Invisible Borders – Persisting Scars

How the Apartheid Homelands Define Contemporary South Africa

Peter Séan Richardson Courtney

Thesis presented in the partial fulfilment
of the requirements for the degree of
MSc Development Studies: Development Economics
at Wageningen University and Research

Supervisor: Asoc. Prof. dr.ir. Maarten Voors

August 2021

To my childhood home, KwaZulu-Natal. May the eponymous unity of KwaZulu and Natal one day reflect the reality of your lands.

ABSTRACT

In this thesis, I utilise a regression discontinuity design (RDD) to examine the long-run and persisting welfare reductions caused by the Apartheid homelands (1948-1994). The homelands were the only areas in Apartheid South Africa where Black African people could reside and own land. The contemporary geographic pattern of welfare reductions caused by the homelands is estimated by identifying a second-best counterfactual population through the RDD estimator. I present a novel improvement to naïve counterfactual identification in spatial RDDs. The results indicate that the homelands have caused a long-run and persisting reduction in education attainment (decreasing the school completion rate by 2.17%), and education inputs (increasing students per teacher by 7.66%), while also doubling the number of schools per square kilometre (controlling for population density). Further, the results show that the homelands have caused long-run population density to double, and erosive agricultural practices have reduced contemporary topsoil quality. In a historical analysis, I highlight the role of the following as likely causes of the contemporary pattern of spatial inequality: the limited size of the homelands, denaturalization, the migrant labour system, parent absenteeism, Apartheid rural policy (including 'influx control' and 'Betterment'), 'Native Law', Bantu education, and property dispossession. I compiled a novel homeland-specific geographic data set to conduct this research. The welfare of contemporary South Africans is significantly reduced by living just 5km on the wrong side of a now non-existent homeland border. Former homeland specific policy is required to contend with this injustice.

Invisible Borders – Persisting Scars

How the Apartheid Homelands Define Contemporary South Africa

Peter Séan Richardson Courtney

August 30, 2021

Word count: 15581

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DBE	Department of Basic Education
EMIS	Education Management Information Systems
OED	Oxford English Dictionary
OLS	Ordinary Least Squares
PoC	People of Colour
PTR	Pupil Teacher Ratio
RDD	Regression Discontinuity Design
SACHED	South African Committee of Higher Education
TA	Traditional Authority
TBVC	Transkei, Bophuthatswana, Venda, Ciskei
TBVCZ	Transkei, Bophuthatswana, Venda, Ciskei, KwaZulu
VOC	*Vereenigde Oost Indische Compagnie*

CHAPTER 1

INTRODUCTION

South Africans live with the consequences of Apartheid as a matter of daily life. The territorial separation of people along ethnic lines is the most socio-economically and politically enduring aspect of Apartheid. Territorial separation was accomplished primarily through the creation of the homelands, the central concern of this thesis. The homelands were the only areas in Apartheid South Africa where black people could legally reside and own land. The historical literature details the harms of the homelands. Yet, it is not sufficient to know that these deprivations once existed. This research seeks to estimate the contemporary magnitudes of deprivations caused by the homelands, more than two decades after the end of Apartheid.

Effective and targeted redress measures require an accurate description of the contemporary magnitude and location of the deprivations caused by Apartheid. With 29.5% of the country residing in the former homelands, in a very real sense the socioeconomic prosperity of the former homelands defines the prosperity of the nation as a whole.[1] As this research shows, living in the former homelands negatively influences livelihoods, from the quality of the soil to the chances of one's child completing high school.

In this thesis, I estimate the contemporary sizes of welfare reductions caused by the homelands more than two decades ago. The contemporary magnitudes of deprivations in three domains are quantified: agriculture (topsoil degradation), population density, and education—education the primary area of investigation. I conducted a literature review of the economic history of the homelands to identify appropriate welfare consequences of the homelands for this research, and the causes of the identified deprivations therein. The historical literature emphasises the deprivations identified as some of the worst and most likely to endure. I additionally provide a geographic/ spatial description of where these deprivations are worst. As such, this research provides a foundation for former homeland specific policy in contemporary South Africa.

This research is situated in the economic literature which estimates the socio-economic persistence of institutions after they cease to exist. Indeed, the estimator of this research replicates, with several important improvements, the RDD estimator of Melissa Dell (2010).

[1] This is particularly true under a liberal egalitarian or Rawlsian framework of justice (e.g. the South African Constitution), where the prosperity of the worst-off matters most. The residents of the former homelands are statistically some of the worst off in the country.

Here, Dell estimates whether the colonial Mita mining institution has lead to differences in household income and child stunting in contemporary Peru; finding that the Mita has negatively impacted both indicators. Much of this body of research specifically estimates whether colonial institutions have had a negative or positive impact on postcolonial livelihoods. Although Dell finds a negative impact of the Mita, she also finds a positive impact of the colonial Dutch Cultivation System on welfare indicators in contemporary Indonesia (Dell and Olken, 2020).

These findings align with the research showing that there are broadly two sorts of colonial institution, extractive and productive (or inclusive), with long term negative impacts found for the former and positive impacts for the latter (Robinson and Acemoglu, 2012). The homelands were not a colonial institution *per se*, nor do they fit either the productive or extractive label neatly. Nonetheless, the results align with exploitative institutions having persisting effects. For an introduction to the institutional persistence literature see: Acemoglu, Johnson, and Robinson (2001, 2002), Cagé and Rueda (2016), Donaldson (2018), Huillery (2009), Jedwab, Kerby, and Moradi (2017), and Valencia Caicedo (2019).

There are two primary requirements for the empirical approach of this research. The first is to only identify deprivations *caused* by the former homelands themselves. The second is to quantify the contemporary magnitudes of these deprivations in democratic South Africa. These requirements suggest the use of a quasi-experimental method to reduce the impacts of selection and endogeneity on causal identification. Consequently, I use a geographic regression discontinuity design (RDD) to quantify the causal impact of the former homelands on contemporary geographic patterns of topsoil quality, population density, and measures of education. The empirical strategy includes multiple bandwidths and specifications to transparently report specification sensitivity. The benchmark estimate is that of the shortest bandwidth (5 or 10km), with Thiessen polygon fixed effects, as described in Chapter 3.

Topsoil quality is proxied for by the content of nitrogen and organic carbon in the topsoil. Education outcomes are proxied for by an approximated school completion rate. Education inputs are proxied for by an approximated number of students per teacher, or classroom size. Finally, school accessibility is proxied for by the number of schools per square kilometre, controlling for population density.

The RDD estimator estimates that the homelands have caused a more than doubling of the population density within the former homelands in contemporary South Africa. The large magnitude of this finding is remarkable as democratic South Africa is characterised by significant rural urban migration, yet the rural homelands remain populous (Collinson and Kok, 2006). Further, it is found that the homelands have reduced the content of topsoil nitrogen in the contemporary former homelands by at least 2.19%. This effect size is substantial enough that at the margin there are likely many areas that are no longer arable due to this reduction in soil quality.

Surprisingly, it is found that the homelands have caused a more than double of the number of schools per square kilometre in the former homelands, after controlling for population density. However, for reasons explained, this finding does not directly imply greater access to schools. Turning to school inputs, I estimate that the homelands have caused classroom size, or students per teacher, to increase by 7.66% within the former homelands. This implies significantly less attention per student, caused by simply studying within a former homeland. Finally, it is estimated that the homelands have reduced the school completion rate by 3.6% in the former homelands.

The estimator of this research does not identify the causes of the persisting pattern of deprivation found. There are two aspects to the causes of this persisting deprivation. The first aspect is what initially caused the relative deprivation in the homelands. These historical causes are described in detail in Chapter 2. Here, the following are shown to be the most likely causes of the relative deprivation of the homelands: the limited size of the homelands, denaturalization, the migrant labour system, parent absenteeism, Apartheid rural policy (including 'influx control' and 'Betterment'), 'Native Law', Bantu education, and property dispossession. These were all directly caused by the Apartheid regime and localised to the homelands. The second aspect is why these deprivations have persisted. Here the persistence of Apartheid institutions into democracy, such as the 'traditional leadership' (for example the Ingonyama Trust) is likely a substantial factor. Yet, as this research motivates, it is primarily the lack of targeted support for these highly disadvantaged areas which has led to this persisting pattern of inequality.

That Apartheid caused substantial harms to black people is well analysed in the historical literature. The persistence of Apartheid deprivations is similarly highly studied. However, there are few studies which attempt to quantify these deprivations in a causal or quasi-experimental framework and none which test the homeland specific hypotheses identified above. This is the primary contribution to the literature. In the course of estimating these persisting deprivations, this research makes several subsidiary contributions to the literature. The first, a methodological contribution to the geographic RDD literature, is the use of Thiessen polygons to identify appropriate counterfactuals more precisely. The second is the novel spatial data set I compiled for this research. These data are used for both estimation and to produce maps of the relationships between the former homelands and contemporary patterns of population density, school density, classroom size, and school completion rates; this mapping is likewise novel in the literature.

The primary limitation of this work is that it is not possible to estimate the variables of interest in a counterfactual world in which Apartheid never happened. However, a second-best counterfactual is identified: the land immediately outside the former homelands. Here a trade-off is made between external validity (or generalisability to the entire homeland or country) in favour of internal validity of the estimator identifying only the homeland caused effects. This is likely to significantly downward bias the generalisable results as per Chapter 3 Section 3.3.

The econometric analysis is significantly limited by the sort of data utilised as the samples for both topsoil nitrogen content and population density were derived from raster maps created with imputed estimates. This is a limitation as the estimator does not account for the uncertainty in these estimates. As such, the standard errors for both the nitrogen and population estimates are not accurate. However, the primary education section uses school level observations, and thus standard errors remain robust in that section. Finally, robustness checks with standard errors clustered at the municipal and homeland levels are conducted and reported in the Appendix.

Chapter 2 provides an economic history of the homelands. This chapter justifies the selection of the deprivations analysed in this thesis and explores how the Apartheid homeland structures caused the deprivations estimated. Chapter 3 describes the method and data used to determine the persistence of the deprivations caused by the homelands in contemporary South Africa. The foci of this chapter are the various threats to validity and identification, and how these were addressed. This chapter further provides the novel contribution to regression discontinuity methodology. Chapter 4 is the results chapter. The results for the overpopulation and topsoil degradation hypotheses are provided and explored in Section 4.2 and 4.2 respectively. Section 4.4 provides the education results.

This thesis was written in the hope that this relatively understudied group of people, those living in the former homelands, can gain consideration for their particular plight. As urban bias continues to have a pervasive hold, the truly worst off—–the rural poor—–are often rendered non-existent in political, academic, and social justice discourses. That living on the wrong side of a now non-existent border can have such harmful effects on one's life is an injustice with which South Africa must contend.

CHAPTER 2

ECONOMIC HISTORY OF THE HOMELANDS

"Of all the manifestations of inequality and oppression under apartheid, none was as stark or potentially as enduring, as the territorial separation of people along racial lines" – Edward Patrick Lahiff (1997: 10)

2.1 INTRODUCTION

Grand Apartheid[1] in South Africa categorically transcended segregation. The division of South Africans by race was insufficient for the National Party[2] (NP). Grand Apartheid instead sought to strip black South Africans of their citizenship, replacing it with an ethnically and geographically defined nationality. This was achieved by the NP legally designating land for black occupation, known as the homelands or bantustans, the central concern of this thesis. The NP thus created a pattern of race, embedded in the geography of South Africa, mirroring access to opportunity, in what would come to be known as spatial Apartheid. This thesis seeks to determine the extent to which Apartheid's geographically determined scarcity and suffering has persisted 25 years into democracy. This chapter is an economic history of the homelands, necessary to contextualise the econometric analyses which follow.

The legacy of spatial Apartheid endures and may be the greatest challenge facing South Africa's young democracy. To this day, the question of land redistribution predominates South African politics (Kepe and R. Hall, 2018); the migrant labour system distorts family life and labour (Rogan, Lebani, and Nzimande, 2009); and traditional leaders control much of South Africa, impacting the livelihoods of millions (Mazibuko, 2014).

Removing the citizenship of black South Africans (denaturalization) was an ideological priority for the NP. Denaturalization reinforced the standing legal requirement for black people to carry passbooks, or domestic passports, in white South Africa. Both were part of the project of 'separate development', a cornerstone of Apartheid ideology. Its logical zenith was the creation of nationally independent homelands. With insidious rhetoric, the NP compared the creation of the homelands to decolonisation, granting the 'native' people sovereignty over their land (Geldenhuys, 1981: 24).

[1] Grand Apartheid encompasses the expansive restrictions on the political, land, and basic human rights of black people. Petty Apartheid refers to the segregation of facilities.

