

The first chapter provides some of the first evidence on the labour market impacts of female internal migration. Merging detailed migration data from censuses and labour market data from labour force surveys, I exploit substantial time-variation in female migrant inflows into over 200 districts. To identify causal effects, I make use of the unique history of South Africa to construct a plausibly exogenous shift-share instrument for female migrant concentration based on earlier male migration flows from reserves during the Apartheid period. I find that this migration increases the employment and hours worked of high-skilled women (due to substitution in household work) and leads to a reduction in the employment of low-skilled female non-migrants (due to increased competition).

The second chapter examines how minimum wage legislation influences the labour market impacts of productivity shocks. Merging district-level high resolution weather data with high frequency data from South Africa's labour force survey, I examine how the introduction of an agricultural minimum wage affects resilience to weather shocks in the short term using a difference-in-differences approach. I find strong evidence that the substantially increased (and inflexible) wage bill after the minimum wage law leads agricultural employers to retrench formally employed workers in the wake of

negative shocks (whereas agricultural employment was more resilient in the pre-law period).

In the third and final chapter, I estimate the magnitude of agglomeration externalities in South Africa using a unique geo-coded panel micro-dataset where workers are tracked as they move across the country. The few studies on developing countries to date have estimated much higher agglomeration elasticities than those found in developed countries, but these studies have generally been unable to control for sorting on unobservables or to work with the ideal geographic units. Employing individual fixed effects and an instrumental variable constructed from a novel dataset on historical population settlements, my preferred estimate for regional wage elasticity is approximately 0.03 - in line with estimates for developed countries.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

I.I Introduction

The broad focus of this thesis is spatial and labour economics in developing countries. I examine topics which are of particular relevance to the development trajectories of these countries - the functioning of labour markets; the characteristics and impacts of internal migration; the resilience of agricultural employment to weather shocks; and the productivity benefits of cities.

While China and India - and to a lesser extent, Brazil, Indonesia and Colombia - have received attention in the burgeoning spatial economics literature focusing on developing countries, South Africa has been relatively neglected. This is despite the fact that the country is a particularly interesting case given its peculiar history of separate development along race lines and restricted internal migration. The first and third chapters of this thesis make use of the quasi-experimental setting created by Apartheid policies.

The work that has gone into this thesis has been very data-intensive. South Africa has a wealth of rich micro-datasets - although they often require a considerable amount of cleaning - that remain under-exploited. It is one of the very few developing countries which carries out a panel household survey. There are also numerous recently-digitised censuses from the Apartheid era.[1] The labour force surveys - avail-

[1]It does however require some work to match districts over time especially in the absence of shapefiles for some of the older censuses.

able since 1993 annually, bi-annually or quarterly depending on the period - are a very useful resource (and have been consolidated into the Post-Apartheid Labour Market Series with consistent cross-entropy weights).[2] I also make use of hard copies of historical censuses and high resolution satellite weather data.

I.II Overview of the thesis

Chapter 1: 'The labour market impacts of female internal migration: Evidence from the end of Apartheid'

The first chapter studies the labour market impacts of female internal migration in South Africa, examining the heterogeneous effects on non-migrants of different skill groups. With a dramatic increase in the rate of urbanisation in the last several decades, policy-makers in South Africa, as in many other developing countries experiencing rapid urbanisation, have expressed concerns that cities have been unable to cope with the influx of migrants. There is a particular concern that rural migrants may undercut low-skilled non-migrants in the labour market. Despite its importance, there have been few studies of the labour market impacts of internal migration in developing countries. No published studies have specifically examined the economic impacts of female internal migration. This is despite the fact that female migrants in developing countries tend to be very differently distributed across sectors as compared to men, and thus their arrival in receiving regions would be expected to have different effects on the non-migrant population.

Whereas many migration papers focused on developing countries are limited by data availability, I merge large sample migration data from South African censuses with detailed labour force survey data, and exploit substantial time-variation in female migrant inflows into over 200 districts. To identify the causal effects of migration on labour market outcomes, I make use of the unique history of South Africa to con-

[2]An important discovery of my PhD was that it is possible to determine lower levels of geography in labour force surveys from unique identification codes by comparing them against lists of district codes from the national statistics agency. These can then be matched with census data.

struct a plausibly exogenous shift-share instrument for female migrant concentration based on earlier male migration flows from reserves during the Apartheid period. In reference to recent literature on migration shift-share instruments, I argue that South Africa - due to the natural experimental setting created by Apartheid era policies - is one of the very few settings where an instrument using historical migration pathways can plausibly capture exogenous variation in migration flows.

I show that female internal migration in South Africa increased the employment and hours worked of highly-educated women (but not men) in the post-Apartheid period. I demonstrate that this was due to the fact that female migrants found employment as domestic workers, particularly in well-educated households, thereby freeing women from household labour and allowing them to pursue employment or increase their work hours. However, it was not the case that all subsections of the non-migrant population benefited from this female in-migration and I show evidence that low-skilled women experienced reduced employment. This suggests that female migrants, who were relatively well-educated though lacking in experience, substituted for low-skilled non-migrants in production.

Chapter 2: 'How does minimum wage legislation influence the labour market impacts of productivity shocks?'

My second chapter studies how reducing labour market flexibility by introducing a sectoral minimum wage influences the effects of productivity shocks on labour market outcomes. I focus on agricultural productivity shocks driven by weather variation. Developing countries are particularly vulnerable to fluctuations in weather affecting agricultural productivity, due to a large share of their labour forces working in this sector and the important role that agricultural production plays for the economic lives of the poor. There is also much recent interest in the impacts of minimum wage laws in developing countries. However, there is a missing perspective on how minimum wage legislation could affect resilience to productivity shocks.

I merge district-level high resolution weather data with high frequency data from

South Africa's labour force survey to form a panel database covering the period 2001-2007. I focus on an agricultural minimum wage introduced in 2003, which resulted in a substantial increase in the median wage (it was set at the 70$^{\text{th}}$ percentile of the prevailing wage distribution). I examine the influence of this law on the effects of weather shocks on labour market outcomes, exploiting variation in the gap between minimum wage levels and pre-law agricultural median wages across districts in a difference-in-differences approach.

Whereas weather shocks have no significant effect on agricultural employment in the pre-law period, in the post-law period, holding the pre-law wage gap at its mean level, a one standard deviation increase (decrease) in rainy season soil moisture relative to the local mean results in a 7.1% increase (decrease) in agricultural employment among the total labour force relative to its mean level within six months. When I focus only on the effects of below-average soil moisture (recoding above-average soil moisture values as zero) on agricultural employment, it is clear that the effects of negative weather shocks are driving the effects I find when I look at the interaction between my continuous weather measure and the minimum wage law. This shows that the minimum wage substantially weakens the resilience of agricultural employment to reduced rainy season moisture in the short term. I find evidence that after negative shocks occur, some of the displaced agricultural workers transition into unemployment while others transition into employment in other industries.

Chapter 3: 'Agglomeration economies in a developing country: Evidence from geo-coded micro-panel data in South Africa'

My third chapter examines the magnitude of agglomeration externalities in South Africa - that is, the effect of the agglomeration of workers or people on the productivity of workers proxied by their wages. The few studies on agglomeration in developing countries have found agglomeration elasticities often of 8% to 20%, considerably higher than the elasticities averaging 3% estimated in rigorous developed country studies. In theory, underdeveloped factor markets in developing countries may enhance the productivity benefits of cities relative to those in advanced countries. However, due to data constraints, agglomeration studies focused on developing countries are not able to include all the endogeneity controls used in the developed country papers and are often forced to work with geographical units that are far removed from local labour markets.

I examine the relationship between urban population and nominal wages in South Africa for the period 2008 - 2016 using all five rounds of the National Income Dynamics Survey, a nationally-representative panel study. I map individuals to a geographic level that approximates local labour markets as closely as possible. Since I have access to an individual-level panel of workers, my paper is able to go further than the other developing country papers in controlling for unobserved characteristics that may be correlated with location choices - a critical feature of the state-of-the-art empirical research on agglomeration effects. To control for simultaneity bias (or omitted variable bias), I construct an instrument from historical population data (specifically, from a hard copy of the 1960 census) that have not been used before in the South African literature. Owing to South Africa's unique history of controlled migration and forced reallocation of the population across space, there is a strong case to be made for the exclusion restriction for this instrument.

In my preferred specification including both individual fixed effects and a historical population instrument, I estimate an elasticity of wages to city population of approximately 3%, which is robust to a large number of tests. I find that using individual effects to control for sorting on unobservables reduces the estimated agglomeration elasticity for South Africa by 18-37% (depending on the specification) - in line with effect sizes in the developed country literature. I argue that these results provide suggestive evidence that the very high agglomeration elasticities found for other developing countries may in part be due to mismeasurement. I also examine the comparative effects of human capital externalities and agglomeration externalities, finding that agglomeration externalities seem to be much more important in predicting wages.

Chapter 1

The labour market impacts of female internal migration: Evidence from the end of Apartheid

1.1 Introduction

Most studies in the economics literature on international and internal migration, and their impacts, focus on male migrants and effects on male natives/non-migrants (Dustmann et al., 2016). However, a gendered focus on migration is important (Pfeiffer et al., 2007). Women may migrate for different reasons to men and the effects of migrant networks may be gender-specific (Davis and Winters, 2001; Enchautegui, 1997). Since women tend to be distributed differently across economic sectors as compared to men, their arrival in receiving regions would be expected to have different effects on the non-migrant population. For the same reason, female non-migrants are likely to be affected by in-migration differently to men. This paper focuses on measuring the labour market impacts of female internal migration in South Africa, examining the heterogeneous effects on non-migrants of different skill groups.

South Africa presents a particularly interesting case for the study of internal migration in developing countries. For over 80 years the mobility of South Africa's black population was strictly controlled and many were forced to live in native reserves in rural areas called 'homelands'. This reflected the needs of the white-controlled government for cheap - predominantly male - migrant labour to support mining and industrial development but also political anxiety about permanent rural-to-urban migration of this disenfranchised population (Turok, 2012). After the Second World War, political considerations dominated and increasingly draconian controls were imposed to limit black urbanisation in order to sustain political domination. While the Apartheid regime officially ended in 1991 (with the first democratic national election in 1994), the most important migration restrictions were withdrawn in 1986. This constituted a large shock to the system, resulting in a spike in the rate of internal migration. Since women had very few opportunities to leave homelands during Apartheid, the female migrant supply shock was particularly large.

With the dramatic increase in the rate of urbanisation, policy-makers in South Africa, as in many other developing countries experiencing rapid urbanisation, have expressed concerns that cities have been unable to cope with the influx of migrants. There is a particular concern that rural migrants may undercut low-skilled non-migrants in the labour market. Despite its importance, there have been few studies of the labour market impacts of internal migration. No published studies have specifically examined the economic impacts of female internal migration.

I examine the changing nature of female migration and impacts thereof on the labour market outcomes of non-migrants in over 200 receiving districts in South Africa after Apartheid. Internal migration in South Africa has a strong gender dimension. Historically, the Apartheid-era migrant labour system meant that predominantly black African men moved to urban areas without their families. After the abolition of influx controls, many women relocated to join their male partners (Von Fintel and Moses, 2017). The share of women in the total migrant population grew at the same time as female participation in the labour market increased. I first document the surge of female migration that occurred after the end of migration restrictions and examine the characteristics of the migration population. I show that while female migrants had very little work experience, they were relatively well-educated as compared to the population in receiving districts. I then look at the impacts of migration on the non-migrant population in receiving districts between 1996 to 2001, a period during which there was considerable variation in district-level migration inflows. I use an empirical specification that includes district and time fixed effects and examine heterogeneous effects across gender and skill groups.

To identify the causal impact of migration flows, I exploit historical migration pathways - particularly between areas that were native reserves and receiving areas - to predict contemporary migration. I employ a shift-share 'past settlement instrument' that makes use of the tendency for female migrants to settle in places where earlier male migrants from the same sending region already reside. Crucially, given South Africa's peculiar history - including centrally-determined migrant allocation during the Apartheid era and dramatic changes in the spatial economic landscape after the end of Apartheid - there is a strong case to be made that the historical distribution of (black) migrants across geographical areas is unrelated to contemporary economic shocks in the same areas.

Overall, in line with much of the migration impacts literature, I find that female migration has a non-significant effect on the employment and wages of non-migrants. However, dividing the population into skill groups, I find that female migration results in an increase in the employment (and hours worked) of highly educated women and a reduction in the employment of low-skilled female non-migrants. I provide evidence that migrant women supply services that are close substitutes with housework thereby freeing up the labour of high-skilled women. On the other hand, female migrants seem to compete for low-skilled jobs with female non-migrants in receiving economies.

My paper contributes to the literature on internal migration in developing countries (Lall et al., 2006). Following on from work on the impacts of international migration in developed countries, several recent studies (Combes et al., 2015; Kleemans and Magruder, 2018; Strobl and Valfort, 2015) have looked at the economic impacts of internal migration in developing countries using spatial variation of migrant inflows for identification. None of these papers have tried to separate the effects of male and female migration nor have they examined in any detail differential effects on male and female non-migrants. Whereas all of these papers have had to make use of small sample migration data or census data for labour market outcomes, I combine large sample migration data with detailed labour force survey data on labour market outcomes. Only Kleemans and Magruder (2018) have have more than a single cross-section of labour market data to work with. Past studies have found mixed results. Some have found negative economic impacts of internal migration on wages and employment (Kleemans and Magruder, 2018; Strobl and Valfort, 2015) in line with basic theory (Borgas, 1999; Card, 2001). Others, however, have found positive effects, which have been attributed to complementarities in the production function and agglomeration economies (Combes et al., 2015).

A well-known challenge in the migration impact literature is overcoming the endogeneity of migration flows: the fact that migration is likely to be correlated with economic conditions in receiving regions, which will bias estimates of migration effects. Since Card (2001), past settlement instruments (otherwise known as 'enclave-' or 'network instruments') using the lagged geographic distribution of migrants have been employed extensively in the immigration literature, and also in the internal migration literature, albeit to a lesser extent. An important difficulty with this type of instrument is that it can only address the endogeneity of migrant location choices if local economic shocks that attracted the earlier migrants are not also at work during the contemporary period (Borgas, 1999). In recent years, this assumption has been increasingly questioned. Most papers do not even attempt to tell a story of why migration determinants should have changed, merely relying on significant time passing between the contemporary and lagged periods. In situations where migration pathways are highly stable over time (which is very often the case), historical migration pathway instruments may predict current migration flows well but it is highly unlikely that the exclusion restriction holds (Jaeger et al., 2018). In this paper, I make use of the fact that dramatic changes in South African government policy created a substantial shift in the area-of-origin composition of internal migrants in South African regions between the Apartheid and post-Apartheid periods.

My paper also adds to the growing literature that tries to exploit exogenous variation resulting from Apartheid to study the development of South Africa (Bakker et al., 2020; Biavaschi et al., 2018; Mariotti, 2015). One shortcoming of most of these papers is that they have assumed that all Apartheid era policies ended abruptly in 1994 with the first democratic elections. A more careful appraisal of the historical record shows that the Apartheid regime was gradually dismantled from the mid-1980s to 1991. In the case of migration restrictions (which also applied for a time to black immigrants from other countries), the most important of these actually ended with the repeal of the Pass Laws Act in 1986 - and not with democratision in 1994 - following which black people were free to move to cities (Ogura, 1996). A few recent papers have tried to measure the effects of international migration on labour market outcomes in South Africa (Biavaschi et al., 2018; Broussard, 2017). These papers have used data from the 1996 census to construct historical migration paths. They have also used census data for native labour outcomes, which may not be ideal for this purpose (Ardington et al., 2006), and have used larger geographical units, often containing several labour markets. These papers have found negative impacts of immigration on the employment and incomes of natives. There have been no nationally representative studies of internal migration in South Africa,[1] nor has there been any previous attempt to measure the economic impacts of internal migration in South Africa.

Lastly, my study contributes to the emerging literature on gender differences in local labour markets. Most studies just assume away gender. To date there have been a few studies examining the impacts of female immigration from low-income countries on native labour market outcomes in advanced countries (Barone and Mocetti, 2011; Cortes and Tessada, 2011). These studies have generally found that this immigra-

[1]Several limited studies (Naidoo et al., 2008; Oosthuizen and Naidoo, 2004; Rogan et al., 2008) have focused on profiling internal migration to particular metropolitan areas (usually Cape Town or Gauteng) in South Africa and on trying to analyse push and pull factors contributing to migration.

tion has increased the hours worked of high-skilled/high-income women and have highlighted the channel of substitution in household work. There are several reasons why the impacts of female *internal* migration in a *developing* country context may be different. First, labour markets in developing countries are structurally quite different from those in developed countries. There are typically much higher rates of informality and unemployment/underemployment, and the median worker is much less skilled. Second, the demographics of the internal migrant population may be different from those of the immigrant population given that it is much cheaper to move and migration drivers may be different. Thirdly, labour markets may be thinner in developing countries, which might result in larger labour market responses (Kleemans and Magruder, 2018). My paper also differs from the papers mentioned above in that it not only focuses on the effects of migration on the labour supply of high-skilled/high-income women but also considers the labour market effects on low-skilled non-migrants. I also consider a broader range of non-migrant labour market outcomes including employment and wages.

The paper is structured as follows. In Section 2, I describe the historical background. In Section 3, I present the data. I detail my estimation strategy in Section 4. In Section 5, I report my results. In Section 6, I present various robustness checks and in Section 7, I turn to an examination of channels. Section 8 concludes.

1.2 Historical background

The philosophy underlying separate development in South Africa can be traced back to the early days of its colonial history. Segregationist policies began in 1913, three years after the formation of the Union of South Africa, when the Natives Land Act was passed. This was designed to counter the flow of blacks to urban areas after the devastation caused by the Second Anglo-Boer War (1899-1902) and demarcated black reserves for black ownership and occupation while prohibiting the black population from owning land outside of them (Ogura, 1996). Further legislation in 1923 and 1937 compelled the black population to live in certain areas and prohibited the black population, born outside of cities, from spending more than 14 days (and, later, three days) in cities to seek work.[2] Despite these restrictions, the black population in cities grew, especially with the increased demand for labour from the rapidly developing manufacturing industry around the time of the Second World War (Christopher, 2001).

Restrictions on migration became much stricter after 1948 when the Afrikaner National Party (NP) came to power and implemented its programme of Grand Apartheid ('apart-ness' in Afrikaans), which aimed at compete social and spatial segregation and was supported by significant government resources. Grand Apartheid aimed at moving a large proportion of the black population - all who were not needed as labourers in white urban areas - to native reserves. To this effect, it has been estimated that 3.5 million black people (equivalent to a fifth of the black population in 1980) were forcibly relocated from 'white' areas. These reserves were overcrowded, had extremely high poverty rates, were reliant on meagre subsidies from the national government and had limited industry. Living conditions and the provision of public services in resettlement camps were dismal (Desmond, 1971; Horrell, 1973). Smallholder farming was one of the only activities that could be undertaken but even this was limited by the poor quality of the agricultural land.

A number of important laws were passed. The Group Areas Act of 1950 established residential areas for black people (which would later become 'townships') in metropolitan areas and smaller cities and was meant to strengthen controls on the flows of blacks into cities (Ogura, 1996). The Population Registration Act (1950) assigned a population group to each citizen, which determined their political and social rights. The Pass Laws Act (1952) forced every black African to carry a passbook (similar to a passport) at all times documenting their permission to be in certain areas, without which they were subject to arrest. In 1959, the Promotion of Bantu Self-Government Act formalised the system of homelands, one for each 'ethnic' group.[3]

[2]The relevant laws were the Native (Urban Areas) Act (1923), the Native Laws Amendment Act 1937 and the Native (Urban Areas) Consolidation Act (1945).

[3]By the time they were reincorporated in 1994, South Africa's homelands - their land mass con-

After 1948, the control of migrant labour became much extensive and bureaucratic (Hepple, 1969; Wilse-Samson, 2013). The employment of blacks was subject to the authority of the central labour bureau in Pretoria, which oversaw and controlled the numerous local bureaux. Blacks could not be employed in white areas without the permission of the local labour bureau. Furthermore, no black person could leave a homeland to work or seek work without the authority of the homeland bureaux, which would pool all workseekers in their respective areas. Since there was often a shortage of agricultural labour,[4] blacks were only allowed to leave rural areas if labour supply in the agriculture industry was deemed sufficient (Wilse-Samson, 2013). It was also government policy that each reserve would send the majority of its workers to one specific industry (Leys, 1975; Mariotti, 2015). For example, the Chamber of Mines in South Africa recruited from four out of ten homelands, and predominantly from just one (Mariotti, 2015). This centralised allocation process was independent of the preferences and economic characteristics of the migrants (ibid.). Mobility restrictions in South Africa were very strict, and hundreds of thousands of arrests were made every year under the pass laws.[5]

Also relevant were the significant efforts to decentralise industry from the late 1950s. These included policies both to promote industrialisation in selected white areas bordering homelands ('carrots') and to restrict industrial expansion in metropolitan centres ('sticks'). While ostensibly designed to alleviate diseconomies of scale in metropolitan areas, the main goal was to support the state's system of influx control by providing alternative employment opportunities in areas near Bantustans (Wellings and Black, 1986). Under one scheme, white industrialists were encouraged to locate their factories in white towns near homelands, with black labourers being housed either in satellite townships or in new towns within the reserves. Some 80

stituting 13% of the country's land mass - were home to approximately 20 million black South Africans.

[4] As a share of the economically active black population in 1970, about 40% of blacks were employed in agriculture, forestry and fishing (2.26m people). Involvement in other sectors was as follows: 11% in mining and quarrying (610,000 people); 9% manufacturing (514,000 people); 18% in the services sector (1.1m people) (Lombard and Stadler (1980) quoted in Wilse-Samson (2013)).

[5] One estimate suggests that between 1916 and 1984, some 17.7 million Blacks were arrested, prosecuted, and relocated from White areas under the pass laws (Savage, 1986).

select 'growth points' were established where manufacturing enterprises benefited from infrastructural developments as well as some of the most generous location incentives available anywhere in the world - including subsidised transport, water and power; reduced taxation and controls on wages for black employees (Best, 1971; Rogerson, 1998). Most growth points were located where 'positional, infrastructural and agglomeration advantages were negligible' (Wellings and Black, 1986). Location decisions were instead driven by the need to allocate several to each of the homelands as well as a need to satisfy the demands of influential white constituencies eager to achieve growth point status (Dewar et al., 1984; MacCarthy, 1982). Under the Physical (later Environment) Planning Act, introduced in 1967, controls on the number of black labour intensive strategies were put in place in the metropolitan centers of Pretoria-Witwatersrand-Vereeniging (PWV), Port Elizabeth-Uitenhage, Bloemfontein and Cape Town. The expansion of industrial land and the establishment of industries with a higher than officially approved black-white employee ratio was prohibited in large swathes of these metropolitan areas (Geyer, 1989). This policy as well as the 'border industries' policy were widely criticised for being economically irrational (Wellings and Black, 1986).

As a consequence of all of the above, and importantly for the identification strategy in this paper, black migrants from homelands during Apartheid were not free to choose where to migrate to on the basis of market-based economic incentives (Feinstein et al., 2005; Mariotti, 2015). To a large extent, they could only move to areas where the government decided their labour was needed, and these decisions were usually made on political rather than economic grounds. While it was not the case then that workers were forced to migrate to particular areas in South Africa, their choice set was severely limited: they either had to accept an offer to work in a particular industry in a particular place or they could remain in homelands and try to get by on subsistence farming or remittances from family members. Those who decided to take up job offers would often have to relocate to specially demarcated areas on the periphery of white areas. Often these areas were more than a day's travel from their homeland of origin, and they might only have been able to visit their families on one or two occasions a year (Ogura, 1996). The result was the formation of migrant enclaves across the country.

Also important for this paper were the conditions around female black migration in the pre-1994 period. There were few employment opportunities for black women in cities and towns. Sometimes they could find work as domestic helpers or in factories but the Apartheid regime was mostly interested in employing male migrant labour. Furthermore, to a large extent, the Pass Laws prevented the spouses or children of pass book holders from accompanying them to the urban areas they were employed in (Healy-Clancy, 2017; Von Fintel and Moses, 2017). The majority of black migrant men lived in single sex compounds near the work-sites and were not allowed to have visitors of the opposite sex staying overnight. Following the end of migration restrictions, female migration from homelands dramatically increased (as I explore in the following section). Many women would have followed in the footsteps of their male partners who had been migrants while the migration restrictions were in place (Von Fintel and Moses, 2017). In support of this thesis, in the 2001-2002 HSRC National Migration Survey the main reason cited by female respondents for having migrated between magisterial districts was 'getting married or moving in with a partner' or 'getting separated or divorced' (13% together of respondents) (Wentzel et al., 2006). 'Moving closer to relatives' was ranked as the third most important reason for moving by (8% of) female respondents, and 'having to move with a spouse' was another common reason cited by (7% of) female respondents (ibid.).[6][7]

While the Apartheid regime officially ended in 1991, the most important migration restrictions were withdrawn in 1986 - with the Abolition of the Influx Control Act. With this act, the pass system was dropped and black people were allowed to pur-

[6]'Looking for work' ranked second highest (11%) as the most important reason for moving for female migrants. For male migrants, on the other hand, this was by far the most important reason for moving (23%) (Wentzel et al., 2006, pg.189).

[7]In general migration networks seem to be an important determinant of internal migration in South Africa. In an econometric analysis based on the 2002 HSRC National Migration Study, it was found that the presence of a migrant network in a possible destination area was by far the most important predictor of whether a respondent chose to migrate in the five years before the survey (Wentzel et al., 2006, pg.195).

chase land and housing outside of the homelands. This resulted in a spike in the rate of internal, largely supply-driven migration. Simultaneously, from the early 1990s, tough international sanctions were loosened and then repealed, and the economy was opened up with dramatic consequences in terms of the spatial distribution of economic activity (Imbs, 2013). While the South African economy experienced negative economic growth in the decade before 1994, in the following decade it experienced positive economic growth (Aron et al., 2009). Rapid regional economic divergence in the post-Apartheid period has also been documented (Bosker and Krugell, 2008). Relatedly, Imbs (2013) studies rapid structural transformation in the post-Apartheid period in South Africa, finding that as South Africa opened to international trade, manufacturing and extractive industries waned and services took over, substantially altering the spatial economic landscape of the country.

1.3 Data

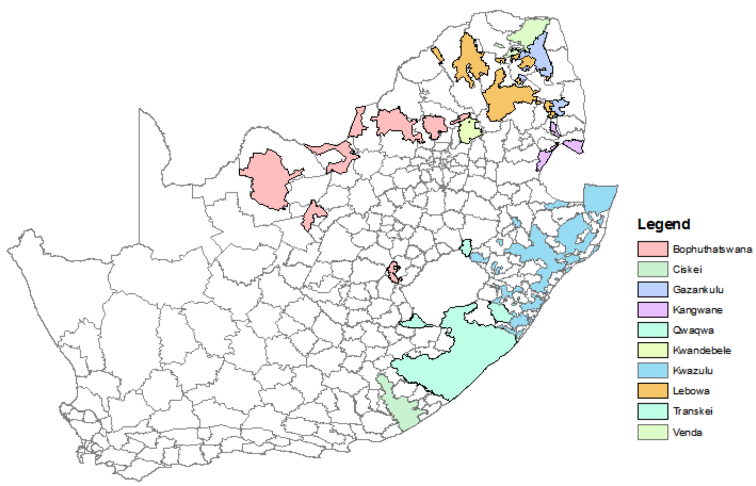
1.3.1 Unit of geography

The main geographic unit of observation in this paper is the magisterial district. There are 354 magisterial districts in South Africa with an average territory of 3447.5 km^2.[8] In 1996, these had an average working age population size of approximately 100,000. Magisterial districts are administrative units but closely approximate labour markets and have been used as such in several recent well-published papers (e.g. Magruder (2010), Magruder (2012)). Each magisterial district corresponds to the jurisdiction of a magistrate's court, the lowest level of the South African court system, and contains at least one sizeable town or city. While regional boundaries in South Africa are known to change quite regularly in recent times, they were remarkably stable during and in the decade after Apartheid.

Since I am focusing on migration from areas classified as homelands during Apartheid (within which black people could move freely) to the rest of South Africa (where black

[8]While magisterial districts still exist, there were replaced by local municipalities as an administrative layer in the census and most surveys after 2001.

