

IZA DP No. 10113

Migration and Urbanisation in Post-Apartheid South Africa

Jan David Bakker
Christopher Parsons
Ferdinand Rauch

July 2016

Migration and Urbanisation in Post-Apartheid South Africa

Jan David Bakker

University of Oxford

Christopher Parsons

University of Western Australia and IZA

Ferdinand Rauch

University of Oxford and CEP

Discussion Paper No. 10113

July 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Migration and Urbanisation in Post-Apartheid South Africa*

Under apartheid, black South Africans were severely restricted in their choice of location and many were forced to live in *homelands*. Following the abolition of apartheid they were free to migrate. Given gravity, a town nearer to the homelands can be expected to receive a larger inflow of people than a more distant town following the removal of mobility restrictions. Exploiting this exogenous variation, we study the effect of migration on urbanisation and the distribution of population. In particular, we test if migration inflows led to displacement, path dependence, or agglomeration in destination areas. We find evidence for path dependence in the aggregate, but substantial heterogeneity across town densities. An exogenous population shock leads to an increase of the urban relative to the rural population, which suggests that exogenous migration shocks can foster urbanisation in the medium run.

JEL Classification: R12, R23, N97, O18

Keywords: economic geography, migration, urbanisation, natural experiment

Corresponding author:

Christopher Parsons
Economics (UWA Business School)
University of Western Australia
35 Stirling Highway
Crawley WA 6009
Australia
E-mail: christopher.parsons@uwa.edu.au

* We thank Daniel de Kadt and Melissa Sands who shared census data from 1991. We are grateful to Matteo Escudé, James Fenske, Doug Gollin, Vernon Henderson, Leander Heldring, Daniel Kaliski, Daniel de Kadt, Lu Liu, Chris Roth, Ludvig Sinander, Andrea Szabo, Tony Venables, Helene Verhoef, Johannes Wohlfart and seminar participants at Oxford and ABCA 2016 for useful comments and discussion. We gratefully acknowledge the support of an Africa Research Program on Spatial Development of Cities at LSE and Oxford funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development.

1 Introduction

Consider the following thought experiment. What happens if a policy maker exogenously increases the population of a random town? There are three possibilities, the town: may remain at the increased size, may revert back to its original level or it may find itself on a new growth trajectory and grow even further. Which of the three reactions is observed may vary with the original size of the town as well as the magnitude of the shock. In answering this question we provide insights into the nature and evolution of urbanised areas, the size and location of cities and the distribution of population in space.

In this paper, we use the abolition of mobility restrictions as a natural experiment, to study the effect of domestic migration on the distribution of population in post-apartheid South Africa. Under the apartheid regime, the black South African population was severely restricted in its mobility. Large parts of the population were forced to live in so-called homelands and townships. En-route to the democratic transition in 1994, these restrictions were lifted and in June 1991 black South Africans could move freely for the first time since the foundation of the state of South Africa in 1910. Substantial internal migration flows resulted, which led to increased urbanisation during the 1990s and 2000s (see Figure 1 below).

The location of the homelands resulted from a long historical process beginning in the 18th century (Lapping, 1986). It is therefore plausible to assume that their location, conditional on covariates, is quasi-random with respect to economic conditions today. Assuming the subsequent migration outflows from the homelands behaved according to gravity, we are able to exploit the exogenous effect of this positive migration shock to identify the subsequent effects of increased internal mobility on the distribution of population in South Africa. In other words, assuming migration costs increase with distance, *ceteris paribus*, a town physically located nearer to a homeland is assumed to have received a larger inflow of previously mobility-restricted black migrants.

A town experiencing an exogenous migration shock can theoretically evolve in just three ways. First, the town's population could shrink back to the initial population level, i.e. mean-revert. Such a reaction would be consistent with an optimal urban network of relative city sizes, where relative sizes might be driven by location fundamentals. Secondly, the town's population could simply remain at the new population level and not adjust endogenously to the shock. In this case, the distribution of city sizes would be path dependent. Thirdly, the city could grow further. This would be consistent with a theory of multiple equilibria, where an initial population shock moves the town onto a new population trajectory growth path from one equilibrium to another.

We provide quasi-experimental evidence that the aggregate population distribution is highly path dependent. Distinguishing between rural and urban areas we find that population levels in areas with high initial population densities experience further agglomeration, i.e. there is endogenous in-migration after the shock. In rural areas, we cannot reject path dependent behaviour. This suggests that a positive exogenous population shock generates a ‘Matthew effect’ (‘those who have will be given’), as densely populated areas gain population relative to sparsely populated areas from an exogenous migration shock. We further investigate this heterogeneity by looking at how the effect varies with both the initial population density and the reduced form size of the shock. We find that for a given initial density a bigger shock leads to higher endogenous population growth. This is consistent with the idea that a significant shock is required to push a locality out of its current and into a new equilibrium.

Modifying the optimal city size model by Henderson (1974) we infer the shape of the agglomeration function from our empirical findings. In our version, agents’ utility is equal to the difference between the agglomeration function and the congestion cost curve. Assuming that the congestion cost curve is increasing and convex in population, we are able to relate our empirical results to the functional form of the agglomeration curve. The path-dependent evolution of the population level suggests that the agglomeration curve has a slope similar to the congestion cost curve as a function of population density. This implies that an exogenous increase in population does not significantly affect the utility level of incumbents and suggests that there are infinitely many equilibria for the distribution of population in space. The agglomeration effects in the densely populated areas suggest that there are potentially multiple equilibria in urban environments.

These results have important policy implications. Since positive population shocks have a permanent effect on the distribution of population in space, temporary policies that induce out-migration from some areas may influence the long-run distribution of city sizes. Such policies also affect the long-run distribution of population between rural and urban areas. Since the population increases more in urban than in rural areas, policies that induce migration also foster urbanisation. This suggests that policy makers are able to promote urbanisation and engineer the growth of metropolitan areas through encouraging higher levels of migration. Conversely, policies that impede migration will hinder urbanisation.

The remainder of this paper is organised as follows. Section 2 discusses the related literature, while section 3 details the historical development of South Africa thereby

providing evidence for the quasi-random location of the homelands. Section 4 introduces the theoretical thought experiment that serves as a framework for the empirical analysis presented in Section 5. The heterogeneous responses to a positive population shock are discussed in Section 6, while section 7 concludes.

2 Related literature

Our paper adds to the growing literature that exploits the exogenous variation resulting from apartheid policy to study the development of South Africa after 1994, see for example de Kadt and Sands (2016), de Kadt and Larreguy (2016) and Dinkelman (2011, 2013). We are the first to use this natural experiment to study the causal economic effect of internal migration and how it effects the distribution of population in space. This paper, thereby, adds quasi-experimental evidence to the literatures that examine the determinants of the uneven distribution of population in space and the relationship between city size and population growth.

After Auerbach (1913) observed that the size distribution of cities follows a power law, there have been many attempts to explain this persistent empirical regularity (often referred to as Zipf's Law, after Zipf (1949)). Following the theoretical work by Gabaix (1999) who showed that Zipf's Law emerges naturally if cities have equal relative growth rates (Gibrat's Law), an extensive empirical literature on the distribution of population has developed. Davis and Weinstein (2002) find that the dispersion of population is highly path-dependent and argue that this is due to the importance of location fundamentals. They show that even after massive destruction and death caused by the nuclear bombs, Hiroshima and Nagasaki rapidly returned to their pre-war growth path. Studies by Brakman et al. (2004) for post-war Germany and by Miguel and Roland (2011) for Vietnam arrive at similar results. These studies though are unable to distinguish between the effects of path-dependence due to state-dependence (i.e. the presence of factors of production) and the importance of the initial natural advantage.

Bleakley and Lin (2012) therefore analyse path dependence by studying the development of towns that experienced a negative shock to their fundamentals, namely portage cities that lost their natural trading advantage due to technological progress. This historical development allows the authors to disentangle state-dependence and the natural advantage effect, since continuing agglomerations cannot be explained by natural advantage. The main result of the study is that these former portage cities maintained their historical importance, a finding consistent with recent results from

local positive population shocks from German refugees after World War II that were highly persistent and could not be explained by location fundamentals (Schumann, 2014).

We contribute to this literature in several ways. First, our study is the first to analyse a large scale and indeed *positive* population shock. The studies of war destruction find evidence of path-dependence but cannot isolate whether this is driven by natural fundamentals, sunk investments, social networks or gains from agglomerations. Bleakley and Lin (2012) provide evidence that it is not driven by location fundamentals, but cannot distinguish between other factors. Since we analyse a positive population shock in which incoming migrants have neither social networks nor private sunk investments, we are able to isolate the effect of gains from agglomeration. Secondly, we provide evidence from a credible natural experiment that is well-identified and are able to draw upon a much larger sample in estimation in comparison with most studies in this literature. Thirdly, we are the first to provide evidence from Africa, a region that is amongst the most rapidly urbanising regions in the world, the continent in which such policies related questions matter most. Fourthly, many of the previous studies focus solely upon urban areas, whereas we are able to look at both rural and urban areas and the differences between the two.