[2] The ethnic nationalist Apartheid government of South Africa from 1948 to 1994.

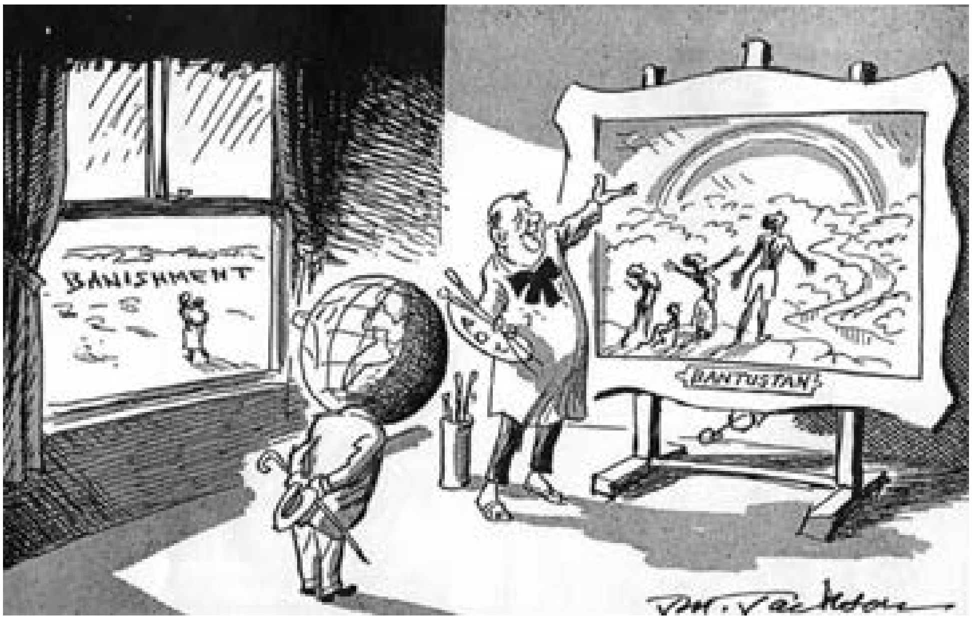

Figure 2.1: Cartoon by J.H. Jackson (1959)

Denaturalization implied the alienation of all positive rights[3] and claims against the NP government. As the NP had scant *de facto* respect for sovereignty (evidenced by its military incursions into bordering nations), the *de jure* sovereignty of the homelands implied no real negative rights against the NP either. Indeed, the four homelands which attained 'independence' (only recognised by the South African state), continued to existentially rely on the Apartheid state. Thus the homelands relied on the South African state for everything from budgetary finance[4] to external trade (most homelands were landlocked, none had ports), thus refuting any possibility of *de facto* sovereignty.

By 1991, 47% of the South African population formally resided within land designated as homelands (C. Cooper et al., 1994). Today 29.5% of South Africans reside within the former homelands.[5] Had all the homelands attained independence at their creation, the white population would have been a plurality in the remainder of South Africa. Indeed, the balkanization of South Africa along ethnic lines was intended foremost to make a minority out of each black African ethnic group, rather than a unitary black nationalist identity. A wave of Afrikaner nationalism had brought the NP to power. The NP was thus intimately familiar with the

[3]Positive rights oblige the state to act. For example, the right to health care obligates the government to provide public health care. Negative rights prohibit the government from action. For example, the right to free speech prohibits the government from prosecuting someone for criticising the state.

[4]Geldenhuys (1981): "Of the total homeland budgeted revenue of R 184 million in 1978/79, only R441 million consisted of revenue from own homeland sources, with a further R59 million being a balance brought forward from the previous year. This meant that South Africa provided an amount of R684 million, or some 58% of the total homeland revenue. At present, some 10% of South Africa's national budget is allocated to homelands."

[5]Author's calculation from Tatem (2015).

'dangers' of nationalist identity formation.

To understand the nature of life within the homelands, and how the deprivation of these areas has persisted, it is crucial to understand the historical forces which led to this "highest stage of separate development" (Geldenhuys, 1981). This chapter is primarily an economic history of the homelands. Yet, one must keep in mind that the economic (in)viability of the homelands bears only derivatively on the innate injustice of the homelands and the 'divide-and-rule' ethnic nationalism that brought the homelands into existence.

2.2 THE ORIGINS AND FUNCTIONS OF THE HOMELANDS

The origins of Apartheid and the homelands lie in the settler colonisation of South Africa. Foremost was the aim to geographically separate people by race and ethnicity. However, there was an opposing incentive to employ cheap (i.e. "non-white") labour. This led to the creation of internal passports for people of colour (PoC) labourers. These were necessary for PoC to access economically active white-only areas. Hence, the passing of pass laws provides the natural starting point to the eventual creation of the homelands.

The VOC [6] is said to have required their slaves to identify themselves with passes as early as 1709 (History, 2011). Yet, South African Union documents (U. o. South Africa, 1922) report that "the earliest reference to pass provisions in the Cape appears to be in the Proclamation of the Earl of Macartney, dated the 27th of June, 1797, which aimed at excluding all natives from colonial territory and directed farmers and others employing natives to discharge them".[7] The pass laws entailed the economically costly expulsion of PoC and increased ethnic homogeneity.

Over the next century, the Afrikaners (of mostly Dutch heritage) and the British would unevenly settle South Africa. Expansion brought with it the military domination of local people, culminating in the Anglo Zulu War of 1879. Defeated, yet retaining much of their socio-political and economic systems, black Southern Africans would continue to retain some of their ancestral lands, mostly in the eastern half of the country (Welsh, 1973: 29) (Thompson, 2008: 109).

The homelands would eventually lie in the eastern half of the country. This fact is in line with the NP's claim that the land chosen for the homelands was selected on the grounds that it was the ancestral lands of the various "native" ethnic groups (implicitly, the land not

[6] The Dutch East India Trading Company.

[7] Determining which of these opposing claims is true has not been trivial. Both are adequate academic sources. Yet, I cannot corroborate either claim. SA History does not directly cite their claim, but are otherwise generally considered reputable. Union reports are excellent historical documents, but there may have been political motive in ascribing the first pass law to a British Earl.

completely conquered or owned by white people). Later scholars and popular opinion would come to contest this in favour of a ‘marginal lands’ hypothesis: that the land was selected both for its poor agricultural productivity as well as its distance from productive centres in white South Africa (Houghton, 1987; Levin and Weiner, 1991; van Zyl and van Rooyen, 1991). The latter half of this hypothesis is certainly true. The average distance from each discrete area of land, which comprise the homelands, to the nearest city of the 20 most populated cities, is 186km.[8] This is the shortest geodesic, or ‘straight line’, distance. Yet, with typically poor road infrastructure, actual travel distances were significantly longer (Butler, Rotberg, and Adams, 1978).

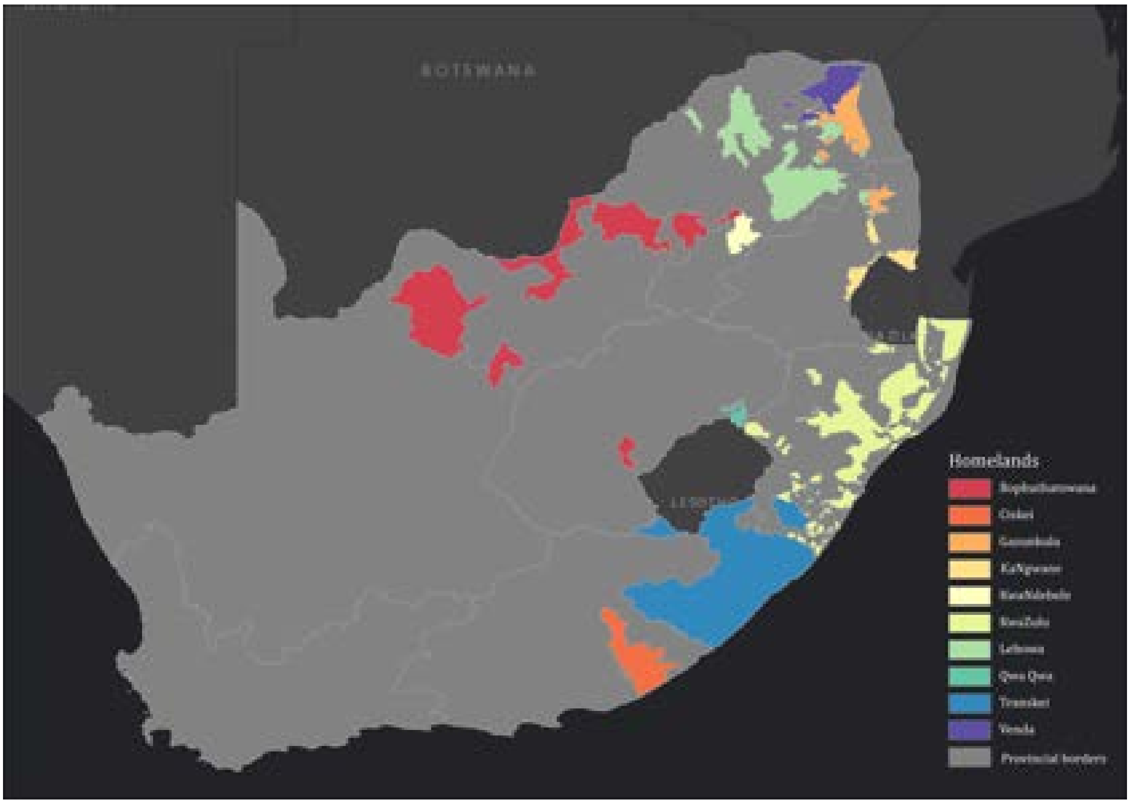

Figure 2.2: The 10 homelands, comprised of 78 non-contiguous units of land, covering 13,7% of South Africa’s land area. Source: Author

In all four settler polities,[9] the drive to expand capitalist production (both industrial and agricultural) required significantly more labour than was supplied by the precapitalist demand for employment. This led the colonial regimes to atrocious interventions to induce the population to part with their labour. An example is Natal’s implementation of a hut tax of 10 shillings per hut in 1857 (T. G. o. South Africa, 1908). Taxes such as the hut tax were often payable only in colonial currency: leading to increased labour in markets delineated by capital, establishment of the economic value and broad tender acceptance of the currency, and

[8]Author’s calculation.

[9]The settler polities were the colonial settler political units which were comprised of the British Colonies of the Cape and Natal and the Afrikaner Traansvaal Republic and Orange Free State.

the bureaucratic 'legibility' or enumeration of the local population.[10] These were all crucial aspects to the successful geographic partitioning of people along ethnic lines.

The precursor to the homelands were the native reserves, such as those created by Sir Shepstone's Native Reserves policy in Natal (N. Nattrass and J. Nattrass, 1990). The rationale for the reserves was explicitly economic: the generation of a surplus supply of labour (ibid.). As decreed by Earl Grey in 1849:

> *It would be difficult or impossible to assign to the natives such locations of an extent sufficient for their support... I regard it on the contrary as desirable that these people should be placed in circumstances in which they find regular industry necessary for their subsistence* (quoted in Van der Horst (1971)).

The commodification or proletarianization of black labour was perceived to require depriving the black population of the means to exist on precapitalist modes of production (primarily subsistence agriculture).[11] This was accomplished through the near total restriction of land rights, relegating the black population to the limited lands designated as native reserves and subsequently the homelands. Severely restricted property holdings combined with fertility rates above replacement, inevitably led to high levels of population, as detailed in Chapter 4.

The drive to induce indigenous labour participation reached its highest levels during the 'Mineral Revolution' (Worden, 2011). The Mineral Revolution began in the late 19^{th} century with the discovery of the largest gold reserves in the world in Witwatersrand (in the Afrikaner Transvaal Republic) and diamonds in Kimberley (in the British Cape Colony) (Norman, 2006). A near inexhaustible demand for labour in the mines produced a migrant labour system which pulled millions of migrant labourers from the rural periphery in what Crush, Jeeves, and Yudelman (1991: 2) describe as "one of the key distinguishing features of South African industrialisation".

South Africa's unique system of migrant labour was directly caused by the homelands due to their distance to productive centres in white South Africa, such as mines, forcing labourers to migrate for employment as they could not legally reside in white South Africa. The homelands

[10] Bureaucratic legibility refers to the identifiablity of individuals, and thus ultimately the taxability and control of a population. The extreme degree of control implied by hut and pol taxes led to many rebellions in Africa, notably the Bambatha Rebellion of 1906 (Stuart, 2013).