Figure 1.1: South African homelands and magisterial districts



people could not move freely when the pass laws were in place), I exclude magisterial districts from my sample of receiving districts that overlapped with homeland areas in the Apartheid period. This leaves a sample of over 200 receiving districts. In the case of most homelands, there was a clear continuation between former homeland boundaries and current magisterial districts. In the case of a few homelands (Kwazulu in particular), some work is required to determine which magisterial districts should be considered as part of former homeland areas. I have used EOG DMSP nighttime light data from 1991 and 2010 to identify population clusters and to determine whether roughly at least 20% of the population of a magisterial district in 1991 was settled on an area within a former homeland area - in which case, I consider this as a 'homeland magisterial district'.[9]

1.3.2 Migration & regional controls

Contemporary migration data come from the 10% sample of the 1996 and 2001 censuses, which cover 3.6 million and 3.7 million individuals respectively. These censuses contain questions on the duration of time at current residence and the

[9]The number of magisterial districts is higher in 1996 than in 1985 but this was mostly just due to some cases where adjacent magisterial districts were combined. Moreover, four homelands were considered independent in 1985 and only after 1994 were new magisterial districts boundaries drawn for these areas.

place of previous residence. Focusing only on migration from former homelands, I calculate migrant stocks in receiving districts in 1996 and 2001.[10] The large size of the samples means that I can create an accurate measure of internal migration at a relatively fine spatial scale. For the construction of the historical pathways instrument, I use data from the full-count 1985 census, which provides information on the place of birth of individuals.[11]

1.3.3 Labour market outcomes & individual characteristics

For the main analysis, I use South African labour force surveys available (annually, biannually or quarterly, depending on the period) from 1996 to 2017 and including on average approximately 25,000 households per wave. These contain much more detailed and consistent information than the census on labour force characteristics (including exact wages earned and more detailed occupation and industry information). While most of these surveys do not provide data on migration, several waves spread throughout the series do include this information, which allows non-migrants to be distinguished from migrants. This is important so that I can exclude in-migrants from the sample. My estimates can then be interpreted as a treatment effect of migration on the non-migrant labour force instead of a compositional change in the labour force with a migrant influx. I pool data from LFS 1996 and LFS 1997[12] to merge with migration data from the 1996 Census and I merge the September 2003 wave of the LFS with the 2001 Census.

I restrict my analysis to individuals between 18 and 64 years old and exclude re-

[10]The 1996 Census provides information on migration at any time during a person's life (the questions were 'What was your district of previous residence?' and 'Which year did you move?').The 2001 Census only provides information on migration in the past 5 years (the questions were 'Did you move from another district in the last 5 years?' 'Where from?' 'Which year did you move?'). I calculate the (imperfect) migration stock in 2001 by adding together the 1996 stock and 1996-2001 flow.

[11]In fact, information is only provided on whether people were born in a homeland (with the particular homeland listed) or whether they were born in (the rest of) South Africa. However, this is sufficient information for my identification strategy.

[12]Actually these survey waves were part of the October Household Survey series but they function as a labour force surveys since they contain detailed information on labour force characteristics, which is directly comparable to that in later surveys. I pool data from two waves here because earlier waves surveyed fewer people. In Section 6, I show that just using data from LFS 1997 does not substantially change my results.

Table 1.1: Summary statistics for recent migrants

	1985		1996		2001	
	Male & female					
	Migrants					
		SD		SD		SD
Age	33.5	12.0	31.4	9.4	30.1	9.1
Avg. years of educ.	4.8	3.8	7.5	4.1	9.0	3.8
Gr. 12 or higher (%)	3.3	17.7	19.1	39.3	37.2	48.3
Female (%)	32.7	46.9	41.6	49.3	46.8	49.9
Female/male employ.			0.68		0.74	
	Non-migrants					
		SD		SD		SD
Age	33.0	13.1	36.2	11.1	35.8	10.9
Avg. years of educ.	7.3	4.1	8.1	4.1	8.5	4.1
Gr. 12 or higher (%)	20.3	40.2	23.9	42.6	33.0	47.0
Female (%)	49.7	50.0	45.0	49.8	46.0	49.8
Female/male employ.			0.80		0.83	
	Female only					
	Migrants					
		SD		SD		SD
Age	33.2	12.7	31.3	9.2	30.0	8.8
Avg. years of educ.	5.0	3.9	7.6	4.1	9.1	3.8
Gr. 12 or higher (%)	4.2	20.0	19.0	39.1	38.2	48.6
	Non-migrants					
		SD		SD		SD
Age	33.0	13.2	35.4	10.9	35.0	10.7
Avg. years of educ.	7.3	4.0	8.3	4.1	8.6	4.0
Gr. 12 or higher (%)	19.2	39.4	24.9	43.2	34.5	47.5

Data from Census 1985, Census 1996 & Census 2001 tired and disabled individuals, and individuals in full-time education. I refer to this group as the 'working age population'. I convert all hours worked and earnings to a monthly amount. If a worker reports working more than 84 hours a week, I recode hours worked as missing. I drop 8 outliers[13] and workers earning less than the first percentile of the wage distribution as I am concerned about under-reporting. In terms of relevant controls, educational attainment is defined for the following five categories: No education, primary education, incomplete secondary education, complete secondary education and tertiary education. I also include controls for age, age squared and race. I define three skill groups for non-migrants: the low-skilled category includes individuals that have at most completed Grade 9 of high school; the middle-skilled category includes individuals who have completed Grade 10 or Grade 11; and the high-skilled category includes individuals that have completed at least Grade 12 (including those with tertiary education qualifications). High school graduation rates are relatively poor in South Africa and completing Grade 12 (the final year of school) gives a substantial advantage in the labour force. Many high school students drop out in Grade 10 or Grade 11, after which they are still eligible to enter state-sponsored technical colleges.

For my summary statistics, I have used South African census data, which contain information on labour market outcomes and individual characteristics.[14]

1.3.4 Summary statistics & discussion

Table 1.1 shows some summary statistics for (recent) migrants and non-migrants for 1985, 1996 and 2001. For the purpose of this descriptive analysis, for 1996 and 2001, migrants are defined as working age individuals who moved from homelands - or former homelands - to the rest of South Africa in the five years prior to the census. Non-migrants are defined as working age individuals in migrant-receiving districts who did not move in the five years prior to the census. I have also included

[13] As suggested by the authors of the Post-Apartheid Labour Market Series, which standardises all waves of the labour force survey (including the October Household Survey).

[14] I prefer not to use this data for my main analysis since employment is not defined consistently across waves and there is only information on income categories.

information from 1985 for comparative purposes, however here 'migrant' is defined as a working age individual who was born in a former homeland area but is now living in a non-homeland area and 'non-migrant' is defined as a working age individual who was born in a non-homeland area.[15]

The average age of the migrant population was 4-6 years lower than that of the non-migrant population in 1996 and 2001 with the gap seeming to widen over time. During Apartheid, migrants had much lower levels of education than non-migrants. However, by 1996, migrants had dramatically closed this gap. In 1996, for both men and women, the difference in average years of education between migrants and non-migrants was less than one year, even if the proportion of migrants that had completed at least high school (Grade 12) was still 4-5 p.p. lower than that for non-migrants. By 2001, however, migrants had on average an extra 0.5 years of education than non-migrants and a higher proportion of them had completed high school. Remarkably, the proportion of female migrants that had completed at least Grade 12 more than doubled between 1996 and 2001 (with a similar increase for males). While the fact that migrants from rural homelands were more educated than non-migrants by 2001 may seem surprising at first, it is important to put this finding into context. The migrant population from homelands/former homelands was made up entirely of black people who were systematically provided with less (and poorer quality) education than their white counterparts during Apartheid. Education reform towards the end of Apartheid dramatically improved educational opportunities for black youth. Thus, the fact, as discussed above, that the migrant population is much younger than the non-migrant population goes a long way towards explaining how it could be that that migrating population from rural homelands in 2001 was more educated than the non-migrant population.[16]

[15] The 1985 census data file only provides detailed information on place of birth of South Africans for those born in homelands. Aside from this, I have information only on whether a person was born in South Africa or not. This is sufficient information for constructing my instrument even if it is not ideal for this descriptive analysis.

[16] While not directly comparable because they have focused on migrants from both rural and urban areas, other studies on migration in developing countries have also found that migrants are slightly more educated than non-migrants (Kleemans and Magruder, 2018; Strobl and Valfort, 2015).

In 1985, before the end of migration controls, the female share of migrants from homeland areas was very low at 33%. By 1996, this share had gone up dramatically to 42% of the migrant population and by 2001 women made up 47% of the migrant population. Also interesting is that while the ratio of female to male employment increased marginally for non-migrants between 1996 and 2001, there was a much larger increase in this ratio (from a lower base) for migrants from former homelands over the same period.

Between 1996 and 2001, female migrant concentration - i.e. the share of the female labour force (migrant and non-migrant, employed and unemployed, foreign-born and nationals) in non-homeland districts who were black female migrants from homeland areas - went from 5.5% to 7.4% (an increase of 35%). I am particularly interested in the nature and implications of this surge in female migration after the end of migration restrictions. If skill level is defined by level of education, then it appears that by 2001, female migrants from rural South Africa were relatively highly-skilled. However, one also needs to consider that these female migrants had little or no work experience before they migrated. According to data on previous work experience in the 1996 Census, of the female migrants in labour force in 1996, only 16% had ever worked before. Of those who had been employed before, 60% had been domestic workers.

Table 1.2 compares the occupational involvement of male and female migrants from homelands in 1996. A standout finding is that female migrants were twice as likely as male migrants to work in elementary occupations. This was driven by the fact that a large proportion (44.16%) of female migrants worked as domestic helpers. Women were also much more likely than men to work as teaching professionals and as clerks. On the hand, men were much more likely than women to work as service workers and as labourers in mining, construction and manufacturing. The fact that female migrants were involved in a completely different set of occupations to their male counterparts suggests that the effects of their arrival on local labour markets

Table 1.2: Female and male migrant occupations in 1996

		1996	1996
		Fem. migrants	**Male migrants**
Occ 1-digit			
3	Technicians and associate professionals	4.1	3.2
4	Clerks	7.0	3.1
5	Service workers and shop and market sales workers	7.9	12.9
9	Elementary occupations	58.6	29.1
Occ 2-digit			
23	Teaching professionals	7.0	2.1
24	Business and admin professionals	0.7	3.1
32	Health associate professionals	0.3	3.8
41	General and keyboard clerks	3.3	1.4
91	Cleaners and helpers	48.5	10.4
92	Agricultural, forestry and fishery labourers	6.3	7.9
93	Labourers in mining, construction, manufacturing	4.4	10.8
Occ 3-digit			
223	Nursing	3.0	0.2
913	Domestic and related helpers, cleaners	44.2	5.3

Data from Census 1996. Focusing on select occupation groups where there is a notable difference in rates of participation between male and female migrants.

were likely to have been very different too. The analysis above also suggests that there was some educational downgrading among the female migrant population.

1.4 Empirical strategy

1.4.1 Baseline estimation

My analysis focuses on the effects of female in-migration on the probability of employment and wages of non-migrants (where a non-migrant is defined as a respondent who has not moved from a different district in the 5 years before the survey) in destination districts. My analysis is run at the individual level using:

$$y_{idt} = \beta f_{dt} + \lambda X_{idt} + \gamma_t + \delta_d + \epsilon_{idt} \tag{1.1}$$

where y_{idt} is the outcome of interest for non-migrant i in district d at time t. Employment is a binary variable equal to 1 if a person is employed or equal to 0 if a person is unemployed, unable to find work or a house-wife/house-husband.[17] Monthly wages

[17] Rather than working with a narrow definition of unemployment, in many developing countries it makes more sense to work with a broad definition of unemployment which also includes discour-

(adjusted for inflation with the base period being December 2017) are calculated for all workers. In some specifications, I examine effects on hours worked i.e. monthly hours.[18] I follow Card (2001) and Card (2007) in standardising female migration inflows by the population of receiving districts since the average and standard deviation of the change in migration may be proportional to the total population of the district, potentially inducing a spurious correlation between labour market outcomes and migration (cf. Peri and Sparber (2011)).[19] The natural logarithm of female migrant concentration is represented by f_{dt}, that is:

$$f_{dt} = \frac{m_{dt}^{fem}}{p_{dt}^{gen}} \tag{1.2}$$

where m_{dt}^{fem} is the working-age female migrant stock from former homeland areas at time t and p_{dt}^{gen} is the total working-age population of the gender group in question at time t. The vector X_{idt} includes individual controls for gender, age (and its square), education levels and race. I include district fixed effects and year fixed effects.[20] β is the parameter of interest, which estimates the impact of the share of female migrants at time $t-1$ on the outcome of interest at time t. This coefficient estimates the

aged job-seekers, since official definitions of job-seeking may be inappropriate in contexts where there are high rates of informality (ILO, 1982). In the South African case, it has been argued that the StatsSA definition of 'searching for employment' is too restrictive and it has been shown that there is very little to distinguish the 'searching unemployed' from the 'non-searching unemployed' (Posel et al., 2014).

[18] I suspect hours worked may be measured with some error so I do not include this outcome measure in my main estimations.

[19] I have also experimented with using migration levels as the variable of interest and adding district population size as a control. The results are substantially unchanged.

[20] My fixed effects estimation with two periods is equivalent to a first-differences estimation i.e.

$$\Delta y_d = \beta \Delta p_d + \lambda X_d + \epsilon_d$$

with district weights equal to the number of observations in each district (multiplied by the cross entropy weights described in the text). Because I include regional and time fixed effects, changes in migrant share between 1996 and 2001 should be interpreted as flows of migrants in this period. My 'non-migrant population' includes some people who would have migrated at some point in their lives but not in the past five years. Using district fixed effects avoids any bias resulting from the fact that levels of migrant shares and levels of labour market outcomes may be spatially correlated because of common fixed influences, which could lead to a positive or negative statistical correlation between migration and economic outcomes, even if there are no real effects of migration.

total effects of migration, taking into account the indirect effects of migration through complementarities across skill groups and across capital and labour (Dustmann et al., 2016).[21] Errors are clustered at the district level to allow for correlation between individuals within district-level labour markets. For all estimations, I have used the recommended Statistics South Africa cross-entropy weights, which make all waves within the Post-Apartheid Labour Market 1994-2017 Series directly comparable to one another by producing a consistent set of totals for each wave (Branson and Wittenberg, 2011).[22]

1.4.2 IV approach

The number of migrants a receiving district receives may be correlated with the economic conditions of that district i.e. the number of internal migrants m_{dt} could be correlated with ϵ. OLS estimates may have an upward bias if female migrants choose to move to districts experiencing positive labour market shocks. However, if female migrants choose to stay away from cities with a high cost of living, OLS estimates may have a downward bias. Measurement error will also push OLS coefficients towards zero. I try to overcome these possible sources of bias by making use of a 2SLS approach.

My instrument uses the tendency of new female migrants to follow in the footsteps of earlier male migrants from the same sending region. As much research has shown, migrant networks are an important consideration in the location choices of prospective migrants since these networks assist with the job search process and assimilation into the new environment (Munshi, 2003). The instrument uses the 1985 distribution of migrants from a given reserve to allocate the new wave of female migrants from that same area in the post-Apartheid period. Formally, the instrument for the change in female migration in district d during time period t can be written as:

[21] Note that the total but not the group-specific inflows of migrants can be considered quasi-random. My specification has the added advantage that identification of β does not require the pre-allocation of migrants to skill groups based on their observable characteristics, thus avoiding the problem of misclassification when that arises when such observable characteristics are used to assign migrants into skill groups in which they do not compete with natives because of migrant downgrading (Dustmann et al., 2016).

[22] I have also run all my regressions without any weights and the results are substantially unchanged.

$$\sum_r \frac{M^{male}_{rd1985}}{M^{male}_{r1985}} \times M^{fem}_{rt} \quad (1.3)$$

where r are areas that were classified as reserves under Apartheid as in the 1985 census; $\frac{M^{male}_{rd1985}}{M^{male}_{r1985}}$ represents the proportion of all migrants from area r included in the 1985 census who were settled in receiving district d; and M^{fem}_{rt} represents the total number of female migrants from an area previously classified as a reserve r at time t in the contemporary period. The instrument is therefore a weighted average of the female migration rates from former homeland areas (the 'shift') with weights that depend on the distribution of earlier male migrants in 1985 (the 'shares').

Table 1.3: First stage regression: impact of historical male migration on contemporary female migrant shares

	(1)	(2)
Instrument	0.512***	0.511***
	(0.070)	(0.071)
District FE	Y	Y
Year FE	Y	Y
Individual controls	N	Y
First stage K-P F-stat	52.1	52.4
Shea R^2	0.24	0.24
Clusters	216	216
Observations	71,867	71,867

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects. Column 2 includes individual controls for age, education and race.

The key identifying assumption is that unobserved factors determining the distribution of migrants before the end of migration restrictions are unrelated to contemporary economic shocks (e.g. Goldsmith-Pinkham et al. (2018)). This assumption is likely to hold for several reasons (for more information on these points, see Section 2):

1. *Black people who were born in homelands could not freely choose their migration paths to non-homeland areas.* During the Apartheid era, due to a centralised allocation process for black migrants and strict pass laws, black people in homelands could not choose where to move to. Their migration options were severely constrained: they could either accept jobs to work in specific industries in specific places or remain in homelands.

2. *Centralised migrant allocation decisions were to a large extent made on political grounds.* It was South African government policy that each reserve would send the majority of its workers to one specific industry. Furthermore, decentralisation efforts brought black migrants to white areas near homelands that, were it not for government incentives, would have had much less industry, and prevented many migrants from moving to metropolitan areas because of the restrictions placed on black labour intensive industries there.

3. *Many locations where there was substantial migrant labour demand during Apartheid did not have have substantial labour demand afterwards.* The end of Apartheid constituted a structural break in the economy of South Africa, and the distribution of economic activity changed substantially between the Apartheid and post-Apartheid periods. 'Border industries' that were established around homelands were dismantled towards the end of Apartheid. With the opening up of the economy, extractive and manufacturing industries waned and so the economies of locations specialised in these industries suffered.

4. *Migration pull factors are/were substantially different for men and for women.* By instrumenting female migration with historical male migration, I also add another layer of exogeneity to my historical instrument and guard against the possibility of serial correlation in migration flows (Barone and Mocetti, 2011). Female migrants in the period under study were employed in a very different set of occupations and industries as compared to male migrants during Apartheid.

If migrant allocation during Apartheid were completely orthogonal to labour market

Table 1.4: The effects of female internal migration on the labour market outcomes of all non-migrants (male & female) in receiving areas: OLS & 2SLS

	Pr(emp)		Ln(wages)	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Ln(Migrant share)	-0.010	-0.037	0.125***	0.123
	(0.019)	(0.038)	(0.043)	(0.100)
Individual controls	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
K-P first stage F-stat		52.8		45.4
N	62,151	61,256	27,583	27,409

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects, and individual controls for age, education, sex and race.

conditions (at least those created by market forces) in receiving areas, then point (2) above would be sufficient to make the instrument valid. Of course, migrant allocation did respond to labour market conditions to some extent, and so (3) is important too. Share exogeneity is a sufficient condition for identification (Borusyak et al., 2018; Jaeger et al., 2018), so it is not necessary to demonstrate shift exogeneity in my setting. However, for the effects of female migration to be separately identified from the effects of male migration, it is important that pull/push factors behind migration differ for men and women in the post-Apartheid period. As shown in Section 3, the distribution of female migrants across economic sectors is very different from that of male migrants in South Africa. I also provided evidence from a nationally-representative migration study from 2001 showing that the ranking of possible reasons for moving across district borders in terms of importance was very different for men as compared to women.

The power of the instrument depends on how strong was the tendency of female migrants from reserves in the post-Apartheid era to cluster in the enclaves generated by Apartheid migration allocation processes. Table 1.3 reports first-stage regression results. Column 2 reports results including individual controls. The correlation of the change in the instrument with the actual change in female migrant share in the contemporary period exhibits a very significant coefficient and a large F-stat of 52.4, showing that the instrument is strong.

1.5 Results

I first examine the effects of female internal migration on the total population of non-migrants including both men and women. Table 1.4 shows the results of my analysis using labour force data for non-migrant characteristics and labour market outcomes.[23] The OLS results show that female migrant share has an almost-zero and non-significant association with the employment of non-migrants and a significant positive association with monthly wages. With the IV, the coefficient for the employment effect remains non-significant though the (negative) coefficient is slightly larger. Looking at the effect on wages, the positive coefficient on the female migrant share variable remains similar in magnitude but becomes non-significant with the inclusion of the instrument. For the estimations in this table, and for all those below, the first-stage F-statistics are always well above 10 (Stock and Yogo, 2005). The non-significant effects on the overall non-migrant population in receiving districts are in line with much of the existing migration literature (e.g. Card (1990)).

In Table 1.5, I split the sample into men and women. With the IV, the employment effects of female migration are non-significant though the negative coefficient on the variable of interest is larger for male than female non-migrants. With the IV, the effect on wages of non-migrants is non-significant for both men and women.

Table 1.6 examines effects on female non-migrants of different skill groups.[24] While the hours worked variable is likely measured with some error, I include it here to make my analysis more comparable to the earlier international studies examining

[23] All South African LFSs seem to survey more women than men, which is why the number of observations for women is higher than those for men in the tables in this section (of course survey weights will correct for this).

[24] When replicating this table for male non-migrants I find no significant effects. I thus choose to focus the rest of my analysis on female non-migrants though in Section 7 I present results for high-skilled non-migrant men in a discussion on channels.

Table 1.5: The effects of female internal migration on the labour market outcomes of male & female non-migrants in receiving areas: OLS & 2SLS

	Male				**Female**			
	Pr(emp)		Ln(wages)		Pr(emp)		Ln(wages)	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Ln(Migrant share)	-0.046**	-0.046	0.123***	0.143	0.017	-0.031	0.107*	-0.016
	(0.023)	(0.042)	(0.046)	(0.093)	(0.023)	(0.040)	(0.057)	(0.122)
Individual controls	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
K-P first stage F-stat		56.1		57.0		59.3		35.8
N	28,966	28,637	16,341	16,248	33,184	32,618	11,238	11,158

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects, and individual controls for age, education and race.

the effects of female migration. Turning to the 2SLS estimations in Panel 2, I find a strong positive and significant effect on the employment of high-skilled women. A 10% increase in female migrant share is associated with a 1.75 p.p. increase in the employment of high-skilled women. I also find a strong effect on the intensive margin: a 10% increase in female migrant share is associated with a 1.63 pp increase in the hours worked of high-skilled women. Focusing on the effects on low-skilled female non-migrants, I find that a 10% increase in female migrant share is associated with a 1.6 p.p reduction in the employment of this subgroup. This suggests there is some competition at the bottom of the economic ladder (Card, 2001). In Panel 3, I run the same 2SLS regressions as above on the sample of black and coloured (i.e. historically disadvantaged) female non-migrants. The effect on employment for the high-skilled section of this sub-group is even larger than for high-skilled female non-migrants in general, though the effect on hours worked is non-significant and smaller. Clearly, the positive effects of female migration are not limited to the white population. Again all IV estimations in Table 1.6 yield non-significant wage effects with very large standard errors.

There are several possible reasons for why I do not find a significant negative effect on wages in the results above. First, wages in South Africa may have been downwardly rigid due to high reservations wages (due in part to a generous social welfare system)

Table 1.6: The effects of female internal migration on the labour market outcomes of female non-migrants of different skill groups in receiving areas: OLS & 2SLS

	High-skilled			**Medium-skilled**			**Low-skilled**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable:	Pr(emp)	Ln(wages)	Ln(hours)	Pr(emp)	Ln(wages)	Ln(hours)	Pr(emp)	Ln(wages)	Ln(hours)
Panel 1: OLS estimates									
Ln(migrant share)	0.113***	0.101	0.007	0.030	0.134	-0.024	-0.040	0.094	-0.038
	(0.040)	(0.066)	(0.042)	(0.046)	(0.155)	(0.047)	(0.027)	(0.077)	(0.032)
Individual controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	8,108	2,950	2,773	6,488	2,001	1,878	18,567	6,209	5,634
Panel 2: 2SLS estimates									
Ln(migrant share)	0.175***	-0.123	0.163**	0.012	0.180	-0.045	-0.160***	0.057	-0.103
	(0.056)	(0.177)	(0.069)	(0.074)	(0.241)	(0.115)	(0.053)	(0.166)	(0.078)
Individual controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
K-P first stage F-stat	30.4	20.2	20.9	41.3	26.9	26.8	70.4	38.8	39.9
N	8,030	2,947	2,771	6,420	1,996	1,874	18,148	6,147	5,581
Panel 3: 2SLS estimates for the black & coloured population									
Ln(migrant share)	0.253***	-0.233	0.070	-0.025	0.298	-0.059	-0.160***	0.067	-0.104
	(0.085)	(0.161)	(0.077)	(0.098)	(0.270)	(0.104)	(0.054)	(0.167)	(0.079)
Individual controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
K-P first stage F-stat	37.8	22.7	28.5	46.6	28.9	27.8	71.7	38.5	39.5
N	4,775	1,605	1,499	5,150	1,643	1,541	17,294	5,990	5,433

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects, and individual controls for age, education and race.

and a powerful trade union movement which keeps wages relatively high in the formal sector.[25] Another possibility is that at the end of Apartheid, the South African economy had a high potential for growth. During the period of restricted migration, firms in cities were short of labour. An improved allocation of factors of production after Apartheid may have boosted productivity, thereby offsetting the effects of migration on wages.

1.6 Robustness checks

1.6.1 Province-year fixed effects

As a robustness check, Panel 1 of Table 1.7 looks at the effects of female migration on high-skilled and low-skilled female non-migrants but includes an additional set of fixed effects for province-years to capture possible unobserved time-varying factors

[25] Boustan et al. (2010) also find no effect on non-migrant wages of internal migration in the US in the early 1900s and put this down to wage rigidities.

at the provincial level that could potentially bias results. Results are similar in magnitude and retain statistical significance though the precision of the estimates is sometimes reduced as this specification sacrifices a lot of cross-province variation. The strength of the instrument is reduced with the additional fixed effects but is still well above conventional thresholds.

Table 1.7: The effects of female internal migration on the labour market outcomes of female non-migrants of different skill groups in receiving areas: robustness checks for 2SLS estimates

	High-skilled			**Low-skilled**		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Pr(emp)	Ln(wages)	Ln(hours)	Pr(emp)	Ln(wages)	Ln(hours)
Panel 1: 2SLS estimates with province-year FEs						
Ln(migrant share)	0.159**	-0.058	0.260***	-0.170***	-0.025	-0.099
	(0.074)	(0.181)	(0.091)	(0.064)	(0.199)	(0.090)
Individual controls	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
K-P first stage F-stat	20.3	16.6	16.4	41.3	28.3	27.0
N	8,030	2,947	2,771	18,148	6,147	5,581
Panel 2: 2SLS estimates with (instrumented) control for male migrant share						
Ln(migrant share)	0.339***	-0.597	0.190	-0.193*	0.333	-0.351*
	(0.114)	(0.509)	(0.260)	(0.112)	(0.307)	(0.192)
Ln(male migrant share)	-0.168	0.476	-0.026	0.039	-0.313	0.261*
	(0.109)	(0.441)	(0.218)	(0.105)	(0.259)	(0.154)
Individual controls	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
K-P first stage F-stat	22.7	19.8	12.7	15.5	22.0	18.2
A-P F-stat for add. instr.	14.8	26.9	16.2	43.5	57.5	50.0
N	8,029	2,947	2,771	18,104	6,122	5,557
Panel 3: 2SLS estimates using sample without data from 1996						
Ln(migrant share)	0.150**	-0.044	0.158**	-0.146***	0.168	-0.105
	(0.065)	(0.164)	(0.078)	(0.055)	(0.154)	(0.085)
Individual controls	Y	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
K-P first stage F-stat	27.7	19.0	19.2	16.4	28.7	30.7
N	6,762	2,437	2,310	14,656	4,878	4,549

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects, and individual controls for age, education and race.

1.6.2 Controlling for male migrant share

One concern could be that my results are driven by male migration from former homeland areas, which could be correlated with female migration. Panel 2 of Table 1.7 replicates the main 2SLS regressions from Table 1.6 but includes an additional control for male migrant share from former homeland areas. This control is instrumented with a shift-share instrument analogous to the one used before but using male migration from former homeland areas as the shift part of the instrument. This control is non-significant in most regressions. For the employment regressions, the coefficients on female migrant share increase in magnitude. The significant effect on hours worked for high-skilled female non-migrants becomes non-significant though the coefficient on the variable of interest remains similar in magnitude.

Table 1.8: The effects of female migration on the size of the non-migrant labour force: OLS & 2SLS

	Ln(non-migrant LF share)			
	Male		**Female**	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Ln(migrant share)	0.001	-0.007	0.001	0.000
	(0.004)	(0.008)	(0.003)	(0.006)
District FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
K-P first stage F-stat		35.2		33.0
Observations	438	418	438	418

Analysis at district level using Census 1985, 1996 & 2001 data. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects.

1.6.3 Using different data

In my main analysis, I have combined LFS waves in 1996 and 1997 to merge with migration data from the 1996 census. In Panel 3 of Table 1.7, I check that my main results hold if I drop the LFS 1996 wave and use only the 1997 wave. The results for the employment and hours worked regressions are virtually unchanged, though, as to be expected, the effects are slightly less precisely estimated.