Our paper also speaks to the existing literature on the relationship between population growth and the size of cities (e.g. Black and Henderson (2003), Eeckhout (2004), Duranton (2007), and Rossi-Hansberg and Wright (2007)). The majority of empirical studies find urban systems tend to obey Gibrat's Law and that city size is uncorrelated with population growth, while others do find departures from Gibrat's Law even for cities (Soo (2007), González-Val et al. (2013), and Holmes and Lee (2010)). Michaels et al. (2012) emphasise the importance of structural transformation for urbanisation. In their long-run study of population growth in the United States from 1880 to 2000, they find that areas with high initial population density obeyed Gibrat's Law, i.e. subsequent population growth was uncorrelated with initial population density. The increasing dispersion in the distribution of population was driven by heterogeneous growth rates for areas with intermediate densities. Up to a threshold of 7 people per km², population growth rate and initial population density are negatively correlated. For areas with an initial density of 7-55 people per km², they are positively correlated. They show that their result is driven by the differences in agriculture's initial share in production and structural transformation that shifted employment away from agriculture. Those areas that did not experience a shift away from agriculture stagnated in terms of population growth.

Our contribution to this research is two-fold. Most importantly, there are no studies in this line of research that examine exogenous population shocks and Gibrat's Law. Secondly, we are able to show how areas react differently depending upon initial population densities, thereby identifying rural and urban differences in the aftermath of an exogenous population shock.

3 Historical Background

Around two-thirds of South Africa's total population live in urban areas, making it one of the most urbanised countries in Africa. The process of urbanisation in South Africa fundamentally differs from other African countries however. In the second half of the 20th century, it was shaped by the apartheid policy of the National Party government (1948-1994). Apartheid - literally meaning "apart-ness" - was by its very nature a spatial concept (Christopher, 2001). The government aimed to completely separate the black and non-black populations. Policies spanned from installing two town hall bathrooms to segregating city quarters and creating native reserves, the so-called homelands (or 'bantustans') that were to become independent states for the black population.

Segregation and mobility restrictions imposed on the black population had a long tradition in South Africa dating back to at least the 18th century (Lapping, 1986). The support for apartheid policies in the run-up to the 1948 elections, especially among poor white South Africans, resulted from the increasing black urbanisation rate during the preceding decades. These dynamics derived from the expansion of manufacturing and labour shortages resulting from World War II (Ogura, 1996). It was generally believed that the problem of white poverty was linked to increasing black urbanisation. The Native Economic Commission (1930-32) provides an example as it explicitly names black urbanisation as a cause for greater levels of unemployment among low-skilled white people (Beinart, 2001, p.122). One of the main goals of the apartheid policies was therefore to prevent and reverse black urbanisation, or to put it in the words of the Stallard Commission (1922): *'The Native should only be allowed to enter urban areas, [...], when he is willing to enter and to minister to the needs of the white man, and should depart therefrom when he ceases so to minister.'* (Feinstein, 2005, p.152). The policies that took shape after 1948 were therefore unique in aiming to achieve complete spatial and social segregation and were achieved by mobilising significant government resources and displacing large numbers of black South Africans.

In order to control the movement of the black population, the government restricted

blacks' rights to own land and their legal ability to settle where they wished. The literature distinguishes two dimensions of separation, 'urban apartheid' and 'grand apartheid' (Christopher, 2001). Urban apartheid aimed at creating separate quarters for the black population that was allowed to stay permanently in urban areas. Grand apartheid rather aimed at moving the majority of the black population - that was not needed as labourers in white urban areas - to native reserves.

The three main measures to implement 'grand' and 'urban apartheid' were the Group Areas Act (1950), the Pass Laws Act (1952) and the Population Registration Act (1950). The latter assigned a population group to each citizen, which largely defined an individual's political and social rights. The Group Areas Act assigned a native reserve to each black population group and enabled the government to remove people that were not living in the area assigned to their population group. In order to control the population flow and black urbanisation flows in particular, the government relied on a pass system. The Pass Laws Act forced every black African to carry an internal passport at all times.¹ If a black African could not present his passport documenting his permission to be in a certain region, he was subject to arrest.

[Table 1: Population distribution across areas]

These strictly enforced laws significantly constrained the distribution of population in space as well as the process of urbanisation. According to the Surplus People Project (1985),² the South African government forcefully relocated at least 3.5 million people between 1960 and 1983. Additionally, several hundred thousand arrests were made every year under the pass laws (Beinart, 2001, p.158f). Table 1 displays the share of the black population living in urban and rural areas within South Africa and the homelands from 1950 to 1980. While the proportion living in urban areas in South Africa stayed roughly constant over the three decades, the proportion living in rural areas decreased by around 15%, while the homelands experienced a commensurate increase. These movements resulted in densely populated areas in the homelands that can be defined as urban in terms of population densities, but not in terms of public service delivery or industrial development. This 'dislocated urbanisation' (Beinart, 2001), driven by government decisions instead of economic fundamentals, provides evidence of the substantial impact that the apartheid policies had on the distribution of population. Overall, while apartheid policies failed to reverse the level of urbanisation of

¹It built on pre-apartheid legislation such as the Natives Urban Areas Act from 1923 and Natives Urban Areas Consolidation Act from 1945 that forced every black man in urban areas to carry passes at all times.

²The Surplus People Project was a non-governmental organisation that documented the forced removal through the apartheid government.

black South Africans, they were able to stop the trend towards increasing urbanisation driven by economic growth and instead channel urbanisation dynamics away from (white) cities towards the homelands.

Table 2 shows the share of the population living in urban areas during apartheid by population group. The three non-black population groups were already far more urbanised in 1951 and by 1991 around 90% of the non-black population lived in urban areas. The black population was predominantly living in rural areas in 1951 and urbanised until 1991, but remained significantly less urbanised than the other three population groups. As previously emphasised, this urbanisation was heavily influenced by government policies that kept the black population out of urban areas in 'white' South Africa and engineered urbanisation in the homelands. During the 1990s, urbanisation rapidly increased (see Figure 1). Since the non-black population was almost entirely urbanised in 1991, this is evidence of large domestic migration flows of the black population.

[Figure 1: Urbanisation in South Africa over time]

Given the historical context, two main concerns arise regarding the proposed research design, which uses distance to the nearest homeland as an instrument for migration. First, that the location of homelands is non-random and that these could have for instance been placed nearer to large industrial centres to serve as labour reservoirs. Secondly, that the constraint on internal mobility was binding.

The homelands established under apartheid (see Figure 2) were confined to areas designated as native reserves under the Native Land Act in 1913. This land comprised 7% of the overall area of South Africa and was largely inhabited by the black population, as the government was unwilling to expropriate white farmers. Hence the land allocation in 1913 failed to transfer large tracts of land between the different population groups and merely legally consolidated the distribution of land that had emerged predominantly through the European conquest of African land (Neame, 1962, p.40f). Since land was largely conquered for agricultural purposes, the African land reserves were of relatively low quality.

In 1913, South Africa was predominantly an agricultural economy with just two important industries - gold mining around Johannesburg and diamond mining around Kimberley. These industries established a system of migrant labour. Both found it optimal to change their entire workforce on a regular basis - every three to six months - and wanted workers' families to remain in reserves. This allowed firms to pay low

wages since the workers' families were supposed to find alternative work in the reserves (e.g. subsistence farming) which also constituted the (very low) opportunity cost of the worker. Additionally, they were able to send sick or injured workers back to the reserves where their tribe would take care of them (Lapping, 1986, p.26). This suggests that there was no need for specifically located labour reservoirs when the homelands were established. Therefore, no significant economic considerations appeared to have motivated the location of homelands, except for perhaps agricultural factors.

[Figure 2: Map of homelands]

The 1936 Land Act and subsequent Government initiatives aimed at consolidating native territories to make them viable as independent states. There were no attempts to relocate them for economic reasons. One possible economic reason would be the proximity of cheap labour. Instead of relocating the homelands, the government created black townships such as Soweto to serve as labour reservoirs. If a homeland was conveniently located, many inhabitants commuted to work in white cities (KwaMashu and Umlazi in the homeland KwaZulu provide an example). There were therefore no incentives to relocate homelands as other ways to increase the pool of cheap labour proved more convenient.

A second concern when analysing the switch from the constrained equilibrium for the black population under apartheid to the unconstrained equilibrium, is whether this constraint was binding. There are several observations that suggest that the constraint was indeed binding and that the switch to an unconstrained equilibrium was a significant shock to the distribution of population. First, the homelands were much poorer than other parts of South Africa. In 1985, GDP per capita in the homelands varied between 600 and 150 Rands, an order of magnitude below the 7,500 Rand estimated for the rest of South Africa (Christopher, 2001, p.93). Secondly, while more than 90% of whites and Indians lived in urban areas in 1986, less than 60% of blacks did, and we observe a large jump in urbanisation starting in the 1990s. Thirdly, while keeping blacks out of the urban areas was one of the major goals of apartheid policy, the absolute level of black population in 'white' urban areas increased nevertheless. This suggests that strong urban attraction pulled blacks into urban areas, while apartheid reduced the rate of urbanisation (Feinstein, 2005, p.157).

4 Theoretical Foundation

This section describes the three hypotheses that guide our empirical analysis and different implications of these hypotheses for the shape of the agglomeration curve in a modified Henderson (1974) model.

Hypothesis 1: *The population of a region shrinks below its new, back to its original population level (**Mean Reversion**).*

The population of the region will shrink back to its initial size if it is part of an optimal network of constant relative city sizes. Such an optimal network emerges if location fundamentals are the main drivers of the location of cities as argued by Davis and Weinstein (2002). In this case, migrants displace incumbents.

Hypothesis 2: *A region remains at its new population size. Population does not change endogenously after the exogenous shock (**Path Dependence**).*

The population level of a region does not change if urbanisation and the distribution of population are highly path dependent (Bleakley and Lin (2012), Michaels and Rauch (2013)), which can arise for various reasons. First, path dependence can be caused by sunk investment such as housing (Glaeser and Gyourko, 2005) that can explain why the distribution of population in Japan has been highly path dependent over millennia (Davis and Weinstein, 2002). Since migration after apartheid constitutes a positive as opposed to a negative population shock in the destination area however, this reasoning cannot apply. In our setting, path dependence can arise if the utility of incumbents is unaffected by the increase of population, or if there are high costs of moving for incumbents relative to gains from migration.