[11] How much welfare can be derived from precapitalist subsistence agriculture is an interesting but separate question. No matter how little can be derived from household production for household consumption, the explicit ends of reducing self sufficiency is sure to have been at least in part effective.

thus became ‘reserves of migrant labour’, redoubling the pre-industrial racial order with a highly oppressive system of male-only mining compounds, typically in a state of violence and privation (Vosloo, 2020).

The relative prosperity of the mineral revolution heightened demand for agricultural product. This led to a general acceptance by the white population that restricting black land rights was required to both ensure the flow of cheap labour to the mines as well as to restrict ‘unfair competition’ to the emerging class of white commercial farmer (N. Nattrass and J. Nattrass, 1990). As Colin Bundy (1979: 115) puts it:

> *“Both the farmer and the mine-owner perceived in the late nineteenth century the need to apply extra-economic pressure to the African peasantry; to break down the peasant’s ‘independence’, increase his wants, and to induce him to part more abundantly with his labour, but at no increased price.”*

However, as per Wolpe (1972), the labour-inducing oppression of black people (primarily accomplished through the restriction of land rights) was limited by a countervailing end of the NP: wage subsidisation. At the household level, agricultural production in the reserves subsidised the wages of the migrant labourer while the extended family provided welfare services that would otherwise be costly, such as housing and childcare. Thus, “African redistributive economies” allowed capitalists to remunerate their migrant workers below the real cost of “reproduction” (Lahiff, 1997: 12).

The household was trapped by the insufficiency of both the penurious wage of migrant household members and the artificially limited agricultural production of the household, thus forced into participating in both. The economic incentive to reduce wages acted to prevent the total restriction of rights in what Wolpe would characterise as a defining feature of segregation and a primary difference with Apartheid, a distinction shortly explained.

The migrant labour system, designed to entrench economic servitude, has had long-lasting social consequences. Although this chapter is an economic history, it would be remiss not to mention the broader social consequences of the migrant labour system. By 1990, up to 80% of men between 25 and 50 were absent from any given homeland (N. Nattrass and J. Nattrass, 1990: 521). They were absent in order to migrate to their places of employment in white South Africa as they could not legally reside in white South Africa. Absenteeism combined with extreme levels of violence in the mining compounds led to a breakdown of social order in the homelands. As per N. Nattrass and J. Nattrass (1990: 521):

"High rates of outmigration reduce the domestic labour supply, increase the burden on the women remaining behind, upset social relations, and hamper production and investment decisions as these are usually tightly controlled by men". "

Consequently, the migrant labour system has been identified as causing many of South Africa's deepest social ills, such as the staggering number of children raised by one or fewer parents (65.6%) (K. Hall and Sambu, 2019) and the extremely high rate of gender-based violence (Elder, 2003). These harms were entrenched by the self-perpetuating nature of the system, as migration undermined the homeland economy leading to a further reliance on migration and "the creation of a cycle of dependency and underdevelopment" (N. Nattrass and J. Nattrass, 1990)."

The first law governing the reserves after the Union of South Africa (1910)[12] was the *1913 Native Land Act*. This Act formalised the reserves (not yet homelands), designating only 7% of the country as reserves, while simultaneously proscribing the sale of all land to black people outside of the reserves. The law additionally abolished sharecropping and other farm tenancy unless the labourer worked a minimum of 90 days of compulsory labour a year.

This pushed the equilibrium further away from the wage subsidising functions of the reserves. Colin Bundy (1979: 213) believed this was done with "the intent to inhibit the process of class differentiation within the reserves and prevent the emergence of either a class of black commercial farmers or a landless proletariat, each of which posed its own threat to the system of racial segregation and migrant labour". Liberal commentators often perceived these laws to be economically irrational, even from the perspective of the NP (Wellings and Black, 1986). Yet, an economic system based on the exploitation of the masses has as its first principle the preservation of the socio-political order.[13]

The first laws to formally create political, judicial, and administrative structures in the reserves were the *1920 Native Affairs Act* and the *1927 Native Administration Act*. These laws established the legal standing of a separate and subsidiary legal system, Native Law, such that "The Minister [of Native Administration] may authorize any native chief or headman recognized or appointed [by the Governor General] to hear and determine civil claims arising out of native law" (Section 12 1a).

Thereafter, chieftaincy gained legal recognition under the NP. Yet, "one should not be mis-

[12]Which united the Afrikaner and British settler polities through the *South Africa Act of 1909*.

[13]For a somewhat revisionist take on the homelands, see the first chapter of Ally and Lissoni (2017).

led by the nomenclature [of chieftaincy] into thinking of this as a holdover from the precolonial era" Mamdani (1996: 23). Native law was a crucial step towards the creation of independent homelands, accompanied by the executive consolidating control of important aspects of 'native administration', once under the authority of parliament. It also created a system of governance in near perfect opposition to the tenets of contemporary institutional economics, as per Mamdani (1996: 23):

> *The authority of the chief thus fused in a single person all moments of power: judicial, legislative, executive, and administrative. This authority was like a clenched fist, necessary because the chief stood at the intersection of the market economy and the nonmarket one. The administrative justice and the administrative coercion that were the sum and substance of his authority lay behind a regime of extra-economic coercion, a regime that breathed life into a whole range of compulsions: forced labour, forced crops, forced sales, forced contributions, and forced removals.*

Poor governance structures have a well-studied direct bearing on the economic prosperity of people (Chong and Calderon, 2000). The legacy of Native Law persists under the democratic Constitution of South Africa, Chapter 12, enshrined as Customary Law. Indeed, these institutions have recently been legislatively strengthened by the *Traditional and Khoi-San Leadership Act 3 of 2019*. Thus, when determining the persistence of the effects of Apartheid socio-political structures on the welfare of the current residents of the former homelands, as this thesis attempts, one must not envisage post-Apartheid politics and governance *de novo*. Much of the persistence found reflects the persistence of the institutions themselves, rather than the lag between the dismantling of pernicious institutions and economic liberation.

Further, a commonly accepted foundation of economic liberty, property rights, is greatly circumscribed by many contemporary traditional leaders and Chapter 12 institutions, such as the Ingonyama Trust (see the Appendix, Figure 6.1 for a map of the land currently held by the Ingonyama Trust). Under the Ingonyama Trust (the Zulu monarchy), land is held in common and administered by the trust, in what has been described as a neo-feudalist system of landholding (Mazibuko, 2014). With the Ingonyama Trust continuing to hold roughly 30% of KwaZulu-Natal and almost all of the former KwaZulu homeland, it is likely that the effects of the Trust will significantly influence the estimates of the persistence of the effects of Apartheid in former KwaZulu. This is perhaps why the pattern of low quality schooling is most clear in contemporary KwaZulu-Natal. Quoting from Basic Education (2005):

Experience from a village education project in the Maputaland area of north-eastern KwaZulu-Natal shows the political difficulties created when traditional power relations are disturbed. The education project began formally in 1989 although its roots lay in a process of community development stretching back to 1978. The project was managed by a democratically oriented development committee as part of a broad donor funded community development programme. The project was able to leverage expertise from universities and NGOs into various aspects of the project. The education programme included:

- *A resources centre with books, videos, magazines and newspapers*
- *Four full-time personnel*
- *A school support programme An 'out-of-school' matriculation programme (in partnership with SACHED)*
- *A literacy programme*
- *A recreational (films, discos) and sports programme.*

The education programme worked closely with work and skills development projects in agriculture, aquaculture, horticulture, healthcare and social welfare, the development of village infrastructure and skills training and production units. Over a five year period, this integrated community development programme developed a strong support base but was unable to win the support of the traditional authority structures. This lead to the closure of the whole development project, including the education programme. The key issue to emerge from this example is that of governance. Who own and controls the project? i.e. The TA or the development committee?

The *Native Trust and Land Act of 1936* increased the land in the reserves to 13.7% of South Africa, where it would remain, with minor changes, until the end of Apartheid. The act was the last major legislation governing the reserves before the beginning of Apartheid. The act established the Native Trust (later the Development Trust) which was responsible for acquiring the land for the expansion of the reserves.

The Trust's land acquisition was exceptionally arbitrary, often involving drawing a border around "Black spots"—areas populated by black people in white South Africa—without any contiguity to the land already designated as homelands. Two examples are the incorporation of the KwaMashu and Umlazi townships into the KwaZulu homeland. This led to the extreme fragmentation of the homelands, with KwaZulu being comprised of 42 isolated fragments (Figure 2.4).

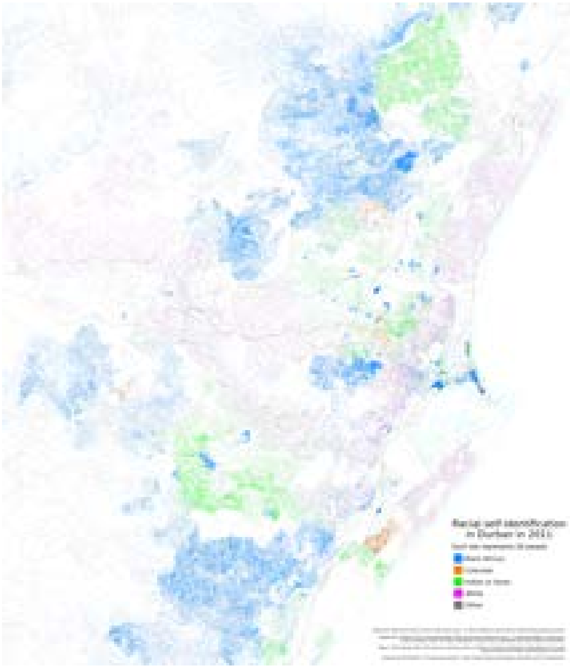

Figure 2.3: The persisting pattern of race in Durban, KwaZulu-Natal. (Firth, 2013)

The colours of the dots represent the self identified races of the Census 2011 takers. The "Black spot" of KwaMashu is the blue (designating "Black African") area beneath the green area (Indian, Phoenix) at the top of the map. Umlazi is the blue area at the very bottom of the map.

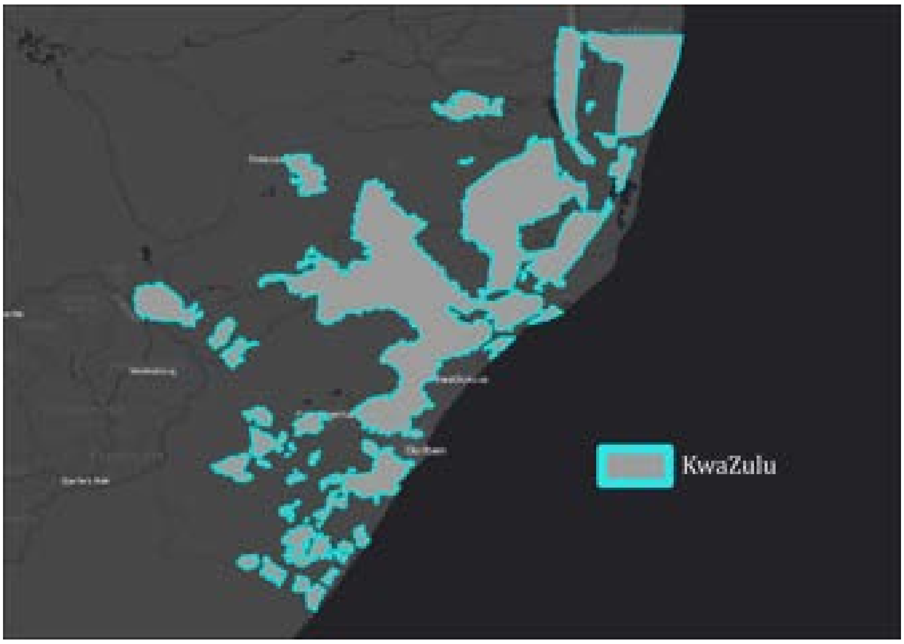

Figure 2.4: The 42 fragments of the former KwaZulu homeland. Source: Author from Malinda (2015).

The Native Trust was responsible for the development of the already overcrowded and poverty-stricken reserves and preventing what was perceived as an imminent ecological crisis induced by erosion (Letsoalo and Rogerson, 1982). This fear led to the programme of 'Betterment', which aimed to prevent soil erosion through, inter alia, controlling the number of cattle in the reserves. Betterment was exceptionally hated, leading to outbreaks of violent opposition, such as in Sekhukhuneland (1958) and Pondoland (1960) (Mbeki, 1964: 111).