1.6.4 Potential serial correlation

In recent work, Jaeger et al. (2018) point out that if the spatial distribution of migrant inflows is stable over time, historical migration instruments are likely correlated with ongoing responses to previous shocks. Focusing on the United States in the post 1970s period, they show that the shift-share instrument might conflate the long and the short run effects of immigration because there is a very high serial correlation (between 0.95 and 0.99) in the sending-destination migration patterns over time. Directly investigating the correlation between historical and post-Apartheid migration shares for each district in my sample, I find that this number is in the order of 0.56-0.57. According to Jaeger et al. (2018), this value is sufficiently low for the shift-share instrument to be unlikely to conflate the long and the short run effects of migration.[26] These results also provide some evidence for the exclusion restriction for my instrument.[27]

1.6.5 Potential relocation

An important source of potential bias in my paper is that workers may relocate from high-migration to low-migration receiving districts. If this would happen, it would bias upwards my estimates of the labour market effects of female migration (Borgas, 1999).

To determine the size of the relocation response - if any - I follow Card (2007) and Peri and Sparber (2011) and estimate the following regression:

$$q_{dt} = \beta f_{dt} + \gamma_t + \delta_d + \epsilon_{dt} \tag{1.4}$$

[26] This correlation coefficient is very low for the literature. Tabellini (2020) constructs a shift-share instrument from immigration data in the US in 1900 to predict immigration between 1910 and 1930. He argues that since the correlation in immigration flows between sending and receiving regions in the two periods was 0.7, there is no reason to be concerned about serial correlation.

[27] Ideally, I would test to see whether pre-trends in employment and hours worked in the period before 1985 and the end of migrations restrictions were uncorrelated with female migration flows in the contemporary period but, unfortunately, I do not have good data on labour market outcomes prior to 1985.

The non-migrant share of the labour force is represented by q_{dt}, that is:

$$q_{dt} = \frac{N_{dt}^{tot}}{p_{dt}^{gen}} \tag{1.5}$$

where the share of non-migrants is the share among the labour force (also including migrants) who have not moved between districts in the last 5 years. Since this regression is estimated with district fixed effects, it is a very demanding in terms of controlling for pre-determined trends of non-migrant growth (Peri and Sparber, 2011). In this specification, the coefficient on female migrant share indicates the elasticity between female migrant concentration in districts and the size of the non-migrant labour force (i.e. a coefficient of zero on female migrant share indicates no spillovers between districts).

Table 1.9: The effects of female internal migration on the labour market outcomes of high-skilled male non-migrants in receiving areas: OLS & 2SLS

	Pr(emp)		Ln(hours)	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Ln(Migrant share)	-0.026	-0.039	-0.009	-0.033
	(0.024)	(0.045)	(0.033)	(0.044)
Individual controls	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
K-P first stage F-stat		30.5		45.5
N	7,847	7,797	3,658	3,643

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects, and individual controls for age, education, sex and race.

In Table 1.8, I show results of the effects of female migration on the size of the male and female labour force in receiving districts. The OLS results show that female migrant share has a non-significant almost-zero association with the non-migrant labour force for both men and women with a point estimate of 0.001 in both regressions. With the inclusion of the IV, the point estimate becomes negative for male non-migrants but remains close to zero and non-significant. For non-migrant women, the point estimate is exactly zero but non-significant. In summary, I find no evidence that female migration led to the displacement of the labour force already present in receiving districts.

1.7 Channels

I have found that female migration results in an increase in the labour supply of high-skilled women. Given that many female migrants work as domestic helpers, one hypothesis is that migrants are substituting for high-skilled women in domestic activities (Barone and Mocetti, 2011; Cortes and Tessada, 2011), allowing the latter to work more and spend more time at work. However, the impact of female migration might also go through other channels. For example, migrants might be complementary to high-skilled non-migrant women in the productive sector, thus increasing the possibilities for employment for the latter. The increase in the number of migrant women might, for example, expand the need for coordination and management activities in offices and factories (Barone and Mocetti, 2011). To investigate this further I perform a couple of checks below.

In South Africa there is a very unequal distribution of domestic activities between men and women. According to the nationally-representative Time Use Survey in 2000, women spent on average three times as much time as men on housework (including cleaning, cooking, childcare, etc.) (Budlender et al., 2001). It is unlikely then that an increase in supply of domestic workers would affect high-skilled men as much high-skilled women. On the other hand, if the impact of female migration were operating through interactions in the productive sector, a positive effect on men's employment and hours worked would be expected. Table 1.9 examines the effects of female migration on the employment and hours worked of high-skilled male non-migrants. With the IV, I find non-significant effects on employment and hours worked which are negative in sign and small in magnitude. These results provide strong evidence for the channel involving substitution in domestic work.

Table 1.10: The effects of female migration on the hourly wages of female domestic workers relative to female workers in other industries: OLS & 2SLS

	Ln(wages)		
	(1)	(2)	(3)
	OLS	2SLS	2SLS
Ln(migrant share)	0.102**	0.097	0.109
	(0.051)	(0.104)	(0.116)
Ln(migrant share)#domestic work	-0.015	-0.113*	-0.114*
	(0.027)	(0.065)	(0.066)
Domestic work	-0.744***	-1.103***	-1.106***
	(0.115)	(0.255)	(0.259)
Individual FE	Y	Y	Y
District FE	Y	Y	Y
Year FE	Y	Y	Y
Province-year FE	N	N	Y
K-P first stage F-stat		5.4	5.3
A-P F-stat for add. instr.		11.1	10.7
Observations	10,063	9,994	9,994

Analysis at individual level using labour force survey data (1996 OHS, 1997 OHS & Sept 2003 LFS) and census data (1985, 1996, 2001) on migration. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects, and individual controls for age, education and race.

To further test my hypothesis, in Table 1.10, I examine the effects of female migration on the wages of female domestic workers relative to female workers in other industries. Domestic workers are identified on the basis of industry and occupation codes.[28] I include district fixed effects and year fixed effects. With the IV, I find that a 10% increase in female migrant share reduces domestic worker wages by 1.1% relative to wages in other sectors (significant at the 10% level). The result is robust to the inclusion of region-year fixed effects.

1.8 Conclusion

This paper examines the effects of female internal migration on the labour market outcomes of non-migrants in South Africa, exploiting substantial variation in the

[28]I use the 'domestic work' industry identifier in the Post-Apartheid Labour Market Series, which was created on the basis of detailed industry and occupation codes to ensure consistency across waves (PALMS, 2020).

number of female migrants received by districts in the post-Apartheid period. I argue that South Africa is one of the very few settings where an instrument using historical migration pathways can plausibly capture exogenous variation in migration flows. Using the natural experimental setting created by Apartheid era policies, I show that female internal migration in South Africa increased the employment and hours worked of highly-educated women in the post-Apartheid period. I demonstrate that this was due to the fact that female migrants found employment as domestic workers, particularly in well-educated households, thereby freeing women from household labour and allowing them to pursue employment or increase their work hours. However, it was not the case that all subsections of the non-migrant population benefited from this female in-migration and I have shown evidence that low-skilled women experienced reduced employment. This suggests that female migrants (who were relatively well-educated though lacking in experience) may have substituted for low-skilled non-migrants in production.

I do not find any evidence that labour reallocation biased my results. I also have shown that my results are robust to including region-year fixed effects and an instrumented control for male migration from homeland districts.

From a policy perspective, several effects of female migration are worth considering. Female migrants are likely to improve the welfare of high-skilled non-migrant women by providing affordable and flexible substitutes to household production (my analysis also shows that historically disadvantaged groups also benefit from this). This paper has also shown that female migration leads to a moderate decline in employment for low-skilled women. With access to longitudinal data, future research could examine the employment trajectories of low-skilled non-migrants who lose out on employment to migrant women.

Bibliography

Ardington, C., Lam, D., Leibbrandt, M., and Welch, M. (2006). The Sensitivity to Key Data Imputations of Recent Estimates of Income Poverty and Inequality in South Africa. *Economic Modelling*, 23(5): 822–835.

Aron, J., Kahn, B., and Kingdon, G. (2009). South African Economic Policy Under Democracy: Overview and Prospects. In *South African Economic Policy under Democracy*, pp. 1–27.

Bakker, J. D., Parsons, C., and Rauch, F. (2020). Migration and urbanization in post-apartheid South Africa. *The World Bank Economic Review*, 34(2): 509–532. Available at: https://doi.org/10.1093/wber/lhy030.

Barone, G. and Mocetti, S. (2011). With a Little Help from Abroad: the Effect of Low-Skilled Immigration on the Female Labour Supply. *Labour Economics*, 18(5): 664–675.

Best, A. C. (1971). South Africa's Border Industries: The Tswana Example. *Annals of the Association of American Geographers*, 61(2): 329–343.

Biavaschi, C., Facchini, G., Mayda, A. M., and Mendola, M. (2018). South–south migration and the labor market: Evidence from South Africa. *Journal of Economic Geography*, 18(4): 823–853.

Borgas, G. J. (1999). The Economic Analysis of Immigration. In *Handbook of Labor Economics*, volume 3, pp. 1697–1760. Elsevier, Amsterdam.

Borusyak, K., Hull, P., and Jaravel, X. (2018). Quasi-experimental shift-share research designs. *NBER Working Paper 24997*.

Bosker, M. and Krugell, W. (2008). Regional income evolution in South Africa after Apartheid. *Journal of Regional Science*, 48(3): 493–523.

Boustan, L. P., Fishback, P. V., and Kantor, S. (2010). The effect of internal migration on local labor markets: American cities during the Great Depression. *Journal of Labor Economics*, 28(4): 719–746.

Branson, N. and Wittenberg, M. (2011). Re–weighting South African National Household Survey Data to create a consistent series over time: A cross-entropy estimation approach. *SALDRU Working Paper No. 54*.

Broussard, N. H. (2017). Immigration and the labor market outcomes of natives in developing countries: A case study of South Africa. *Economic Development and Cultural Change*, 65(3): 389–424.

Budlender, D., Chobokoane, N., and Mpetsheni, Y. (2001). *A survey of time use: How South African women and men spend their time*. Pretoria: Statistics South Africa.

Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *Industrial and Labor Relations Review*, 43(2): 245–257.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics*, 19(1): 22–64.

Card, D. (2007). How Immigration affects U.S. Cities. *Centre for Research and Analysis of Migration Discussion Paper 11/07*.

Christopher, A. J. (2001). *The Atlas of Changing South Africa*. London and New York: Routledge.

Combes, P.-P., Démurger, S., and Li, S. (2015). Migration externalities in Chinese cities. *European Economic Review*, 76: 152–167.

Cortes, P. and Tessada, J. (2011). Low-Skilled Immigration and the Labor Supply of Highly Skilled Women. *American Economic Journal: Applied Economics*, 3(3): 88–123.

Davis, B. and Winters, P. (2001). Gender, Networks and Mexico-US Migration. *Journal of Development Studies*, 38(2): 1–26.

Desmond, C. (1971). *The Discarded People.* London: Penguin Press.

Dewar, D., Todes, A., and Watson, V. (1984). *Industrial decentralization policy as a mechanism for regional development in South Africa: Its premises and record.* Occasional Paper No 17, University of Cape Town Urban Problems Unit.

Dustmann, C., Schönberg, U., and Stuhler, J. (2016). The Impact of Immigration: Why Do Studies Reach Such Different Results? *Journal of Economic Perspectives*, 30(4): 31–56.

Enchautegui, M. E. (1997). Welfare Payments and Other Economic Determinants of Female Migration. *Journal of Labor Economics*, 15(3): 529–554.

Feinstein, C. H. et al. (2005). *An Economic History of South Africa: Conquest, Discrimination, and Development.* Cambridge, Cambridge University Press.

Geyer, H. S. (1989). Apartheid in South Africa and industrial deconcentration in the PWV area. *Planning Perspective*, 4(3): 251–269.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik Instruments: What, When, Why, and How. *NBER Working Paper No. 24408.*

Healy-Clancy, M. (2017). Women & Apartheid. In *Oxford Research Encyclopedia.* DOI: 10.1093/acrefore/9780190277734.013.184.

Hepple, A. (1969). *Workers Under Apartheid.* International Defence and Aid Fund, London.

Horrell, M. (1973). *The African Homelands of South Africa.* Johannesburg: South African Institute of Race Relations.

ILO (1982). Resolution Concerning Statistics of the Economically Active Population, Employment, Unemployment and Underemployment. In *Thirteenth International Conference of Labour Statisticians.*

Imbs, J. (2013). The Premature De–Industrialization of South Africa. *The Industrial Policy Revolution II: Africa in the 21st Century, Palgrave McMillan*, pp. 529–540.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. *NBER Working Paper Series*.

Kleemans, M. and Magruder, J. (2018). Labour market responses to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*, 128(613): 2032–2065.

Lall, S. V., Selod, H., and Shalizi, Z. (2006). Rural-urban migration in developing countries: A survey of theoretical predictions and empirical findings. *World Bank Policy Research Working Paper 3915*. Washington, DC.

Leys, R. (1975). South African Gold Mining in 1974: 'The Gold of Migrant Labour'. *African Affairs*, 74(295): 196–208.

Lombard, J. A. and Stadler, J. J. (1980). *The Role of Mining in the South African Economy: A Study prepared for the Chamber of Mines of South Africa*. Bureau of Economic Policy and Analysis, University of Pretoria.

MacCarthy, C. (1982). Industrial decentralization: Reflections on the new initiatives. *South African Journal of Economics*, 50(3): 238–252.

Magruder, J. R. (2010). Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa. *American Economic Journal: Applied Economics*, 2(1): 62–85.

Magruder, J. R. (2012). High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa. *American Economic Journal: Applied Economics*, 4(3): 138–66.

Mariotti, M. (2015). Fathers' Employment and Sons' Stature: The Long-Run Effects of a Positive Regional Employment Shock in South Africa's Mining Industry. *Economic Development and Cultural Change*, 63(3): 485–514.

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the US Labor Market. *Quarterly Journal of Economics*, 118(2): 549–599.

Naidoo, N., Leibbrandt, M., and Dorrington, R. (2008). Magnitudes, Personal Characteristics and Activities of Eastern Cape Migrants: A Comparison with Other Migrants and with Non-migrants using Data from the 1996 and 2001 Censuses. *Southern African Journal of Demography*, pp. 3–38.

Ogura, M. (1996). Urbanization and Apartheid in South Africa: Influx Controls and Their Abolition. *The Developing Economies*, 34(4): 402–423.

Oosthuizen, M. and Naidoo, P. (2004). Internal Migration to the Gauteng Province. *Development Policy Research Unit Working Paper 04/88.* University of Cape Town.

PALMS (2020). South Africa - Post Apartheid Labour Market Series 1993-2019. https://doi.org/10.25828/gtr1-8r20.

Peri, G. and Sparber, C. (2011). Assessing inherent model bias: An application to native displacement in response to immigration. *Journal of Urban Economics*, 69(1): 82–91.

Pfeiffer, L., Richter, S., Fletcher, P., and Taylor, J. E. (2007). *Gender in Economic Research on International Migration and Its Impacts: A Critical Review.* In Morrison, A., Schiff, M. & Sjöblom, M. (eds.), International Migration. New York: Palgrave Press.

Posel, D., Casale, D., and Vermaak, C. (2014). Job Search and the Measurement of Unemployment in South Africa. *South African Journal of Economics*, 82(1): 66–80.

Rogan, M., Lebani, L., and Nzimande, N. (2008). Internal Migration and Poverty in Kwazulu-Natal: Findings from Censuses, Labour Force Surveys and Panel Data. *South African Labour and Development Research Unit Working Paper No 30.* Cape Town, SALDRU, University of Cape Town.

Rogerson, C. M. (1998). Restructuring the post-Aapartheid space economy. *Regional Studies*, 32(2): 187.

Savage, M. (1986). The Imposition of Pass Laws on the African Population in South Africa 1916-1984. *African Affairs*, 85(339): 181–205.

Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression, Ch. 5. in J.H. Stock and D.W.K. Andrews (eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg. Technical report.

Strobl, E. and Valfort, M.-A. (2015). The Effect of Weather-Induced Migration on Local Labor Markets. Evidence from Uganda. *The World Bank Economic Review*, 29(2): 385–412.

Tabellini, M. (2020). Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration. *Review of Economic Studies*, 87(1): 454–486.

Turok, I. (2012). *Urbanisation and development in South Africa: Economic imperatives, spatial distortions and strategic responses*. London, England: International Institute for Environment and Development.

Von Fintel, D. and Moses, E. (2017). Migration and gender in South Africa: Following bright lights and the fortunes of others? *Regional Science Policy & Practice*, 9(4): 251–268.

Wellings, P. and Black, A. (1986). Industrial Decentralization under Apartheid: The Relocation of Industry to the South African Periphery. *World Development*, 14(1): 1 – 38.

Wentzel, M., Viljoen, J., and Kok, P. (2006). *Contemporary South African migration patterns and intentions, In P. Kok, D. Gelderblom, J. Oucho, & J. van Zyl (eds.),*. Migration in South and Southern Africa. Dynamics and determinants. Cape Town: HSRC Press.

Wilse-Samson, L. H. (2013). *Structural change and democratization: Evidence from rural Apartheid*. Mimeo, Columbia University, New York City. Available online: http://www. columbia.edu/ lhw2110/wilsesamsonapartheid.pdf.

Chapter 2

How does minimum wage legislation influence the labour market impacts of productivity shocks?

2.1 Introduction

Developing countries are particularly vulnerable to weather shocks affecting agricultural productivity, due to a large share of their labour forces working in this sector and the important role that agricultural production plays for the economic lives of the poor. With climate change increasing the frequency of these shocks, numerous studies have tried to measure their impacts on labour markets (Henderson et al., 2017; Jayachandran, 2006; Jessoe et al., 2018; Townsend, 1994). Separately, there has been much recent interest in the effects of minimum wage legislation in developing country labour markets (Bhorat et al., 2014; Dinkelman and Ranchhod, 2012; Gindling and Terrell, 2007; Neumark et al., 2006). It has been theorised that these laws may have stronger effects than in developed countries due in part to larger instituted wage increases for low-skilled workers and to weaker employment protection laws. However, these two strands of research have not been brought together, and there is a missing perspective on how minimum wage legislation could affect resilience to productivity shocks.

Agriculture has historically accounted for a substantial share of employment of low-skilled workers in South Africa. However, there has been a trend towards commercialisation and mechanisation in the last three decades, during which time over one million agricultural jobs have been shed (Liebenberg and Johann, 2013). Agricultural employment is also highly variable due to currency and, in particular, weather fluctuations (BFAP, 2016). There is anecdotal and descriptive evidence of droughts causing massive destruction in the agricultural economy (BFAP, 2016; Vogel and Drummond, 1993). However, to date, no studies of this phenomenon have been undertaken.

In 2003 a national agricultural minimum wage was adopted in South Africa. This was set at the 70$^{\text{th}}$ percentile of the prevailing wage distribution, leading to a substantial increase in the median wage. Several papers have studied the impacts of this law (Bhorat et al., 2014; Conradie, 2004; Murray and Van Walbeek, 2007), though most have been small-scale case studies. In the single national study, (Bhorat et al., 2014) find evidence of a positive effect on the income of agricultural workers but a negative effect on their employment (with no significant effect on hours worked). The authors also note the possibility of an increase in the employment of casual workers in place of permanent employees.

In this paper I focus on the interaction between this law and the effects of weather-induced productivity shocks on agricultural labour market outcomes in the short term. It is not *a priori* clear how the introduction of a minimum wage, assuming this results in a substantial increase in agricultural wages, should affect the relationship between productivity shocks and agricultural employment. One possibility is that if the minimum wage causes an immediate scaling back of employment on farms and perhaps the substitution of low-wage labour with capital, a consolidated agricultural labour force might become more resilient to negative shocks in the post-law period. On the other hand, by making the wage bill less affordable for employers and/or reducing the flexibility of employers to adjust wages to shocks (Franklin and

Labonne, 2019), the minimum wage could increase the susceptibility of agricultural employment to negative shocks.

My empirical approach makes use of a difference-in-differences strategy to statistically test for the labour market effects of the interaction between the agricultural minimum wage law and weather shocks. I exploit before-after variation in the application of the law and pre-existing cross-sectional variation related to the intensity of the law, together with weather variation across time and space. I match a high resolution moisture index called the Standardised Precipitation Evapotranspiration Index (SPEI) - which considers the joint effects of precipitation, potential evaporation and temperature, and when calculated on short time-scales is closely related to soil moisture - to district boundaries and compute rainy season SPEI Z-scores based on long-run district patterns. I use 13 waves of a biannual labour force survey (LFS) from September 2001 to September 2007 to capture reported employment, wages and hours of work over time at a relatively high frequency, and also match these data to districts. I then examine the relationship between individual labour market outcomes in a particular district and wave and localised SPEI Z-scores in the previous rainy season controlling for individual characteristics, district and time fixed effects and district-specific linear time trends. I focus on these short-run effects since the data are well suited to this and because doing so means there is little time for workers to relocate across districts (Franklin and Labonne, 2019).

Whereas weather shocks have no significant effect on agricultural employment in the pre-law period, in the post-law period, holding the pre-law wage gap at its mean level, a one standard deviation increase (decrease) in rainy season SPEI relative to the local mean results in a 7.1% increase (decrease) in agricultural employment relative to its mean level among the total labour force within six months. When I focus only on the effects of below-average SPEI (recoding above-average SPEI values as zero) on agricultural employment, it is clear that the effects of negative weather shocks are driving the effects that I find when I look at the interaction between my continuous weather measure and the minimum wage law. This shows that the minimum wage substantially weakens the resilience of agricultural employment to reduced rainy season moisture in the short term.[1] I find evidence that after negative shocks occur, some of the displaced agricultural workers transition into unemployment while others transition into employment in other industries. Negative weather shocks also appear to lead to a small increase in hours worked in agriculture wage labour after the minimum wage is introduced (there is no significant effect before), most likely because employers are making their remaining workers work more to compensate for the retrenchments. I do not detect substantive effects of weather shocks on wages either before or after the minimum wage, although when I only focus on negative weather shocks, I find some evidence of a small negative effect on wages in some areas in the pre-period. Lastly, once the minimum wage is in place, negative weather shocks appear to lead to the casualisation of the agricultural labour force. The overall picture suggests then that the agricultural minimum wage, by substantially reducing the affordability of the wage bill of employers in the industry and therefore making it more difficult for them to absorb shocks, leads employers to retrench formally employed workers in the wake of negative productivity shocks.

This paper contributes to the literature on economic and other impacts of negative environmental shocks in developing countries (Dell et al., 2014; Henderson et al., 2017; Jessoe et al., 2018). In South Africa, studies have looked at the effects of droughts on long-term health outcomes (Dinkelman, 2017) and on internal migration over the long run (Mastrorillo et al., 2016), but surprisingly there has been no attempt to measure the economic impacts of droughts (or dry shocks). More specifically, my paper contributes to work on the impact of agricultural productivity shocks on local economic activity. Some recent papers have focused on short term impacts of these shocks, looking at the effect of agricultural productivity shock-induced reallocation on industrial outcomes in the Indian context (Colmer, 2018; Santangelo, 2019), whereas older work has tended to focus on long-run changes in agricultural productivity (Bustos et al., 2016; Henderson et al., 2017; Hornbeck, 2012).

[1] Below-normal rainy season SPEI (moisture) can result in droughts depending on the size of the deviation, the institutional capacity and the topography of the district (Sekhri and Storeygard, 2014).

My paper also contributes to the literature on the labour market effects of minimum wages (Card and Krueger, 1994). It adds to the relatively small literature on the effects of minimum wages in developing countries (Bhorat et al., 2014; Dinkelman and Ranchhod, 2012; Gindling and Terrell, 2007; Neumark et al., 2006). For several reasons effects are expected to be different in developing countries. Typically the size of the wage increase is larger in developing countries and there are more low-income workers who stand to be affected. These workers are likely to have much weaker employment protection and they may have limited alternatives in other sectors. Laws may not apply to workers in the informal sector and even in the formal sector, compliance and enforcement of laws might be expected to be more relaxed in developing countries (Neumark et al., 2007). My paper highlights an unexplored and important channel for minimum wage laws having a negative impact on employment in developing countries - that is, they may reduce the ability of employers to absorb economic shocks.

Lastly, my paper also contributes to the literature on labour market policy and local productivity shocks in developing countries. Adhvaryu et al. (2013) and Chaurey (2015) examine employment responses to local demand shocks using state-level variation in firing costs in India. Colmer (2018) uses the same variation in labour regulations to test the extent to which labour reallocation mitigates the consequences of temperature-driven reductions in the demand for agricultural labour. Santangelo (2019) examines whether the introduction of a large-scale rural workfare programme affects the response of local economies to agricultural productivity. No papers have, however, focused on the influence of minimum wage legislation on the effects of productivity shocks.

The paper is structured as follows: Section 2 describes the agricultural industry in South Africa and the characteristics of the new law introduced in 2003. After describing the data and presenting summary statistics in Section 3, I discuss results in Section 4. In this section, I examine the wage, earnings, employment and hours of work effects of weather shocks and then consider effects of weather shocks interacting with the minimum wage law using difference-in-differences regressions. In Section 5, I consider heterogeneous impacts on workers.

2.2 Context

2.2.1 The agricultural sector

As in many other developing countries, until the 1990s industrial agriculture in South Africa was heavily subsidised and protected. After the end of Apartheid, the government quickly deregulated and liberalised the sector. Over the last 30 years, a substantial share of smallholder farms in the commercial sector have consolidated into larger farms, many of which are oriented towards the export market. At the same time, subsistence farming continues to take place, mostly notably in former Apartheid homelands ('Bantustans').

Despite the agriculture sector's relatively small contribution to GDP - it contributed only 2.33-3.88% of annual GDP between 2001 and 2007 (World Bank, 2020) - and the fact that industrialisation of the sector has resulted in a downward trend in agricultural employment, the sector still employs a substantial number of mostly low-skilled workers.[2] During my period of study, 2001-2007, just over one million people were employed in the agricultural sector as their primary activity, the vast majority of whom had not completed high school. However, the sector also indirectly benefits or involves a wider group of people. According to the 2011 Census, in addition to workers who counted agriculture as their primary activity, an additional one million people were casually involved in agriculture at the time of the census (Liebenberg and Johann, 2013). Historically many agricultural workers have lived on the farms where they are employed with their extended families. A recent study showed that close to 600 000 households with over 2 million people lived on farms in 2015 (Ferrer and Visser, 2015). Furthermore, a large number of jobs are created in industries with backward and forward linkages to the sector.

[2] During my period of study, 2001-2007, agricultural workers represented approximately 19% of the workforce with fewer than ten years of education.

2.2.2 Drought trends & impacts

Droughts are important for agricultural output in South Africa (Baudoin et al., 2017).[3] While all areas of production are affected by droughts, field crop production is particularly volatile due its greater share of dry-land production (BFAP, 2016). For example, maize, the country's staple crop, is rain-fed and limited water availability reduces maize output by interrupting growth at several points in the growing season (Le Roux et al. (2009), cited in Dinkelman (2017)). Since the 1960s, the frequency of extreme heat events has accelerated in South Africa (Kruger and Sekele, 2013). Over the same period, interannual rainfall variability has increased and droughts have become more intense and widespread in South Africa (Fauchereau et al., 2003). This pattern is likely to continue into the future if global emissions are not reduced (Flatø et al., 2017).

2.2.3 The 2003 agricultural minimum wage

The agricultural minimum wage was officially introduced in March 2003. Until this time the agricultural sector had been barely unionised and reported the lowest wages of any sector in the country (Bhorat et al., 2014). In addition to setting a legal wage floor, the new law also defined conditions of employment for the agriculture sector that included maximum working hours and the establishment of a written employment contract for employees. According to Sectoral Determination No. 75 of 1997, the law was to apply to 'the employment of farmworkers in all farming activities in South Africa'. The law was intentionally vague about what was entailed by 'farming activities'.[4]

Importantly, the minimum wage set for the agriculture industry in South Africa resulted in changes to low-skilled worker wages that are several times greater those

[3]There is some anecdotal evidence that severe flooding may affect agriculture in South Africa, but this seems to be much less important than droughts (Dinkelman, 2017).

[4]The exact wording was: 'Without limiting its meaning, 'farming activities' includes primary and secondary agriculture, mixed farming, horticulture, aqua farming and the farming of animal products or field crops excluding the Forestry Sector.'

brought about by minimum wages in developed country contexts. The median district experienced a 43% increase in agricultural monthly wages for full-time workers after the minimum wage was implemented. Two separate wage levels were prescribed for workers according to how urbanised their region was. In 2003 over 80% of farm workers were earning less than the urban minimum, and over 60% were earning less than the rural minimum (Bhorat et al., 2014). To support implementation of the new legislation, labour inspectors were tasked with enforcement activities, visiting farms, reviewing worker contracts and interviewing a sample of workers. However, the fact that some farms were very remote made this task quite difficult and so compliance was not perfect.