Hypothesis 3: *The population of a region grows by more than the initial exogenous shock (**Agglomeration**).*

In a world of multiple equilibria due to increasing returns - in line with benchmark models of economic geography (e.g. Krugman (1991)) - a city can grow further if the initial migration shock induces a shift between equilibria. If population exogenously increases beyond a threshold that fosters the development of new industries, this may alter the economic structure of an area significantly. Production technologies might also be affected which might increase labour demand and/or wages. Multiple equilibria arise because increasing returns create an incentive for agents to co-locate.

These three hypotheses can be conceptualised in a modified version of the Henderson (1974) model. In the model, agents derive utility from locating in a certain area. In equilibrium, there cannot be any gains from mobility such that utilities of all agents have to be equal across locations. In the original Henderson model utility is derived from consumption, in this version utility stems from the difference between the agglomeration ($A(N)$) and the congestion cost ($C(N)$) curves, which are both functions of population density (N).³ Spatial utility in region i is defined as: $U(N^i) = A(N^i) - C(N^i)$. The agglomeration curve summarises the consumption gains from a greater number of varieties as well as higher wages resulting from productivity gains due to agglomeration effects. The congestion cost curve is determined by rents and commuting costs. The population allocation equilibrium is determined by the indifference condition that the spatial utility from locating in a certain area has to be equalised across all K areas: $U(N^1) = \dots = U(N^K)$. When assuming a particular functional form for one of the two functions, we can infer characteristics of the shape of the other function from the three hypotheses outlined above. There are no intuitive guidelines on the shape of the agglomeration curve as a function of population density $A(N)$. The agglomeration function could be non-monotonic, as new industries could emerge that replace other industries when the population level crosses some threshold. This could result in significant changes in the structure of the local economy. For the congestion cost curve $C(N)$ on the other hand, given finite space, it is plausible to assume that it is increasing ($C'(N) > 0$) in population density, convex ($C''(N) > 0$) and tending to infinity after a certain population density threshold has been reached ($\lim_{N \rightarrow \bar{N}} C(N) = \infty$).

Given these assumptions on the congestion cost function, different shapes for the agglomeration function follow from the three hypotheses outlined above. The definition of equilibrium implies that the utility across locations has to be equal before the shock hits. The population movements as a reaction to the exogenous shock then again have to equalise the utility across locations to attain a new equilibrium.

[Figure 3: Three graphs as illustrations of theory]

In the empirical analysis, all areas are treated by a population shock with varying intensity depending upon geographical proximity to the homelands. The utility level in the initial equilibrium is denoted by $U(N_0)$. $U(N_S)$ denotes the utility level after the shock, and $U(N_1)$ is the utility level of the new equilibrium after agents have adjusted their location decisions. By the definition of equilibrium, $U(N_0)$ and $U(N_1)$ have to

³In the context of this stylised model, we use changes in the population level and changes in population density interchangeably, since we consider a fixed amount of space.

be equal across all locations (treated and control), while $U(N_S)$ is not related to an equilibrium and can therefore vary across locations.

The population level mean reverts (Panel A in Figure 3⁴) if the utility at the new population level $U(N_S)$ is below the utility in the initial equilibrium $U(N_0)$ that can be attained in the untreated areas. Agents will move from the treated areas to the control areas until the utilities are equalised across both types of locations $U(N_1^T) = U(N_1^C)$, which leads to a reduction of population below N_S . For the shape of the agglomeration function, this implies that its slope has to be locally smaller than the slope of the congestion cost function.

The evolution of population is path dependent (Panel B) if the utility at the population level after the shock is equal to the utility in the initial equilibrium $U(N_0) = U(N_S)$. This implies that there are no gains from moving between control and treatment areas and therefore no endogenous adjustment of location decisions takes place, such that the new population level is an equilibrium population level: $N_S = N_1$. Since the difference between agglomeration and congestion function at N_0 is equal to the difference at N_S , the slopes of the two functions between N_0 and N_S have to be equal. If this property holds globally, then there exist infinitely many equilibria of the spatial distribution of population.

In the case of agglomeration (Panel C), agents move from control areas to treated areas. This implies that gains from migration exist such that the utility level after the shock has to be greater than the utility in the initial equilibrium: $U(N_0) < U(N_S)$. Utility cannot be strictly increasing in population between N_0 and N_S however, because that would imply the existence of gains from migration at N_0 . The existence of such gains would contradict the definition of equilibrium, such that N_0 cannot be an equilibrium. For N_0 to be an equilibrium therefore, utility has to be non-monotonic, implying the existence of multiple equilibria for the spatial distribution of population. In order for the utility function to be non-monotonic, the slope of the agglomeration function has to be non-monotonic.⁵

⁴Note that the graphs only display the evolution of population in treated areas.

⁵Note that the functional form displayed in Panel C is just one example of a broad class of possible agglomeration functions. In particular, it is not necessary for the slope of $A(N)$ to be locally negative for the existence of multiple equilibria.

5 Empirical Analysis

5.1 Data

In order to empirically test the three hypotheses, we make use of a unique geographically referenced South African census dataset. It contains observations for the years 1991, 1996, 2001 and 2011 at the ward level and hence bridges across the democratic transition in 1994. This dataset consists of two parts. First, it contains publicly available census data aggregated to the ward level for the censuses in 1996, 2001 and 2011 provided by Statistics South Africa. This allows us to distinguish between the short-, medium- and long-run effects of the exogenous population shock. Secondly, it contains data from the last census under the apartheid government in 1991. De Kadt and Sands (2016) matched a partial enumerator area map from the census in 1991 with the 100% sample of the individual level census data made available by DataFirst at the University of Cape Town and aggregated it to the 2011 ward level. This last census was implemented in March 1991. This timing is crucial as the Native Land Act, the Population Registration Act and the Group Areas Act were repealed in June 1991. The Pass Laws Act was already repealed in 1986. While identification would be cleaner if data from before 1986 were available, this is not a major impediment to the identification strategy. Given that the Group Areas Act and the Population Registration Act were still in place, the black population was still severely constrained in its choice of residence until June 1991.⁶

[Table 2: Summary statistics around here]

5.2 Identification

Distance to the nearest homeland is implemented as an instrument for migration flows in order to causally identify the effect of migration on population distribution. The validity of the instrument relies on the conditional quasi-random allocation of homelands, which has been argued for extensively in Section 2. The assumption may be violated for areas adjacent to the homelands however, as they are likely to be affected

⁶The data from 1991 does not cover the entirety of South Africa. One general drawback of the dataset is that it does not cover the homelands. This does not affect the analysis since we only look at areas outside the homelands. Another more relevant drawback is that there are a few areas that are not covered within South Africa (see Figure 5 in the appendix). This is due to two reasons. First, Statistics South Africa only has a partial map of the census enumeration areas in 1991. Therefore, part of the census data cannot be geographically referenced. Secondly, due to violent turmoil at the time, some areas could not be visited by enumerators and no data are available on a granular level. This is potentially beneficial for our analysis since we exclude areas with high racial tension, which could otherwise bias our results. So, while this reduces the number of observations and therefore the statistical power in the empirical analysis, our parameter estimates will remain consistent.

by economic spillovers from the neighbouring homeland in a variety of ways that are not related to the cost of out-migration from the homelands. To adjust for this problem, we exclude areas within 10 km from the homelands as a robustness check to ensure that the estimates are not driven by local economic spillovers.

For our instrument to be informative people need to migrate according to gravity because the cost of migration increases in distance, which would imply that a town located nearer to the homelands *ceteris paribus* receives more migrants than a town further away. The gravity assumption is a common assumption in the migration literature⁷ and the informativeness can be tested empirically by looking at the partial correlation between instrument and endogenous variable. The informativeness of distance as an instrument crucially depends upon the level of fixed effects chosen, which affect the variation in the data. As shown in Table 6 in the appendix, the informativeness of the instrument decreases almost monotonically in the granularity of the fixed effects.⁸ A trade-off therefore exists between accounting for local unobservables and retaining sufficient identifying variation to ensure the implementation of an informative instrument. Province level fixed effects are chosen in order to maintain sufficient identifying variation while accounting for potential differences in policies at the province level.⁹

5.3 Estimation

To estimate the causal effect of migration on the distribution of population, the following system is estimated using two-stage least squares (2SLS):

$$BlackPopGr_{i,t} = \alpha_2 + \log(distance_i)\pi + \mathbf{X}_{i,1991}\gamma_2 + \delta_p + v_m \quad (1)$$

$$\Delta N_{i,t-1991} = \alpha_1 + \widehat{BlackPopGr}_{i,t}\beta + \mathbf{X}_{i,1991}\gamma_1 + \delta_p + \epsilon_m \quad (2)$$

where $\Delta N_{i,t-1991}$ denotes overall or non-black absolute population growth in ward i between 1991 and t . Controls are included for population groups, population density, education, gender ratio, employment and income ($\mathbf{X}_{i,1991}$) and province-level fixed effects (δ_p) (see Table 3). The errors are clustered at the municipality level.¹⁰ The ward level is the lowest geographical level that can be tracked consistently over time

⁷For example, Peri (2012) uses distance to the Mexican border as an instrument for the intensity of migration to different US states.

⁸This is intuitive, since for example when using municipality level fixed effects, the identifying variation of the instruments explains in which part of Johannesburg migrants are going to settle. This is likely to be uncorrelated with distance to homelands especially for urban areas.