The NP claimed that culling cattle was necessary to improve milk yield, the genetic blood stock, and pasturage, further claiming culled stock fetched a fair price when sold at auction (Beinart and C. Bundy, 1980: 300). Yet, for the rightful owners of the cattle, these arguments fell on deaf ears with the perception that the auctions "provide[d] a captive market for speculators and (white) farmers" where prices are "often determined at an artificially low level" (Yawitch, 1982: 12). Yet, it is possible that this brutal regime was effective in reducing erosion and thus the quality of the soil. I test this hypothesis in Chapter 4 Section 4.2, where I find it is likely that the homelands have reduced topsoil fertility.

2.3 THE BEGINNING OF APARTHEID, RESERVES BECOME HOMELANDS

The election of the NP in 1948, the beginning of Apartheid, marks the beginning of the shift of the reserves into self-governing homelands or bantustans. Yet, the word bantustan was coined earlier. The first usage I can find is from the South African Institute of Race Relation's Fourteenth annual report, 1942: "Some speakers have referred to it [the reserves] as "Bantustan" — but it is to be compared, not with Pakistan, but with Utopia or with Plato's republic" (SAIRR, 1942). The word is a portmanteau of 'Bantu', the large linguistic group, and '-stan' the suffix for land in the Persian group of languages.

At first, 'Bantustan' was used by the NP before it was co-opted as a term of disparagement. Indeed, even the first usage above uses the word disparagingly, a practice adopted by liberationists, such as Biko (1978).[14] Soon the NP used the term 'Homeland' nearly exclusively. I use the word 'homeland' throughout this thesis not in support of the NP's usage, but because the residents of the homelands preferred the homelands to be called as such (Lahiff, 1997: 9).

In the progression from segregation to Apartheid and Separate Development, Wolpe identifies not merely greater intensity in the project of segregation and wage subsidisation, but an

[14] I follow Biko in not capitalising 'black' and 'white' for the adjectives describing people. The NP capitalised the terms, using them as proper nouns, to bolster their essentialist notion of race. I do not capitalise bantustan or homeland either, as per Biko.

entirely new paradigm of oppression:

> *"The practice and policy of Separate Development must be seen as the attempt to retain, in a modified form, the structure of the "traditional" societies, not, as in the past* [under segregation]*, for the purposes of ensuring an economic supplement to the wages of the migrant labour force, but for the purposes of reproducing and exercising control over a cheap African* ***industrial*** [emphasis added] *labour force in or near the 'homelands', not by means of preserving the pre-capitalist mode of production but by the political, social, economic and ideological enforcement of low levels of subsistence... under circumstances in which the conditions of reproduction (the redistributive African economy in the reserves) of that labour force is rapidly disintegrating"* (Wolpe, 1972: 450).

World War II induced a period of industrial expansion, drawing black workers to white-only cities (Levin and Weiner, 1991: 88). Black industrial employment, which the NP associated with a rise in the militancy of black workers, occurred alongside the collapse in subsistence agriculture (primarily due to the pre-Apartheid land laws detailed above).[15] The equilibrium between wage subsidisation and labour inducement was thus broken, leading the NP to double down on 'influx control' and the 'three rural pillars of Apartheid' (communal land tenure, tribal administration, and betterment (Hendricks et al., 1990)) in an attempt to maintain control over the black population. Failure to maintain subsistence outside of the tightly controlled migrant labour system led to ever more brutal forms of NP oppression. This characterises a fundamental shift, from segregation to Apartheid, according to Wolpe.

Between 1950 and 1980, 3.5 million people were forcibly removed from their homes, most "repatriated" to the homeland which legally corresponded to their ethnicity (Platzky and Walker, 1985). In this time, the Apartheid regime passed a slew of repressive legislation on the path to creating nationally independent homelands.

The *Bantu Authorities Act of 1951* expanded the powers of the chieftaincy and created "territorial authority status", first received by Transkei in 1957, the highest from of authority provided for in the Act (Geldenhuys, 1981: 5). The *Promotion of Bantu Self Government Act of 1959* created a tiered system of Tribal, Regional, and Territorial authorities and was

[15] Partly corroborated by estimates placing the real value of subsistence agriculture in the 1980's at less than 10% of total, with many homelands producing less than 1% of their income from subsistence agriculture (N. Nattrass and J. Nattrass, 1990: 520).

the first to legislatively tie ethnicity with specific homelands, creating eight "separate national units" (Hill, 1964: 15). The *Transkei Constitution Act of 1963* granted self-government to the Transkei. Transkei gained nominal independence in 1976, followed by Bophuthatswana in 1977, Venda in 1979, and Ciskei in 1981, the so-called TBVC states. Independence was 'nominal' because only the South African Republic recognised their independence.

Figure 2.5: The flags of the homelands (Source: Unknown)

The homelands continued to substantially subsidise the real cost of the migrant labour system, both before and after nominal independence, although no longer primarily through precapitalist modes of production. The system both artificially reduced revenue collection while burdening the homelands with most public costs. Migrants typically only remitted between one fifth and one quarter of their income to their families in the homelands (J. Nattrass, 1976). Yet, this income comprised between 45 and 60 percent of the total product (or GNP) of the homelands (Southern Africa, 1987).

Consequently, the vast majority of the earnings of the legal residents of the homelands were taxed in the Republic where it was spent. Low real taxation combined with public expenditures still falling to the homeland governments (such as schooling and retirement) demonstrates the parasitism of the migrant labour system on the homelands. Thus, by 1987, post nominal independence of the TBVC homelands, transfers from the Republic comprised more than 50% of the homeland budgets (Geldenhuys, 1981). This ensured the homelands were existentially dependent on the NP, guaranteeing a high degree of control.

2.4 CONCLUSION

It is thus evident that Apartheid had profound effects on the lands designated as homelands.As this thesis shows, these harms have persisted into democracy. How democratic South Africa contends with spatial injustice is dependent on a firm knowledge of exactly how persistent the patterning of subjugation and impoverishment has been. While acknowledging the historicity of current oppression is vital, it does not follow that a reversal of the programmes that led to this oppression will undo anything. Yet, backward looking considerations are important for (non-consequentialist) justice and locating and understanding the forces which continue to impinge on the prosperity of South Africa's people.

CHAPTER 3

METHOD AND DATA

The most profound challenges to South Africa's development and democracy can be found in its rural hinterlands. These areas, systematically and intentionally deprived of the most basic resources under apartheid, continue to lag behind the rest of the country in the post-apartheid era. Nelson Mandela quoted in Nelson Mandela Foundation (2005).

3.1 INTRODUCTION

This chapter describes the method and data I used in this research. The chapter begins with the identification framework I used to determine the causal impact of the creation of the homelands on contemporary socioeconomic indicators in South Africa. Thereafter, I identify various threats to identification and validity. The chapter further describes how the estimation strategy accounts for each of the threats to validity identified. Further, I note where these threats pose a limitation to the estimation strategy. Section 3.5 provides a description of a novel solution implemented to improve the comparability of counterfactual observations. Finally, the chapter ends with a description of how I compiled the novel geographic data set used in this thesis.

3.2 IDENTIFICATION FRAMEWORK

Selection and endogeneity are two important threats to causal identification. These threats have led to a shift in inferential methods in the social sciences towards quasi-experimental methods, as with RDDs. Quasi-experimental methods replace true experimentation when control and treatment groups cannot be randomly assigned under the administration of the researcher. The aim of both natural and quasi-experimental methods is to identify a counterfactual or control population that is likely to be as identical as possible to the treatment population, except for not being treated. Here, the best available counterfactual is the areas immediately outside the homelands. The effective treatment population is the current residents of the former homelands. For the case of topsoil nitrogen content, the treatment population is all topsoil within the former homelands.

This research, and indeed any research, cannot determine a perfectly sound counterfactual of what would have occurred had Apartheid or the homelands not existed at all. To illustrate this: a family dispossessed of their townhouse in Johannesburg and moved to Transkei, and a family living on a Cape farm and moved to Transkei, would have had very different lives had Apartheid never occurred. Yet, it would clearly be infeasible to create an average counterfactual that accounts for these divergent individual counterfactuals. Nonetheless, this chapter motivates why the counterfactual population identified is an ideal second-best.

I estimate the causal effect of the establishment of the homelands, the treatment, on various dependent variables. A naïve approach might be to compare the homelands with the rest of the country, or the nearest province. However, this approach would likely suffer from selection bias, a form of endogeneity. It is possible that the lands selected to be homelands were chosen for a characteristic that is either a dependent variable, or a variable which correlates with a dependent variable. Consequently, finding a difference between the homeland and the rest of the country could simply reflect where the homelands were chosen to lie, rather than the effects of the sociopolitical institutions that were the homelands.

This is a specific instance of an identification failure due to endogeneity. In essence, the dependent variable has a causal influence on the independent variable of interest. As the independent variable of interest is a dummy of whether the observation lies in the homeland or not, the relevant from of endogeneity in geographic RDDs is selection, i.e. a factor correlated with where the observation lies. Therein, as per the example above, if under the naïve approach one found that the homelands had higher nitrogen levels, and the homelands were chosen to lie in places with high rainfall, and rainfall is positively correlated with topsoil nitrogen levels, then one could be detecting the effects of the rainfall on where the homelands lie, rather than the effects of the homelands on nitrogen levels.

The RDD solution is to compare observations which lie close to the treatment cut-off, i.e., as close as possible to either side of the homeland border, and determine whether there is a discontinuity in the dependant variable at the border. This works in the geographic context as distance is a continuous variable and, as per Tobler (1970), the "first law of geography is everything is related to everything else, but near things are more related than distant things". Thus, finding sudden discontinuities that lie at the border of the homeland suggests that it is the border itself which is creating these discontinuities.

Continuing with the rainfall & nitrogen correlation example, suppose that the homelands *were* selected for higher rainfall. Nevertheless, immediately on either side of the homeland border, it is very unlikely that rainfall will change discontinuously (i.e., fall harder immediately to one side of the border) even if on average, over the entire province, there is a difference in rainfall. Consequently, if it is found that nitrogen changes discontinuously at the border, that effect is not being driven by rainfall's correlation with nitrogen.

As such, the closer to the border the samples are taken, the lower the chance of selection effects reducing causal identification. Therein, continuing with the nitrogen example, one might sample the nitrogen levels only one metre on either side of the border. A narrow bandwidth reduces the influence of non-comparable observations distant from the homeland, e.g. the distant lower average rainfall. Yet, it is now unlikely that one's sample size would be large enough to discover statistically significant discontinuities along the border, the issue of statistical power.

Yet, if one sampled hundreds of kilometres on either side of the border, it is possible that the selection for higher rainfall will be detected rather than the true changes in nitrogen, the issue of covariate discontinuity. The sampling distance from the border is known as the bandwidth. Bandwidth selection must thus balance the issue of power with the issue of covariate discontinuity. This paper reports three to four bandwidths for robustness: 50km, 25km, 10km, and 5km. The 10km bandwidth is taken as the benchmark specification as it is typically significant and has the lowest chance of bias from broader discontinuities. In Section 4.4, 5km results are taken as the benchmark.

There are two temporal components of the treatment. The first is the direct harms that occurred in these areas during Apartheid and which have persisted. For topsoil nitrogen, the direct harms would include overgrazing that occurred during Apartheid, as contained to the homelands. The second are events which occurred after the end of Apartheid, yet are geographically contained to the lands of the former homelands. These post Apartheid effects can be due to anything from the persistence of correlated covariates such as population, to the continuation of geographically defined political structures—such as 30% of KwaZulu-Natal remaining under the control of the Zulu traditional leadership (see Figure 6.1).

3.3 VALIDITY

The two internal validity tests for RDDs are continuity of covariates and density tests. The density test tests whether there is selection into or out of the treatment group determined by the eligibility threshold (the homeland borders), a form of endogeneity. For the nitrogen data, it is not possible for there to be a greater density of observations at or on either side of the eligibility threshold as the raster data cover the country uniformly. For the remaining dependent variables, I utilise the `rdrobust` Stata function for the density test. No dependent variable failed the density test.

The next internal validity requirement is covariate continuity. Covariate continuity requires that the covariates which correlate with the dependent variable do not change discontinuously at the threshold. Keeping with the rain example, this requires that rain is not discontinuous at the border of the homeland for reasons already explained. Fortunately, rainfall is continuous at the 10km and 25km levels, yet not at the 50km level, as per Figure 3.1. The discontinuity at 50km is to be expected as features start to differ the further apart they are, as per Tobler's first law of geography. Unfortunately, the slope and elevation covariates are not continuous through the border at the 10km and 25km bandwidths. Although these discontinuities are a significant limitation, it is also an accurate reflection of the placement of the homelands' borders as some follow geographic features, such as contour lines and rivers. This further explains why rainfall is discontinuous at the 50km bandwidth as topology (mountains) and geographic features, such as rivers, influence rainfall.