2.3 Data

2.3.1 Labour market data

The labour market data for this study are drawn from 13 waves of the South African Labor Force Survey (LFS) conducted between September 2001 and September 2007. These LFS surveys are biannual rotating panel surveys, conducted in February/March and September each year and include detailed data on work and unemployment experiences of 60,000 to 70,000 working-age individuals living in 30,000 households. In each wave, 20% of households interviewed in the previous wave are rotated out of the survey entirely.[5] The chosen sample includes four waves prior to and including the legislation's effective date (March 2003) and nine afterwards. While there are three earlier waves of data going back to March 2000, the baseline sample was redrawn for the September 2001 round and so I start my analysis with data from this round (following Dinkelman and Ranchhod (2012)).[6] All 13 waves are pooled and treated as repeated cross sections over time.

[5]There is a panel data component of the LFS survey, but this is not well maintained (and also not made publicly available), and, following others, I choose not to use it because of serious concerns about the representativeness and quality of the panel data set of workers (cf. Dinkelman and Ranchhod (2012) for a longer discussion).

[6]Results are substantially unchanged when I include the other waves but to ensure the most accurate estimates, I choose to work without them.

The sample of workers includes all urban and rural employed and unemployed men and women aged 18 to 64. I identify agricultural workers in each wave of the LFS using the the South African Standard Classification of Occupations (SASCO) codes, as well as the three-digit International Standard Industrial Classification (ISIC) industry codes. In addition to using the full sample of labour force participants, for my agricultural employment analysis I also use a restricted labour force sample including only people who have similar demographic characteristics to agricultural workers i.e. they are aged between 20 and 50 and have ten or fewer years of education. Since it is very possible that some agricultural workers may work without pay (for example, as unpaid family members), I do not exclude workers without pay from my employed sample.

The labour force surveys give geographic information only for provinces, of which there are nine in the country. It is possible to work out magisterial districts (an administrative layer) for the period September 2001 to January 2003 based on the unique identification codes for respondents. However, for the period September 2003 to September 2007, the unique identification codes give information only about local municipalities, which were not defined in the earlier period. District municipalities were constant over the period of study and both magisterial districts and local municipalities fit neatly into district councils. These are therefore used as my geographic units of analysis is this paper.[7]

2.3.2 Weather measures

While most economic studies of weather shocks have tended to focus only on precipitation, the importance of taking temperature into account is increasingly being

[7]While it is possible that some of the larger district councils could be seen as including more than one local labour market, district councils have been used as the geographic unit of analysis in several well-published papers (e.g. Bhorat et al. (2014)) and it is even quite common for papers to use provincial units, which is the highest administrative layer in the country (e.g. Dinkelman and Ranchhod (2012); Magruder (2010)). For my particular analysis, larger geographic units may be preferable since the effects of weather shocks may not be captured at high resolutions because a drought in a limited area may not have enough of an impact on local labour market outcomes when there is smoothing across agricultural markets (Harari and Ferrara, 2018).

recognised (Challinor et al., 2014; Colmer, 2018; Henderson et al., 2017). The impact of rainfall on the growing cycle of a plant depends also on the capacity of the soil to retain water. Indeed drought is usually characterised by a combination of abnormally low or failing rainfall and abnormally high temperature causing unusually high evaporation.

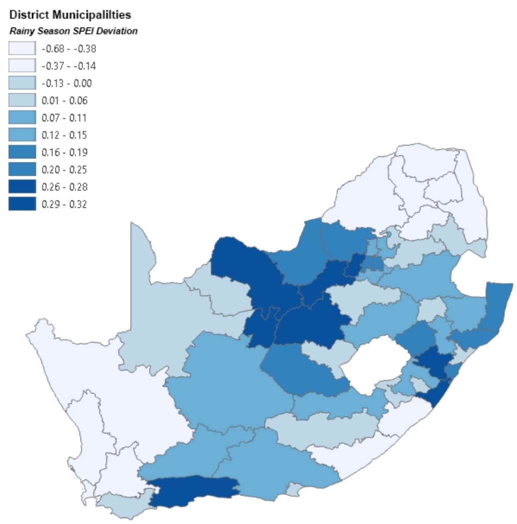
I use the Standardised Precipitation Evapotranspiration Index (SPEI), which has been shown to track dry and wet conditions better than traditional indices like the Standardised Precipitation Index (SPI) or Palmer Drought Severity Index (PDSI) (Vicente-Serrano et al., 2010, 2012). The SPEI is based on a monthly climatic water balance (precipitation minus potential evapotranspiration, called PET) and is expressed as a standardised Gaussian variate with a mean of zero and a standard deviation of one.[8] The index uses the monthly difference between precipitation and PET. However, unlike other water balance-based drought indices such as the Palmer Drought Severity Index, the SPEI does not rely on the water balance of a specific soil system (ibid.). On short time-scales, the SPEI is closely related to soil moisture, while at longer time-scales, the SPEI can be related to groundwater and reservoir storage (ibid.).

I use the Centre for Environmental Data Analysis High Resolution SPEI Dataset for Africa (Peng et al., 2020a) for the period 1981-2016, which has a 0.05° resolution. This is calculated based on precipitation estimates from the satellite-based Climate Hazards group InfraRed Precipitation with Station data (CHIRPS) and potential evaporation estimates by the Global Land Evaporation Amsterdam Model (GLEAM) (Peng et al., 2020b). I match this gridded dataset to district council boundaries and calculate mean SPEI values for each district for every month from 1981 to 2016. Following Harari and Ferrara (2018), I use cumulative monthly SPEI during the rainy season as these are the months in which rain-fed agriculture is most heavily

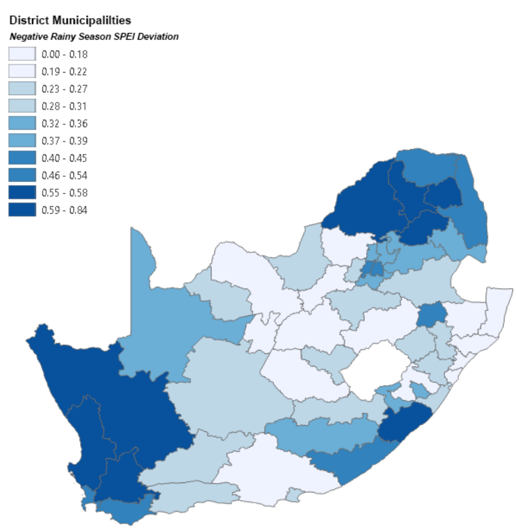
[8]The method to calculate SPEI is based on the original Standardised Precipitation Index (SPI) calculation procedure. The SPI is calculated using monthly (or weekly) precipitation as the input data whereas the SPEI uses the monthly (or weekly) difference between precipitation and PET. This represents a simple climatic water balance which is calculated at different time scales (Vicente-Serrano et al., 2010).

Figure 2.1: Variation of *Rainy Season SPEI Deviation* and variation of *Negative Rainy Season SPEI Deviation* across district municipalities

(a) Average *Rainy Season SPEI Deviation* over whole period



(b) Average *Negative Rainy Season SPEI Deviation* over whole period



Rainy Season SPEI Deviation refers to rainy season SPEI z-scores and *Negative Rainy Season SPEI Deviation* refers to the transformed rainy season SPEI z-score variable focusing only on negative values across district municipalities. Data from CEDA High Resolution SPEI Dataset for Africa for the period September 2001 to September 2007.

impacted by variations in weather in South Africa (Mastrorillo et al., 2016). The rainy season lasts from November to March in most of the country, except for districts located in the Western Cape province where it runs from May to August (Botai et al., 2016).

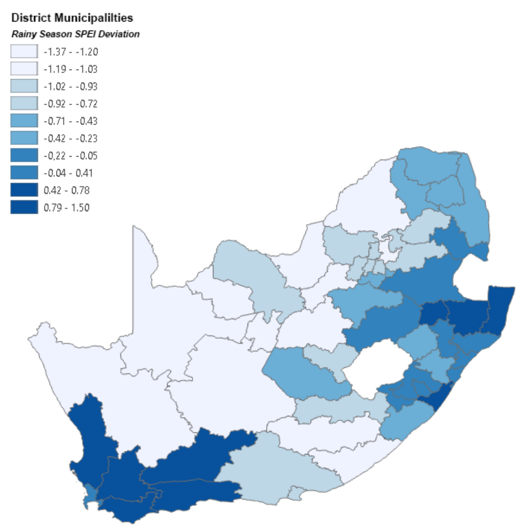
My key weather indicator, called *Rainy Season SPEI Deviation*, is the standardised total monthly SPEI for each rainy season in a particular district i.e. the deviation of total rainy season monthly SPEI from the long-term district average, normalising it by the district-specific standard deviation (Jayachandran, 2006; Santangelo, 2019). Standardising the rainy season SPEI by the district-specific standard deviation controls for the fact that some districts might have very high SPEI standard deviations and thus are more likely in each period to experience large deviations from the average.

From the available evidence, above-normal SPEI does not hurt agricultural productivity in South Africa (Dinkelman, 2017). Therefore higher values of *Rainy Season SPEI Deviation* correspond to more favourable conditions for agriculture. While I am mostly interested in deviations of rainy season SPEI below the local mean, I use the full distribution of the Z-score variable to increase power. As long as dry SPEI shocks and wet SPEI shocks have symmetrical effects, this approach is most efficient.[9] However, as a check, in the appendix I reproduce all tables using an alternative one-sided (low) SPEI variable *Negative Rainy Season SPEI Deviation* where I recode positive values as zero and take absolute values of the negative values (cf. Dinkelman (2017); Mastrorillo et al. (2016)). I generally only refer to these tables in any detail when the results diverge from those when *Rainy Season SPEI Deviation* is used. Figure 2.1 shows variation in average rainy season weather measure values across district municipalities over the period September 2001 to September 2007, and Figure 2.2 shows variation in rainy season weather measure values across district municipalities at a time when South Africa was experiencing a drought.

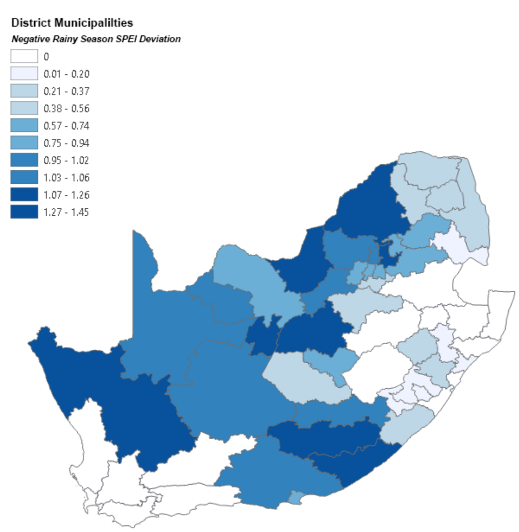
[9] This approach is also widely used (e.g. Jayachandran (2006)).

Figure 2.2: Variation of *Rainy Season SPEI Deviation* and variation of *Negative Rainy Season SPEI Deviation* across district municipalities during a drought

(a) *Rainy Season SPEI Deviation* in the rainy season prior to March 2007



(b) *Negative Rainy Season SPEI Deviation* in the rainy season prior to March 2007



Rainy Season SPEI Deviation refers to rainy season SPEI z-scores and *Negative Rainy Season SPEI Deviation* refers to the transformed rainy season SPEI z-score variable focusing only on negative values. Data from CEDA High Resolution SPEI Dataset for Africa for the rainy season prior to the labour force survey wave of March 2007.

2.3.3 Minimum wage data

The minimum wage law was promulgated in December 2002 and came into effect on the 1st of March 2003. In line with earlier studies (Bhorat et al., 2014), September 2003 is treated as the first wave where the direct impacts of the law should become evident. Two separate wage levels were prescribed for full-time farm workers, according to geographic location: a higher minimum wage (R800 per month) for those working within urbanised municipal areas classified as Area A, and a lower wage (R650 per month) for predominantly rural areas classified as Area B.[10] To evaluate which minimum wage applied to each individual, I assigned individuals to geographic areas by matching geographic information available in the LFS to areas A and B listed in the sectoral minimum wage schedules.

2.4 Descriptive statistics

Table 2.1 presents summary statistics for various samples before and after the introduction of the minimum wage during my period of study from September 2001 to September 2007. The agricultural industry in South Africa employed around 9% of the total labour force (including employed and unemployed) in the period before the minimum wage and about 7% of the labour force in the period of study. This reduction in agricultural employment occurred even as the overall employment rate increased. Around 68% of agricultural workers were employees, while the rest were self-employed - a ratio which is constant over my period of study. As expected, there is a significant increase (of almost 30%) in agricultural real wages between the pre- and post-law periods. Likely as a direct result of the law, the share of agricultural workers with written contracts went up significantly between the pre and post-law periods from 37% to 54%. The fact that agricultural hours worked dropped slightly in the post-law period might also be indicative of the effects of increased regulation in the sector. Employers may have reduced demand at the intensive margin to com-

[10]The equivalent US dollar values today are around USD 134 and USD 110. In 2015 rands, the values are R1570 and R1276.

ply with the 45 hours per week maximum set out in the minimum wage schedule, or simply to afford the higher wage.

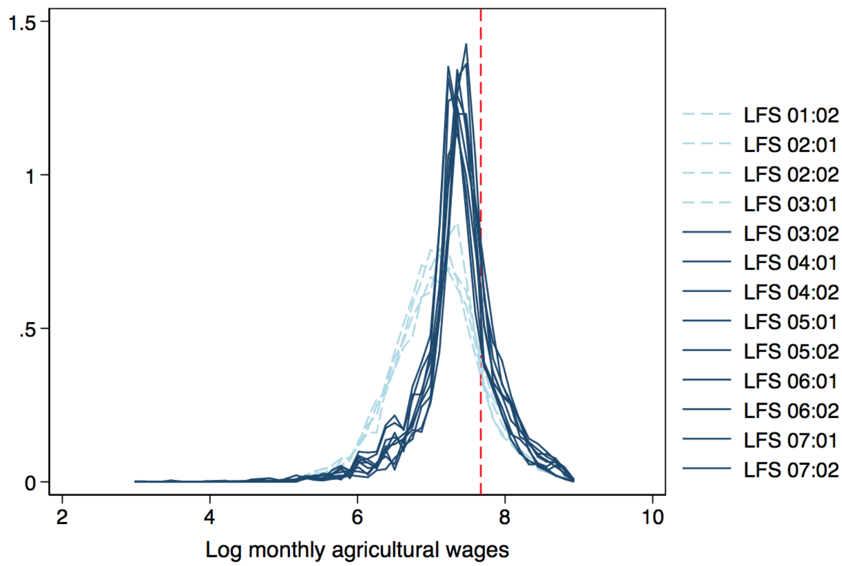
Agricultural workers are not highly educated: during my period of study, the average worker had six years of education. Approximately two-thirds of workers were male, and 76% were African. Their mean age was 38.

Table 2.1: Summary statistics

	(1) N	(2) Pre-law mean (SD)	(3) Post-law mean (SD)	(4) Post-pre diff.	(5) P value of the diff.
1. Conditional on being in the labour force:					
Employed at all	395,212	0.67	0.71	0.05	0.00
		(0.49)	(0.44)		
Employed as an agricultural worker	395,212	0.09	0.07	-0.02	0.01
		(0.29)	(0.25)		
2. Conditional on being in the restricted labour force:					
Employed at all	188,028	0.66	0.70	0.04	0.00
		(0.48)	(0.45)		
Employed as an agricultural worker	188,028	0.12	0.11	-0.01	0.02
		(0.33)	(0.30)		
3. Conditional on being an agricultural worker:					
In wage employment	47,465	0.67	0.68	0.01	0.78
		(0.44)	(0.48)		
Hours worked per month (total)	47,465	171.23	167.48	-3.74	0.41
		(73.35)	(79.05)		
Hours worked per month in wage jobs	33,520	201.3	194.65	-6.7	0.00
		(50.75)	(54.52)		
Real wages (ZAR)	33,520	1377.79	1779.91	402.12	0.00
		(855.28)	(1036.56)		
Full-time worker	33,520	0.96	0.95	-0.00	0.36
		(0.19)	(0.21)		
Proportion with a job contract	33,520	0.37	0.54	0.17	0.00
		(0.46)	(0.51)		
4. Characteristics of agricultural workers:					
Female	47,465	0.37	0.36	-0.01	0.72
		(0.45)	(0.49)		
Years of education	47,465	5.70	6.37	0.66	0.00
		(3.87)	(4.35)		
Age	47,465	38.15	38.21	0.06	0.81
		(11.55)	(12.65)		
African	47,453	0.76	0.76	-0.01	0.74
		(0.39)	(0.44)		

Data from labour force surveys (Sept 2001-Sept 2007). Panel 1 includes the entire labour force; Panel 2 is restricted to the labour force between the ages of 20-50 and with 10 or fewer years of education; Panel 3 and 4 are restricted to agricultural workers. All statistics are weighted and the standard errors of differences and p-values are calculated taking those weights and district-level clustering into account. The pre-law period includes all waves between September 2001 and March 2003 and the post-law period includes all waves between September 2003 and September 2007. A full-time worker is someone who reports at least 27 hours of work a week. Real wages are obtained from deflating nominal wages to 2015 Rands using the CPI. Note that rows may not add up due to rounding.

Figure 2.3: Distribution of (log) agricultural wages



Note: The vertical line is the level of the full-time minimum wage in 2003. Focusing on the period Sept 2001 - Sept 2007.

2.5 Empirical strategy

My analysis proceeds in two parts. I first exploit exogenous variation in productivity due to rainy season SPEI realisations across districts and over time to look at the relationship between labour market outcomes and local weather shocks over my period of study. The second part examines heterogeneous effects before and after the agricultural minimum wage and across areas where the minimum wage had varying levels of effective intensity.

I examine the local labour market effects of deviations of rainy season SPEI from the local mean using the following:

$$y_{idt} = \beta_1 C_{dt-1} + X_{idt} + \delta_d + \gamma_t + \theta_{dt} + \epsilon_{dt} \tag{2.1}$$

where y_{idt} is the outcome of interest for individual i in district d at time t and C_{dt-1} is the total rainy season SPEI Z-score in the previous rainy season. X_{idt} includes individual controls for gender, education, age and race to tackle the possibility that district composition changes. I include district fixed effects (δ_d) to capture any time-invariant district characteristics that could affect my outcomes. Wave fixed effects (γ_t) remove biannual shocks common to all districts.[11] I also include linear time trends, to allow dependent variables to trend differently across districts. Standard errors are clustered at the district level to capture serial correlation. I use the suggested Post-Apartheid Labour Market Series (PALMS) weights, ensuring comparability of waves over time.[12]

I estimate the effects on (i) employment, (ii) (log) hours worked, and (iii) (log) wages. I examine effects in the agricultural sector as well as across all sectors (including agriculture). The latter results allow me to check whether the effects of weather shocks and of weather shocks interacting with the agricultural minimum wage are limited to the agriculture sector or whether these shocks impact more broadly.[13] When considering agricultural employment, I carry out estimations using the full labour force sample as well as a restricted labour force sample where I consider only workers and unemployed people with similar demographics to agricultural workers (see Section 3.1 for more information).[14] I also look to separate out transitions between agricultural employment and unemployment, and agricultural employment and employment in other sectors. Details of all dependent variables used in my analysis can be found in Table 2.A.1 in the appendix.

The second part of the analysis tests heterogeneous effects of the agricultural minimum wage. The first step is to use the same empirical specification above but with an interaction term $C_{dt-1}\#POST_t$ where $POST_t$ is a time dummy for before and after the minimum wage is introduced.[15] This will correctly identify the effects of

[11] I also experiment with year fixed effects and the results are substantially unchanged. I prefer to use wave fixed effects to minimise the likelihood of migration across districts between periods.

[12] Results are almost identical without weights.

[13] Since stronger effects are expected in the agriculture sector, if I do not find this to be the case, this could also be seen as a sign of endogeneity.

[14] The latter approach is commonly used in papers examining employment effects on low-skilled workers in developing countries (e.g. Dinkelman and Ranchhod (2012)) since it is unrealistic that someone could transition between agricultural work and, for example, high-skilled office work.

[15] Note that the main effect of $POST_t$ will drop out because of the time fixed effects.

the minimum wage if there were no idiosyncratic shocks in addition to the law that affected agricultural workers in the post period. However, since employment, wage and hours worked variables tend to fluctuate with general economic conditions, it seems likely that other time-varying factors could influence some of these outcomes post-law, thereby confounding the effects of the law. For this reason, in the second step, I go beyond the simple pre-post comparison and investigate whether there are larger changes in my outcomes in places where the new wage floor is more binding. Using variation in the 'bite' of the minimum wage across districts will also increase the precision of my estimates. I therefore estimate the following:

$$\begin{aligned} y_{idt} = \quad & \beta_1 C_{dt-1} + \beta_2 (C_{dt-1} \times POST_t) + \beta_3 (C_{dt-1} \times WG_{d,0}) + \beta_4 (WG_{d,0} \times POST_t) + \\ & \beta_5 (C_{dt-1} \times WG_{d,0} \times POST_t) + X_{idt} + \delta_d + \gamma_t + \theta_{dt} + \epsilon_{idt} \end{aligned} \tag{2.2}$$

where $WG_{d,0}$ is a constructed variable identifying the cross-sectional variation in the wage gap between district councils in the pre-law period.[16] Following (Lee, 1999), I define the district-level wage gap as:

$$WG_{d,0} = log[min(W_d*)] - log[median(W'_d)] \tag{2.3}$$

where W_d* is the initial minimum wage for workers working at least 35 hours a week in district d and W'_d is the median real agricultural wage in the pre-wage period for workers working at least 35 hours a week.[17][18]

Under the assumption that the wage gap measure (and therefore the difference-in-differences term) is orthogonal to the error in (2),[19] the following holds: Parameter

[16] Note that the main effects of $POST_t$ and $WG_{d,0}$ drop out because of the district and time fixed effects.

[17] I use all waves in the pre-period instead of just the wave before to improve the accuracy of the estimates of median agricultural wages.

[18] Sectoral Determination No. 75 of 1997 specified that full-time workers working at least 35 hours a week should receive the minimum monthly wages discussed in Section 3.3.

[19] Based on a review of the Department of Labour documentation on setting the agricultural minimum wage, it appears that the potential impact of the wage on employment had minimal if any role in the policy decision. For example, in DoL (2001) it states that '...a minimum wage cannot be opposed purely on grounds of its adverse effect on employment' (cited in Garbers et al.

β_1 represents the average effects of *Rainy Season SPEI Deviation* across all districts across my period of study. Parameter β_2 represents the average difference in outcomes between pre and post-law periods as a result of the effects of my weather measure. Parameter β_3 represents the average effect of my weather measure in areas with larger than average wage gaps across my entire period. Parameter β_4 represents the average difference in outcomes between pre-law and post-law periods for my sample in areas with larger than average wage gaps across the entire period. Lastly, parameter β_5 represents the average difference in outcomes as a result of the interaction between my weather measure and the law for my sample in areas with larger than average wage gaps across the entire period. To show the overall effects of *Rainy Season SPEI Deviation* in different scenarios, I estimate average marginal effects for my weather measure for high and low wage gap areas in pre and post-law periods.

Lastly, I turn to heterogeneous effects of the interaction between the agricultural minimum wage and weather shocks. I re-estimate (2) for the following dependent variables: (i) the probability of being an employee conditional on working; and (ii) the probability of having a written contract conditional on being an employee (see Table 2.A.1 in the appendix for more details on these variables).

2.6 Results

Table 2.2 presents the results of Equation 1, looking at the effects of *Rainy Season SPEI Deviation* on the probability of employment. From Column 2, I estimate that a one standard deviation increase (decrease) in rainy season SPEI from the local mean is associated with a 3.3% increase (decrease) in agricultural employment relative to its mean level in the labour force (the effect on overall employment is non-significant). The effect is 80.0% stronger when considering a restricted labour force sample with demographics similar to those of agricultural workers in Column 3. From Columns

(2015)). This gives credence to the assumption that the size of the wage gaps used in this study are exogenous.

4 and 5, an increase (decrease) in rainy season SPEI is associated with movement from (into) unemployment and movement from (into) other employment sectors into (from) agriculture, though movement from (into) the former category is stronger.

Table 2.A.6 in the appendix replicates Table 2.2 but focuses only on *Negative Rainy Season SPEI Deviation*. I do not find a significant effect of my weather measure on agricultural employment in Column 2 but I do when I restrict the labour force sample in Column 3. This may however just be due to a lack of power. The coefficients on $NegSPEI\_RS\_Dev_{t-1}$ in Column 2 and 3 are either equal to or only slightly larger than the coefficients on the continuous weather measure indicating that both positive and negative shocks are important for the relationship between my continuous weather measure and agricultural employment.

Table 2.2: The effects of standardised rainy season SPEI on the probability of employment

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res. employ.)
SPEI_RS_Dev	0.0033	0.0024**	0.0063**	0.0141***	0.0073***
	(0.002)	(0.001)	(0.002)	(0.004)	(0.003)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean SPEI_RS_Dev	0.0638	0.0638	0.0450	0.0166	0.0596
SD SPEI_RS_Dev	0.9640	0.9640	0.9418	0.9211	0.9545
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.3 replicates Table 2.2 but uses hours worked and wages as the dependent variables. Standardised rainy season SPEI has a weak negative association with hours worked in agricultural wage employment in Column 3. In the corresponding table focusing only on *Negative Rainy Season SPEI Deviation* (Table 2.A.7 in the appendix), in Column 3 this variable has a positive coefficient of approximately the same magnitude and is also significant at the 10% level. Reduced soil moisture is

Table 2.3: The effects of standardised rainy season SPEI on monthly hours worked and wages

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
SPEI_RS_Dev	-0.0035	-0.0128	-0.0106*	-0.0040	0.0062
	(0.0025)	(0.009)	(0.0053)	(0.0052)	(0.0091)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean SPEI_RS_Dev	0.0778	-0.0138	-0.0757	0.0892	-0.0786
SD SPEI_RS_Dev	0.9802	0.8922	0.9079	0.9771	0.9063
Observations	262,769	45,447	33,439	160,973	30,356

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). $SPEI\_RS\_Dev$ is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

therefore positively associated with hours worked in agricultural wage employment. An intuitive explanation for this would be that as agricultural employers let go of some workers in the wake of dry shocks, they look to compensate for this by getting their other workers to work longer hours. In both Table 2.3 and Table 2.A.7, the effects on total hours worked in all sectors, total agricultural hours worked, wages earned in all sectors and agricultural wages are non-significant.

Table 2.A.2 (focusing on employment) and Table 2.A.4 (focusing on hours worked and wages) in the appendix show the results of estimating Equation 1 with the interaction between $POST_t$ and $SPEI\_RS\_Dev_{t-1}$ included as an additional term. Across all estimations the coefficient on the interaction term is non-significant. To check if this is just due to insufficient power, Table 2.A.3 and Table 2.A.5 calculate corresponding average marginal effects of my weather measure before and after the minimum wage. In columns 2 and 3 of Table 2.A.3, it can be seen that deviations in rainy season SPEI above (below) the local mean are associated with a stronger increase (decrease) in agricultural employment in the post-law period. Column 2 of Table 2.A.5 shows that deviations in rainy season SPEI above (below) the local mean are associated with a larger and more precisely estimated decrease (increase) in total

agriculture hours worked (including those worked by employees and the self-employed in agriculture) in the pre-law period compared to the post-law period in Column 2. However, as discussed in Section 5, there could be several other time-varying factors that could influence these variables over my period of study independent of the law. The corresponding results for Negative Rainy Season SPEI Deviation are in the appendix.[20]

Table 2.4: The effects of standardised rainy season SPEI on the probability of employment after vs. before the minimum wage law considering the size of the wage gap

	(1)	(2)	(3)	(4)	(5)
	Employment (rest of LF)	Ag. employ. (rest of LF)	Ag. employ. (rest of res. LF)	Ag. employ. (res. unemploy.)	Ag. employ. (res employ.)
SPEI_RS_Dev	0.0030	0.0010	0.0026	0.0022	-0.0008
	(0.006)	(0.004)	(0.004)	(0.016)	(0.011)
SPEI_RS_Dev#POST	-0.0067	-0.0095	-0.0138	-0.0126	-0.0223
	(0.012)	(0.008)	(0.011)	(0.020)	(0.014)
SPEI_RS_Dev#WG	-0.0036	0.0053	0.0072	0.0143	0.0051
	(0.009)	(0.005)	(0.007)	(0.013)	(0.009)
POST#WG	-0.0119	-0.0246	-0.0167	-0.0337	-0.0143
	(0.026)	(0.020)	(0.021)	(0.038)	(0.030)
POST#SPEI_RS_Dev#WG	0.0078	0.0173**	0.0275**	0.0256	0.0451**
	(0.010)	(0.008)	(0.012)	(0.022)	(0.017)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean *wage gap*	0.7505	0.7505	0.7688	0.7744	0.7630
SD *wage gap*	0.2699	0.2699	0.2807	0.2874	0.2875
Mean SPEI_RS_Dev	0.0638	0.0638	0.0450	0.0166	0.0596
SD SPEI_RS_Dev	0.9640	0.9640	0.9418	0.9211	0.9545
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). $SPEI\_RS\_Dev$ is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.4 shows the results of estimating Equation 2 focusing on employment as the dependent variable with a three-way interaction term including $POST_t$, $SPEI\_RS\_Dev_{t-1}$ and the pre-law agricultural sector wage gap together with an interaction between

[20]Due to potential endogeneity issues, I do not comment on these tables in depth. Table 2.A.8 and Table 2.A.10 in the appendix replicate Table 2.A.3 and Table 2.A.5, respectively, but using *Negative Rainy Season SPEI Deviation*. Again all interaction terms are non-significant. Table Table 2.A.9 and Table 2.A.11 show corresponding average marginal effects results. The results are more or less in line with those obtained for the continuous weather measure, though sometimes weaker.