⁹Provinces are equivalent to states in the US.

¹⁰Using Conley (1999) standard errors to account for spatial correlation yields similar results.

and municipalities are the lowest level of local government. Distance is defined as the distance to the nearest homeland measured from centroid to centroid. Since no measure of domestic migration is available in the census data, black population growth conditional on fixed effects and covariates is used as a proxy for black migration.¹¹ A dummy variable for Cape Town is also included as the municipality is a special case in terms of location, politics and demographics and hence migration patterns. The Western Cape was the only province where the African National Congress did not come first in the general elections in 1994. Until today, it has not achieved the political dominance in the province or the municipality of Cape Town that it has in the rest of the country. In terms of demographics, there is a much higher white and especially coloured population in Cape Town than anywhere else in the country. Most importantly, there is a lot of circular migration from the Eastern and the Northern Cape into Cape Town. These migration dynamics in particular are quite distorting for the identification strategy and accordingly including a dummy for Cape Town significantly increases the predictive power of the instrument.

5.4 Linking theory and the variable of interest in the empirical estimation

In order to test the three competing hypotheses outlined in Section 4, it proves crucial to link the predictions from the hypotheses to the parameter of interest β . If the underlying process was driven by mean reversion, then the effect of the exogenous population shock as measured by β would be less than one and decreasing over time, as the shock dissipates through the urban system. In the case of path dependence, β would be expected to be equal to one in all periods. In the agglomeration scenario, β would be significantly greater than one. If we use non-black instead of overall population growth, the same classification holds but has to be defined relative to 0 rather than to 1.

In order to assign these theoretical interpretations to the estimated parameters, we have to avoid using percentage growth rates in the endogenous variable and in the dependent variables in the second stage. Using percentage growth would make the shock a function of the share of black population that already lives in the area, which makes the interpretation we are looking for impossible. Therefore we instead define our variables of interest as absolute growth rates relative to the overall population in period t where t corresponds to 1996, 2001 and 2011. We incorporate this normalising factor since the size of the population shock should be measured relative to the overall

¹¹As has been used previously as a proxy for migration status, see for example Czaika and Kis-Katos (2009).

population rather than in absolute terms to get a good understanding of its impact.

6 Results

6.1 Baseline results

Table 3 summarises the main results and Table 4 provides further results from different sub-samples as robustness checks. Each cell of the tables summarises one regression.¹² Before moving on to interpreting the estimated coefficients of interest, we will discuss a number of results. The difference of the parameters in Columns (1) and (2) in Table 3 sums up to one. This is a mechanical effect stemming from the relation between the absolute population growth values of black, non-black and overall population growth.¹³ This result holds exactly in the OLS specifications and approximately in the 2SLS estimations (Columns (3) and (4)). We therefore focus on the results for the effects on the non-black population in all other regressions and the subsequent discussion. The Angrist and Pischke (2009) F-statistic of the first stage is well above the rule of thumb threshold of 10 for all specifications for the medium and long horizon.¹⁴ Weak instrument problems only arise for the short period between 1991-1996 and we will not discuss these parameter estimates. The increased explanatory power over the longer time horizons is consistent with the fact that migration decisions only adjust intermittently to a change in policy such as the end of apartheid.

In the OLS regression, we cannot reject the null hypothesis that the coefficient is different from zero in the short-run (1991-1996). For the two subsequent periods on the other hand, the coefficient estimates are well below zero. This suggests that black population growth occurred in areas with low non-black population growth and vice versa. These results should not be assigned a causal interpretation however, since the result could be driven by unobserved shocks that induce black in-migration and white out-migration or vice versa jointly. The baseline results from the two-stage least squares estimation (Columns (3) and (4)) suggest that the coefficient is not different from zero at any horizon such that there is no causal effect from exogenous black migration on aggregate migration decisions of non-black incumbents. This is strong evidence that an exogenous population shock is absorbed without endogenous reaction of the population level. The results suggest that the effect of an exogenous population shock on the aggregate long-run equilibrium of the population distribution is consistent with the

¹²I.e. each cell in the first row of Table 4 summarises the causal partial effect of exogenous migration of black population between 1991 and 1996 on the overall absolute non-black population growth rate between 1991 and 1996.

¹³ $\frac{\Delta P_t}{\Delta B_t} = \frac{\Delta NB_t + \Delta B_t}{\Delta B_t} = \frac{\Delta NB_t}{\Delta B_t} + 1$

¹⁴The corresponding first stage regressions for Tables 3 and 4 are reported in the appendix.

theoretical notion of path-dependence (Hypothesis 2).

[Table 3: Baseline regression results]

In addition to the baseline regressions, we report several regressions based on different sub-samples as robustness checks (Table 4). We include a dummy for Johannesburg in Column (1) as the largest metropolitan area and industrial centre to ensure that it is not driving the results. As outlined above, we exclude areas close to the homelands, since for these distance to the nearest homeland could affect them not only through migration, but also through local economic spillovers (Column (2)). We also exclude areas with a low share of white population in 1991 because the migration restrictions under apartheid might have been less binding for these areas (Columns (3) and (4)). As a further robustness check, we exclude the areas in the upper tail of the distance distribution in Column (5) to ensure that the high number of observations in the upper tail of the distance distribution does not skew the results. In Column (6) we report results using district instead of province fixed effects. We also aggregate wards up to the municipality level and run a separate regression to test whether the results are robust to using a different level of aggregation (Column (7)). These robustness tests using different sub-samples as reported in Table 4 corroborate our baseline findings since none of the coefficients significantly deviates from zero.

[Table 4: Robustness baseline regression results]

Overall, there is strong evidence for path dependence (Hypothesis 2). These results suggest that, in the aggregate, there is no strong evidence for multiple equilibria and a non-monotonic agglomeration curve or mean-reverting behaviour. The evidence in favour of path-dependence is consistent with an agglomeration function that has the same slope as the congestion function or high costs of migration as found by Imbert and Papp (2014) for temporary labour migration in India. This corroborates the dynamics found by Bleakley and Lin (2012) for fall line cities in the US and Michaels and Rauch (2013) for Roman cities in France and Britain.

6.2 Heterogeneity

This Section discusses how the causal effect of migration on population growth varies across two dimensions, the initial population density or level of urbanisation of an area and the size of the exogenous population shock.

First, we are interested in how the causal effect varies with initial population density. In this case, theory does not provide clear guidance. Due to the convexity of the cost curve, the causal effect of migration could be decreasing in initial population density because the additional costs generated by new migrants reduce the utility level of incumbents. At the same time, Michaels et al. (2012) show that long-run population growth in the US is smaller for low initial population densities and increases with population density after a cut-off of 7 people per km². Such a result would be consistent with an agglomeration curve that is much steeper in urban than in rural areas. This could suggest that in densely populated areas, exogenous migration leads to a larger increase in population than in less densely populated areas.

In order to estimate how the effect of a positive population shock varies across initial population densities, we define a dummy variable for high initial population density and for high initial share of urbanised households. The results reported in Table 5 show that there is a positive and significant interaction between high initial density and the population shock. This suggests that areas with high initial density experience a significant endogenous inflow of population as a reaction to the exogenous population shock while others do not. This effect exists in the medium and long-run but is economically stronger in the medium run for population density. It loses significance in the long-run for the share of urbanised households. The results of both specifications indicate that the population dynamics induced by a positive population shock differ between less densely populated rural areas and highly populated urban areas. While we cannot reject path dependence for rural areas, there is a significant and positive effect in urban areas suggesting that an exogenous population shock leads to endogenous immigration. This is evidence for the existence of multiple equilibria within urban areas.

[Table 5: Dummy for high initial population density/share of urban households]

We also add the size of the shock as an additional dimension of heterogeneity. So we estimate how non-black population growth varies with the size of the shock and initial population density. In order to do so we combine the deciles of the two distributions and estimate 100 separate conditional means:

$$\Delta NB_{i,2011} = \sum_{j=1}^{10} \sum_{k=1}^{10} \beta_{j,k} \left[1[\text{if } Popden_{i,1991} \text{ in decile } j] \times 1[\text{if } distance_i \text{ in decile } k] \right] + \gamma' X_{i,1991} + \delta_p + \epsilon_m \quad (3)$$

The $\beta_{j,k}$ s are the coefficients of interest and estimate how the conditional non-black

population growth varies by the deciles of the initial population density distribution and the size of the shock distribution. The size of the shock is measured using the reduced form, i.e. distance to the nearest homeland. While the estimates do not provide the same clear cut causal evidence as the two stage least squares approach they are indicative on how the effect of the exogenous population shock varies with the size of the shock and the initial density. The results displayed in Figure 4 suggest that for a given initial density an increase in the size of a shock results in a higher population growth rate. This is in line with the idea that it requires a substantial shock to switch between equilibria.

[Figure 4: 3d graph]

The fact that the population of more densely populated areas increases relative to less densely populated areas could be interpreted as a ‘Matthew effect’¹⁵ of an exogenous population shock. In general terms, a ‘Matthew effect’ describes a situation where ‘the rich get richer and the poor get poorer’.¹⁶ In this case, areas rich in population gain over-proportionally from a positive population shock.

In the context of the modified Henderson model presented in Section 4, this result suggests that the shape of the agglomeration function is different between urban and rural areas for the relevant population levels. In rural areas the gains from agglomeration are below the increased congestion cost if the population increases exogenously. In urban areas, the gains from an increase in population seem to be equal to the additional costs. The gains from agglomeration therefore seem to be much larger in urban areas than in less densely populated rural areas.