Table 3.1: **Variable Continuity Bandwidth Tests**

Bandwidth:	50km		25km		10km	
Variable	Outside	Diff Inside	Outside	Diff Inside	Outside	Diff Inside
rain (mm)	95.69***	3.551***	98.94***	1.208	102.0***	-0.885
	(0.498)	(0.870)	(0.596)	(0.945)	(0.822)	(1.183)
slope (degrees)	3.023***	1.108***	3.252***	0.851***	3.502***	0.556***
	(0.0522)	(0.0913)	(0.0662)	(0.105)	(0.0974)	(0.140)
elevation (m)	1,116***	-203.5***	1,048***	-135.6***	971.3***	-63.10***
	(7.909)	(13.82)	(9.927)	(15.74)	(14.63)	(21.06)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Outside is the average of the variable outside the homeland at the given bandwidth.
Diff Inside is the difference in the variable within the homeland from the variable outside the homeland (at the given bandwidth).
If the difference is insignificant (rain at 25km and 10km), this shows the variable is likely continuous and vice versa.

Fortunately, the discontinuity of slope and elevation at the homeland borders is not a critical threat to identification as the estimator includes covariates for slope and elevation, effectively controlling for the discontinuities found. Of course, it is possible that there are unobserved discontinuities of variables correlated with the dependent variables which could invalidate the results. This is a significant limitation to any RDD.

All regressions include a population density covariate. As the population density variable is itself dependent on the homelands, the estimates are typically downward biased (in absolute terms). The population covariate is included nevertheless in case there are unobserved variables which have influenced the patterning of population, as discussed in Chapter 4 Section 4.2.

Moreover, there is a mechanistic trade-off between external and internal validity in the RDD estimator. This is so as neither the country or former homelands is sampled randomly to derive the estimates, only the bandwidth distance on either side of the former homeland borders is sampled. As such, the sample is not representative of either the entire country or the entire former homelands. However, this is necessary under the identification framework to only identify the causal aspect of the the creation of the homelands. Nonetheless, the border proximate sampling is likely to downward bias the results in absolute terms as it is typically found that within the former homelands, the closer to the border the better things are and outside the homeland, the further from the homeland, the better things are (see Figure 4.7).

Spillovers are the last consideration. A useful example here is population density, where I find that the former homelands are about 105% more densely populated than the lands immediately surrounding them. It may appear that there are two possible spillovers. The negative spillover might be thought to occur through the movement of people out of the homeland (for example in search of less densely populated land) which would seem to downward bias the estimate. Or inversely, people entering the more densely populated homelands (for example in search of employment), which would seem to upward bias the estimate. Yet, neither of these are true spillovers under the current framework. That is because the estimator is not estimating the effects of the homelands during Apartheid, but rather the *persistence* of these effects into democracy. Thus, should people leave the former homelands after 1994, these people are rightly not included in determining the persistence of the effects of Apartheid.

The real concern with spillovers is if the homeland itself has an influence on the lands immediately surrounding the homeland, or vice versa. An example might be people choosing not to live near a homeland during Apartheid due to the stigmatisation of these areas. This is a true spillover as it reduces the validity of the counterfactual—what would have occurred in the surrounding areas had the homelands never existed. Although spillovers are a potential source of bias, it is a much narrower case than what might be thought under the previous paragraph.

3.4 STANDARD ERRORS

The estimator crucially accounts for spatial autocorrelation. Spatial autocorrelation occurs as near things are systematically more related than distant things, as per Tobler's first law of geography. Spatial autocorrelation is thus the spatial corollary of temporal autocorrelation. The primary mathematical difference is that time is unidimensional, while space has, at least for our purposes, two dimensions of extension. Consequently, if the errors are treated as independent and identically distributed, the significance of the results will tend to be overestimated.

Following Conley (1999), I control for spatial autocorrelation using a method employing generalised spatial two stage seast squares estimation (a general method of moments estimator). There are broadly two options when accounting for proximity. The first option creates a matrix with measures of contiguity, i.e., proximity is determined by whether two observations share an edge. The second employs a matrix of the inverse distance between the centroids of the observations (such that the closer the observation, the greater its weighting).

Both methods effectively 'cluster' nearer observations' errors. The inverse distance method likely reflects the underlying data generating process more accurately as factors such as rainfall and soil quality are continuous without discrete borders. Further, the coefficient estimates employing the contiguity method were significantly different from OLS estimates whereas the inverse distance estimates were not. Consequently, the inverse distance measure was used throughout.

The Stata `spregress` function was used to control for autocorrelation, as per the above. For the `spregress` function to work, each observation must have a unique location. However, in the school data, there are occasionally multiple schools on one site all assigned a single GPS location. As such, spatial autocorrelation was controlled for here through the inclusion of longitude and latitude variables, their polynomials, and their interactions, as per Dell (2010). These covariates effectively create a spatial weighting matrix, as per the `spregress` function.

In the context of sampling design, clustering is required when one has sampled from a population using geographically prescribed or clustered sampling (Abadie et al., 2017). As sampling is done on specific geographic areas, it is likely that clustering will allow for more accurate standard errors. The `spregress` function, used to control for spatial autocorrelation, effectively clusters all errors by the proximity of each observation to all other observations. As such, the function does not take clustering as an option. Nevertheless, for each regression two robustness checks are run with heteroscedastic robust, clustered, OLS regressions, reported in the Appendix.

Following Abadie et al. (2017), the primary robustness check is clustering at the homeland level, the most applicable region of clustered sampling. To do so, for each non-contiguous unit of homeland, a point is placed in the centroid of the unit, used as a seed for Thiessen polygons (the nature of which is explored in Section 3.5). These polygons become the areas under which standard errors are clustered. The second robustness check clusters at the municipal level. This contemporary administrative level might have an unobserved correlation of observations determined by the political structure of municipalities, with locations correlated with the sampling undertaken. For both robustness checks, statistical significance typically remains.

The greatest remaining threat to accurate standard errors is induced by the sort of data employed for the population density and topsoil nitrogen dependent variables. These data are in raster format, i.e., a map of pixels with each pixel an observation or estimate. The population data do not contain standard errors for the estimates. The nitrogen data set does report a confidence interval. However, it is beyond the scope of this thesis to account for these certainty estimates.

Nonetheless, these raster datasets are sampled with a sampling matrix and not directly translated into a set of observations. The sample size is thus determined by the size of the sampling grid, which in turn influences the standard errors. Topsoil nitrogen content is sampled at a rate of 160 pixels or estimates to one observation in the final regression. Of course, this is not a formal control and as such standard errors remain imprecise/ erroneous. All regressions are heteroscedasticity robust.

In order to be transparent about multiple hypothesis testing (and thus the true family-wise error rate), where regressions were run but not included due to insignificance, this will be noted. Overall, I began this project by following a non-parametric estimation strategy testing imputed welfare data from the Demographic and Health Surveys that was not fruitful due to the noisiness of the data. After this strategy was dropped, I have reported all parametric regressions run.

3.5 ONE DIMENSIONAL BANDWIDTH IN A TWO DIMENSIONAL WORLD

In this section, I present a novel improvement to naïve counterfactual identification in spatial RDDs. In an RDD, treatment is a binary function of a known covariate, known as the running variable, effectively the bandwidth of this paper's estimator. In geographic RDDs, the running variable is a function of the distance to the border, such that observations equidistant to the border (or within the same bandwidth), but on opposite sides of the border, are treated as relevant counterfactuals. However, as distance to the border is one dimensional, and geographic space is two dimensional, the distance of two observations to one another, and thus their counterfactual relevance to each other, is not accurately reflected by the bandwidth alone. The estimator must account for the two dimensionality of space, as distance is a measure of relative comparability, as per Tobler (1970)'s first law of geography. The problem is illustrated in Figure 3.1.

Figure 3.1: The single running variable problem

In Figure 3.1, the gird represents the sampling grid ($10km^2$) for the topsoil nitrogen and population density dependent variables, each square is an observation. The polygons atop the grid are two different homelands. Suppose A to E lie in the centre of an observation (from which the distance to the homeland border[1] is measured) and all points are equidistant to the nearest

[1]The homeland borders which overlap with the national border are excluded as there are no measured

homeland border. Under a 'naïve' geographic RDD, each observation outside the border counts as a relevant and equal counterfactual to each observation within the border, so long as they are equidistant to the border. However, it is quite clear that B is a more relevant counterfactual to C than D is to C, as B and C are closer to one another than C is to D (as per Tobler's first law of geography). Likewise, both B and D are more relevant counterfactuals to C than E is to C. Yet, even though E is outside a different homeland, it is equidistant to a homeland border as C. Thus, under the naïve RDD E counts as a counterfactual to C. This is particularly a problem as some homelands are very far from one another.

This research's estimator replicates, with improvement, the RDD estimator used by Melissa Dell (2010). Here, Dell ameliorates this dimensionality problem by segmenting the treatment border and running fixed effects within each segment. Segmentation reduces the total distance between counterfactuals (not the perpendicular distance to the border) as each observation shares a regression intercept with only the observations within the segment in which it lies. A possible segmentation strategy is represented by the black lines in Figure 3.1. As such, C is no longer compared to D as it lies in a different segment, while still being compared to B as it lies in the same segment.

Yet, in Dell (2010), there is only one border, which is not enclosed. Enclosed borders create a problem when there is a homeland within the bandwidth of another homeland. This can be seen as the issue of not double counting (e.g. point B by assigning it as a counterfactual to both homelands in Figure 3.1). Or as an assignment issue: to which homeland is a point in the bandwidths of two homelands to be assigned as the counterfactual? There is also the problem of over-extension, such as E remaining a counterfactual to C if the segment extends indefinitely.

The novel solution I used is to add a point every two hundred kilometres along the homeland borders and then form Thiessen polygons from those points, known as seeds, creating a Voronoi diagram (see Figure 3.2). As per the definition of a Thiessen polygon, all points within the polygon lie closer to the seed which created the polygon than to any other seed. As such, the distance between each seed is bisected by a Thiessen polygon edge. The generated Voronoi diagram is thus a matrix of segments dividing the borders of the homelands approximately every two hundred kilometres. These divisions then divide the space between the homelands roughly evenly, thus preventing double counting. As with bandwidth selection, there is a trade-off between counterfactual comparability and power. This is so as smaller segments identify more similar counterfactual observations, yet reduce the sample size within that segment.

---

counterfactual observations outside the country

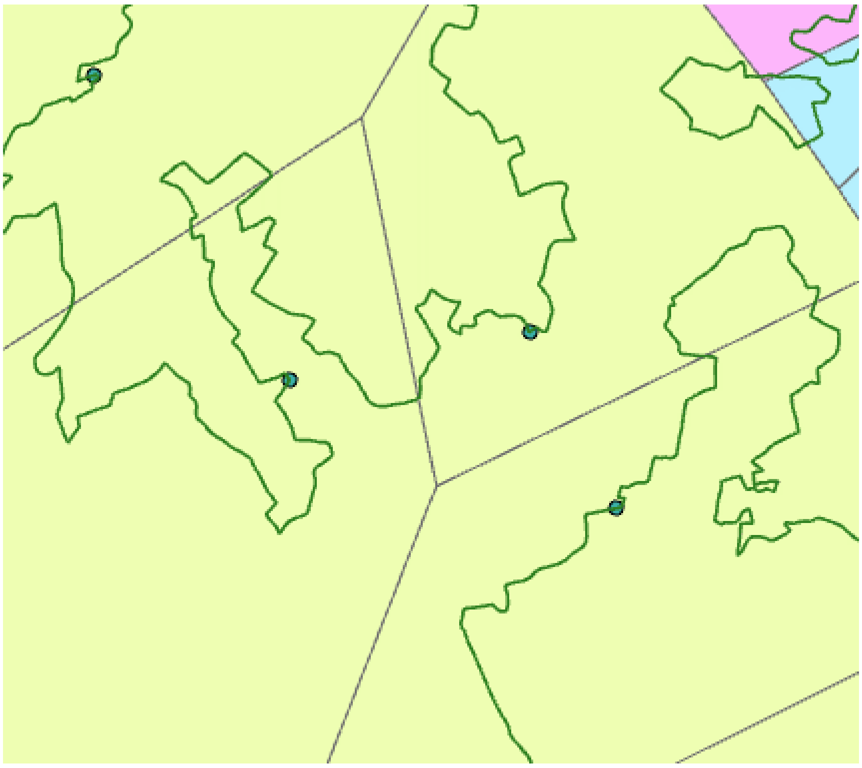

Figure 3.2: Close-up of Thiessen segmentation strategy

Each Thiessen polygon is then assigned to the homeland in which the seed that formed it lies, as per the colour scheme in Figure 3.3. Homeland assignment allows the effects found to be attributed to each homeland, as per its Thiessen polygon creation point. However, a more sophisticated technique of running the estimator for each homeland was conducted for the homeland decomposition for the education section.