Table 2.5: The effects of standardised rainy season SPEI on the probability of employment after vs. before the minimum wage law considering the size of the wage gap

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res employ.)
Avg. marginal effect of SPEI_RS_Dev when:					
WG = low & POST = 0	0.0059	0.0007	0.0020	0.0091	0.0016
	(0.0073)	(0.0049)	(0.0056	(0.0107)	(0.0075)
WG = low & POST = 1	0.0029	-0.0004	0.0017	0.0090	0.0008
	(0.0030)	(0.0018)	(0.0027)	(0.0047)	(0.0035)
WG = med. & POST = 0	0.0049	0.0022	0.0040	0.0132	0.0031
	(0.0067)	(0.0043)	(0.0047)	(0.0088)	(0.0061)
WG = med. & POST = 1	0.0040	0.0057**	0.0114***	0.0204***	0.0152***
	(0.0032)	(0.0024)	(0.0032)	(0.0061)	(0.0041)
WG = high & POST = 0	0.0039	0.0036	0.0060	0.0173	0.0046
	(0.0068)	(0.0041)	(0.0044)	(0.0083)	(0.0056)
WG = high & POST = 1	0.0052	0.0118***	0.0212***	0.0319***	0.0296***
	(0.0042)	(0.0043)	(0.0056)	(0.0099)	(0.0069)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean *wage gap*	0.7505	0.7505	0.7688	0.7744	0.7630
SD *wage gap*	0.2699	0.2699	0.2807	0.2874	0.2875
Mean SPEI_RS_Dev	0.0638	0.0638	0.0450	0.0166	0.0596
SD SPEI_RS_Dev	0.9640	0.9640	0.9418	0.9211	0.9545
Observations	385,177	385,177	184,718	88,436	126,347

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD. After running the regression in Equation 2, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

$SPEI\_RS\_Dev_{t-1}$ and $POST_t$, and interaction between $SPEI\_RS\_Dev_{t-1}$ and the wage gap, and an interaction between $POST_t$ and the wage gap. Since this specification includes multiple interaction effects, it is very demanding and so it is unsurprising that some of the effects are non-significant (even if the marginal effects calculated in the next step are highly significant). Though non-significant, the positive coefficient on $SPEI\_RS\_Dev_{t-1}$ in most specifications implies that above (below) mean SPEI deviations on average increase (decrease) the probability of employment during my period of study. The negative coefficient on the interaction term $SPEI\_RS\_Dev_{t-1}\#POST_t$ implies that in the post-law period, the positive (neg-

Table 2.6: The effects of standardised rainy season SPEI on monthly hours worked and wages after vs. before the minimum wage law considering the size of the wage gap

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
SPEI_RS_Dev	0.0066	-0.0164	0.0111	0.0184	0.0626
	(0.014)	(0.024)	(0.024)	(0.024)	(0.043)
SPEI_RS_Dev#POST	-0.0043	-0.0111	-0.0182	-0.0404	-0.0661
	(0.018)	(0.033)	(0.027)	(0.028)	(0.050)
SPEI_RS_Dev#WG	-0.0212	-0.0231	-0.0312	-0.0190	-0.1085*
	(0.019)	(0.024)	(0.033)	(0.027)	(0.062)
POST#WG	0.0537**	0.2729***	0.0310	0.1119*	0.2396***
	(0.026)	(0.079)	(0.050)	(0.059)	(0.054)
POST#SPEI_RS_Dev#WG	0.0114	0.0335	0.0193	0.0411	0.1023
	(0.024)	(0.037)	(0.038)	(0.037)	(0.064)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean *wage gap*	0.7364	0.7670	0.6926	0.7377	0.6993
SD *wage gap*	0.2744	0.3401	0.3128	0.2631	0.3129
Mean SPEI_RS_Dev	0.0778	-0.0138	-0.0757	0.0892	-0.0786
SD SPEI_RS_Dev	0.9802	0.8922	0.9079	0.9771	0.9063
Observations	255,887	45,386	33,395	157,185	30,320

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

ative) effect of above (below) mean SPEI deviations is smaller than in the pre-law period. This could be the result of areas becoming increasingly resilient to negative weather shocks over time. The positive coefficient on $SPEI\_RS\_Dev_{t-1}\#WG_{d,0}$ in all columns (except Column 1 where the probability of overall employment is the dependent variable) suggests that the average effect of above (below) mean SPEI deviations was to increase (decrease) employment in areas with higher than average wage gaps. The negative coefficient on the interaction term $POST_t\#WG_{d,0}$ implies that the minimum wage law coincided with decreased (agricultural) employment in areas with larger than average wage gaps. Lastly, in most regressions looking at agricultural employment, the three-way interaction is positive and significant, suggesting that for districts where the minimum wage had substantial bite, there is a significant increase (decrease) in employment as a result of the interaction between above (below) mean deviations and the law.

Table 2.7: The effects of standardised rainy season SPEI on monthly hours worked and wages after vs. before the minimum wage law considering the size of the wage gap

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
Avg. marginal effect of SPEI_RS_Dev when:					
WG = low & POST = 0	-0.0032	-0.0263	-0.001	0.0095	0.206
	(0.007)	(0.016)	(0.014)	(0.013)	(0.023)
WG = high & POST = 1	-0.0022	-0.0231*	-0.012*	-0.0114*	-0.006
	(0.004)	(0.013)	(0.006)	(0.006)	(0.014)
WG = med & POST = 0	-0.0090*	-0.0342***	-0.0105	0.0044	-0.0133
	(0.005)	(0.012)	(0.011)	(0.010)	(0.017)
WG = med & POST = 1	-0.0049*	-0.0195*	-0.0153**	-0.0056	-0.0079
	(0.003)	(0.011)	(0.007)	(0.006)	(0.010)
WG = high & POST = 0	-0.0149	-0.0420***	-0.0203	-0.0005	-0.0472
	(0.008)	(0.013)	(0.016)	(0.011)	(0.028)
WG = high & POST = 1	-0.0076**	-0.1560	-0.0190*	0.0003	-0.0098
	(0.004)	(0.012)	(0.010)	(0.010)	(0.012)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean *wage gap*	0.7364	0.7670	0.6926	0.7377	0.6993
SD *wage gap*	0.2744	0.3401	0.3128	0.2631	0.3129
Mean SPEI_RS_Dev	0.0778	-0.0138	-0.0757	0.0892	-0.0786
SD SPEI_RS_Dev	0.9802	0.8922	0.9079	0.9771	0.9063
Observations	255,887	45,386	33,395	157,185	30,320

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD. After running the regression in Equation 2, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.12 in the appendix replicates Table 2.4 using *Negative Rainy Season SPEI Deviation*. Comparing the results in the two tables, it is noticeable that coefficients on the significant three-way interaction terms in all columns in Table 2.A.12 are larger (and, as expected, of opposite sign) than those in Table 2.4. This shows that the interaction between the minimum wage and weather shocks matters more in the case of negative weather shocks.

Table 2.5 shows the average marginal effects (corresponding to the regressions undertaken to create Table 2.4) of standardised rainy season SPEI in the pre-law and post-law periods on employment holding the wage gap variable at high, medium and low levels (see table footnotes for more information). In columns 2 and 3, it can be seen that $SPEI\_RS\_Dev_{t-1}$ has no significant effect on agricultural employment in the pre-law period. However, in the post-law period, $SPEI\_RS\_Dev_{t-1}$ has a strong and significant effect on agricultural employment in areas where the minimum wage had substantial bite. Holding the wage gap at its mean level, a one standard deviation increase (decrease) in rainy season SPEI relative to the local mean results in a 7.1% increase (decrease) in agricultural employment relative to its mean level among the total labour force.[21] This effect is 37.1% larger when considering a restricted version of the labour force in Column 3.

Table 2.A.13 replicates Table 2.5 but focusing on *Negative Rainy Season SPEI Deviation*. As expected, the pattern of marginal effects is the same as in Table 2.5 but the significant coefficients on the marginal effects across all specifications for the post period where the wage gap is either set at its average level or at one standard deviation above its average level, are larger and more precisely estimated. Holding the wage gap at its mean level, a one standard deviation increase in *Negative Rainy Season SPEI Deviation* (really, one standard deviation decrease in rainy season SPEI) relative to the local mean results in a 7.1% decrease in agricultural employment relative to its mean level among the total labour force.[22]

As before, columns 4 and 5 in Table 2.5 and Table 2.A.13 look to break down effects of above (below) mean SPEI deviations on the probability of agricultural employment into transitions from (into) unemployment and transitions from (into) employment in other sectors. In Table 2.5 it appears that both channels are at play

[21] This is calculated using the numbers available in Column 2 of Table 2.5 i.e. 0.0057÷0.0771×0.9640×100=7.1268.

[22] In results not shown, positive weather shocks (using $SPEI\_RS\_Dev_{t-1}$ and recoding negative values as zero) have a weaker but still significant effect (in the opposite direction) on agricultural employment in the post-law period i.e. holding the wage gap at its mean level, a one standard deviation increase in rainy season SPEI relative to the local mean results in a 4.7% increase in agricultural employment relative to its mean level among the total labour force. If the minimum wage (gradually) leads to a more consolidated agriculture workforce during times where there are no weather shocks, it should be expected that hiring would increase in response to positive shocks relative to the situation in the pre-law period.

but the effect is slightly larger in Column 5 than Column 4, suggesting that above (below) mean rainy season SPEI interacting with the agricultural minimum wage is more likely to result in agricultural workers moving (from) into other industries than moving (from) into unemployment. Focusing on *Negative Rainy Season SPEI Deviation* in Table 2.A.13, the effect sizes in Column 4 are larger than in Column 5, especially when holding the wage gap at its mean level. As before, then, there seems to be an asymmetry: negative shocks are more likely to lead to a movement out of agricultural employment and into unemployment than into employment in other industries (which may be, in any case, purely survivalist), whereas positive shocks are more likely to lead to a movement of workers from non-agricultural industries into agriculture than a movement of unemployed people into agriculture.

Table 2.6 replicates Table 2.4 but considers hours worked and wages as the dependent variables. Again many of the terms, including, in this case, the three-way interaction terms, are non-significant. This could be either because there are no statistically meaningful relationships between hours worked and wages, the minimum wage law and my weather measure, or because there is not enough power. Though non-significant, the positive coefficient on $SPEI\_RS\_Dev_{t-1}$ in most specifications (except where total agricultural hours is the dependent variable) implies that above (below) mean rainy season SPEI deviations on average increase (decrease) hours worked and wages during my period of study. The (non-significant) negative coefficient on the interaction term $SPEI\_RS\_Dev_{t-1}\#POST_t$ in all columns implies that in the post-law period, the (negative) positive effect of (negative) $SPEI\_RS\_Dev_{t-1}$ on hours worked and wages is smaller than in the pre-law period. The (non-significant) negative coefficient on $SPEI\_RS\_Dev_{t-1}\#WG_{d,0}$ in all columns (except the first) shows that the average effect of above (below) mean SPEI deviations was to decrease (increase) hours worked and wages in areas with higher than average wage gaps. The positive (and significant in most specifications) coefficient on the interaction term $POST_t\#WG_{d,0}$ implies that the minimum wage law coincided with increased hours worked and wages in areas with larger than average wage gaps. Lastly, in all estimations, the three-way interaction is positive (though non-significant), implying that for districts where the minimum wage had substantial bite, there is an increase (decrease) in hours worked and wages as a result of the interaction between above (below) mean SPEI deviations and the law.

Table 2.A.14 replicates Table 2.6 but focuses on the effects of *Negative Rainy Season SPEI Deviation*. The most notable difference is that the coefficients in Column 5 are larger and are more precisely estimated indicating a stronger relationship between agricultural wages and the interaction between agricultural weather shocks and the minimum wage legislation in the case of negative weather shocks.

Table 2.8: The effects of standardised rainy season SPEI on the probability of wage employment and wage employment with a contract after vs. before the minimum wage law considering the size of the wage gap

	(1) Wage employ. (Self-employ.)	(2) Ag. wage employ. (Ag. self-employ.)	(3) Wage employ. w. contract (Wage emp. w/o contract)	(4) Ag. wage employ. w. contract (Ag. wage employ. w/o contract)
SPEI_RS_Dev	0.0142	-0.0097	0.0394	0.0325
	(0.009)	(0.022)	(0.030)	(0.036)
SPEI_RS_Dev#POST	-0.0021	0.0255	-0.0257	-0.0440
	(0.010)	(0.024)	(0.035)	(0.039)
SPEI_RS_Dev#WG	-0.0214**	-0.0116	-0.0284	-0.0122
	(0.009)	(0.020)	(0.030)	(0.026)
POST#WG	0.0242	0.0970*	-0.0009	0.0151
	(0.030)	(0.056)	(0.053)	(0.049)
POST#SPEI_RS_Dev#WG	0.0062	-0.0274	0.0234	0.0539
	(0.011)	(0.025)	(0.036)	(0.033)
Individual controls	Y	Y	Y	Y
District FEs	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y
Mean dep. var.	0.7762	0.6750	0.6250	0.3230
Mean *wage gap*	0.7391	0.7753	0.7181	0.7753
SD *wage gap*	0.2759	0.3424	0.2653	0.3424
Mean SPEI_RS_Dev	0.0715	-0.0113	0.0751	-0.0113
SD SPEI_RS_Dev	0.9738	0.8902	0.9794	0.8902
Observations	263,513	47,388	203,866	47,388

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). $SPEI\_RS\_Dev$ is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.7 shows the average marginal effects of $SPEI\_RS\_Dev_{t-1}$ on hours worked and wages corresponding to the regressions in Table 2.6. The main result is the effect of the minimum wage on the relationship between $SPEI\_RS\_Dev_{t-1}$ and hours worked in agricultural wage employment relative to their mean level in Column 3. Whereas there is no distinguishable effect of $SPEI\_RS\_Dev_{t-1}$ on hours worked of wage employees in agriculture in the pre-law period, in the post-law period it appears that above (below) mean SPEI deviations lead to a decrease (increase) in hours worked for these workers. Holding the wage gap at its mean level, a one standard deviation increase (decrease) in rainy season SPEI relative to the local mean results in a 2.0% increase (decrease) in wage agriculture hours worked. In the case of SPEI deviations below the local mean, the positive effect on wage hours worked is probably the result of employers making their remaining workforce work harder after letting go of some workers. This effect increases in strength with the size of the wage gap. In columns 1 and 2, I also find that rainy season SPEI Z-scores have a significant negative effect on hours worked in all sectors and in total agricultural hours worked (including both employees and self-employed) in some areas in the pre-law or post-law period but there is no distinct pattern and so it not clear that the minimum wage has any effect. Lastly, in columns 4 and 5, I find that in general rainy season SPEI has no discernible effects on agricultural wages or wages in all industries.

Table 2.A.15 in the appendix replicates Table 2.7 for *Negative Rainy Season SPEI Deviation*. The most notable difference is that in Column 5 it appears that in some areas agricultural wages were negatively affected by below average rainy season moisture in the pre-law period. Specifically, in low wage gap areas (i.e. where the wage gap is held at one standard deviation below its mean level), a one standard deviation decrease in rainy season SPEI relative to the local mean results in a 0.7% decrease in agricultural wages relative to their mean level. In the post-law period, in these same areas, the coefficient on the marginal effect becomes positive and non-significant. This finding suggests that in certain parts of the country prior to the minimum wage law, agricultural wages were adjusted downwards in the wake of negative shocks. In the post-period, there is no adjustment, probably due to the fact

that employers have less room to adjust wages once the law is in place (assuming they are observing it).

2.7 Heterogeneous effects

I now turn to heterogeneous effects on employment of the interaction between the agricultural minimum wage and weather shocks.

Table 2.8 replicates Table 2.4 but examines the probability of wage employment conditional on employment and the probability of having a written contract conditional on having wage employment in the agriculture sector and across all sectors. Most terms are non-significant and coefficients have almost all the same signs in columns 1, 3 and 4 where the probability of wage employment across all industries, the probability of having a contract as a wage employee in all industries and the probability of having a contract as a wage employee in the agricultural sector are the respective dependent variables. Column 2 with agricultural wage employment as the dependent variable displays quite different results.

Though non-significant, the positive coefficient on $SPEI\_RS\_Dev_{t-1}$ in most specifications in Table 2.8 (except where agriculture wage employment is the dependent variable in Column 2 where the coefficient is negative and non-significant) implies that above (below) mean rainy season SPEI on average increases (decreases) the probability of wage employment (except in agriculture) and the probability of having a written contract during my period of study. The (non-significant) negative coefficient on the interaction term $SPEI\_RS\_Dev_{t-1}\#POST_t$ (except where agriculture wage employment is the dependent variable where the coefficient is positive and non-significant) implies that in the post-law period, the positive (negative) effect of above (below) mean rainy season SPEI is smaller than in the pre-law period. As before, this could be related to the fact that areas tend to become increasingly resilient to negative weather shocks over time. The negative coefficient on $SPEI\_RS\_Dev_{t-1}\#WG_{d,0}$ in all columns (which is non-significant except for in Column 1 where the probability of wage employment across all industries is the dependent variable) shows that the average effect of above (below) mean rainy season SPEI was to decrease (increase) the probability of having wage employment and a written contract in areas with higher than average wage gaps. The positive but non-significant coefficient on the interaction term $POST_t\#WG_{d,0}$ in all columns except Column 3 (where it is negative but practically very small and non-significant) implies that the minimum wage law generally coincided with an increased probability of wage employment and employment with a contract in areas with larger than average wage gaps. Lastly, in most estimations the three-way interaction is positive and non-significant (except for in Column 2 where agricultural wage employment is the dependent variable where the coefficient is negative and non-significant), implying that for districts where the minimum wage had substantial bite, there is an increase (decrease) in the likelihood of wage employment and of having a wage contract as a result of the interaction between above (below) mean SPEI deviations and the law.

Table 2.A.16 in the Appendix replicates Table 2.8 but focuses on *Negative Rainy Season SPEI Deviation*. One important difference is that in Column 2 the coefficient on the three-way interaction term is considerably larger suggesting that the relationship between the probability of agricultural wage employment and the interaction between weather shocks and the minimum wage law is stronger when considering negative weather shocks.

Table 2.9 shows the corresponding average marginal effects for rainy season SPEI in the pre-law and post-law periods. In Column 2, holding the wage gap at a high level (at one standard deviation above the mean), it appears that above (below) mean rainy season SPEI deviations reduce (increase) the probability of wage employment Vis-à-vis self-employment among agricultural workers after the minimum wage law is introduced (holding the wage gap at its mean, there is no effect). In many developing country contexts one might expect to find negative employment effects of lower-than-normal soil moisture for agricultural wage workers only, and perhaps some movement from wage agriculture into self-employed/subsistence agri-

Table 2.9: The effects of standardised rainy season SPEI on the probability of wage employment and wage employment with a contract after vs. before the minimum wage law considering the size of the wage gap

	(1)	(2)	(3)	(4)
	Wage employ. (Self-employ.)	Ag. wage employ. (Ag. self-employ.)	Wage employ. w. contract (Wage emp. w/o contract)	Ag. wage employ. w. contract (Ag. wage employ. w/o contract)
Average marginal effect of SPEI_RS_Dev when:				
WG = low & POST = 0	0.0043	-0.0147	0.0265	0.0272
	(0.006)	(0.016)	(0.018)	(0.026)
WG = low & POST = 1	0.0051**	-0.0010	0.0115**	0.0066
	(0.002)	(0.007)	(0.006)	(0.011)
WG = med. & POST = 0	-0.0017	-0.0187	0.0190	0.0231
	(0.005)	(0.012)	(0.013)	(0.019)
WG = med. & POST = 1	0.0009	-0.0144	0.0102**	0.0209**
	(0.003)	(0.008)	(0.004)	(0.009)
WG = high & POST = 0	-0.0076*	-0.0227	0.0115	0.0189
	(0.004)	(0.012)	(0.012)	(0.014)
WG = high & POST = 1	-0.0033	-0.0278**	0.0089	0.0352***
	(0.005)	(0.012)	(0.006)	(0.013)
Individual controls	Y	Y	Y	Y
District FEs	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y
Mean dep. var.	0.7762	0.6750	0.6250	0.3230
Mean *wage gap*	0.7391	0.7753	0.7181	0.7753
SD *wage gap*	0.2759	0.3424	0.2653	0.3424
Mean SPEI_RS_Dev	0.0715	-0.0113	0.0751	-0.0113
SD SPEI_RS_Dev	0.9738	0.8902	0.9794	0.8902
Observations	263,513	47,388	203,866	47,388

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD. After running the regression in Equation 2, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

culture. However, as detailed in Section 2.1, South Africa is an unusual case where a large proportion of agricultural workers live on farms with their extended families. Since the extended family members are often involved in self-employed agriculture, when the wage worker loses his or her job and the household is forced to move off the farm, many self-employed agricultural workers may lose their employment too. In the special case of South Africa then, the agricultural minimum wage may have negative effects that spill over to self-employed workers.

In the corresponding marginal effects table for *Negative Rainy Season SPEI Deviation* (see Table 2.A.17 in the appendix), it can be seen that the coefficients on the marginal effects in Column 2 are larger and more precisely estimated showing that negative shocks are driving the patterns seen in the previous table.[23]

In Column 4 of Table 2.8, it can be seen that above (below) mean rainy season SPEI deviations increase (reduce) the probability of an agricultural worker having a written contract in the post-law period. This is a notable finding since, in general, the proportion of agricultural wage workers with written contracts went up in the post law period. It seems that the interaction between the agricultural minimum wage and lower-than-normal soil moisture led to the casualisation of the agricultural labour force in the post-law period. Holding the minimum wage gap at its mean level, I estimate that a one standard deviation increase (decrease) in rainy season SPEI relative to the local mean results in a 5.7% increase (decrease) in the probability that an agricultural employee has a written contract relative to the mean level of this dependent variable. It can be deduced from Table 2.A.17 and results not shown where I focus only on the effects of positive shocks that both positive and negative shocks are driving this relationship.

2.8 Conclusion

This paper examines how the introduction of a sectoral minimum wage affects resilience to productivity shocks in a developing country context by looking at the interaction between an agricultural minimum wage and agricultural productivity shocks in South Africa. Exploiting substantial variation across districts in pre-law median agricultural wages and variation in weather across districts and over time, I determine that the minimum wage increases the susceptibility of agricultural employment to low soil moisture i.e. rainy season SPEI deviations below the local mean. I do not find effects on agricultural wages and only find marginal effects on hours worked

[23]In results not shown, when I focus only on positive shocks, there is no significant effect of this weather measure in the post-period on the probability of wage agricultural employment.

in the agricultural sector. The evidence best supports a story where the dramatic increase in the agricultural wage bill as a result of the law makes it less affordable for employers to keep on their workforces in the wake of negative productivity shocks, which results in retrenchments.

There are strong arguments to be made that the agricultural minimum wage in South Africa was long overdue and there was a very real need to improve workers' conditions in this sector. However, it also seems that the minimum wage was set at a very high level in relation to the prior wage distribution of agricultural workers and it is unsurprising that there could have been some negative implications for employment. This paper highlights one important mechanism for how agricultural minimum wages can have negative employment effects. As policymakers are increasingly concerned about how to increase resilience in developing countries in the face of climate change, the paper also shows that they should be aware that agricultural minimum wages interacting with negative weather shocks may lead to many vulnerable workers losing their jobs and to incentives changing around the provision of permanent contracts.

Bibliography

Adhvaryu, A., Chari, A. V., and Sharma, S. (2013). Firing Costs and Flexibility: Evidence from Firms' Labour Adjustments to Shocks in India. *Review of Economics and Statistics*, 95(3): 725–740.

Baudoin, M.-A., Vogel, C., Nortje, K., and Naik, M. (2017). Living with drought in South Africa: lessons learnt from the recent El Niño drought period. *International Journal of Disaster Risk Reduction*, 23: 128–137.

BFAP (2016). *Policy Brief on the 2015/2016 Drought*. Bureau for Food and Agricultural Policy (BFAP), Pretoria.

Bhorat, H., Kanbur, R., and Stanwix, B. (2014). Estimating the impact of minimum wages on employment, wages, and non-wage benefits: The case of agriculture in South Africa. *American Journal of Agricultural Economics*, 96(5): 1402–1419.

Botai, C. M., Botai, J. O., Dlamini, L. C., Zwane, N. S., and Phaduli, E. (2016). Characteristics of droughts in South Africa: a case study of Free State and North West provinces. *Water*, 8(10): 439.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural Productivity and Structural Transformation: Evidence from Brazil. *American Economic Review*, 106(6): 1320–65.

Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4): 720–793.

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S., Smith, D., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4(4): 287–291.

Chaurey, R. (2015). Labor Regulations and Contract Labor Use: Evidence from Indian Firms. *Journal of Development Economics*, 114: 224–232.

Colmer, J. (2018). Weather, labor reallocation and industrial production: evidence from india. *American Economic Review: Applied Economics (conditionally accepted)*.

Conradie, B. (2004). *Revisiting Casualisation on Fruit Farms in the Western Cape*. Centre for Social Science Research (CSSR) Working Paper No. 177, University of Cape Town.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, 52(3): 740–98.

Dinkelman, T. (2017). Long-run Health Repercussions of Drought Shocks: Evidence from South African Homelands. *The Economic Journal*, 127(604): 1906–1939.

Dinkelman, T. and Ranchhod, V. (2012). Evidence on the impact of minimum wage laws in an informal sector: Domestic workers in South Africa. *Journal of Development Economics*, 99(1): 27–45.

DoL (2001). Determination of employment conditions in South African agriculture. Government Gazette No. 22648.

Fauchereau, N., Trzaska, S., Rouault, M., and Richard, Y. (2003). Rainfall variability and changes in Southern Africa during the 20th century in the global warming context. *Natural Hazards*, 29(2): 139–154.

Ferrer, S. and Visser, M. (2015). *Farm workers' living and working conditions in South Africa: Key trends, emergent issues, and underlying and structural problems*. International Labour Organization, Pretoria.

Flatø, M., Muttarak, R., and Pelser, A. (2017). Women, weather, and woes: The triangular dynamics of female-headed households, economic vulnerability, and climate variability in South Africa. *World Development*, 90: 41–62.

Franklin, S. and Labonne, J. (2019). Economic Shocks and Labor Market Flexibility. *Journal of Human Resources*, 54(1): 171–199.

Garbers, C., Burger, R., and Rankin, N. (2015). The impact of the agricultural minimum wage on farmworker employment in South Africa: A fixed effects approach. *Biennial Conference of the Economic Society of South Africa*.

Gindling, T. H. and Terrell, K. (2007). The effects of multiple minimum wages throughout the labor market: The case of Costa Rica. *Labour Economics*, 14(3): 485–511.

Harari, M. and Ferrara, E. L. (2018). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics*, 100(4): 594–608.

Henderson, J. V., Storeygard, A., and Deichmann, U. (2017). Has climate change driven urbanization in Africa? *Journal of Development Economics*, 124: 60–82.

Hornbeck, R. (2012). The Enduring Impact of the American Dust Bowl: Short- and Long-run Adjustments to Environmental Catastrophe. *American Economic Review*, 102(4): 1477–1507.

Jayachandran, S. (2006). Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. *Journal of Political Economy*, 114(3): 538–575.

Jessoe, K., Manning, D. T., and Taylor, J. E. (2018). Climate Change and Labour Allocation in Rural Mexico: Evidence from Annual Fluctuations in Weather. *The Economic Journal*, 128(608): 230–261.

Kruger, A. and Sekele, S. (2013). Trends in extreme temperature indices in South Africa: 1962–2009. *International Journal of Climatology*, 33(3): 661–676.

Le Roux, N. et al. (2009). *Seasonal maize yield simulations for South Africa using a multi-model ensemble system*. PhD thesis, Unpublished thesis manuscript, University of Pretoria, April.

Lee, D. S. (1999). Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage? *Quarterly Journal of Economics*, 114(3): 977–1023.

Liebenberg, F. and Johann, K. (2013). Statistics on Farm Labour in South Africa.