These heterogeneous gains from agglomeration arise in a simple two-sector economic geography model. The agricultural sector produces food using a fixed endowment of land and labour under a technology with decreasing returns to labour.¹⁷ The industrial sector, consisting of manufacturing and services, produces consumption goods using capital and labour with external agglomeration economies. Labour is perfectly mobile across sectors and locations. Areas with low population density are predominantly agricultural, while urban areas are predominantly industrial. If an exogenous population

¹⁵‘For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath.’ Matthew 25:29, (American Bible Society, 1999).

¹⁶Similar dynamics have been discovered in various fields such as the philosophy of science (Merton, 1968), education (Adams, 1990) and individual career dynamics (Petersen et al., 2011). In economics it is well established in the new new trade literature where big firms gain over-proportionally from trade liberalisation (Mrázová and Neary, 2013).

¹⁷Capital could be included as an additional factor in production but does not change the resulting dynamics and is therefore omitted.

shock hits both urban and rural areas, the marginal product of labour decreases in rural areas and generates displacement effects because the real wage decreases. This dynamic arises naturally from the assumption that there is only a fixed amount of land available for agricultural production. In urban areas, an increase in the labour force generates higher investment in capital (assuming a constant real interest rate equal to the world interest rate). Therefore, the marginal product of labour does not fall and might even increase due to external economies of scale. This generates agglomeration effects or a path-dependent evolution of population in urban areas.

A similar result emerges in a standard model used in the migration literature (e.g. Borjas (1999) and Kremer and Watt (2006)) that distinguishes between low- and high-skilled labour used in production in urban areas. The production in rural areas only uses low-skilled labour and the fixed amount of land as inputs with the same technology as above. In urban areas, low- and high-skilled labour are used as complements in production with a constant returns to scale technology. In this framework, the population shock we analyse in the data is best approximated by an increase of unskilled labour, since the apartheid government only provided a bare minimum of schooling to the black population (Feinstein, 2005, p.159f). In the model, an increase in unskilled labour increases the wage for high-skilled labour and the rents for capital. If the supply of capital is elastic, this leads to an increase in capital and an inflow of skilled workers such that all factor prices return to their initial equilibrium values. Therefore, an exogenous increase in the number of unskilled workers attracts skilled workers such that the population level of urban areas experiences agglomeration and a shift towards a new equilibrium.

7 Conclusion

We study the effect of an exogenous migration shock generated by the abolition of migration restrictions for the black population on the distribution of population in South Africa. There are three ways in which an area can react to an exogenous population shock that arise from different theories describing the distribution of population in space. The population level of an area could mean revert towards its initial level, it could remain at the new population level (path dependence) or it could grow further, i.e. agglomerating population, suggesting the existence of multiple equilibria. Using a modified Henderson model and assuming an increasing and convex congestion cost curve, we are able to infer the shape of the agglomeration function from these different predictions.

The empirical results suggest that in the aggregate, the reaction of the population level to an exogenous population shock is consistent with path dependence. For the modified Henderson model, where the spatial utility is equal to the difference between the agglomeration and congestion cost curve, path dependence implies that the two curves have a similar slope such that many, possibly infinitely many equilibria of the distribution of population in space exist. This has important policy implications. Since the population level of a region behaves according to path dependence, a temporary policy measure that induces migration can permanently change the distribution of population.

Additionally, we find that the reaction of an area to an exogenous population shock varies with the initial population density. In rural areas with low initial population densities, the effect of an exogenous population shock is significantly smaller than in urban areas with high population densities. In urban areas, the dynamics of the population level are consistent with agglomeration. We provide evidence that for a given initial population density a bigger exogenous population shock leads to more endogenous immigration and therefore makes the transition to a new equilibrium more likely. In the context of the modified Henderson model, this result shows that the agglomeration curve in rural areas is much more concave than in urban areas and also suggests that its slope is not monotone. These results are consistent with a simple economic geography model where production in rural areas features decreasing returns to labour due to a fixed endowment of land usable for agricultural purposes. A steeper agglomeration function in urban areas also emerges in a standard model from the migration literature that features complementarities between low- and high-skilled labour in urban, but not in rural areas. If an exogenous population shock hits both rural and urban areas, these different dynamics increase the share of the population living in cities. Exogenous migration, thus, generates urbanisation. From a public policy perspective, this is of vital importance because it suggests that urbanisation can be engineered by public policies that induce migration.

References

- ADAMS, M. J. (1990): *Beginning to Read: Thinking and Learning about Print*. MIT Press, Cambridge.
- AMERICAN BIBLE SOCIETY (ed.) (1999): *The Holy Bible, King James Version*. American Bible Society.
- ANGRIST, J. D., AND J. S. PISCHKE (2009): *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.
- AUERBACH, F. (1913): “Das Gesetz der Bevölkerungskonzentration,” *Petermanns Geogr Mitt*, 59, 74–76.
- BEINART, W. (2001): *Twentieth-Century South Africa*. Oxford University Press.
- BLACK, D., AND V. HENDERSON (2003): “Urban evolution in the USA,” *Journal of Economic Geography*, 3(4), 343–372.
- BLEAKLEY, H., AND J. LIN (2012): “Portage and Path Dependence,” *The Quarterly Journal of Economics*, 127(2), 587–644.
- BORJAS, G. J. (1999): “The economic analysis of immigration,” *Handbook of Labor Economics*, 3, 1697–1760.
- BRAKMAN, S., H. GARRETSEN, AND M. SCHRAMM (2004): “The strategic bombing of German cities during World War II and its impact on city growth,” *Journal of Economic Geography*, 4(2), 201–218.
- CHRISTOPHER, A. J. (2001): *The Atlas of Changing South Africa*. Routledge.
- CONLEY, T. G. (1999): “GMM estimation with cross sectional dependence,” *Journal of Econometrics*, 92(1), 1–45.
- CZAIKA, M., AND K. KIS-KATOS (2009): “Civil Conflict and Displacement: Village-Level Determinants of Forced Migration in Aceh,” *Journal of Peace Research*, 46(3), 399–418.
- DAVIS, D. R., AND D. E. WEINSTEIN (2002): “Bones, Bombs, and Break Points: The Geography of Economic Activity,” *American Economic Review*, 92(5), 1269–1289.
- DE KADT, D., AND H. LARREGUY (2016): “Agents of the Regime? Traditional Leaders and Electoral Clientelism in South Africa,” *working paper*.
- DE KADT, D., AND M. SANDS (2016): “Segregation drives racial voting: New evidence from South Africa,” *working paper*.

- DINKELMAN, T. (2011): “The Effects of Rural Electrification on Employment: New Evidence from South Africa,” *American Economic Review*, 101(7), 3078–3108.
- (2013): “Mitigating Long-run Health Effects of Drought: Evidence from South Africa,” Working Paper 19756, National Bureau of Economic Research.
- DURANTON, G. (2007): “Urban evolutions: The fast, the slow, and the still,” *The American Economic Review*, pp. 197–221.
- EECKHOUT, J. (2004): “Gibrat’s law for (all) cities,” *American Economic Review*, pp. 1429–1451.
- FEINSTEIN, C. (2005): *An Economic History of South Africa*. Cambridge University Press.
- GABAIX, X. (1999): “Zipf’s Law for Cities: An Explanation,” *The Quarterly Journal of Economics*, 114(3), 739–767.
- GLAESER, E. L., AND J. GYOURKO (2005): “Urban Decline and Durable Housing,” *Journal of Political Economy*, 113(2), 345–375.
- GONZÁLEZ-VAL, R., L. LANASPA, AND F. SANZ-GRACIA (2013): “New evidence on Gibrat’s law for cities,” *Urban Studies*, 51(1), pp. 93–115.
- HENDERSON, J. V. (1974): “The Sizes and Types of Cities,” *The American Economic Review*, 64(4), pp. 640–656.
- HOLMES, T. J., AND S. LEE (2010): “Cities as six-by-six-mile squares: Zipf’s law?,” in *Agglomeration Economics*, pp. 105–131. University of Chicago Press.
- IMBERT, C., AND J. PAPP (2014): “Short-term Migration and Rural Workfare Programs: Evidence from India,” *Job Market Paper, University of Oxford*.
- KREMER, M., AND S. WATT (2006): “The globalization of household production,” *Weatherhead Center For International Affairs, Harvard University*.
- KRUGMAN, P. (1991): “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 99(3), 483–99.
- LAPPING, B. (1986): *Apartheid: A History*. Grafton Books.
- MERTON, R. K. (1968): “The Matthew effect in science,” *Science*, 159(3810), 56–63.
- MICHAELS, G., AND F. RAUCH (2013): “Resetting the Urban Network: 117-2012,” *CEP Discussion Paper No 1248*.