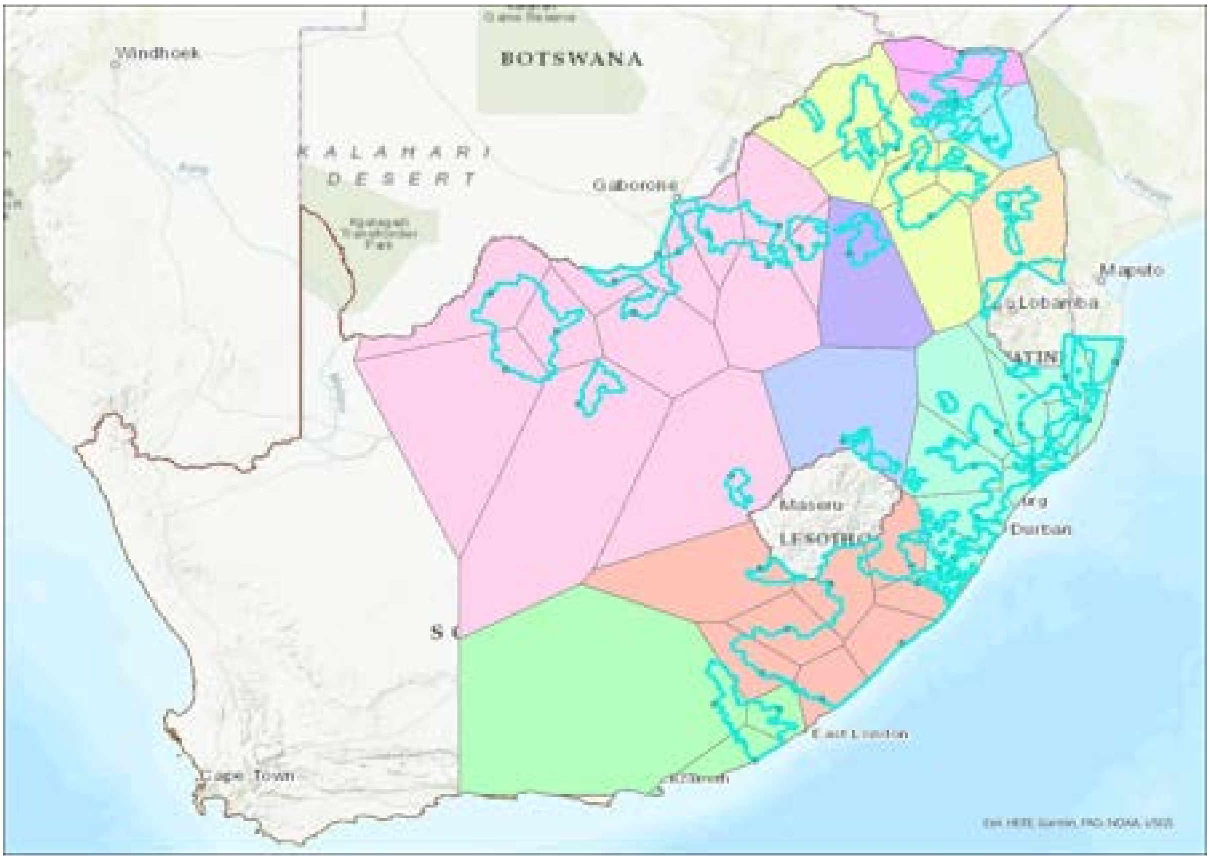

Figure 3.3: National Thiessen segmentation strategy. Observations within a given segment share a regression intercept.

To my knowledge, border seeded Thiessen polygons are a unique solution to the two dimensionality of space while using a one-dimensional running variable. A Scopus search of `"regression discontinuity" and "Thiessen"` returned no applicable results. Alix-Garcia et al. (2013) and Asher, Garg, and Novosad (2020) use Thiessen polygons in the context of an RDD. However, this was not to segment existing borders but to create borders around points and use those borders to create the running variable. An optimisation solution is provided by Keele and Titiunik (2015). The implementation of this solution was beyond the scope of this research.

3.6 ESTIMATION FRAMEWORK

The estimand of the RDD estimator employed is a local average treatment effect (LATE) as the analysis is looking at the treatment of the subset of the population that complied with the treatment, i.e. the non-counterfactual observations are within the homelands.

This research replicates the multiple specification paradigm of Dell (2010: 1882)[2]. As such, for each dependent variable, multiple specifications are run to ensure robustness to covariate

[2]The table replicated from (Dell, 2010) can be found in the Appendix, Figure 6.2.

inclusion, polynomial orders of covariates, and to transparently report specification sensitivity. Of these specifications, the specification which includes Thiessen polygon fixed effects at the shortest bandwidth is the benchmark, this is typically specification 7 at 10km.[3] Running multiple regressions requires the preponderance of estimates to be significant as a multiple hypothesis correction is not performed. Yet, when significance is found across all specifications, the true statistical significance of the result is higher than the p-value implies for any one specification. As per Gelman and Imbens (2019), specifications with running variable polynomials higher than order two were not replicated. The generalised equation of the estimator is given in Equation 3.1:

$$C_{isb} = \alpha + \gamma home_{sb} + X'_{isb}\beta + f(geographic\ location) + \Phi_{ib} + \varepsilon_{isb} \tag{3.1}$$

C_{isb} is the outcome variable for observation i, in bandwidth b in Thiessen segment s, *home* is a dummy variable for whether the observation lies within the homeland (1) or outside the homeland (0). X'_{isb} is a vector of covariates. $f(geographic\ location)$ represents the RD location functions, the autocorrelation control, longitude and latitude, and clustering. Φ_{ib} is the Thiessen segment fixed effects, as described above. A full table of variables and their specification inclusion can be found in the Appendix, Table 6.9. Replication files can be found in the Online Appendix.

3.7 DATA

This section describes how the novel data set used in this research was compiled. For each variable, Table 6.9 contains the variable label, a description, the source, and the specification inclusion of that variable. The GIS spatial data, the .do table replication files, and the Stata datasets used are available in the Online Appendix.

I created the primary variables created in a geographic information systems programme[4], required to handle the geographic dimensionality of the data. First, I set the geodetic datum of all data sets to the `Hartebeesthoek94` datum, required for a distance accurate map projection of South Africa.[5] Factor variables for the various locations of each observation were created using `Selection by Location`. The factor variables describe each observation's location in the following areas: the homeland, the province (excluding the homelands), the municipality, the

[3]The Stata code for this specification at each bandwidth is as follows: `spregress y_variable home dhome_nb dhome_nb2 lon lat lon2 lat2 dcity rural slope pop_dens tcity rain i.seg, gs2sls errorlag(S) force`. Variable descriptions can be found in Table 6.9.

[4]ArcGIS Pro

[5]This was done using the `Define Projection` function in ArcGIS Pro.

homeland centroid Thiessen polygon, and the homeland border Thiessen polygon. The utility of these polygons is explained in Section 3.5.

Using `Calculate Geometry`, I added variables for the longitude and latitude of the centroid of each observation, calculated in decimal degrees. Using the `Near` function, I created variables for the distance from the centroid of each observation to the nearest: city, coast, homeland border, homeland border excluding the national borders, and to each of the TBVCZ[6] borders.

I generated several variables which average the values of a variable in the proximity of each observation using the `Zonal Statistics` function. For the school observations, I created a polygon with a 5km[7] radius from each school using the `Buffer` function. The generated variables are thus an average of a variable beneath each polygon. The following variables were created using this method: top soil nitrogen content, population density, elevation, rainfall, and slope. In turn, slope was created from the elevation data using the `Slope` function.

After these variables were generated, I completed auxiliary calculations in Stata to generate the remaining variables. I calculated the natural logs of variables, polynomials of variables, and interactions terms of variables. The number of teachers per student at the school level was likewise created in Stata. As too was the school completion rate, pseudo-code for the creation of which can be found in the Appendix, Section 6.6.1. A description of these variables, their data labels, their specification inclusion, and data source can be found in Table 6.9. The dependent variables of this thesis are as follows: topsoil nitrogen, population density, schools per square kilometre, classroom size, and school completion rate.

3.8 CONCLUSION

In this chapter, I have motivated why the empirical strategy followed is appropriate for the hypotheses investigated. The chapter began with a description of the utility of quasi-experimental methods, showing how they mitigate selection effects and endogeneity. This was followed by a description of how the RDD estimates approximate the causal effect of the treatment through identifying a second-best counterfactual, the lands immediately outside the homelands. Here, the trade-off between counterfactual suitability (determined by the narrowness of the bandwidth) and sample size is explored. I describe the two RDD internal validity tests, covariate continuity and density tests, in Section 3.3. Here I presented a major limitation of this work, covariate discontinuity, and how this limitation is addressed. Thereafter, I presented the follow-

[6] Transkei, Bophuthatswana, Venda, Ciskei, KwaZulu

[7] 5km was chosen as a reasonable travel distance to a school.

ing issues pertaining to accurate standard error estimation: spatial autocorrelation, clustered sampling, raster data certainty, heteroscedasticity, and the family-wise error rate. This was followed by Section 3.5 describing a novel improvement to counterfactual identification in geographic RDDs. The method sections end with the estimation framework and the estimator equation. Lastly, Section 3.7 provides a description of how the novel geographic data set used in this work was compiled.

CHAPTER 4

RESULTS

4.1 INTRODUCTION

In this chapter, I provide the results for the three areas of investigation: population density, Section 4.2; topsoil degradation, Section 4.2; and education, Section 4.4. The chapter begins with population density, where I find that the homelands have likely induced a doubling of the population density in the former homelands. The population section is followed by the topsoil degradation results as topsoil quality was likely reduced in part by high population density, as explored in that section. Here I find that the homelands have reduced the topsoil nitrogen content by 2.19%. The education results show that the homelands have worsened school education outcomes and inputs while potentially having improved school accessibility.

4.2 POPULATION

The overcrowding of the homelands was a numerical inevitability. The black population of South Africa comprised 68.6% and 76% of the nation in 1946 and 1990 respectively (Chimere-Dan, 1992). Yet, only 13.7% of the nation was designated as homelands, the only legal places of residence for black South Africans. Consequently, by 1991, 47% of the country *de facto*[1] resided on land designated as homelands (C. Cooper et al., 1994). The homeland residents included approximately 3.5m people who were forcibly dispossessed of their property and relocated to the homelands (Platzky and Walker, 1985). As such, in 1985, the population density of the homelands was 83.95 (people per square kilometre) and only 18.48 for the rest of the (calculated from Tapson (1985: 237)). This large average difference has persisted. In 2015, the average population density of the former homelands was 87.38 while for the rest of the nation it was only 36.92. 29.5% of South African's continue to reside in the former homelands. The hypothesis:

> *Apartheid homeland policy has led to a long-run and persisting overcrowding in the former homelands.*

This hypothesis tests the causal aspect of the homelands' overcrowding. It is plausible that these lands would have had a higher population density than the rest of the country even if Apartheid had not occurred. For example, I find in Section 4.2 that the homelands have higher

[1]The *de facto* estimates are important as the *de jure* numbers were significantly inflated as functionally all black people were legally designated to a homeland which ostensibly corresponded to their ethnicity.

soil fertility, as measured by nitrogen content, than the rest of the nation on average. Higher agricultural potential could have induced a greater density of settlement. However, the RDD, as per Chapter 3, should shed some light on the component of the overcrowding that is due to Apartheid policy as the natural endowments of the areas are unlikely to change discontinuously at the border. Nonetheless, even without such a rigorous estimator, it can be perceived from a heat map of population density alone the effect of the homelands, as per Figure 4.1