University of Pretoria Department of Agricultural Economics, Extension and Rural Development Working Paper.

Magruder, J. R. (2010). Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa. *American Economic Journal: Applied Economics*, 2(1): 62–85.

Mastrorillo, M., Licker, R., Bohra-Mishra, P., Fagiolo, G., Estes, L. D., and Oppenheimer, M. (2016). The influence of climate variability on internal migration flows in South Africa. *Global Environmental Change*, 39: 155–169.

Murray, J. and Van Walbeek, C. (2007). Impact of the Sectoral Determination for Farm Workers on the South African Sugar Industry: Case Study of the Kwazulu-Natal North and South Coasts. *Agrekon*, 46(1): 94–112.

Neumark, D., Cunningham, W., and Siga, L. (2006). The Effects of the Minimum Wage in Brazil on the Distribution of Family Incomes: 1996–2001. *Journal of Development Economics*, 80(1): 136–159.

Neumark, D., Wascher, W. L., et al. (2007). Minimum Wages and Employment. *Foundations and Trends in Microeconomics*, 3(1-2): 1–182.

Peng, J., Dadson, S., Hirpa, F., Dyer, E., Lees, T., Gonzalez Miralles, D., Vicente-Serrano, S. M., and Funk, C. (2020a). *High resolution Standardized Precipitation Evapotranspiration Index (SPEI) dataset for Africa*. Centre for Environmental Data Analysis. doi:10.5285/bbdfd09a04304158b366777eba0d2aeb.

Peng, J., Dadson, S., Hirpa, F., Dyer, E., Lees, T., Gonzalez Miralles, D., Vicente-Serrano, S. M., and Funk, C. (2020b). A pan-African high-resolution drought index dataset. *Earth System Science Data*, 12: 753–769.

Santangelo, G. (2019). Firms and Farms: The Local Effects of Farm Income on Firms' Demand. *Cambridge Working Paper in Economics 1924*.

Sekhri, S. and Storeygard, A. (2014). Dowry deaths: Response to weather variability in India. *Journal of Development Economics*, 111: 212–223.

Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica*, pp. 539–591.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, 23(7): 1696–1718.

Vicente-Serrano, S. M., Zouber, A., Lasanta, T., and Pueyo, Y. (2012). Dryness is accelerating degradation of vulnerable shrublands in semiarid Mediterranean environments. *Ecological Monographs*, 82(4): 407–428.

Vogel, C. and Drummond, J. (1993). Dimensions of drought: South African case studies. *GeoJournal*, 30(1): 93–98.

World Bank (2020). World Development Indicators. Washington, D.C.: The World Bank. http://data.worldbank.org/data-catalog/world-development-indicators, (Accessed 10 July 2020).

2.A Additional Tables

Table 2.A.1: Definitions of dependent variables

Dependent variable	**Type**	**Definition**
Employment (rest of labour force)	Binary	1 if an individual is employed; 0 if unemployed.
Agricultural employment (rest of labour force)	Binary	1 if an individual is employed in agriculture; 0 if employed in another sector or unemployed.
Restricted agricultural employment (rest of restricted labour force)	Binary	1 if an individual is employed in agriculture; 0 if employed in another sector or unemployed. (Restricted to people between 20 and 50 years old with less than 10 years of education.)
Restricted agricultural employment (restricted unemployment)	Binary	1 if an individual is employed in agriculture; 0 if unemployed. (Restricted to people between 20 and 50 years old with less than 10 years of education.)
Restricted agricultural employment (restricted employment)	Binary	1 if an individual is employed in agriculture; 0 if employed in another sector. (Restricted to people between 20 and 50 years old with less than 10 years of education.)
Log (hours total)	Continuous	Natural log of monthly hours worked in any sector.
Log (agricultural hours total)	Continuous	Natural log of monthly hours worked in agriculture (either as an employee or as self-employed).
Log (agricultural hours wage employment)	Continuous	Natural log of monthly hours worked in agriculture as an employee.
Wage employment (self-employment)	Binary	1 if an individual is an employee; 0 if in self-employment.
Agriculture wage employment (agriculture self-employment)	Binary	1 if an individual is an employee in an agriculture; 0 if in self-employment in agriculture.
Wage employment with contract (wage employment without contract)	Binary	1 if an individual is an employee with a contract; 0 if an employee without a contract.
Agricultural wage employment with contract (agricultural wage employment without a contract)	Binary	1 if an individual is an employee in agriculture with a contract; 0 if an employee in agriculture without a contract.

Table 2.A.2: The effects of standardised rainy season SPEI on the probability of employment after vs. before the minimum wage law

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res. employ.)
SPEI_RS_Dev	0.0030	0.0010	0.0026	0.0122	0.0007
	(0.006)	(0.004)	(0.004)	(0.008)	(0.006)
SPEI_RS_Dev#POST	0.0004	0.0017	0.0046	0.0025	0.0082
	(0.007)	(0.005)	(0.006)	(0.011)	(0.008)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean SPEI_RS_Dev	0.0638	0.0638	0.0450	0.0166	0.0596
SD SPEI_RS_Dev	0.9640	0.9640	0.9418	0.9211	0.9545
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.3: The effects of standardised rainy season SPEI on the probability of employment after vs. before the minimum wage law

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res employ.)
Avg. marginal effect of SPEI_RS_Dev when:					
POST = 0	0.0030	0.0012	0.0026	-0.0027	-0.0160
	(0.006)	(0.004)	(0.004)	(0.005)	(0.012)
POST = 1	0.0033	0.0027*	0.0072***	0.0027	0.0002
	(0.003)	(0.002)	(0.003)	(0.003)	(0.007)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean SPEI_RS_Dev	0.0638	0.0638	0.0450	0.0166	0.0596
SD SPEI_RS_Dev	0.9640	0.9640	0.9418	0.9211	0.9545
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. After running the regression in Equation 1 with the extra interaction term, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.4: The effects of standardised rainy season SPEI on monthly hours worked and wages after vs. before the minimum wage law

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
SPEI_RS_Dev	-0.0096	-0.0266**	-0.0105	0.0062	-0.0081
	(0.007)	(0.013)	(0.012)	(0.011)	(0.023)
SPEI_RS_Dev#POST	0.0074	0.0189	-0.0001	-0.0123	0.0186
	(0.008)	(0.014)	(0.013)	(0.013)	(0.026)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean SPEI_RS_Dev	0.0778	-0.0138	-0.0757	0.0892	-0.0786
SD SPEI_RS_Dev	0.9802	0.8922	0.9079	0.9771	0.9063
Observations	262,769	45,447	33,439	160,973	30,356

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.5: The effects of standardised rainy season SPEI on monthly hours worked and wages after vs. before the minimum wage law

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
Avg. marginal effect of SPEI_RS_Dev when:					
POST = 0	-0.0096	-0.0266**	-0.0105	0.0062	-0.0081
	(0.007)	(0.013)	(0.012)	(0.011)	(0.023)
POST = 1	-0.0023	-0.0077	-0.0106*	-0.0061	0.0106
	(0.003)	(0.010)	(0.006)	(0.006)	(0.010)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean SPEI_RS_Dev	0.0778	-0.0138	-0.0757	0.0892	-0.0786
SD SPEI_RS_Dev	0.9802	0.8922	0.9079	0.9771	0.9063
Observations	262,769	45,447	33,439	160,973	30,356

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *SPEI_RS_Dev* is the district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. After running the regression in Equation 1 with the extra interaction term, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.6: The effects of negative deviations of rainy season SPEI on the probability of employment

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res. employ.)
NegSPEI_RS_Dev	-0.0043	-0.0024	-0.0070**	-0.0182***	-0.0077
	(0.003)	(0.002)	(0.003)	(0.006)	(0.005)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean NegSPEI_RS_Dev	0.3593	0.3593	0.3591	0.3651	0.3582
SD NegSPEI_RS_Dev	0.5218	0.5218	0.5217	0.5253	0.5222
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.7: The effects of negative deviations of rainy season SPEI on monthly hours worked and wages

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
NegSPEI_RS_Dev	-0.0003	0.0031	0.0099*	-0.0026	-0.0179
	(0.004)	(0.008)	(0.005)	(0.009)	(0.015)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean NegSPEI_RS_Dev	0.3607	0.3639	0.4006	0.3568	0.4016
SD NegSPEI_RS_Dev	0.5221	0.5603	0.5796	0.5099	0.5796
Observations	262,769	45,447	33,439	160,973	30,356

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.8: The effects of negative deviations of rainy season SPEI on the probability of employment after vs. before the minimum wage law

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res employ.)
NegSPEI_RS_Dev	-0.0050	0.0044	0.0041	-0.0043	0.0091
	(0.008)	(0.006)	(0.008)	(0.012)	(0.011)
NegSPEI_RS_Dev#POST	0.0014	-0.0086	-0.0143	-0.0184	-0.0213
	(0.009)	(0.009)	(0.011)	(0.017)	(0.016)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean NegSPEI_RS_Dev	0.3593	0.3593	0.3591	0.3651	0.3582
SD NegSPEI_RS_Dev	0.5218	0.5218	0.5217	0.5253	0.5222
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.9: The effects of negative deviations of rainy season SPEI on the probability of employment

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res employ.)
Avg. marginal effect of NegSPEI_RS_Dev when:					
POST = 0	-0.0050	0.0044	0.0041	-0.0029	-0.0079
	(0.008)	(0.006)	(0.008)	(0.007)	(0.017)
POST = 1	-0.0036	-0.0042	-0.0102*	-0.0027	0.0014
	(0.004)	(0.003)	(0.005)	(0.005)	(0.010)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean NegSPEI_RS_Dev	0.3593	0.3593	0.3591	0.3651	0.3582
SD NegSPEI_RS_Dev	0.5218	0.5218	0.5217	0.5253	0.5222
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. After running the regression in Equation 1 with the extra interaction term, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.10: The effects of negative deviations of rainy season SPEI on monthly hours worked and wages after vs. before the minimum wage law

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
NegSPEI_RS_Dev	-0.0014	-0.0010	-0.0033	-0.0029	-0.0177
	(0.010)	(0.015)	(0.018)	(0.018)	(0.033)
NegSPEI_RS_Dev#POST	0.0013	0.0055	0.0168	0.0004	-0.0002
	(0.013)	(0.019)	(0.022)	(0.023)	(0.037)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean NegSPEI_RS_Dev	0.3607	0.3639	0.4006	0.3568	0.4016
SD NegSPEI_RS_Dev	0.5221	0.5603	0.5796	0.5099	0.5796
Observations	262,769	45,447	33,439	160,973	30,356

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.11: The effects of negative deviations of rainy season SPEI on monthly hours worked and wages after vs. before the minimum wage law

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
Avg. marginal effect of NegSPEI_RS_Dev when:					
POST = 0	-0.0014	-0.001	-0.0038	-0.0029	-0.0199
	(0.010)	(0.015)	(0.018)	(0.018)	(0.029)
POST = 1	-0.0000	0.0045	0.0134*	-0.0025	0.0158
	(0.005)	(0.010)	(0.007)	(0.011)	(0.016)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean NegSPEI_RS_Dev	0.3607	0.3639	0.4006	0.3568	0.4016
SD NegSPEI_RS_Dev	0.5221	0.5603	0.5796	0.5099	0.5796
Observations	262,769	45,447	33,439	160,973	30,356

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. After running the regression in Equation 1 with the extra interaction term, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.12: The effects of negative deviations of rainy season SPEI on the probability of employment after vs. before the minimum wage law considering the size of the wage gap

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res employ.)
NegSPEI_RS_Dev	-0.0255*	-0.0104	-0.0064	-0.0098	-0.0113
	(0.015)	(0.014)	(0.019)	(0.032)	(0.024)
NegSPEI_RS_Dev#POST	0.0213	0.0322*	0.0338	0.0219	0.0540*
	(0.018)	(0.018)	(0.024)	(0.039)	(0.028)
NegSPEI_RS_Dev#WG	0.0160	0.0100	0.0039	-0.0039	0.0117
	(0.015)	(0.015)	(0.022)	(0.034)	(0.029)
POST#WG	-0.0139	-0.0022	0.0088	-0.0226	0.0276
	(0.027)	(0.019)	(0.022)	0.039	(0.030)
POST#NegSPEI_RS_Dev#WG	-0.0163	-0.0530**	-0.0638**	-0.0561	-0.1002***
	(0.021)	(0.023)	(0.030)	(0.047)	(0.036)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean *wage gap*	0.7505	0.7505	0.7688	0.7744	0.7630
SD *wage gap*	0.2699	0.2699	0.2807	0.2874	0.2875
Mean NegSPEI_RS_Dev	0.3593	0.3593	0.3591	0.3651	0.3582
SD NegSPEI_RS_Dev	0.5218	0.5218	0.5217	0.5253	0.5222
Observations	394,858	394,858	187,909	89,428	128,573

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). $NegSPEI\_RS\_Dev$ is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.13: The effects of negative deviations of rainy season SPEI on the probability of employment after vs. before the minimum wage law considering the size of the wage gap

	(1) Employment (rest of LF)	(2) Ag. employ. (rest of LF)	(3) Ag. employ. (rest of res. LF)	(4) Ag. employ. (res. unemploy.)	(5) Ag. employ. (res employ.)
Avg. marginal effect of NegSPEI_RS_Dev when:					
WG = low & POST = 0	-0.0178	-0.0055	-0.0045	-0.0117	-0.0057
	(0.0090)	(0.0011)	(0.0092)	(0.0176)	(0.0122)
WG = low & POST = 1	-0.0044	0.0113	-0.0018	-0.0171	0.0007
	(0.0044)	(0.0040)	(0.0055)	(0.0068)	(0.0066)
WG = med. & POST = 0	-0.1350*	-0.0028	-0.0034	-0.0128	-0.0023
	(0.0071)	(0.0044)	(0.0056)	(0.0114)	(0.0077)
WG = med. & POST = 1	-0.0045	-0.0105***	-0.0187***	-0.0343***	-0.0247***
	(0.0058)	(0.0039)	(0.0047)	(0.0091)	(0.0055)
WG = high & POST = 0	-0.0092	-0.0001	-0.0023	-0.0139	0.0011
	(0.0072)	(0.0047)	(0.0072)	(0.0118)	(0.0104
WG = high & POST = 1	-0.0045	-0.0221***	-0.0355***	-0.0516***	-0.0501***
	(0.0085)	(0.0058)	(0.0068)	(0.0144)	(0.0077)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	0.6997	0.0771	0.1099	0.2578	0.1608
Mean *wage gap*	0.7505	0.7505	0.7688	0.7744	0.7630
SD *wage gap*	0.2699	0.2699	0.2807	0.2874	0.2875
Mean NegSPEI_RS_Dev	0.3593	0.3593	0.3591	0.3651	0.3582
SD NegSPEI_RS_Dev	0.5218	0.5218	0.5217	0.5253	0.5222
Observations	385,177	385,177	184,718	88,436	126,347

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD. After running the regression in Equation 2, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.14: The effects of negative deviations of rainy season SPEI on hours worked and wages after vs. before the minimum wage law considering the size of the wage gap

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
NegSPEI_RS_Dev	-0.0236	-0.0329	-0.0491	-0.0079	0.1694**
	(0.029)	(0.041)	(0.050)	(0.047)	(0.076)
NegSPEI_RS_Dev#POST	0.296	0.0713	0.0592	0.0249	0.1726*
	(0.039)	(0.061)	(0.058)	(0.054)	(0.088)
NegSPEI_RS_Dev#WG	0.0294	0.0576	0.0608	0.0195	0.2138**
	(0.034)	(0.042)	(0.066)	(0.051)	(0.097)
POST#WG	0.0722	0.2927***	0.0404	0.1147	0.2897***
	(0.045)	(0.104)	(0.071)	(0.073)	(0.084)
POST#NegSPEI_RS_Dev#WG	-0.0327	-0.0743	-0.0442	-0.0440	-0.1958*
	(0.051)	(0.067)	(0.075)	(0.066)	(0.100)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean *wage gap*	0.7364	0.7670	0.6926	0.7377	0.6993
SD *wage gap*	0.2744	0.3401	0.3128	0.2631	0.3128
Mean NegSPEI_RS_Dev	0.3607	0.3639	0.4006	0.3568	0.4016
SD NegSPEI_RS_Dev	0.5221	0.5603	0.5796	0.5099	0.5796
Observations	255,887	45,386	33,395	157,185	30,320

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). $NegSPEI\_RS\_Dev$ is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD.

Table 2.A.15: The effects of negative deviations of rainy season SPEI on hours worked and wages after vs. before the minimum wage law considering the size of the wage gap

	(1) Log (hours total)	(2) Log (ag. hours total)	(3) Log (ag. hours wage emp.)	(4) Log (wages)	(5) Log (ag. wages)
Avg. marginal effect of NegSPEI_RS_Dev when:					
WG = low & POST = 0	-0.0100	-0.0084	-0.0260	0.0014	-0.0868**
	(0.146)	(0.026)	(0.027)	(0.026)	(0.043)
WG = high & POST = 1	0.0045	0.0312**	0.0164*	0.0055	0.0101
	(0.007)	(0.015)	(0.009)	(0.011)	(0.023)
WG = med & POST = 0	-0.0019	0.0112	-0.0070	0.0065	-0.0199
	(0.009)	(0.016)	(0.016)	(0.017)	(0.027)
WG = med & POST = 1	0.0036	0.0256**	0.0216**	-0.0009	0.0158
	(0.0059)	(0.012)	(0.008)	(0.013)	(0.016)
WG = high & POST = 0	0.0061	0.0308*	0.0120	0.0117	0.0470
	(0.011)	(0.016)	(0.026)	(0.017)	(0.037)
WG = high & POST = 1	0.0026	0.0199	0.0268**	-0.0074	0.0214
	(0.0099)	(0.017)	(0.012)	(0.020)	(0.017)
Individual controls	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y	Y
Mean dep. var.	5.1381	5.0626	5.2625	8.1133	7.2439
Mean *wage gap*	0.7364	0.7670	0.6926	0.7377	0.6993
SD *wage gap*	0.2744	0.3401	0.3128	0.2631	0.3128
Mean NegSPEI_RS_Dev	0.3607	0.3639	0.4006	0.3568	0.4016
SD NegSPEI_RS_Dev	0.5221	0.5603	0.5796	0.5099	0.5796
Observations	255,887	45,386	33,395	157,185	30,320

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD. After running the regression in Equation 2, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Table 2.A.16: The effects of negative deviations of rainy season SPEI on the probability of wage employment and wage employment with a contract after vs. before the minimum wage law considering the size of the wage gap

	(1)	(2)	(3)	(4)
	Wage employ. (Self-employ.)	Ag. wage employ. (Ag. self-employ.)	Wage employ. w. contract (Wage emp. w/o contract)	Ag. wage employ. w. contract (Ag. wage employ. w/o contract)
NegSPEI_RS_Dev	0.0143	-0.0024	-0.0585	-0.0280
	(0.019)	(0.035)	(0.046)	(0.060)
NegSPEI_RS_Dev#POST	-0.0065	-0.0252	0.0627	0.0644
	(0.021)	(0.040)	(0.055)	(0.069)
NegSPEI_RS_Dev#WG	0.0192	0.0094	0.0285	-0.0026
	(0.018)	(0.032)	(0.048)	(0.042)
POST#WG	0.0182	0.0671	0.0342	0.0576
	(0.033)	(0.065)	(0.076)	(0.065)
POST#NegSPEI_RS_Dev#WG	0.0061	0.0559	-0.0458	-0.0720
	(0.022)	(0.041)	(0.062)	(0.055)
Individual controls	Y	Y	Y	Y
District FEs	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y
Mean dep. var.	0.7762	0.6750	0.6250	0.3230
Mean *wage gap*	0.7391	0.7753	0.7181	0.7753
SD *wage gap*	0.2759	0.3424	0.2653	0.3424
Mean NegSPEI_RS_Dev	0.3587	0.3616	0.3589	0.3616
SD NegSPEI_RS_Dev	0.5273	0.5604	0.5270	0.5604
Observations	263,513	47,388	203,866	47,388

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). $NegSPEI\_RS\_Dev$ is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race.

Table 2.A.17: The effects of negative deviations of rainy season SPEI on the probability of wage employment and wage employment with a contract after vs. before the minimum wage law considering the size of the wage gap

	(1)	(2)	(3)	(4)
	Wage employ. (Self-employ.)	Ag. wage employ. (Ag. self-employ.)	Wage employ. w. contract (Wage emp. w/o contract)	Ag. wage employ. w. contract (Ag. wage employ. w/o contract)
Average marginal effect of NegSPEI_RS_Dev when:				
WG = low & POST = 0	-0.0053	0.0016	-0.0456*	-0.0291
	(0.012)	(0.023)	(0.027)	(0.004)
WG = low & POST = 1	-0.0091**	0.0006	-0.0037	0.0041
	(0.004)	(0.009)	(0.008)	(0.016)
WG = med. & POST = 0	-0.0000	0.0049	-0.0380**	-0.030
	(0.008)	(0.017)	(0.018)	(0.323)
WG = med. & POST = 1	-0.0021	0.0229**	-0.0082	-0.0214*
	(0.005)	(0.010)	(0.008)	(0.013)
WG = high & POST = 0	0.0052	0.0081	-0.0304*	0.0309
	(0.007)	(0.016)	(0.016)	(0.020)
WG = high & POST = 1	0.0049	0.0453***	-0.0128	-0.0469***
	(0.008)	(0.017)	(0.013)	(0.015)
Individual controls	Y	Y	Y	Y
District FEs	Y	Y	Y	Y
Period FEs	Y	Y	Y	Y
District-period trends	Y	Y	Y	Y
Mean dep. var.	0.7762	0.6750	0.6250	0.3230
Mean *wage gap*	0.7391	0.7753	0.7181	0.7753
SD *wage gap*	0.2759	0.3424	0.2653	0.3424
Mean NegSPEI_RS_Dev	0.3587	0.3616	0.3589	0.3616
SD NegSPEI_RS_Dev	0.5273	0.5604	0.5270	0.5604
Observations	263,513	47,388	203,866	47,388

Analysis at individual level using labour force survey data (Sept 2001-Sept 2007). *NegSPEI_RS_Dev* is the transformed district rainy season SPEI deviation (in the previous period) from the long run local mean normalised by its standard deviation where positive values are recoded as zero and absolute values are taken of negative values. Standard errors clustered at the district level in parenthesis. Significance levels: *p<0.1, **p<0.05, ***p<0.01. With district and year fixed effects and linear district-period time trends, and individual controls for age, education, sex and race. *Low wage gap* = mean wage gap - 1 SD; *med. wage gap* = mean wage gap; *high wage gap* = mean wage gap + 1 SD. After running the regression in Equation 2, average marginal effects are calculated using the *margins* command in Stata with the *dydx()* option specified to obtain the derivative with respect to the weather measure and with the option *vce(unconditional)* specified to ensure the correct clustering of standard errors.

Chapter 3

Agglomeration economies in a developing country: Evidence from geo-coded micro-panel data in South Africa

3.1 Introduction

Are agglomeration effects much higher in developing than developed countries or are they just overestimated? The few studies on agglomeration in developing countries have found agglomeration elasticities often of 8% to 20% (Chauvin et al., 2017; Combes et al., 2017), considerably higher than the elasticities averaging 3% estimated in rigorous developed country studies. In theory, underdeveloped factor markets in developing countries may enhance the productivity benefits of cities relative to those in advanced countries. However, due to data constraints, agglomeration studies focused on developing country papers are not able to include all the endogeneity controls used in the developed country papers and are often forced to work with geographical units that are far removed from local labour markets. In this study, I make use of a unique geo-coded panel micro-dataset for South Africa, which allows me to rigorously investigate most of the first order issues addressed in the developed country literature.

As extensively theorised in urban economics, the productivity benefits of cities derive from reduced transport costs, easier access to consumers and suppliers, technology and knowledge spill-overs and improved matching between workers and firms (see reviews in Ciccone and Hall (1996); Duranton and Puga (2004)). Since according to neoclassical theory, in competitive labour markets, wages are equal to the marginal product of labour, agglomeration economies should result in higher wages in larger cities as compared to smaller urban areas or rural areas (Combes et al., 2008; Glaeser and Mare, 2001).[1] The identification challenge is that there are several other reasons why wages might be higher in large cities (Combes et al., 2010). Highly skilled workers may sort into larger cities because they are more specialised in skill-intensive industries or because these workers stand to benefit more from agglomeration economies. Another possibility is that favourable local endowments (including geographical amenities, public capital or institutions) could facilitate greater productivity, resulting in higher wages for workers and further agglomeration.[2]

Assuming all the endogeneity issues mentioned above were controlled for, why might agglomeration elasticities still be higher in developing countries? One theory is that the higher trade costs in developing countries means that the proximity to suppliers of intermediate goods becomes more important (Duranton, 2014). Another related theory is that large cities have a much higher level of connectivity to the outside world relative to other parts of a developing country (Glaeser and Xiong, 2017). Technology entering through cities may give firms there a productivity boost relative to those in less well-connected places.

I examine the relationship between urban population and nominal wages in South Africa for the period 2008-2016 using all five rounds of the National Income Dynamics

[1]While the first studies of agglomeration economies estimated production functions using cross-sectional aggregate data, more recent studies tend to rather use micro-data, either at the firm or worker level, to estimate production or wage functions. The advantages of micro-level data are that they can better represent firm optimising behaviour posited in economic theory, they can limit aggregation bias linked to unobserved heterogeneity and they can provide more variation (Melo et al., 2009).

[2]There is also some evidence that firms have less monoposony power in cities, which could lead to higher wages (Manning, 2010). This is, however, arguably related to there being thicker labour markets in cities, which is a foundational tenet of agglomeration theory (Proost and Thisse, 2019).

Survey, a nationally-representative panel study. I map individuals to a geographic level that approximates local labour markets as closely as possible. To control for possible bias caused by the sorting of workers of different skills, I first include a battery of controls for observed ability. However, since I have access to an individual-level panel of workers, my paper is able to go further than the other developing country papers in controlling for unobserved characteristics that may be correlated with location choices - a critical feature of the state-of-the-art empirical research on agglomeration effects (Combes and Gobillon, 2015).[3] I am effectively able to 'follow' workers across the country to observe how their wages change as the employment activity around them changes. To control for simultaneity bias (or omitted variable bias), I construct instruments from historical population data (specifically, from a hard copy of the 1960 census) that have not been used before in the South African literature. Owing to South Africa's unique history of controlled migration and forced reallocation of the population across space, there is a strong case to be made for the exclusion restriction for this instrument.

My paper contributes to the literature in several ways. Firstly, there have been few rigorous studies of agglomeration effects in developing countries, despite numerous reviews having highlighted that it would be very useful to compare results from developing countries to the robust results obtained for more developed countries (Duranton, 2014; Henderson, 2005; Overman and Venables, 2005). There have been some attempts at measuring agglomeration elasticities in developing countries with cross-sectional data. Combes et al. (2017) estimate agglomeration externalities of 9 to 10% for China. In another paper, Chauvin et al. (2017) compare population/population density elasticities between the US, Brazil, China and India. While Brazil's population elasticity is measured around 5% (similar to that in the US), the

[3]Note that I focus on measuring static agglomeration economies in South Africa. While panel data has been used in advanced country studies to estimate dynamic agglomeration economies (i.e. the effects of agglomeration on productivity growth) (e.g. Roca and Puga (2017)), my series is unfortunately not long enough (and the dates between waves too close to one another) to precisely estimate these effects (cf. Henderson and Kriticos (2018)).

authors find much higher elasticities for India and China (around 8%).[4] They also find an even higher density elasticity for India of 19%. In a related paper, Hering and Poncet (2010) examine the relationship between individual wages and market access in China, estimating an elasticity of 14%. In several of these papers, the authors admit that they may be overestimating elasticities but are constrained by the data available to them. In one particularly careful study, Duranton (2016) estimates an elasticity of wages with respect to city population of about 5% for Colombia using national household surveys, though this paper is also limited by the non-availability of panel data.

Secondly, I contribute to the spatial economics literature on South Africa. Several studies have looked at spatial inequality in South Africa at the regional level. Topics have included the determinants of economic growth at the subnational level (Naudé and Krugell, 2006) and the evolution of regional GDP (Msulwa and Turok, 2013). More directly related to the topic at hand, in an attempt at a firm-level analysis of agglomeration in South Africa, Krugell and Rankin (2012) use cross-sectional World Bank Enterprise Survey data from 2003 and 2007. The paper is largely descriptive and limited by the fact that the authors only have data for four South African cities at the city level. Most of these previous studies have been carried out at the regional or industry level and have resorted to using crude measures of output or regional income generated by market research consultancies.

Thirdly, I contribute to the debate on spatial development in policy-making circles in South Africa. For the most part, post-Apartheid policy economic policy has paid little attention to subnational impacts and regional differentiation. What attention has been paid to spatial aspects of development has been almost exclusively focused on closing geographical disparities in output and employment (Robbins, 2015; Todes and Turok, 2018). Knowing more about the benefits from and costs of cities in South Africa is important, so that policy can be devised on the basis of systematic evidence.