- MICHAELS, G., F. RAUCH, AND S. REDDING (2012): “Urbanization and structural transformation,” *Quarterly Journal of Economics*, 127(2), 535–586.
- MIGUEL, E., AND G. ROLAND (2011): “The long-run impact of bombing Vietnam,” *Journal of Development Economics*, 96(1), 1 – 15.
- MRÁZOVÁ, M., AND J. P. NEARY (2013): “Selection effects with heterogeneous firms,” *CEPR Discussion Paper No. DP9288*.
- NEAME, L. E. (1962): *The History of Apartheid*. Pall Mall Press.
- OGURA, M. (1996): “Urbanization and apartheid in South Africa: Influx controls and their abolition,” *The Developing Economies*, 34(4), 402–423.
- PERI, G. (2012): “The Effect Of Immigration On Productivity: Evidence From U.S. States,” *The Review of Economics and Statistics*, 94(1), 348–358.
- PETERSEN, A. M., W.-S. JUNG, J.-S. YANG, AND H. E. STANLEY (2011): “Quantitative and empirical demonstration of the Matthew effect in a study of career longevity,” *Proceedings of the National Academy of Sciences*, 108(1), 18–23.
- ROSSI-HANSBERG, E., AND M. L. WRIGHT (2007): “Urban structure and growth,” *The Review of Economic Studies*, 74(2), 597–624.
- SCHUMANN, A. (2014): “Persistence of Population Shocks: Evidence from the Occupation of West Germany after World War II,” *American Economic Journal: Applied Economics*, 6(3), 189–205.
- SOO, K. T. (2007): “Zipf’s Law and urban growth in Malaysia,” *Urban Studies*, 44(1), 1–14.
- SURPLUS PEOPLE PROJECT (1985): *The Surplus People*. Ravan Press, Johannesburg.
- TUROK, I. (2012): *Urbanisation and Development in South Africa: Economic Imperatives, Spatial Distortions and Strategic Responses*. International Institute for Environment and Development.
- ZIPF, G. K. (1949): *Human behavior and the principle of least effort*. Addison-Wesley Press.

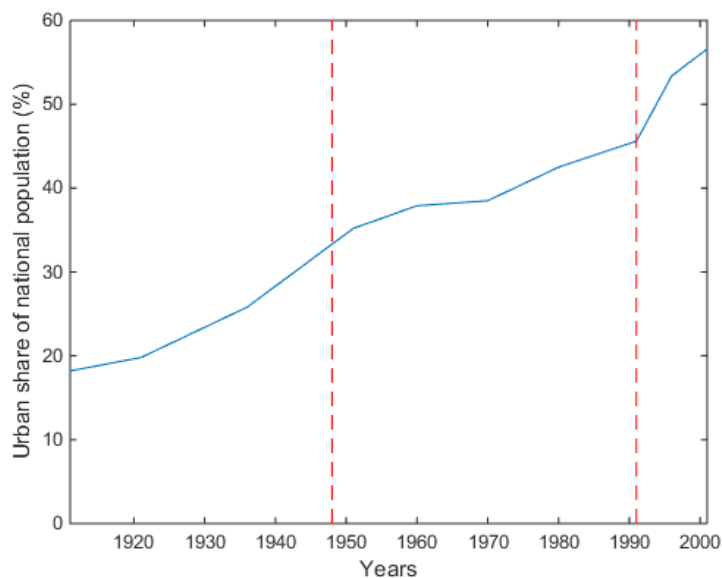
A Tables and figures

Table 1: DESCRIPTIVE STATISTICS ON THE POPULATION DISTRIBUTION

Distribution of black population across area types				Share of urbanised population across population groups				
Year	Urban	Rural	Homelands	Year	White	Coloured	Indian	Black
1950	25.4	34.9	39.7	1951	78	65	78	27
1960	29.6	31.3	39.1	1960	84	68	83	32
1970	28.1	24.5	47.4	1980	88	75	91	49
1980	26.7	20.6	52.7	1991	91	83	96	58

Source: Surplus People Project (1985, p.18)

Figure 1: URBAN SHARE OF THE NATIONAL POPULATION (%), 1911-2001



Data from Turok (2012), red vertical lines mark the apartheid regime of the National Party (1948-1991)

Figure 2: HOMELANDS (BANTUSTANS) ESTABLISHED UNDER APARTHEID

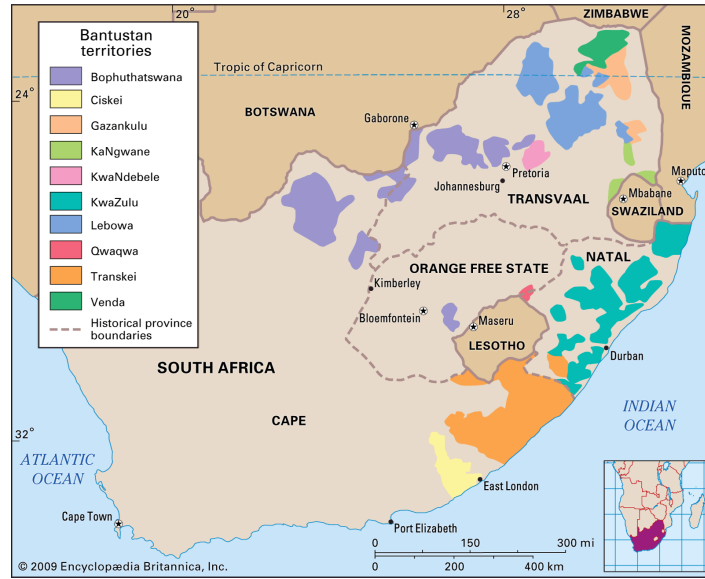


Table 2: SUMMARY STATISTICS OF INCLUDED VARIABLES

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
EXCLUDED INSTRUMENT					
log distance	2,093	4.092	1.564	0.0529	6.746
ENDOGENOUS VARIABLES					
Δ Black Population (1991-1996)	2,093	-1.081	20.75	-615.4	0.200
Δ Black Population (1991-2001)	2,093	0.0310	0.0381	-0.200	0.100
Δ Black Population (1991-2011)	2,093	0.0179	0.0207	-0.168	0.0499
DEPENDENT VARIABLES					
Δ Total Population (1991-1996)	2,093	-1.787	29.81	-829	0.200
Δ Total Population (1991-2001)	2,093	0.0348	0.0417	-0.358	0.1000
Δ Total Population (1991-2011)	2,093	0.0214	0.0213	-0.168	0.0500
Δ Nonblack Population (1991-1996)	2,093	-0.709	18.51	-773.8	0.192
Δ Nonblack Population (1991-2001)	2,093	0.00380	0.0211	-0.294	0.0960
Δ Nonblack Population (1991-2011)	2,093	0.00354	0.00989	-0.0669	0.0425
PROVINCE FIXED EFFECTS					
Eastern Cape	2,093	0.100	0.301	0	1
Free State	2,093	0.0994	0.299	0	1
Gauteng	2,093	0.172	0.378	0	1
KwaZulu-Natal	2,093	0.145	0.352	0	1
Limpopo	2,093	0.0674	0.251	0	1
Mpumalanga	2,093	0.102	0.302	0	1
North West	2,093	0.0726	0.260	0	1
Northern Cape	2,093	0.0717	0.258	0	1
Western Cape	2,093	0.170	0.375	0	1