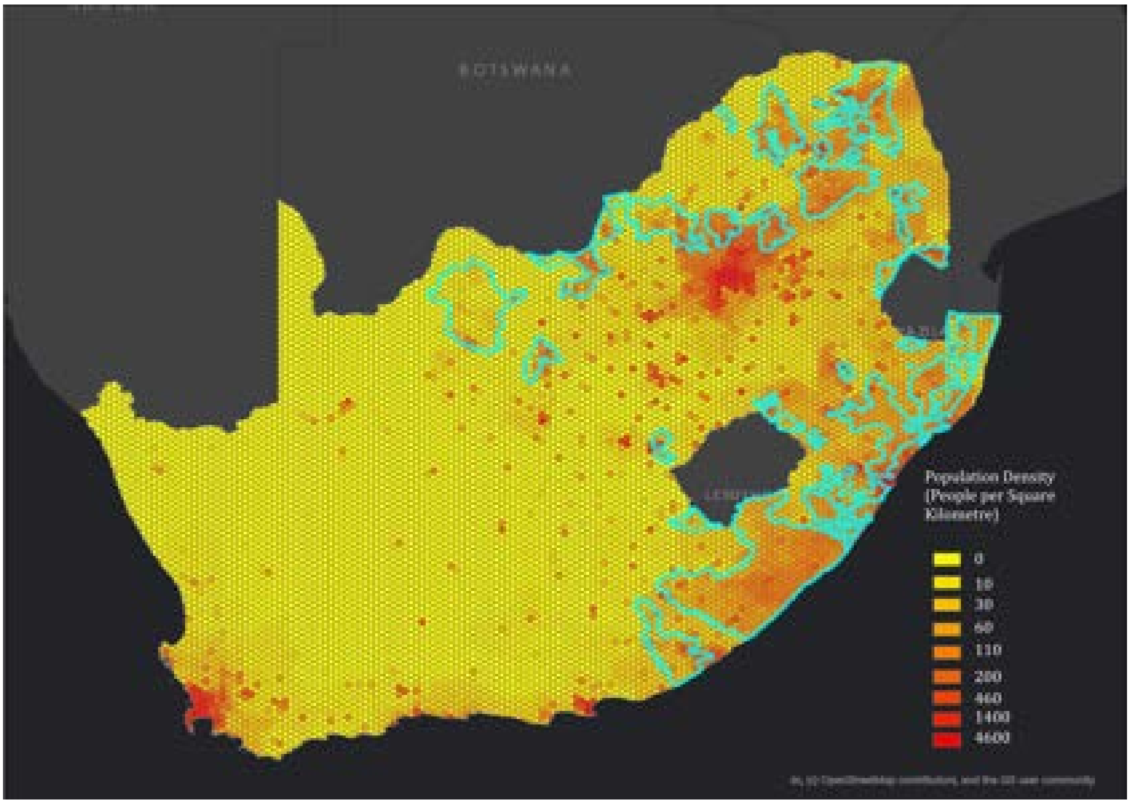

Figure 4.1: Hexagonal heat map of population density in South Africa with the homeland borders. Source: Author

Population patterning is a highly path dependent process. The former homeland areas could have had higher or lower population densities for a number of reasons which cannot be controlled for here. For example, areas of high black population density in white South Africa (labelled 'Black spots' by the NP) were often incorporated into the homelands, which would upward bias the estimates due to endogeneity. Yet, even here, the high population density in 'Black spots' was likely caused by Apartheid policy. This is so as Apartheid policy made it illegal for black people to purchase land in white South Africa. As such, only the limited lands with low legal oversight could be populated by black people in white South Africa, concentrating the density of people. Further, illegal occupation favours high population density as protection from the state security forces is likely improved by density. Consequently, high population density preceding the creation of the homelands can likely still be attributed to Apartheid.

The population of the former homelands has been allowed to move freely post-Apartheid.[2] Yet, this analysis is measuring the contemporary or persisting effects of the homelands, and thus allows for post-1994 movement of people (the first year black people could legally live outside of the homelands). Nonetheless, it is more likely that the flow of people will be out of more densely populated areas and into less populated areas in search of agricultural opportunity. This would downward bias an estimate of what the population density was during Apartheid, if that were the question at hand. The results of the RDD are reported in Table 4.1.

[2]At least in the legal sense. Long distances to the cities, poor transport infrastructure, and expensive modes of transport remain substantial impediments to the free movement of people with significant implications for employment. See Ardington, Case, and Hosegood (2009).

Table 4.1: **Population Density (people/km^2) RDD Specification Tests**

Specification / Bandwidth:	10km	25km	50km
1) Linear polynomial in distance to boundary			
Homelands	21.10***	26.61***	31.50***
	(7.249)	(6.404)	(6.432)
2) Quadratic polynomial in distance to boundary			
Homelands	19.03**	22.33***	22.19***
	(7.407)	(6.402)	(6.412)
3) Ordinary least squares			
Homelands	21.12***	26.03***	24.44***
	(7.321)	(6.234)	(6.280)
4) Linear polynomial in lat and long			
Homelands	22.40***	27.55***	29.03***
	(7.305)	(6.213)	(6.140)
5) Quadratic polynomial in lat and long			
Homelands	21.10***	26.60***	31.50***
	(7.249)	(6.404)	(6.432)
6) Interacted quadratic polynomials in lat and long			
Homelands	21.08***	26.69***	30.47***
	(7.057)	(6.425)	(6.422)
7) Thiessen segment fixed effects (Benchmark)			
Homelands	**17.55*** **	22.63***	25.54***
	(6.677)	(6.448)	(6.563)
8) Log-linear with Thiessen segment fixed effects			
Homelands	**0.7189*** **	0.812***	0.815***
	(0.0378)	(0.351)	(0.0353)
Observations	5112	3734	2202

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Spatial autocorrelation robust.
Homelands is a dummy variable of the lands of the *former* homelands, within the given bandwidth.
The bandwidth is the sampling distance on either side of the former homeland border.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.9
Regressions 4,5,6 include location polynomials to control for spatial autocorrelation.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.

Following the benchmark specification (7), the RDD estimator finds, at a 10km bandwidth, that the homelands are 17.55 people per square kilometre denser. Turning to the log-linear result (8), the estimate of 0.7189 shows that the homelands are approximately 105.22% (after the appropriate log transformation[3]) more densely populated than the area 10km outside the homelands. The remainder of the results corroborate the benchmark finding, with estimates typically higher than the benchmark specification. All estimates are significant at the 1% level. Both the effect size and significance illustrate the extreme degree of overcrowding and its persistence in democracy. Again, the raster data source did not provide certainty estimates, as such, the significance levels are likely inflated. Further, the size of the sampling matrix was arbitrarily defined at 10km^2 which dictated the number of matrix elements (observations), and in turn the size of the standard errors.

4.3 TOPSOIL DEGRADATION

> *"Because of their size and situation, the Reserves never had other than an extremely low productive potential. The shortage of land was acute, particularly in view of the non-capitalist forms of production practised. By the 1920s already the Reserves were predominantly characterised by overcrowding, overstocking, and overgrazing. Reserves could be distinguished at sight by their bareness; desert conditions were developing rapidly in many parts of the Reserves."* Molteno (1977: 18).

In this section, I test whether the homelands have led to a long-run and persisting degradation of the topsoil in the former homelands. Topsoil nutrient levels (in this case nitrogen content) are reduced by overgrazing, erosion, and unimproved crop agriculture, and improved by the use of fertilisers and nitrogen fixing crops. The literature suggests that the overpopulation of the areas (corroborated in Section 4.2), and moreover, the purported cattle overstocking and overgrazing (Molteno, 1977: 18), has induced erosion leading to reductions in topsoil quality. These results have important welfare implications as many rural South African's continue to depend on agriculture. This section supports the hypothesis that the homelands' topsoil has become degraded. The hypothesis:

> *The homelands have led to a long-run and persisting degradation of the topsoil in the former homelands.*

The empirical strategy is to use topsoil nitrogen content, one of the three macronutrients essential for plant growth, as a proxy for topsoil fertility. Essentially, this chapter finds that the topsoil immediately outside the homelands contains more nitrogen than the topsoil immediately within the homeland. This implies the soils are indeed degraded as it is likely only the homeland

[3] $((e^{\beta}) - 1) \times 100$

itself which could create a discontinuity exactly at the homeland border. This result is robust to all robustness checks as per Chapter 3.

Topsoil nitrogen content is a time variant factor and most of the data were collected between 2000 and 2015 (Hengl et al., 2017). Consequently, the reduction in topsoil quality may have occurred after the end of Apartheid. However, time variance does not break the causal chain from the establishment of the homelands to contemporary patterns of soil quality as it is the geographically persisting effects of the homelands which is being detected. Nonetheless, it is worth noting that the programme of 'betterment' (which, *inter alia*, sought to reduce soil erosion primarily through culling cattle) ended at the end of Apartheid (see Chapter 2 Section 2.2). It is thus possible that this programme was effective (at large social cost) while still finding the result which is found, i.e., that the homelands led to large reductions in current topsoil quality.

The historical literature strongly supports the hypothesis that the overcrowding of the homelands led to livestock overstocking and thus to the erosion of the topsoil (N. Nattrass and J. Nattrass, 1990: 528; Molteno, 1977: 18; De Wet, 1995; Levin and Weiner, 1997). However, this literature is historical and thus does not measure whether these harms have persisted into democracy. Further, these were observational descriptive assertions, none of the literature employs a rigorous empirical technique. Lastly, there is some scientific literature using remote sensing to determine erosion and land degradation in the former homelands (Seutloali, Dube, and Mutanga, 2017; Sepuru and Dube, 2018; Wessels et al., 2004). However, this literature is not comparative and certainly does not seek to determine the effect of the establishment of the homelands on topsoil degradation. As such, the question at hand remains novel in the literature.

Table 4.2: **Topsoil Nitrogen (cg/kg) RD Specification Tests**

Specification/ Bandwidth	50km	25km	10km
1) Linear polynomial in distance to boundary			
Homelands	-7.358***	-5.983***	-3.676***
	(0.890)	(0.921)	(1.061)
2) Quadratic polynomial in distance to boundary			
Homelands	-8.208***	-6.207***	-3.829***
	(0.894)	(0.910)	(1.054)
3) Linear polynomial in lat and long			
Homelands	-7.165***	-6.468***	-4.129***
	(0.733)	(0.746)	(0.833)
4) Quadratic polynomial in lat and lon			
Homelands	-5.775***	-5.105***	-3.456***
	(0.658)	(0.654)	(0.707)
5) Interacted quadratic polynomial in lat and lon			
Homelands	-4.503***	-5.032***	-3.454***
	(-0.628)	(0.654)	(0.746)
6) Thiessen Segment fixed effects (Benchmark)			
Homelands	-6.816***	-5.944***	**-4.454*****
	(0.676)	(0.643)	**(0.657)**
7) Log-linear with Thiessen Segment fixed effects			
Homelands	-0.0313***	-0.0257***	**-0.0217*****
	(0.0029)	(0.00292)	**(0.00318)**
Observations	5112	3734	2202

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Spatial autocorrelation robust
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.9

Table 4.2 provides the results for the reduction in topsoil nitrogen caused by the homelands. The benchmark specification (6), at 10km, shows that a reduction in topsoil nitrogen of 4.454cg/kg can likely be attributed to the homelands. Looking at the log-linear specification

(7), this amounts to an approximately 2.19% reduction in topsoil nitrogen. These effect sizes may not seem large. However, at the margin, there are likely many areas that are no longer arable due to this reduction in nitrogen (and the other soil nutrients this proxies for). As agriculture remains vital to the livelihoods of rural residents, this result has important welfare considerations.

This result is particularly robust as the homelands themselves are slightly higher in nitrogen than the country as a whole (see Figure 6.3 a national map and histogram comparing topsoil nitrogen). Consequently, it is evident from the sign of the coefficients alone that the RDD methodology is likely working as intended, creating an appropriate counterfactual rather than simply comparing the averages of the homelands with another arbitrary geographic entity. The Appendix provides the following robustness checks: clustering at the homeland centroid Thiessen polygon level (Table 6.3), clustering at the municipal level (Table 6.4), homeland segment fixed effects (Table 6.2) and full regression output at 10km (Table 6.5), all of which support the primary findings. The signs and coefficients (of the non-interacted coefficients) are all fairly similar, varying from -8.2 to -3.45 and significant at the one percent level.

However, there are important limitations to this result. This section is critically limited by not accounting for the uncertainty in the raster data used for the topsoil nitrogen content observations. As such, the standard errors do not accurately reflect the underlying uncertainty in the data. However, as per Chapter 3, 160 raster observations were aggregated into a single observation for the analysis undertaken above— arbitrarily reducing the sample size and thus significance found. Yet, as this is not a formal control for the uncertainty in the raster data, a robustness exercise is undertaken. As such, a different ISRIC source of actual samples is used such that the significance level found is entirely valid (Batjes et al., 2017). Unfortunately, there are only 114 observations for nitrogen, and the average distance to the homelands of these points is 134km. As such, the nitrogen data-set is unusable.