[4] In Chauvin et al. (2017), China's population elasticity is statistically insignificant, so this estimate is only tentative.

In my preferred specification including both individual fixed effects and a historical population instrument, I estimate an elasticity of wages to city population of approximately 3%, which is robust to a large number of tests. I find that using individual effects to control for sorting on unobservables reduces the estimated agglomeration elasticity for South Africa by 18-37% (depending on the specification) - in line with effect sizes in the developed country literature (Baum-Snow and Ferreira, 2015). I argue that these results provide suggestive evidence that the very high agglomeration elasticities found for other developing countries may in part be due to mismeasurement. Controlling for market access - the urban employment population of surrounding cities - has little effect on the elasticities I estimate. Once I control for individual fixed effects, I do not find evidence of heterogeneous effects for workers based on skill level, participation in the informal/formal economy or gender. I also examine the comparative effects of human capital externalities and agglomeration externalities, finding that agglomeration externalities seem to be much more important in predicting wages.

The paper is organised as follows: Section 2 describes the spatial economy in South Africa. Section 3 describes the data and the empirical approach. Section 4 presents results with robustness checks and examines heterogeneous effects. Section 5 concludes.

3.2 Context

South Africa is renowned as having one of the highest levels of inequality in the world. Inequality has remained stubbornly high since the end of Apartheid in 1994 (Leibbrandt et al., 2016). In 2015, the Gini coefficient based on income was estimated to be 0.68 and the Gini coefficient based on expenditure was estimated to be 0.64 (Statistics SA, 2017). The poorest 60% of the population earn about 10% of the total income of the population (Leibbrandt et al., 2016). It is important to know how much of these disparities in income can be explained by agglomeration effects. While many of the previous studies on the spatial economy of South Africa have relied on questionable data, there is at least preliminary evidence that the country has a very high level of spatial economic unevenness (Turok, 2012). It has been estimated, for example, that 20% of cities and towns produce 82% of South Africa's GDP (Krugell and Naude, 2005).

Much of the regional inequality is probably driven by uneven spatial sorting of skills and by agglomeration forces favouring localised growth poles. However, South Africa also has an unusual history of restricted urbanisation (at least for the majority black population) and spatial segregation along race lines with development being deliberately focused on predominantly white urban centres.

From the discovery of diamonds and gold in South Africa in the 17th and 18th century until the end of the 20th century, urbanisation of the black population was restricted by pass laws (put in place by the whites-only national government), which required that black mine workers carry a pass at all times and prevented them from bringing their families to cities. Then with the 1913 Native Land Act, black people were prevented from owning land outside rural 'reserves' in the former native African states. The reserves (known as 'Bantustans') were mostly arid and without resources, and they were far away from the main economic centres. After World War II, Apartheid (Afrikaans for 'separateness') became official state policy and the government increased influx controls. It forcibly relocated approximately 3.5 million black people from white areas to Bantustans and hundreds of thousands of people were arrested every year for disobeying pass laws (Ogura et al., 1996). The 1950 Group Areas Act prescribed the racial composition of every residential area, so that black people who had formal jobs in cities were forced to live in informal settlements (known as 'townships') on their outskirts (Christopher, 2001).

The end result of these policies was that there was considerable spatial divergence between the economic centres of the country and the dense rural reserves, and within cities between inner city areas and the surrounding informal settlements (Turok, 2012). While today, approximately two-thirds (65%) of South Africa's total population of 55 million live in urban areas (World Bank, 2019), the urbanisation level might have been much higher in the absence of Apartheid and the earlier system of spatial segregation (Krugell and Naude, 2005). Under-urbanisation and high transport costs are likely to have undermined agglomeration economies.

Restrictions on migration were abandoned in 1986 before the election of the first democratic government in 1994, which hastened rural-urban migration. International sanctions were lifted, opening up the economy to globalisation forces. As elsewhere, this has led to the growth of economic activities in areas connected to international trade (i.e. major cities and harbours). These developments have put added pressure on politicians to try to spread jobs and wealth to rural areas neglected during Apartheid and now by global economic forces. A variety of rural development schemes have or are in the process of being tried, including public works employment projects and industrial decentralisation initiatives (Todes and Turok, 2018). While up until recently the national government seemed oblivious to market forces favouring cities, it has recently acknowledged their significance, and the difficulty of working against them (Farole and Sharp, 2017).

While there are some historical peculiarities in South Africa that have affected the process of urbanisation, the country is still comparable to other developing countries for which agglomeration externalities have been estimated. While this paper does not aim to comprehensively compare the factor markets in South Africa against those in China, India and Brazil, the markets would not seem to be very different. For example, all four countries receive similar scores for the quality of their trade and transport-related infrastructure.[5] While per capita GDP in India is approximately one-third of that in Brazil, per capita income levels in China, Colombia and South Africa are very similar, half way between the two extremes.[6]

[5]In the World Development Indicators for 2012 (chosen because 2012 is the midpoint of my period of study), South Africa received 3.79 for this indicator, while Brazil received 3.07, China received 3.59 and India received 3.08 (World Bank, 2019).

[6]In 2012, the official per capita GDP (constant 2011 PPP $) in Brazil was $15,215, in China was $11,115, in Colombia was $11,996, in India was $4817, and in South Africa was $12,253.

3.3 Data & econometric approach

3.3.1 Data

The paper makes use of the South African National Income Dynamics Survey (NIDS). NIDS is South Africa's first nationally representative panel study. The survey provides detailed information on income, education and employment as well as many other variables. There are approximately 7900 households per wave. The survey revisits the same households each wave and tracks household members even when they migrate to other parts of the country.[7]

Due to the so-called 'modifiable unit area problem', the choice of geographic level can affect statistical results (e.g. Briant et al. (2010)). It is therefore preferable to work with regions that closely approximate local labour markets.[8] The secure version of NIDs provides the exact geo-coordinates of households, which can then be matched to local municipalities.[9][10] Local municipalities are the contemporary equivalent of magisterial districts (phased out as administrative units in 2001), which have been used in several papers (focusing on earlier periods) to approximate local labour markets (e.g. Dinkelman and Ranchhod (2012); Magruder (2012)). In 2011 South African municipalities covered on average 4669 square kilometres and the average population of local municipalities was 221,241.

Local municipality boundaries from the 2011 census were used, since this was the last

[7]In terms of attrition, 78% of the individuals who were interviewed in Wave 1 were successfully interviewed in Wave 4. This is a relatively low attrition rate, especially when compared to similar panel surveys in developing countries. Survey refusal is just as likely as non-contact as a cause of attrition (Brophy et al., 2018). Since my main specification (with individual fixed effects) gets identification from movers, this means that attrition in NIDS is less of a problem for my paper than it would otherwise be.

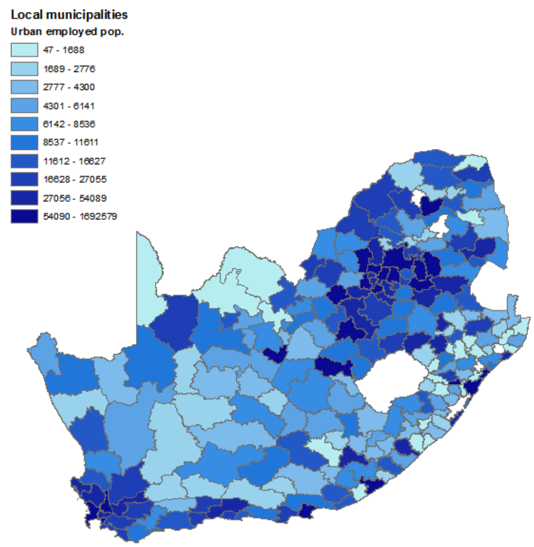
[8]One potential disadvantage of the NIDS dataset is that it only provides information on where people live, rather than where they work, yet much of the existing agglomeration literature is concerned with the link from work place size to wages. In practice, this should not be a major problem, since local municipalities are large enough such that places where people live are also likely to be places where people work.

[9]In order to preserve anonymity, the publicly available version of NIDs provides limited information on the geographic location of respondents – down to the district municipality level.

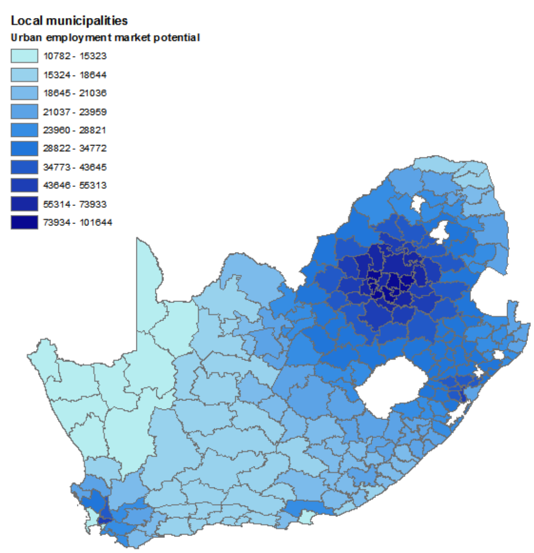
[10]This was carried out using the 'gpsbound' command using the 2011 Census shape-files.

Figure 3.1: Variation of urban employment population and market potential across local municipalities in 2011.

(a) Urban employment population



(b) Urban employment market potential



The market potential measure is calculated as the inverse-weighted sum of urban employment population over all municipalities other than the municipality in question (Harris, 1954). Data from Census 2011.

time a full national census was carried out. This ensures that population information and other regional information are as accurate as is possible. There were 234 local municipalities in the 2011 geographic frame. Eight of these were classified as metropolitan municipalities: City of Johannesburg, City of Tshwane (Pretoria), City of Cape Town, Ekurhuleni (East Rand), Nelson Mandela Metro (Port Elizabeth), Ethekweni Metro (Durban), Manguang Metropolitan Municipality and Buffalo City (East London). Metropolitan municipalities include the largest cities in South Africa as well as surrounding towns. There were also 226 non-metropolitan local municipalities, which encompass (sometimes multiple) smaller cities and towns as well as rural areas.

I focus for the most part on the urban portion of local municipalities as defined in the 2011 Census, though, for comparison, I also present results where the main variable of interest is based on total municipal population including that in rural areas. Since agglomeration benefits are most likely to come from being surrounded by people who are economically active, my main specifications focus on the employed population.[11] Two possible candidates for my main measure of agglomeration then include urban municipal employment population and urban municipal population density. As explained in Chauvin et al. (2017), both measures are useful. If there is much arbitrary variation in how tightly municipal borders are drawn around agglomerations, then population will be a better measure than density. On the other hand, if local municipalities sometimes include multiple towns or cities, density may be the preferred measure of agglomeration. As to be expected, given the high correlation between population and density, I find that the results are very similar with the two different specifications. For my main results, I use urban municipal employment population together with a control for (log) municipal land area in keeping with the specification in Duranton (2016). Summary statistics for local municipalities can be found in Table 3.A.1 of the Appendix.

[11]The distinction between employed and unemployed is particularly important in South Africa, where unemployment levels are very high (around 25-35% depending on the definition of unemployment used.)

This paper makes use of pre-tax monthly wages (deflated to a 2013 base year using the South African CPI) that individual workers received from their primary occupations.[12] Since there are some concerns over self-reported working hours, the analysis mainly focuses on the total wage for workers who reported working fewer than 15 hours per week (though I also report results with hourly wages as the dependent variable, which are never significantly different from my main results). I include both men and women in my sample, though I also examine the separate effects on each gender in Section 4.5. I include both wages in both the formal and informal sectors so as to avoid the potential selection issues associated with only focusing on formal sector employees and because it is important to consider informal sector employment in developing countries like South Africa. Since the earnings of the informally employed may be differentially distorted across municipalities, I estimate agglomeration elasticities separately for workers who are formally and informally employed in Section 4.5. I drop individuals with no labour income and the 1% of workers with the lowest and highest wages. I also drop public sector workers whose wages are not set by market mechanisms. Individuals with missing values for age (16 observations) are also dropped. These restrictions leave a base sample of 4518 workers contributing 12070 wage observations over the years 2008-2016 for empirical analysis. For the fixed effects analysis (described below), identification is from movers - those workers that moved across local municipality boundaries during the period in question, earning wages in both sending and receiving municipalities - which include 529 workers contributing 1472 observations.

Table 3.1 shows descriptive statistics for the sample. Comparing the sample of non-movers workers (Column 1) to the sample of movers (Column 4) shows that movers tend to be slightly younger, better-educated and higher-earning (as has been found to be the case in other countries cf. D'Costa and Overman (2014)), but that these changes are not particularly large.[13] Columns 2 and 3 also compare non-movers living

[12]I do not consider income from the self-employed (here I depart from Duranton (2016), whose sample includes self-employed in both the formal and informal sectors) as I am concerned that income may get confounded with profits.

[13]It is important to note here that this exercise only shows that movers and non-movers are similar

Table 3.1: Descriptive statistics for base sample

	Non-movers (observed at least twice)						Movers	
	Total		High UEP munic.		Low UEP munic		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Monthly gross wages	6737.6	7602.6	7874.0	8411.6	5812.4	6734.4	7669.8	8308.4
Female (%)	48.7	50.0	49.0	50.0	48.7	50.0	37.1	48.3
Avg. age	37.9	10.6	37.0	10.6	38.6	10.6	33.8	9.2
Age 16-29 (%)	26.3	44.1	30.0	45.8	23.4	42.4	37.8	48.5
Age 30-43 (%)	42.7	49.5	42.2	49.4	43.1	49.5	45.7	49.8
Age 44-57 (%)	26.9	44.3	23.9	42.7	29.2	45.5	14.9	35.7
Age 58-70 (%)	4.1	19.8	3.9	19.4	4.2	20.1	1.5	12.1
Black (%)	57.2	49.5	64.2	47.9	51.4	50.00	65.9	47.4
Coloured (%)	32.6	46.9	25.8	43.7	38.2	48.6	3.1	17.4
Asian/Indian (%)	6.0	23.7	4.1	19.7	7.5	26.3	3.1	17.4
White (%)	3.4	18.22	4.96	21.7	2.2	14.6	2.8	16.5
Avg. years of educ.	9.8	3.4	10.5	2.8	9.2	3.7	10.5	3.1
Grade 1-5 (%)	6.7	25.0	4.0	19.6	8.9	28.4	4.3	20.2
Grade 6-8 (%)	15.0	35.7	11.4	31.7	17.9	38.4	9.2	29.0
Grade 9-11 (%)	33.2	47.1	35.4	47.8	31.4	46.4	32.2	46.7
Grade 12 (%)	34.8	47.6	40.5	49.1	30.2	45.9	43.8	49.6
Technikon diploma (%)	0.5	7.3	0.6	7.7	0.5	7.0	0.3	5.2
University degree (%)	5.4	22.5	6.2	24.1	4.7	21.1	6.7	25.1
Trade union (%)	29.8	45.7	31.8	46.6	28.2	45.0	26.8	44.3
Informal (%)	26.8	44.3	23.6	42.5	29.3	45.5	28.2	45.0
Married (%)	33.7	47.3	33.5	47.2	33.9	47.3	24.6	43.1
Observations	12070		5387		6633		1472	
Individuals	4518		2067		2451		529	

in the 30 local municipalities with the highest urban employment population to those living in other municipalities. Mean gross wages are 26.3% higher in the municipalities with highest urban employment population (R7874) as compared to those in the other municipalities (R5812). To investigate the possible sources of this wage rate differential, statistics are also reported on various other individual characteristics of the workers in the sample. White and black workers make up a larger proportion of the population in high urban employment population municipalities than in other municipalities, whereas 'coloured' (mixed race) workers (a historically disadvantaged group) and Asian/Indians are mostly found in lower urban employment population municipalities. A higher proportion of workers in high urban employment population municipalities have completed secondary school (Grade 12) and higher degrees than in the other municipalities. Working in the informal sector is slightly more common

when it comes to observable characteristics. There may still be differences when it comes to unobservable characteristics.

amongst workers in lower urban employment population municipalities. This basic comparison suggests that there are some systematic differences in workers' characteristics across space, which need to be accounted for in any analysis of regional wages.

I also compare statistics between movers who are captured in the 30 highest urban employment population municipalities versus movers in the rest (see Table 3.A.2 in the Appendix). The patterns of differences between high and lower urban employment population locations for movers are very similar to those described above for the sample of non-movers. This mitigates concerns about sample selection issues between movers and non-movers (cf. D'Costa and Overman (2014)).

3.3.2 Econometric approach

I estimate variants of the following regression:

$$\ln w_{it} = \beta_1 + \ln E_{imt}\beta_2 + X_{it}\beta_3 + \gamma_t + \delta_d + \epsilon_{it} \tag{3.1}$$

The outcome variable of the Mincerian wage function is the natural logarithm of worker i's total wages from primary activities in a municipality m at time t. The main variable of interest is $\ln E_{imt}$, the natural logarithm of the total number of workers in each worker's municipality of residence.[14] The other explanatory variables are as follows: X_{it} is a vector of individual characteristics (age, race, marital status, education, whether the worker is in formal or informal employment and trade union membership). γ_t is a time effect to control for macro level changes in wage rates that are common to all individuals.

The first step of the approach is to run a basic pooled OLS estimation for the relationship between wage rates and the number of workers in municipalities of residence. However, an immediate concern is that higher wages in large cities could then just reflect a higher concentration of skilled workers. Typically industries demanding

[14]For the main analysis, population information is from the 2011 National Census.

high skills - such as universities and multinational companies - locate in large cities. Large cities also offer consumption amenities which may be particular attractive to highly educated workers. It is therefore important to include controls for individual worker characteristics that could affect both sorting and wages. While this provides a decent first approximation of regional wage elasticity, it is not able to take into account unobserved individual ability, which is unlikely to be distributed evenly and is therefore likely to bias coefficients (Glaeser and Mare, 2001). To try to tackle this issue an individual-level fixed effects estimation is carried out. This removes that part of omitted ability which is worker-specific and time-invariant, and identification then comes off movers.

While using fixed effects is seen as the best possible approach in the absence of random allocation (D'Costa and Overman, 2014), there may be concerns with the representativity of the population of movers (Combes et al., 2008). However, as discussed in Section 3.1, I find no substantial differences between the mover sample and my full sample. Another possible concern with focusing on movers is that there may be endogeneity at the individual level when workers' location choices are based on the exact wage that they get in particular places, typically when they receive job offers associated with known wages (Combes and Gobillon, 2015). However, in a developing country like South Africa, it is much less likely than in developed countries that people migrate in response to concrete wage offers. In support of this argument, in the 2001-2002 HSRC National Migration Survey, the most important reason male migrants cited for moving across districts was 'looking for work', whereas female migrants moved mainly because of 'getting married or moving in with a partner' or 'getting separated or divorced' (Wentzel et al., 2006). Very few migrants reported moving to take up a job.

The above concerns endogeneity at the individual level but there could also be endogeneity at the local level. As mentioned, it could be the case that the presence of certain endowments - better institutions and technology - in large cities could directly facilitate greater productivity, resulting in higher wages for workers and further agglomeration (Combes et al., 2010). This could either be seen as a reverse causality problem or an omitted variable problem - with the variable being the amenity that determines both agglomeration and wages (Combes and Gobillon, 2015).[15]

The paper therefore uses an instrumental variable strategy examining the impact of historical population levels in a two stage least squares (2SLS) estimation.[16] In order to control for potential simultaneity bias, an instrument should be able to predict worker population but should otherwise be uncorrelated with the productivity differences in the model. Long lags of population or population density have been used since Ciccone and Hall (1996). The rationale for this is that the factors which determined population location in the past may be different from the factors determining population location today (satisfying the exclusion restriction for a valid instrument). At the same time, population patterns tend to be relatively persistent over time partly as a result of the durability of infrastructure and housing (satisfying the relevance condition for a valid instrument). This paper instruments municipal-level urban employment populations with estimated municipal urban populations in South Africa in 1960.

Information on the historical populations of towns and cities in South Africa was obtained from the hardcopy of the 1960 national census. The geo-coordinates of these towns and cities were used to map them to 2011 local municipality census boundaries. Unfortunately, no information was available for 36 (mostly low urban employment population) current-day municipalities in 1960, so the 2SLS estimations are run on a slightly smaller sample of observations (11,440 instead of 12,070). Table 3.A.3 in

[15]Of course it is also possible that local amenities could bias estimates of regional wage elasticity downwards e.g. if consumption amenities make the price of land higher without increasing productivity (Combes and Gobillon, 2015).

[16]To address endogeneity at the local level, another possible approach is to include time-invariant local fixed effects in specifications estimated on panel data to deal with missing local variables that are constant over time. However, apart from this requiring a long panel and much time-variation in population/density over time, there are other drawbacks, including an inability to deal with time-varying omitted variables as well as reverse causality (for further discussion cf. Combes and Gobillon (2015)). Duranton (2016) tries to include regional fixed effects in his regressions for Colombia, finding that while fixed effects estimates do not not contradict cross-sectional estimates, they do not provide enough precision to be useful.

Appendix replicates Table 3.1 focusing on this slightly reduced sample, finding very similar descriptive statistics to those for the full sample.

While a historical population instrument is expected to be relevant, showing that it passes the exclusion restriction is not quite as straightforward. However, the South African economy over the 2008-2016 period was much changed from what it was in 1960, suggesting that the historical determinants of population location are not the same as those during my period of study. As explained in the background section, in 1960 separationist policies in South Africa were in full swing, and the majority black population was restricted to certain areas of the countries and prevented from participating properly in the economy. Whereas since the turn of the century, South Africa has undergone rapid deindustrialisation after liberalising its economy, the country was industrialising in the 1950s (Carmody, 2002). Furthermore, as is the case in most other countries that have reached middle-income status, production techniques used in manufacturing, agriculture and service industries today are dramatically different from what they were in 1960.

Lastly, this paper also considers the effects of market access on productivity. Better access to outside markets makes local firms more productive by allowing them to access a broader variety of goods at a cheaper price and increasing the demand for their output (Krugman and Venables, 1995). This can lead to an increase in the marginal product of their labour and therefore local wages. This can be captured by a market potential variable. This paper uses the Harris (1954) market potential definition, which takes the inverse-distance weighted sum of employment population over all municipalities other than the municipality in question.[17]

[17] Having data on trade between municipalities, as well as data on trade between municipalities and foreign countries, would also be very useful (e.g. Fally et al. (2010)), but this is not currently available for South Africa.

3.4 Results

3.4.1 Baseline results

Table 3.2 presents basic pooled OLS regression results. In absence of any other control, the specification in Column 1 shows an elasticity of 0.079 between wages and urban employment population. This indicates that in the cross-section of South African municipalities between 2008 and 2016, a doubling of urban employment population is associated with 7.9% higher wages. Column 2 adds individual characteristics. This reduces the coefficient on log urban employment population to 0.043, implying that about 45% of the relationship between wages and urban employment population is explained by observed worker characteristics. This elasticity is on the extreme lower end of the range of estimates for regional wage elasticity in developing countries. With the exact same specification, Duranton (2016) estimates an elasticity of 0.054 between municipal urban population and wages for Columbia. This figure is substantially lower than the population elasticities - around 0.08 - reported for Chinese and Indian cities (Chauvin et al., 2017; Combes et al., 2017). Column 3 includes a control for (log) municipal land area. The coefficient on urban employment population marginally decreases to 0.040. This is my preferred OLS specification. The rest of the columns in this table are devoted to establishing the robustness of this finding and to exploring possible heterogeneous effects.

In Column 4, after controlling for industries and occupations as well as the land area control, the regional wage elasticity drops slightly to 0.038.[18] Relative to the specification of Column 3, Column 5 adds market access, which reduces the coefficient on (log) urban employment population to 0.032. This suggests that part of the relationship between wages and (log) urban employment population is explained by the fact that municipalities with larger urban employment populations also have access to larger markets. Column 6 duplicates Column 3 but uses total urban population

[18]Industries and occupations are arguably endogenous to municipal size, which is why they are not included in the main regression.

Table 3.2: The effects of agglomeration externalities on wages: Pooled OLS

	(1) Only UEP	(2) Indiv.	(3) Area	(4) Ind.+occ.	(5) MA	(6) UP	(7) TP	(8) Non-lin.	(9) HW	(10) YOE
log UE pop.	0.079***	0.043***	0.040***	0.038***	0.032***			-0.028	0.040***	-0.059***
	(0.014)	(0.007)	(0.007)	(0.005)	(0.007)			(0.054)	(0.007)	(0.017)
log UE pop.^{2}								0.003		
								(0.002)		
log area			-0.028**	-0.034***	-0.004	-0.031**	-0.030**	-0.028**	-0.036***	-0.028**
			(0.013)	(0.011)	(0.016)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)
log MA					0.115***					
					(0.033)					
log urban pop.						0.041***				
						(0.007)				
log total pop.							0.049***			
							(0.010)			
log UE pop.*YOE										0.009***
										(0.002)
YOE										0.019
										(0.017)
R-squared	0.109	0.580	0.580	0.639	0.584	0.580	0.581	0.581	0.572	0.539
Observations	12,070	12,070	12,070	11,507	12,070	12,070	12,253	12,070	11,672	11,979

Robust standard errors clustered at the local municipality level in parentheses. Columns 2-10 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-10 include a control for the (log) land area of local municipalities. Year fixed effects included in all columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.01

as the measure of agglomeration. This also has very little effect on the elasticity. Column 7 also duplicates Column 3 but uses total population (from both urban and rural parts of municipalities) as the agglomeration measure. The coefficient on (log) total population is 0.049 - a 23% increase in the size of the coefficient relative to that obtained with my preferred specification in Column 3. Not having data on the urban employed population, some of the other agglomeration economies studies use total population as their measure of agglomeration. This analysis suggests that this might account, in part, for the inflated elasticities they estimate. Column 8 adds a quadratic term for urban employment population. The coefficients on (log) urban employment population and its square are non-significant showing that linear form does better in approximating the relationship between my agglomeration measure and wages. Column 9 uses hourly wages instead of wages as the dependent variable, which, reassuringly, has no effect on the coefficient on (log) urban employment population. Column 10 includes an interaction between the education of individual workers and (log) urban employment population. This interaction term is positive and significant at the 1% level, while the coefficient on (log) urban employment population is negative (while remaining significant at the 1% level). This suggests that there are higher returns to agglomeration for skilled workers and that very low-skilled workers experience negative returns to agglomeration. While Duranton (2016) did not find that education made any difference to agglomeration externalities in Colombia, the result here is in line with the literature on the US (cf. Bacolod et al. (2010); Wheeler (2001)).

3.4.2 Individual fixed effects

Table 3.3 duplicates Table 3.2 but includes individual fixed effects to control for potential sorting on unobservables. The inclusion of individual fixed effects reduces the coefficient on urban employment population, or variants thereof, in all specifications. Looking at the main result in Column 3, the coefficient on (log) urban employment population is 0.033 - 18% lower than the corresponding OLS estimate. However, the individual fixed effects have a larger effect (reducing the wage elasticity by 37%) when total population is used as the measure of agglomeration. This suggests that studies using total population as the measure of agglomeration but not individual fixed effects might be substantially overestimating agglomeration elasticities. Using hourly wages instead of gross wages as the dependent variable results in no significant change in the elasticities estimated. Lastly, with the inclusion of individual fixed effects, the interaction between education and (log) urban employment population is non-significant and the coefficient is 0.00 suggesting that there are no higher returns to urban agglomeration for the highly educated once unobserved individual ability is accounted for.

In international studies, including individual fixed effects typically reduces regional wage elasticity by about 30-50%, suggesting that there is positive sorting of high fixed effect (unobserved ability) individuals into more urbanised areas (Baum-Snow and Ferreira, 2015). In South Africa, using my favoured specification (in Column 3), fixed effects reduces the coefficient on urban employment population by 18-37% depending on the agglomeration measure used.