Figure 3: MODIFIED HENDERSON MODEL WITH GAINS FROM AGGLOMERATION AND CONGESTION COSTS

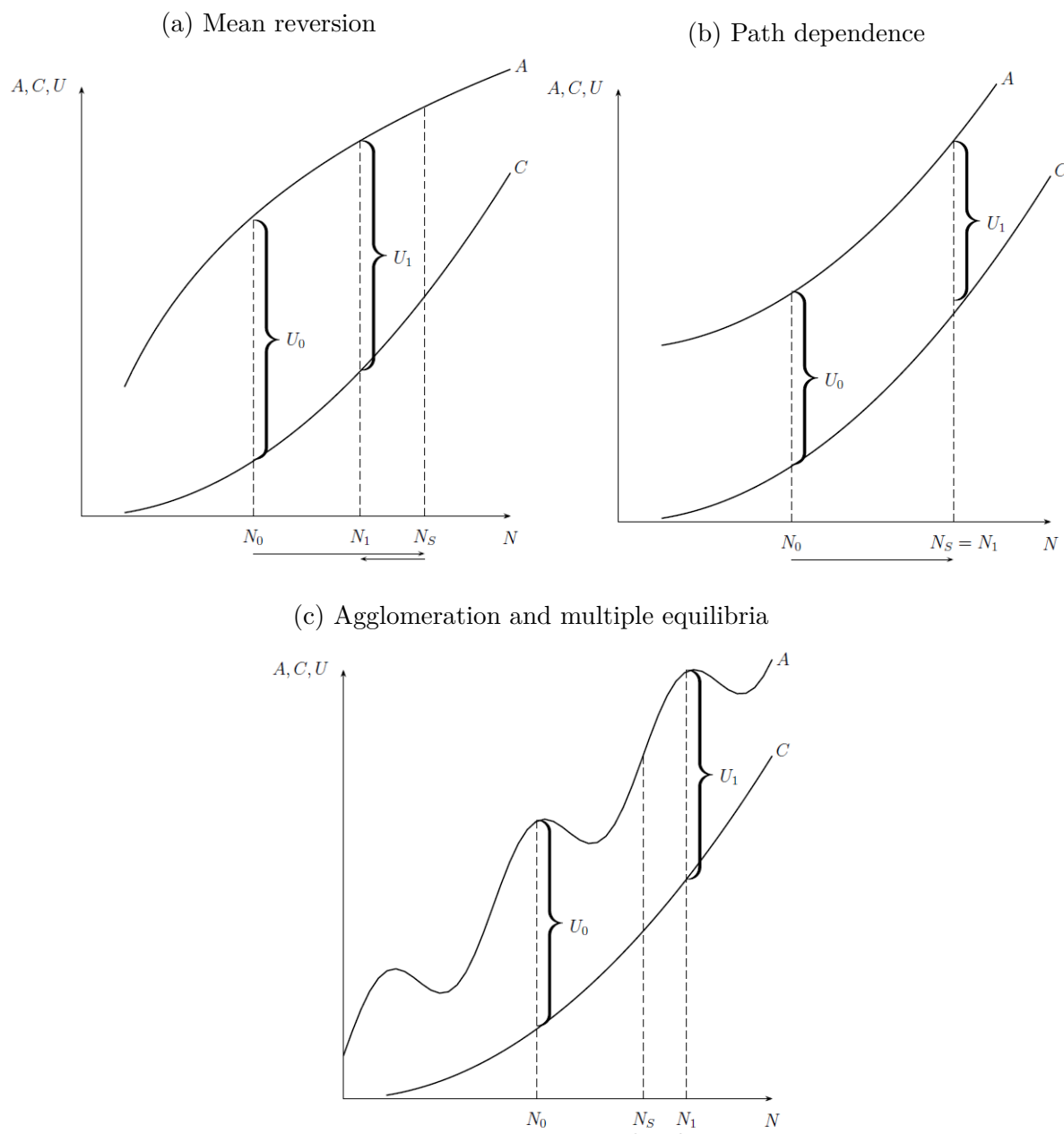


Table 2 - continued

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
CONTROL VARIABLES (FROM 1991 CENSUS IN LOGS)					
Male share	2,093	0.506	0.0826	0	1
Population group ratio	2,093	0.275	0.321	0	1
Population density	2,093	4.647	2.478	0.0148	10.40
Total population	2,093	8.282	1.197	0.693	10.45
Black population	2,093	6.712	1.997	0	9.944
Employed	2,093	7.261	1.306	0	9.635
Unemployed	2,093	4.977	1.389	0	8.183
Not economically active	2,093	7.665	1.258	0	9.925
No schooling	2,093	6.864	1.154	0	9.317
Some primary schooling	2,093	6.807	1.188	0	9.135
Finished primary school	2,093	5.414	1.168	0	8.255
Some secondary schooling	2,093	6.792	1.333	0	9.585
Finished secondary school	2,093	5.952	1.578	0	9.404
Higher education	2,093	3.442	1.899	0	8.283
No income	2,093	7.618	1.250	0	9.932
Income: R1-499	2,093	3.655	1.391	0	7.349
Income: R500-699	2,093	3.266	1.285	0	6.852
Income: R700-999	2,093	3.775	1.246	0	6.952
Income: R1000-1499	2,093	4.610	1.246	0	7.594
Income: R1500-1999	2,093	4.604	1.228	0	7.489
Income: R2000-2999	2,093	5.188	1.302	0	7.856
Income: R3k-4k	2,093	5.111	1.280	0	7.953
Income: R5k-6k	2,093	4.706	1.259	0	8.084
Income: R7k-9k	2,093	4.786	1.396	0	8.357
Income: R10k-14k	2,093	4.874	1.498	0	9.125

Table 3: OLS AND 2SLS BASELINE REGRESSIONS

	(1)	(2)	(3)	(4)
	OLS		2SLS	
	Population growth	Nonblack growth	Population growth	Nonblack growth
<i>Panel A: Population growth rates (1991-1996)</i>				
Δ Black Population	1.126 (0.0814)	0.127 (0.0815)	0.360 (0.862)	-0.640 (0.862)
FS AP F-Stat	-	-	2.52	2.52
<i>Panel B: Population growth rates (1991-2001)</i>				
Δ Black Population	0.899* (0.0393)	-0.101* (0.0393)	1.061 (0.115)	0.0614 (0.115)
FS AP F-Stat	-	-	29.98	29.98
<i>Panel C: Population growth rates (1991-2011)</i>				
Δ Black Population	0.895*** (0.0236)	-0.105*** (0.0236)	0.993 (0.0873)	-0.00709 (0.0873)
FS AP F-Stat	-	-	42.95	42.95
Province fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2093	2093	2093	2093

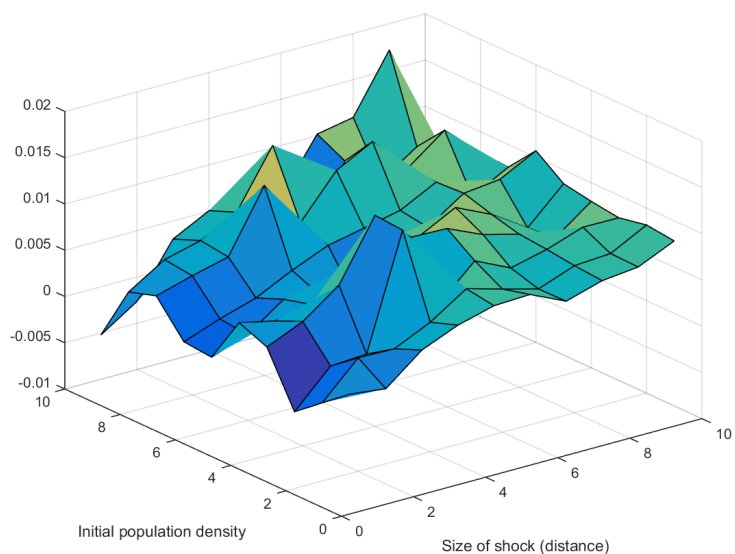
Notes. This Table displays estimates of equation (2) in the text. Each cell presents estimates from a separate regression. The baseline sample consists of all wards inside South Africa for which 1991 data is available. The standard errors are clustered on the municipality level. There are 201 clusters. Columns (1) and (2) are estimated using OLS and columns (3) and (4) are estimated using 2SLS where the natural log of distance to the nearest homeland is used to instrument for absolute black population growth in the relevant time period divided by the overall population. The outcome variable is absolute overall or non-black population growth in the relevant time period divided by overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. There are nine provinces for which fixed effects are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are significantly different from zero at the 90% level of confidence are marked with a *; at the 95% level, a **; and at the 99% level, a *** in columns (2) and (4). In columns (1) and (3) they denote statistical significant departures from one. Standard errors in parentheses.

Table 4: 2SLS REGRESSIONS USING DIFFERENT SUB-SAMPLES

	(1) Dummy for Johannesburg	(2) Drop within 10 km	(3) Drop < 5% white	(4) Drop < 10% white	(5) Drop distance ≥ 6	(6) District fixed effects	(7) Municipality level
<i>Panel A: Population growth rates (1991-1996)</i>							
Δ Black Population	-0.640 (0.861)	-0.385 (0.442)	22.37 (32.85)	19.75 (25.53)	-0.655 (0.853)	-1.652 (1.926)	-0.174*** (0.0576)
FS AP F-Stat	2.52	1.36	0.25	0.31	2.61	1.67	10.27
<i>Panel B: Population growth rates (1991-2001)</i>							
Δ Black Population	0.0688 (0.112)	0.0661 (0.189)	0.170 (0.144)	0.231 (0.173)	0.00799 (0.108)	-0.0392 (0.0837)	0.0379 (0.154)
FS AP F-Stat	32.23	12.51	30.67	25.55	30.55	34.68	11.17
<i>Panel C: Population growth rates (1991-2011)</i>							
Δ Black Population	-0.00019 (0.0857)	0.0460 (0.131)	0.0486 (0.139)	0.0618 (0.146)	-0.0747 (0.0879)	-0.119 (0.0771)	-0.0742 (0.151)
FS AP F-Stat	44.67	20.69	31.09	28.89	42.81	36.59	12.91
District fixed effects	No	No	No	No	No	Yes	No
Province fixed effects	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2093	1790	1374	1137	1730	2093	203

Notes. This Table displays estimates of equation (2) in the text for different sub-samples. Column headings denote sub-sample used in each specification. Each cell presents estimates from a separate regression. The standard errors are clustered on the municipality level. There are 201 clusters. All columns are estimated using 2SLS where the natural log of distance to the nearest homeland is used to instrument for absolute black population growth in the relevant time period divided by the overall population. The outcome variable is absolute non-black population growth in the relevant time period divided by the overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. There are nine provinces for which fixed effects are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are significantly different from zero at the 90% level of confidence are marked with a *; at the 95% level, a **; and at the 99% level, a ***. 95% confidence intervals are in brackets.

Figure 4: THE EFFECT OF INITIAL DENSITY AND SIZE OF SHOCK FOR POPULATION GROWTH



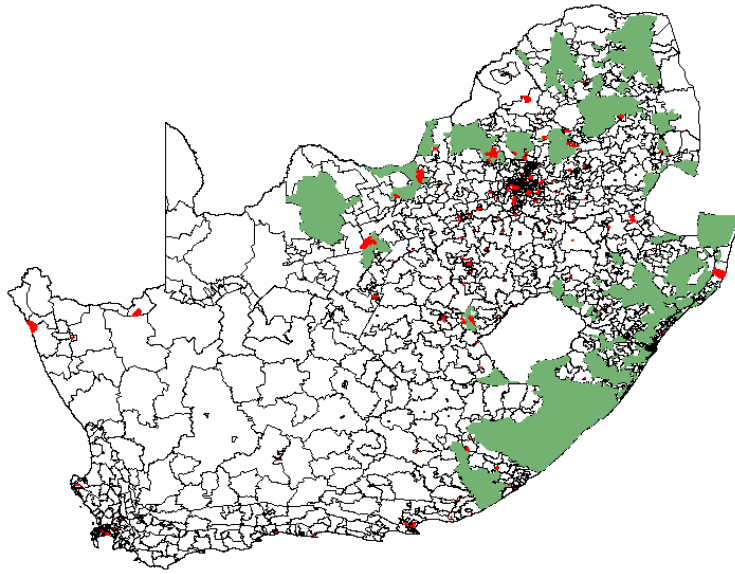
This figure displays the $\beta_{j,k}$ coefficients resulting from estimating equation (3) in the main text. The size of the shock increases along the x-axis. It starts off with the highest decile of the distance distribution going to the decile with the lowest values (i.e. those closest to a homeland). Similarly, the first value on the y-axis corresponds to those wards in the lowest decile of the initial population density distribution while the last one contains the highest decile. The z-axis displays differences in the conditional mean of non-black population growth in the period 1991-2011.

Table 5: HETEROGENEITY WITH RESPECT TO THE INITIAL POPULATION DENSITY AND LEVEL OF URBANISATION

	(1) High population density dummy	(2) High urban share dummy
<i>Panel A: Population growth rates (1991-1996)</i>		
Δ Black Population	-0.897 (1.425)	-7.333 (11.35)
High initial urban share dummy \times Δ Black Population		7.589 (11.56)
High initial population density dummy \times Δ Black Population	0.509 (1.133)	
FS AP F-Stat: Δ Black Population	0.81	0.07
FS AP F-Stat: Urban interaction	-	0.07
FS AP F-Stat: Density interaction	0.45	-
<i>Panel B: Population growth rates (1991-2001)</i>		
Δ Black Population	0.106 (0.138)	0.00167 (0.111)
High initial urban share dummy \times Δ Black Population		0.178** (0.0840)
High initial population density dummy \times Δ Black Population	0.681** (0.284)	
FS AP F-Stat: Δ Black Population	26.54	36.98
FS AP F-Stat: Urban interaction	-	23.70
FS AP F-Stat: Density interaction	18.71	-
<i>Panel C: Population growth rates (1991-2011)</i>		
Δ Black Population	-0.0432 (0.0924)	-0.0138 (0.0827)
High initial urban share dummy \times Δ Black Population		0.0399 (0.0619)
High initial population density dummy \times Δ Black Population	0.348** (0.152)	
FS AP F-Stat: Δ Black Population	44.53	46.50
FS AP F-Stat: Urban interaction	-	27.49
FS AP F-Stat: Density interaction	17.67	-
Province fixed effects	Yes	Yes
Controls	Yes	Yes
Observations	2093	2093

Notes. This Table displays estimates of equation (2) in the text with an additional interaction term. Each column displays one specification. The standard errors are clustered on the municipality level. There are 201 clusters. All columns are estimated using 2SLS. Absolute black population growth divided by the overall population and the same term interacted with a dummy for high population density in 1991 or for high urban share of households are the endogenous variables. Log distance to the nearest homeland and log distance to the nearest homeland times a dummy for high population density in 1991 or high urban share of households are used as instruments for the endogenous variables. An area is defined as having a high initial population density if it is among the 25% most dense areas. An area is defined as having a high urban share if it is among the areas with the 75% highest share of urban households in 1991. The outcome variable is absolute non-black population growth in the relevant time period divided by the overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. There are nine provinces for which fixed effects are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are significantly different from zero at the 90% level of confidence are marked with a *; at the 95% level, a **; and at the 99% level, a ***. 95% confidence intervals are in brackets.

Figure 5: MISSING WARDS



All wards are hollow. Those areas with red borders and fill are missing. The former homelands are green.

Table 6: SUMMARY OF FIRST STAGE REGRESSIONS FOR THE BASELINE SPECIFICATIONS WITH DIFFERENT FIXED EFFECTS

	(1) No FE	(2) Province FE	(3) District FE	(4) Municipality FE
<i>Panel A: Population growth rates (1991-1996)</i>				
log distance	0.308 (0.365)	-0.940 (0.593)	-0.936 (0.723)	-1.148 (0.993)
<i>Panel B: Population growth rates (1991-2001)</i>				
log distance	-0.00733*** (0.000811)	-0.00675*** (0.00123)	-0.00647*** (0.00110)	-0.00323** (0.00144)
<i>Panel C: Population growth rates (1991-2011)</i>				
log distance	-0.00423*** (0.000398)	-0.00354*** (0.000540)	-0.00381*** (0.000629)	-0.00172** (0.000737)
Level of Fixed effects	No	Province	District	Municipality
Controls	Yes	Yes	Yes	Yes
Observations	2093	2093	2093	2093

Notes. This Table displays estimates of equation (1) in the main text. Column headings denote different specification. Each cell presents estimates from a separate regression. The standard errors are clustered on the municipality level. There are 201 clusters. All columns are estimated using OLS where the natural log of distance to the nearest homeland is the variable of interest. The outcome variable is absolute black population growth in the relevant time period divided by the overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. Fixed effects at varying levels are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are statistically significant at the 90% level of confidence are marked with a *; at the 95% level, a **; and at the 99% level, a ***. Standard errors in parentheses.

Table 7: SUMMARY OF FIRST STAGE REGRESSIONS CORRESPONDING TO TABLE 5

	(1)
	Black population growth
<i>Panel A: Population growth rates (1991-1996)</i>	
log distance	-0.940 (0.593)
<i>Panel B: Population growth rates (1991-2001)</i>	
log distance	-0.00675*** (0.00123)
<i>Panel C: Population growth rates (1991-2011)</i>	
log distance	-0.00354*** (0.000540)
Fixed effects	Yes
Controls	Yes
Observations	2093

Notes. This Table displays estimates of equation (1) in the main text. Each cell presents estimates from a separate regression. The standard errors are clustered on the municipality level. There are 201 clusters. All columns are estimated using OLS where the natural log of distance to the nearest homeland is the variable of interest. The outcome variable is absolute black population growth in the relevant time period divided by the overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. There are nine provinces for which fixed effects are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are statistically significant at the 90% level of confidence are marked with a *; at the 95% level, a **; and at the 99% level, a ***. t-statistics in parentheses.

Table 8: SUMMARY OF FIRST STAGE REGRESSIONS CORRESPONDING TO TABLE 6

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dummy for Johannesburg	Drop within 10 km	Drop < 5% white	Drop < 10% white	Drop distance ≥ 6	District fixed effects	Municipality level
<i>Panel A: Population growth rates (1991-1996)</i>							
log distance	-0.947 (0.597)	-0.805 (0.689)	0.0758 (0.152)	0.106 (0.189)	-1.004 (0.622)	-0.936 (0.723)	-0.948 (0.700)
<i>Panel B: Population growth rates (1991-2001)</i>							
log distance	-0.00697*** (0.00123)	-0.00625*** (0.00177)	-0.00777*** (0.00140)	-0.00747*** (0.00148)	-0.00693*** (0.00125)	-0.00647*** (0.00110)	-0.00602*** (0.00137)
<i>Panel C: Population growth rates (1991-2011)</i>							
log distance	-0.00367*** (0.000550)	-0.00395*** (0.000868)	-0.00345*** (0.000619)	-0.00359*** (0.000668)	-0.00345*** (0.000527)	-0.00381*** (0.000629)	-0.00316*** (0.0005)
Level of Fixed effects	Province	Province	Province	Province	Province	District	Province
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2093	1790	1374	1137	1730	2093	203

Notes. This Table displays estimates of equation (1) in the main text. Column headings denote different specification. Each cell presents estimates from a separate regression. The standard errors are clustered on the municipality level. There are 201 clusters. All columns are estimated using OLS where the natural log of distance to the nearest homeland is the variable of interest. The outcome variable is absolute black population growth in the relevant time period divided by the overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. There are nine provinces for which fixed effects are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are statistically significant at the 90% level of confidence are marked with a *, at the 95% level, a **, and at the 99% level, a ***. Standard errors in parentheses.

Table 9: FIRST STAGE REGRESSIONS OF SPECIFICATION WITH LINEAR INTERACTION TERM CORRESPONDING TO TABLE 8

	(1)	(2)	(3)	(4)	
	Δ Black pop growth	Δ Black pop growth \times high population density dummy	Δ Black pop growth	Δ Black pop growth \times high urban share dummy	
<i>Panel A: Population growth rates (1991-1996)</i>					
log distance	(0.727)	-1.250* (0.656)	-0.821 (0.846)	-1.426* (0.841)	-1.388
log distance \times high population density dummy	1.426 (1.130)	1.564 (1.253)	- -	- -	
log distance \times high urban share dummy	- -	- -	0.710 (0.736)	0.815 (0.707)	
<i>Panel B: Population growth rates (1991-2001)</i>					
log distance	-0.00715*** (0.00123)	0.00140*** (0.000345)	-0.00869*** (0.00139)	0.00386** (0.00168)	
log distance \times high population density dummy	0.00166* (0.000975)	-0.00365*** (0.000769)	- -	- -	
log distance \times high urban share dummy	- -	- -	0.00290*** (0.00109)	-0.00861*** (0.00193)	
<i>Panel C: Population growth rates (1991-2011)</i>					
log distance	-0.00367*** (0.000545)	0.000312 (0.000232)	-0.00445*** (0.000705)	0.00176* (0.000907)	
log distance \times high population density dummy	0.000662 (0.000685)	-0.00226*** (0.000524)	- -	- -	
log distance \times high urban share dummy	- -	- -	0.00131** (0.000600)	-0.00465*** (0.001000)	
Province fixed effects	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	
Observations	2093	2093	2093	2093	

Notes. This Table displays estimates of equation (1) in the main text with an additional interaction term. Column headings denote different specifications. The standard errors are clustered on the municipality level. There are 201 clusters. All columns are estimated using OLS where the natural log of distance to the nearest homeland and the same term times a dummy for high initial population density or high initial urban share of households are the variables of interest. The outcome variable are absolute black population growth divided by the overall population times a dummy for high initial population density or high initial share of urban households in 1991 and absolute black population growth divided by the overall population. The relevant time periods are 1991-1996 in Panel A, 1991-2001 in Panel B and 1991-2011 in Panel C. Controls include variables on education, income, population group, population density and employment in 1991. There are nine provinces for which fixed effects are included. The estimated coefficients for the first stage regressions are reported in the appendix. Coefficients that are statistically significant at the 90% level of confidence are marked with a *; at the 95% level, a **; and at the 99% level, a ***. Standard errors in parentheses.