Table 3.3: The effects of agglomeration externalities on wages: Individual fixed effects

	(1) Only UEP	(2) Indiv.	(3) Area	(4) Ind.+occ.	(5) MA	(6) UP	(7) TP	(8) HW	(9) YOE
log UE pop.	0.043***	0.036***	0.033***	0.029***	0.028***			0.027***	-0.038
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)			(0.010)	(0.034)
log UE pop. ^{2}									
log area			-0.054	-0.057*	-0.036	-0.054	-0.070**	-0.052	-0.052
			(0.033)	(0.030)	(0.036)	(0.033)	(0.032)	(0.034)	(0.033)
log MA					0.084				
					(0.054)				
log urban pop.						0.036***			
						(0.010)			
log total pop.							0.031**		
							(0.013)		
log UE pop.*YOE									-0.000
									(0.003)
YOE									0.024
									(0.035)
R-squared	0.431	0.453	0.453	0.466	0.453	0.453	0.453	0.426	0.453
Observations	12,048	12,048	12,048	11,262	12,048	12,048	12,253	11,475	11,918

Robust standard errors clustered at the individual level in parentheses. Columns 2-9 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-8 include a control for the (log) land area of local municipalities. Year fixed effects included in all columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.01

3.4.3 2SLS results

Table 3.4 duplicates Table 3.2 but instruments log municipal employment population (or variants thereof) with historical urban population in 1960. In columns 1-6, it can be seen that instrumenting with historical employment population marginally decreases elasticity estimates. The ancillary tests show that the instruments are strong i.e. the first-stage F-stat is over 1000.[19] In Column 3, with individual and area-level controls, the elasticity is reduced to 0.035 (which is significantly different from the OLS estimate). In Column 6, the results are again little changed by using urban population instead of urban employment population. In Column 7, the coefficient on total municipal population is 0.050, which is substantially larger than the coefficient obtained in Column 3. It appears that the inclusion of the instrument (unlike the inclusion of individual fixed effects) has no effect on this coefficient. In Column 8 the quadratic term remains non-significant and the (non-significant) coefficient on (log) urban employment population increases substantially. In Column

[19] I do not report F-stats for each table in this paper but have checked to make sure that they are all substantially higher than the traditional threshold of 10.

9, using hourly wages instead of gross wages, the coefficient on urban employment population is again unchanged. In Column 10, relative to the OLS specification, the negative coefficient on (log) urban employment is slightly larger, while the coefficient on the interaction with skill level is unchanged.

Table 3.4: The effects of agglomeration externalities on wages: 2SLS

	(1) Only UEP	(2) Indiv.	(3) Area	(4) Ind.+occ.	(5) MA	(6) UP	(7) TP	(8) Non-lin.	(9) HW	(10) YOE
log UE pop.	0.069***	0.038***	0.035***	0.034***	0.028***			7.097	0.035***	-0.066***
	(0.016)	(0.008)	(0.008)	(0.006)	(0.008)			(22.182)	(0.008)	(0.022)
log UE pop.^{2}								0.109		
								(0.463)		
log area			-0.035***	-0.040***	-0.012	-0.037***	-0.031**	0.109	-0.042***	-0.037**
			(0.012)	(0.012)	(0.016)	(0.014)	(0.015)	(0.463)	(0.013)	(0.015)
log MA					0.113***					
					(0.035)					
log urban pop.						0.037***				
						(0.008)				
log total pop.							0.050***			
							(0.012)			
log UE pop.*YOE										0.009***
										(0.002)
YOE										0.020
										(0.023)
R-squared	0.102	0.574	0.575	0.635	0.578	0.574	0.573	.	0.567	0.533
Observations	11,440	11,440	11,440	10,905	11,440	11,440	11,440	11,440	11,084	11,352

Robust standard errors clustered at the local municipality level in parentheses. Columns 2-10 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-10 include a control for the (log) land area of local municipalities. Year fixed effects included in all columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.01

In line with findings from studies elsewhere, the 2SLS regional wage elasticity estimates are lower than the corresponding OLS estimates, suggesting that there was some simultaneity-/omitted variable bias in the relationship between agglomeration and wages in South Africa. With my main specification (Column 3), the coefficient on urban employment population drops by 12.5% with the inclusion of the instrument.

3.4.4 2SLS and individual fixed effects combined

Table 3.5 replicates Table 3.2 but includes both individual fixed effects and the historical instrument. All the coefficients on (log) urban employment population, or variants thereof, are smaller than in Table 3.3 or Table 3.4, except for the coefficient on (log) total population which increases relative to that in Table 3.3 but is smaller than that in Table 3.4. In Column 3, my main specification yields a regional wage elasticity of 0.031, significant at the 1% level. This is my preferred estimate of regional wage elasticity for the paper.[20] In Column 4, the coefficient on (log) market access remains significant at the 5% level but the coefficient on (log) urban employment population is only slightly smaller showing that market access remains a significant determinant of wages but not override the effect of local agglomeration. Lastly in Column 9, the interaction between (log) urban employment and individual education remains non-significant and close to zero. While this is not conclusive evidence, it suggests that high-skilled workers do not enjoy higher agglomeration benefits (in line with the findings from Duranton (2016)).

Table 3.5: The effects of agglomeration externalities on wages: 2SLS and individual fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Only UEP	Indiv.	Pr.+AC	Ind.+occ.	MA	UP	TP	HW	YOE
log UE pop.	0.038***	0.032***	0.031***	0.028**	0.025**			0.018	-0.010
	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)			(0.012)	(0.034)
log area			-0.016	-0.024	0.005	-0.018	-0.016	-0.011	-0.019
			(0.033)	(0.031)	(0.035)	(0.033)	(0.033)	(0.034)	(0.034)
log MA					0.104*				
					(0.055)				
log urban pop.						0.032***			
						(0.012)			
log total pop.							0.046***		
							(0.017)		
log UE pop.*YOE									0.004
									(0.003)
YOE									-0.029
									(0.031)
R-squared	0.427	0.448	0.448	0.460	0.448	0.448	0.447	0.420	0.447
Observations	11,372	11,372	11,372	10,621	11,372	11,372	11,372	10,864	11,247

Robust standard errors clustered at the individual level in parentheses. Columns 2-9 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-9 include a control for the (log) land area of local municipalities. Year fixed effects included in all columns. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.01$

To summarise the empirical results thus far, looking at the relationship between wages and urban employment population in South Africa, endogeneity at both the individual (caused by sorting on observables and unobservables) and local levels (caused by simultaneity bias/omitted variable bias) seems to be at play. Using individual fixed effects, my preferred specification yields an elasticity of 0.033 (17.5% below the corresponding OLS estimate). Including a historical instrument for urban employed population, my preferred elasticity is 0.035 (12.5% below the corresponding

[20] My results are unchanged (and remain significant at the 5% level) when I use two-way clustering at either the individual-year level or the municipality-year level.

OLS estimate). Combining historical instruments and individual fixed effects in the same specification yields a coefficient of 0.031 (22.5% below the corresponding OLS estimate).

3.4.5 Further heterogeneous effects

This paper extends the examination of the possible heterogeneous effects of agglomeration, which has been a major preoccupation of the recent literature. I have already tested if there are non-linearities in the effects of city size, finding no evidence for this. I have also examined the differential effects of agglomeration on workers of different skill groups, finding that larger effects for skilled workers obtained with the OLS and 2SLS specifications disappear when I include individual fixed effects. I now turn to examine the effects of participation in the formal/informal economy and the effects of gender using my specification with both the historical instrument and individual fixed effects.

In Panel 1 of Table 3.A.4 I examine the separate effects of workers either with or without written contract by including an interaction between my measure of agglomeration and a binary variable equal to 1 if a worker is informally employed (without a contract) or 0 if formally employed. In all specifications, I find that the coefficient on the interaction term is positive but non-significant. While my results are inconclusive, higher returns to agglomeration in the informal sector have been found in several other studies (Duranton, 2016; García, 2016). Businesses hiring workers informally would also tend to be informal businesses, whose products are sold locally and whose income is more directly related transportation and local housing costs Duranton (2016).

In Panel 2 of Table 3.A.4 I examine the separate effects on workers of different genders by including an interaction between my measure of agglomeration and a binary variable equal to 1 if a worker is female or 0 if male. With my main specification in Column 3, I find a positive but non-significant coefficient on my interaction term. Like Duranton (2016), I find no statistically significant difference in the size of agglomeration effects between men and women in Colombia.

3.4.6 Human capital externalities

As already mentioned, one theoretical mechanism behind agglomeration economies is spillovers occur between people working in close proximity, thereby enhancing productivity. This is obviously related to the theory that there are benefits to urban workers from being located in more educated places (on top of the private benefits of education). However, human capital externalities differ from the agglomeration effects measured above because they relate to the skill composition of the workforce/population rather than its absolute size. Two separate strands of literature have emerged and there have been few attempts to compare the relative effects of the two variables of interest (Duranton, 2016). In the regressions that follow I include both a measure of agglomeration and a measure of the share of more educated workers.[21] To measure municipal education, I experiment with two measures: first, the share of workers with some university/technical college education, and second, the share of workers that have completed Grade 10. The intention here is not to do a full exploration of the effects of human capital externalities but mainly to assess the robustness of earlier findings.

Table 3.A.5 duplicates Table 3.5 adding the share of population with tertiary education. I run the regressions with individual fixed effects to control for the possibility that places with higher aggregate levels of education also have people with higher levels of unobserved human capital. Adding this control marginally increases the coefficient on urban employment population. Table 3.A.6 duplicates Table 3.5 adding the share of the population that have completed Grade 10.[22] This again

[21]It is possible that some omitted explanatory variable both disproportionately attracts skilled people to an area and increases wages, thus it would be good to have an instrument for area education. One possible instrument is the number of schools per local municipality (cf. Moretti (2004)) but I do not have good data for this in South Africa. Another possibility is to use the lagged demographic structure of municipalities (cf. Moretti (2004)) but this would be difficult to estimate given that local municipalities did not exist prior to 2001.

[22]Many high-school students drop out of school at Grade 10 level (as long as they have completed

marginally increases the coefficient on urban employment population. In both cases, the coefficient on municipal education is mostly non-significant (and negative). In estimations that include both agglomeration and regional education, agglomeration seems to be more important.

When the same regressions as in Table 3.A.5 and Table 3.A.6 are estimated without a measure of agglomeration, interestingly, the coefficients on municipal education are negative and statistically significant (see Table 3.A.7 for the results using the second measure of regional education level).[23]

My results are similar to those found by Duranton (2016) for Colombia, who also finds agglomeration externalities dominate human capital externalities. However, they are very different to those obtained by Chauvin et al. (2017) for Brazil, China and India. The latter paper finds large (elasticities of 3.0-6.7%) and significant coefficients on their aggregate education measure (share of the population with university education) even when they control for regional population. This leads them to speculate that developing world city success may depend on education rather than agglomeration. This argument does not seem to hold in the case of South Africa.

3.5 Conclusion

This paper examines the size of agglomeration effects in South Africa applying rigorous econometric strategies to micro-level panel data from the National Income Dynamics Survey. Previous related studies in South Africa had made use of aggregated regional output measures of poor quality. In order to map respondents to a geographic level that approximates local labour markets as closely as possible, this paper works with the secure version of NIDS, which contains household geo-coordinates.

this grade, they are eligible to enter state-sponsored technical colleges), so this is a good cut-off to use.

[23] I also experiment with using the logged form of these municipal education measures but they remain non-significant in most specifications.

In addition to controlling for an extensive range of observable worker characteristics that could affect wages, this paper exploits the panel nature of the dataset by using individual fixed effects to control for sorting on unobservables. There are no other studies of agglomeration economies in developing countries that have made use of individual fixed effects. To control for potential reverse causality/omitted variable bias in the relationship between municipal employment population and wages, the paper makes use of a historical population instrument constructed from South African census data in 1960. I argue that the historical population instrument is more likely than in other settings to be exogenous due to the peculiar history of separate development in South Africa.

Using an OLS specification, the raw regional wage elasticity in South Africa is estimated to be 7.9%. My preferred specification, including both the instrumental variable and individual fixed effects, yields a regional wage elasticity of 3.1%, which is considerably lower than existing estimates for other developing countries. Including fixed effects turns out to be important, reducing regional wage elasticity by 18-37%, depending on the agglomeration measure used.

Using my full specification including both my historical instrument and individual fixed effects, I do not find significant heterogeneous effects of agglomeration based on gender, skill level or participation in the formal/informal economy. I do however find some evidence that agglomeration externalities matter more than human capital externalities in South Africa.

Bibliography

Bacolod, M., Blum, B. S., and Strange, W. C. (2010). Elements of skill: Traits, intelligences, education, and agglomeration. *Journal of Regional Science*, 50(1): 245–280.

Baum-Snow, N. and Ferreira, F. (2015). Causal Inference in Urban Economics. In Duranton, G, J. H. . W. S., editor, *Handbook of Regional and Urban Economics*, volume 5. Elsevier.

Briant, A., Combes, P.-P., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3): 287–302.

Brophy, T., Branson, N., Daniels, R. C., Leibbrandt, M., Mlatsheni, C., and Woolard, I. (2018). National Income Dynamics Study Panel User Manual. Available from: http://www.nids.uct.ac.za/documentation/overview-documentation/wave-4 (Accessed 09/07/2018).

Carmody, P. (2002). Between Globalisation and (Post) Apartheid: The Political Economy of Restructuring. *South Africa Journal of Southern African Studies*, 28(2): 255–275.

Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98: 17–49.

Christopher, A. J. (2001). *The Atlas of Changing South Africa*. London and New York: Routledge.

Ciccone, A. and Hall, R. (1996). Productivity and the Density of Economic Activity. *American Economic Review*, 86: 54–70.

Combes, P.-P., Démurger, S., and Li, S. (2017). Urbanisation and Migration Externalities in China. *Asian Development Review*, 34(2): 184–200.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2): 723–742.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration economics*, pp. 15–66. University of Chicago Press.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, pp. 247–348. Elsevier.

D'Costa, S. and Overman, H. G. (2014). The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, 48: 168–179.

Dinkelman, T. and Ranchhod, V. (2012). Evidence on the impact of minimum wage laws in an informal sector: Domestic workers in South Africa. *Journal of Development Economics*, 99(1): 27–45.

Duranton, G. (2014). *Growing through cities in developing countries.* World Bank Policy Research Working Paper 6818.

Duranton, G. (2016). Agglomeration effects in Colombia. *Journal of Regional Science*, 56(2): 210–238.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pp. 2063–2117. Elsevier.

Fally, T., Paillacar, R., and Terra, C. (2010). Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics*, 91(1): 155–168.

Farole, T. and Sharp, M. (2017). *Spatial industrial policy, special economic zones and cities in South Africa.* Background paper for the World Bank South African Urbanisation Review.

García, G. A. (2016). *Agglomeration Economies in the Presence of an Informal*

Sector: The Colombian Case. Documentos CEDE 014575, Universidad de los Andes - CEDE.

Glaeser, E. L. and Mare, D. C. (2001). Cities and skills. *Journal of Labor Economics*, 19(2): 316–342.

Glaeser, E. L. and Xiong, W. (2017). Urban productivity in the developing world. *Oxford Review of Economic Policy*, 33(3): 373–404.

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American geographers*, 44(4): 315–348.

Henderson, J. V. (2005). Urbanization and growth. In Durlauf, P. A. . S., editor, *Handbook of Economic Growth*, volume 1, pp. 1543–1591. Elsevier.

Henderson, J. V. and Kriticos, S. (2018). The development of the African system of cities. *Annual Review of Economics*, 10: 287–314.

Hering, L. and Poncet, S. (2010). Market access and individual wages: Evidence from China. *The Review of Economics and Statistics*, 92(1): 145–159.

Krugell, W. and Naude, W. (2005). The Geographical Economy of South Africa. *Journal of Development Perspectives*, 1: 85–128.

Krugell, W. and Rankin, N. (2012). Agglomeration and firm-level efficiency in South Africa. *Urban Forum*, 23(3): 299–318.

Krugman, P. and Venables, A. J. (1995). Globalization and the inequality of nations. *The Quarterly Journal of Economics*, 110(4): 857–880.

Leibbrandt, M., Finn, A., and Oosthuizen, M. (2016). *Poverty, Inequality, and Prices in Post-Apartheid South Africa*. In Growth and Poverty in sub-Saharan Africa (ed. Arndt, C., McKay, A & Tarp, F.) UNU-WIDER Studies in Development Economics.

Magruder, J. R. (2012). High unemployment yet few small firms: The role of centralized bargaining in South Africa. *American Economic Journal: Applied Economics*, 4(3): 138–66.

Manning, A. (2010). The Plant Size-Place Effect: Agglomeration and Monopsony in Labour Markets. *Journal of Economic Geography*, 10(5): 717–744.

Melo, P. C., Graham, D. J., and Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3): 332–342.

Moretti, E. (2004). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1-2): 175–212.

Msulwa, R. and Turok, I. (2013). *Does density drive development?* Urban Transformation Research Report, Johannesburg: Wits University.

Naudé, W. A. and Krugell, W. (2006). Sub-national growth rate differentials in South Africa: an econometric analysis. *Papers in Regional Science*, 85(3): 443–457.

Ogura, M. et al. (1996). Urbanization and Apartheid in South Africa: Influx Controls and Their Abolition. *The Developing Economies*, 34(4): 402–423.

Overman, H. G. and Venables, A. J. (2005). *Cities in the Developing World*. CEPR Discussion Paper, London: Centre for Economic Policy Research.

Proost, S. and Thisse, J.-F. (2019). What Can Be Learned from Spatial Economics? *Journal of Economic Literature*, 57(3): 575–643.

Robbins, G. (2015). *Searching for Space in South Africa's Post–Apartheid Industrial Policy*. Presentation to Economies of Regions Learning Network Conference. Pretoria.

Roca, J. D. L. and Puga, D. (2017). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1): 106–142.

Statistics SA (2017). *Poverty trends in South Africa: An examination of absolute* *poverty between 2006 and 2015.* Report No. 03-10-06. Statistics South Africa, Pretoria.

Todes, A. and Turok, I. (2018). Spatial inequalities and policies in South Africa: Place-based or people-centred? *Progress in Planning*, 123: 1–31.

Turok, I. (2012). *Urbanisation and Development in South Africa: Economic Imperatives, Spatial Distortions and Strategic Responses.* Urbanisation and Emerging Population Issues, Working Paper 8, October. London: Human Settlements Group, International Institute for Environment and Development.

Wentzel, M., Viljoen, J., and Kok, P. (2006). *Contemporary South African migration patterns and intentions.* In P. Kok, D. Gelderblom, J. Oucho, & J. van Zyl (eds.), Migration in south and Southern Africa: Dynamics and determinants. Cape Town: HSRC Press.

Wheeler, C. H. (2001). Search, sorting, and urban agglomeration. *Journal of Labor Economics*, 19(4): 879–899.

World Bank (2019). World Development Indicators. Available from: http://data.worldbank.org/.

3.A Additional Tables

Table 3.A.1: Summary statistics for municipalities

	Total		High UEP munic.		Low UEP munic	
	Mean	SD	Mean	SD	Mean	SD
Total pop. 2011	221241.7	521223.5	944170.2	12222452	114928.7	98717.8
Urban pop. 2011	139100.7	504421.3	840035.7	1205709	36022.0	32891.5
Urban emp. pop. 2011	44272.5	179726.9	282546.1	437677.3	9232.2	9237.6
Urban emp. pop 2011 market access	37228.9	23272.7	48980.6	27440.5	27394.3	12349.2
Area (km^2)	4669.6	4987.8	2882.4	1722.9	6124.7	6164.6
Est. urban pop. 1960	211924.3	348487.1	427507.5	403820	51744.5	36952.3
Est. black urban pop. 1960	95983.9	180346.9	194926.9	220889.6	12380.0	9146.4
% tertiary educated	4.0	2.2	7.3	2.4	3.4	1.7
% completed Grade 10	53.7	57.2	46.1	39.1	54.9	5.1
No. of municipalities	234		30		204	

Table 3.A.2: Descriptive statistics for movers in base sample

	Movers					
	Total		High EP munic.		Low EP munic	
	Mean	SD	Mean	SD	Mean	SD
Monthly gross wages	7669.8	8308.4	8959.8	9022.0	6525.3	7440.4
Female (%)	37.1	48.3	37.7	48.5	36.5	48.2
Age	33.8	9.2	33.3	8.9	34.2	9.4
Age 16-29 (%)	37.8	48.5	40.2	49.1	35.7	48.0
Age 30-43 (%)	45.7	49.8	44.4	49.7	46.9	49.9
Age 44-57 (%)	14.9	35.7	13.3	35.0	15.5	36.2
Age 58-70 (%)	1.5	12.1	1.2	10.7	1.8	13.3
Black (%)	65.9	47.4	71.7	45.1	60.8	48.8
Coloured (%)	27.2	44.5	22.7	41.9	31.3	46.4
Asian/Indian (%)	3.1	17.4	1.45	11.9	4.6	21.0
White (%)	2.8	16.5	3.3	17.9	2.3	15.0
Avg. years of educ.	10.5	3.1	11.2	2.5	9.9	3.5
Grade 1-5 (%)	4.3	20.2	1.9	13.6	6.4	24.5
Grade 6-8 (%)	9.2	29.0	7.4	26.2	10.9	31.2
Grade 9-11 (%)	32.2	46.7	30.2	46.0	34.0	47.4
Grade 12 (%)	43.8	49.6	49.3	0.5	39.0	48.8
Technikon diploma (%)	0.3	5.2	0.43	6.6	0.1	3.6
University degree (%)	6.7	25.1	9.3	29.0	4.5	20.7
Trade union (%)	26.8	44.3	29.1	45.4	24.7	43.2
Informal (%)	28.2	45.0	23.1	42.2	32.7	46.9
Married (%)	24.6	43.0	25.2	43.4	24.1	42.8
Observations	1,472		692		780	
Individuals	529		252		277	

Table 3.A.3: Descriptive statistics for base sample excluding districts where no historical population information

	Movers					
	Total		High EP munic.		Low EP munic	
	Mean	SD	Mean	SD	Mean	SD
Monthly gross wages	6833.3	7691.3	7874.0	8411.6	5870.0	6819.1
Female (%)	48.6	49.9	48.6	50.0	48.6	49.9
Age	37.8	10.7	37.0	10.6	38.6	10.7
Age 16-29 (%)	26.7	44.2	30.0	45.8	23.6	42.5
Age 30-43 (%)	42.6	49.5	42.2	49.4	43.0	49.5
Age 44-57 (%)	26.6	44.2	23.9	42.7	29.1	45.4
Age 58-70 (%)	4.1	19.9	3.9	19.4	4.3	20.3
Black (%)	55.9	49.7	64.2	47.9	48.2	50.0
Coloured (%)	33.2	47.1	25.8	43.7	40.1	49.0
Asian/Indian (%)	6.3	24.3	4.1	19.7	8.4	27.7
White (%)	3.7	18.8	49.6	21.7	2.5	15.5
Avg. years of educ.	9.9	3.2	10.5	2.76	9.3	3.5
Grade 1-5 (%)	6.5	24.6	4.0	19.56	8.8	28.4
Grade 6-8 (%)	15.1	35.8	11.3	31.7	18.5	38.9
Grade 9-11 (%)	33.9	47.3	35.4	47.8	32.6	46.9
Grade 12 (%)	34.8	47.6	40.5	49.0	29.6	45.7
Technikon diploma (%)	0.5	7.5	0.6	7.7	0.5	7.3
University degree (%)	5.4	22.6	6.2	24.1	4.7	21.1
Trade union (%)	29.9	45.8	31.8	46.6	28.2	45.0
Informal (%)	25.8	43.8	23.6	42.5	27.9	44.8
Married (%)	34.2	47.4	33.5	47.2	34.8	47.6
Observations	11,040		5,499		5,941	
Individuals	4,212		2,037		2,175	

Table 3.A.4: The effects of agglomeration externalities on wages: 2SLS and individual fixed effects with interactions

Panel 1: Effects of informal/formal employment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Only UEP	Indiv.	Area	Ind.+occ.	MA	UP	TP	HW
log UE pop.	0.036***	0.031***	0.030***	0.027**	0.024**			0.018
	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)			(0.012)
log UE pop#informal	0.003	0.004	0.004	0.005	0.003			0.000
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)			(0.008)
log MA					0.103*			
					(0.055)			
log MA#informal					0.006			
					(0.026)			
log urban pop.						0.031***		
						(0.012)		
log urban pop.#informal						0.004		
						(0.007)		
log total pop							0.044***	
							(0.017)	
log total pop.#informal							0.005	
							(0.009)	
log area			-0.015	-0.023	0.005	-0.018	-0.016	-0.011
			(0.033)	(0.031)	(0.035)	(0.033)	(0.033)	(0.034)
R-squared	0.434	0.448	0.448	0.460	0.448	0.448	0.447	0.420
Observations	11,372	11,372	11,372	10,621	11,372	11,372	11,372	10,864
Panel 2: Effects of gender								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Only UEP	Indiv.	Area	Ind.+occ.	MA	UP	TP	HW
log UE pop.	0.044***	0.034**	0.033**	0.030**	0.024			0.025
	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)			(0.016)
log UE pop#female	-0.013	-0.006	-0.005	-0.004	0.001			-0.015
	(0.024)	(0.023)	(0.023)	(0.023)	(0.023)			(0.023)
log MA					0.110*			
					(0.066)			
log MA#female					-0.018			
					(0.119)			
log urban pop.						0.035**		
						(0.016)		
log urban pop.#female						-0.005		
						(0.024)		
log total pop							0.049**	
							(0.023)	
log total pop.#female							-0.008	
							(0.034)	
log area			-0.015	-0.023	0.005	-0.017	-0.016	-0.008
			(0.034)	(0.031)	(0.035)	(0.034)	(0.034)	(0.035)
R-squared	0.427	0.448	0.448	0.460	0.448	0.448	0.447	0.421
Observations	11,372	11,372	11,372	10,621	11,372	11,372	11,372	10,864

Panel 1 includes an interaction between my agglomeration measure and a dummy variable = 1 if worker is informally employed and 0 if formally employed. The main effect of *informal* drops out because of the individual fixed effects. Panel 2 includes an interaction between my agglomeration measure and a dummy variable = 1 if worker is female and 0 if male. The main effect of *female* drops out because of the individual fixed effects. Robust standard errors clustered at the local municipality level in parentheses. Columns 2-8 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-8 include a control for the (log) land area of local municipalities. Year fixed effects included in all columns. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.01$

Table 3.A.5: The effects of agglomeration externalities on wages of workers: 2SLS with individual fixed effects including a control for share of population with tertiary education

	(1) Only UEP	(2) Indiv.	(3) Area	(4) Ind.+occ.	(5) MA	(6) UP	(7) TP	(8) HW
log UE pop.	0.055***	0.048***	0.047***	0.044***	0.040***			0.030*
	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)			(0.016)
Prop. tert. edu.	-0.860*	-0.815*	-0.806*	-0.809	-0.680	-0.787*	-1.114**	-0.577
	(0.506)	(0.471)	(0.470)	(0.498)	(0.471)	(0.465)	(0.552)	(0.492)
log area			-0.011	-0.019	0.006	-0.015	-0.011	-0.008
			(0.033)	(0.031)	(0.035)	(0.033)	(0.033)	(0.034)
log MA					0.090			
					(0.055)			
log urban pop.						0.049***		
						(0.016)		
log total pop.							0.079***	
							(0.026)	
R-squared	0.427	0.448	0.448	0.461	0.449	0.448	0.448	0.421
Observations	11,372	11,372	11,372	10,621	11,372	11,372	11,372	10,864

Robust standard errors clustered at the individual level in parentheses. Columns 2-8 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-8 include a control for the (log) land area of local municipalities. All columns include a control for the share of the population with tertiary education. Year fixed effects included in all columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.01

Table 3.A.6: The effects of agglomeration externalities on wages of workers: 2SLS with individual fixed effects including a control for share of population with at least Grade 10 education

	(1) Only UEP	(2) Indiv.	(3) Area	(4) Ind.+occ.	(5) MA	(6) UP	(7) TP	(8) HW
log UE pop.	0.042***	0.036***	0.035***	0.031***	0.030**			0.021*
	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)			(0.012)
Prop. Gr. 10	0.210*	-0.207*	-0.216**	-0.205*	-0.180*	-0.224**	-0.317**	-0.162
	(0.116)	(0.108)	(0.107)	(0.110)	(0.108)	(0.108)	(0.127)	(0.111)
log area			-0.023	-0.029	-0.006	-0.025	-0.027	-0.017
			(0.033)	(0.031)	(0.034)	(0.032)	(0.032)	(0.034)
log MA					0.079			
					(0.057)			
log urban pop.						0.036***		
						(0.013)		
log total pop. dens.							0.054***	
							(0.019)	
R-squared	0.427	0.448	0.448	0.461	0.449	0.448	0.448	0.421
Observations	11,372	11,372	11,372	10,621	11,372	11,372	11,372	10,864

Robust standard errors clustered at the local municipality level in parentheses. Columns 2-8 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. Columns 3-8 include a control for the (log) land area of local municipalities. All columns include a control for the share of the population with at least Grade 10 education. Year fixed effects included in all columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.01

Table 3.A.7: The effects of human capital externalities on wages of workers: 2SLS with individual fixed effects

	(1) Only UEP	(2) Indiv.	(3) Ind.+occ.	(4) HW
Prop. Gr. 10	-0.209***	-0.201***	-0.199***	-0.178***
	(0.076)	(0.070)	(0.068)	(0.077)
R-squared	0.431	0.453	0.466	0.425
Observations	12,253	12,253	11,454	11,668

Robust standard errors clustered at the local municipality level in parentheses. Columns 2-4 include individual controls for age, age squared, educational attainment, marital status, trade union membership, etc. The variable of interest is the share of the population with at least Grade 10 education. Year fixed effects included in all columns. Significance levels: *** p<0.01, ** p<0.05, * p<0.01