Fortunately, the data set also contains a variable for organic carbon, another important measure of soil fertility (Zmora-Nahum et al., 2005). Here, significance is found for most specifications (see Table 4.3). However, this is unfortunately only under a simple OLS regression without a comparison at various bandwidths.[4] OLS is required here as the average distance to a homeland border of the organic carbon observations is 110km (there are 122 observations within the former homelands and 514 outside). Thus, limiting the sample to an appropriate

[4] All other parameters of the estimator are the same, except for the exclusion of the RDD bandwidth comparison

bandwidth for an RDD reduces power beyond usability. This is an equally severe limitation to the indeterminablity of significance in the RDD specifications used prior, as there is no causal identification strategy in OLS. It is possible that the lands were chosen for their poor arability which is reflected in the finding.

Table 4.3: **Organic Carbon OLS (g/kg)**

	Centroid clusters		Municipal clusters	
	(1)		(2)	
	home	border dist	home	border dist
1) Ordinary Least Squares				
Homelands	-3.595***	omitted	-3.595**	omitted
	(1.229)	omitted	(1.466)	omitted
2) Quadratic polynomial in distance to boundary				
Homelands	-3.962***	-1.32e-05**	-3.962***	-1.32e-05*
	(1.136)	(6.42e-06)	(1.431)	(7.11e-06)
3) Linear polynomial in lat and long				
Homelands	-3.684***	omitted	-3.684**	omitted
	(1.248)	omitted	(1.537)	omitted
4) Quadratic polynomial in lat and long				
Homelands	-3.390**	5.63e-06	-3.390**	5.63e-06
	(1.281)	(7.15e-06)	(1.597)	(8.06e-06)
5)Interacted quadratic polynomials in lat and long				
Homelands	-3.615***	3.12e-05***	-3.615**	3.12e-05***
	(1.253)	(8.66e-06)	(1.568)	(1.06e-05)
6) Thiessen Segment fixed effects				
Homelands	-4.015**	6.04e-06	-4.015**	6.04e-06
	(1.598)	(8.72e-06)	(1.949)	(6.13e-06)
7) Log-linear with Thiessen Segment fixed effects				
Homelands	-0.1644**	1.57e-06***	-0.1644***	1.57e-06***
	(0.09058)	(5.50e-07)	(0.0647)	(5.09e-07)
observations	636		636	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Specifications can be found in Table 6.1.

Nonetheless, significant results are found, and the sign of the coefficients corroborate the main findings. Further, with distance to the border often insignificant (especially important for the benchmark specification including Thiessen fixed effects), this suggests that it is only crossing the border itself, the homelands dummy, which is driving the effect and not distance. These results find a a reduction in organic carbon levels in the benchmark specification (6) of 4.015g/kg or approximately 17.87% reduction in organic carbon. This effect size is substantial and points to the same underlying data generating process of soil erosion induced soil nutrient depletion, as found with the RDD nitrogen results.

A natural extension of this work is the use of other soil fertility measurements, such as phosphorous and potassium, to ensure the robustness of the finding. A further robustness exercise would be the use of remote sensing of soil erosion, as per Seutloali, Dube, and Mutanga (2017), to assess the causes of the reduced soil fertility within the homelands. These data might be of sufficiently high spatial resolution for significance to be maintained under the RDD design.

Finally, using data from Gilbert et al. (2018), I test to see whether it is in fact cattle overstocking which is the cause of the soil degradation. Running specification 7, I do not find any statistically significant difference in cattle density at any bandwidth. I report this to be transparent about this work's underlying family-wise error rate.

4.4 BASIC EDUCATION

Foremost among the challenges facing rural South Africa is the task of improving the quality of education. What is often overlooked, however, is the immense, untapped potential of rural communities to take the lead in shaping a better future for themselves. Nelson Mandela in Nelson Mandela Foundation (2005)

4.4.1 Introduction

The Bantu Education Act of 1953 caused much of the the substantial inequity in education outcomes between black and white children in South Africa today. The Act perpetuated a pattern of inequality, but also expanded formal education for black children (Giliomee, 2009). With only 24.5% of black school age children in school in 1948 (the year the NP came to power), the bar for improvement was exceptionally low (Giliomee, 2009). Hendrik Verwoerd,[5] the force behind the act, sought to restructure the education system which he perceived to show black people the "green pastures of the European [while] still... not allow[ing] him to graze there" as "there is no place for [a black man] in the European community above the level of certain forms of labour".[6]

The iniquitous pre-Apartheid pattern of education was perpetuated by Bantu education most grievously through exceptionally low per capita education funding for black students. As per Giliomee (2009): "Per capita spending on black pupils dropped from R17,99 in 1953-1954 to R11,56 in 1962-1963, after which it began to rise." This can be compared to per capita white funding of R76.58 in 1945, R127.84 in 1953, and R144.57 in 1968 (Horrell, 1969: 39). Thus a most fundamental human capital investment was an order of magnitude higher for white children. As investments in human capital tend to be perpetuated through generations, this racial pattern of education inequality has persisted. As such, 78% of white students received a bachelor's pass, compared to only 23% of black students in 2018 (Spaull, 2019: 20). However, the focus of this chapter—the geographic persistence and distribution of education outcomes in relation to the former homelands—is novel in the literature. These results illustrate a geo-

[5]Verwoerd, commonly known as the architect of Apartheid, was born in Amsterdam to Dutch parents sympathetic to the Afrikaner nationalist cause (Kenney, 2016).

[6]These are likely the most oft quoted words on education by Verwoerd. However, they stand in tension with less studied words of Verwoerd: "We shall have to negotiate frequently with [black people] in the future over many issues, including education and politics. It would be better to negotiate with people who are well informed and educated" (Giliomee, 2009) and in the Eiselen Commission, commissioned by Verwoerd: "The intelligence of black children was of [no] special and peculiar a nature as to demand on these grounds a special type of education" and "The Bantu child comes to school with a basic physical and psychological endowment that differs... so slightly from that of the European child that no provision has to be made in educational theory or basic aims (Government, 1951: 131)." This is perhaps Apartheid doublespeak.

graphically determined reduction on the returns to the largest human capital investment the government makes, low cost public education.

The hypothesis:

> *The creation of the former homelands has led to a long-run and persisting reduction in education outcomes, inputs, and accessibility in the former homelands.*

Overall, I find that education inputs and outcomes are likely much lower in the former homelands, but it is possible that access, as measured by distance to the nearest school, is higher in the former homelands. This latter finding has important qualifications, detailed in Subsection 4.4.3.

4.4.2 Descriptive Statistics

Table 4.4: **School Descriptive Statistics**

	National		Homelands		National Excl homeland	
Variable/ Year:	2002	2016	2002	2016	2002	2016
Student/teacher ratio	58.062	31.305	50.177	32.296	70.114	30.248
Observations	[17977]	[24724]	[10867]	[12752]	[7110]	[11972]
Teachers per school	10.987	16.887	10.096	12.638	12.349	21.412
Observations	[17977]	[24724]	[10867]	[12752]	[7110]	[11972]
Students per school	497.219	510.776	421.101	402.990	583.970	626.456
Observations	[22043]	[25283]	[11741]	[13088]	[10302]	[12195]
Completion rate	82.186	94.039	77.666	93.519	87.156	94.6
Observations	[19122]	[22998]	[10014]	[11942]	[9108]	[11056]
Schools per km^2		0.0209		0.0754		0.0123
Observations		[12716]		[12716]		[12716]
Schools per person		0.0005		0.0011		0.0004
Observations		[12716]		[12716]		[12716]
Teachers per person		0.0051		0.0116		0.0040
Observations		[12716]		[12716]		[12716]
Students per km^2		11.084		30.133		8.078
Observations		[12716]		[12716]		[12716]

Homeland refers to the lands of the *former* homelands.
Author's calculations from Basic Education (2016)
Only schools with a 2016 EMIS (school identification number) are included in the 2002 estimates.
The completion rate is a 3,4,5 year average completion rate, dropping schools without the required number of years below the highest grade.

These descriptive statistics should be interpreted with caution. Schools per square kilometre is likely higher for the homelands due to the much higher population density of the homelands.[7] Schools per person is likely higher in the homelands in part due to the younger population demographics in rural areas in South Africa. Students per square kilometre is highly correlated with both population density as well as schools per square kilometre, the latter even after controlling for population density. How much of the latter finding is endogeneity where the presence of a school induces school going and how much of this is the correlation between high population density and youthful demographics is indeterminable as there are no open source geographic demographic data available.

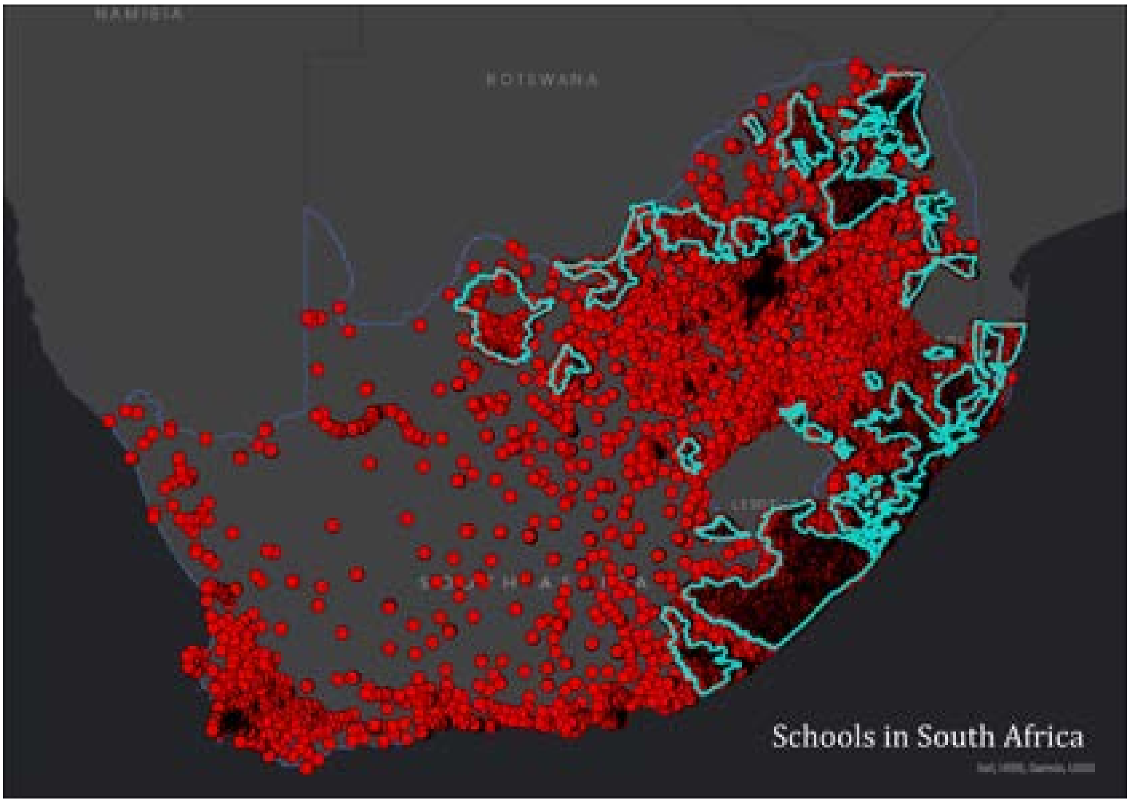

Figure 4.2: A map of the 25,312 schools in South Africa (2016) with the homeland borders

Teachers per person comes with the same youthful population demographics caveat. Where population is a factor, interpretability would be easier if it were the numerator. However, this often required dividing by zero and as such was avoided. Students per teacher, also known as the PTR (pupil teacher ratio), is calculated per school (as is the school completion rate).

[7]In the final regressions, population is controlled for. However, this is likely to downward bias the true estimates as population density is correlated with the homelands, as seen in Chapter 3.

Therein, a school with 500 students and a PTR of 10 is counted equivalently to a school with 10 students and a PTR of 10. As such, the national PTR cannot be calculated from the total student and teacher numbers.

4.4.3 Dependent Variables

There are four dimensions by which this chapter seeks to determine the quality and nature of schooling induced by the homelands. The primary quality indicators are students per teacher (PTR), i.e. classroom size, (a measure of education inputs) and the completion rate (a measure of education outcomes). Both of these indicators are imperfect proxies, employed due to the lack of publicly available educational attainment data at the school level. Indeed, the only publicly available data useful for creating proxies for educational attainment at the school level are the number of students per grade and the number of teachers in the school.

Hence, from these data, only the PTR and completion rate are determinable. The PTR is an imprecise proxy for attainment as it is possible that excellent teachers can teach more students equally effectively. Further, there are several reasons the PTR does not reflect classroom size exactly, including teacher absenteeism and class scheduling. However, it likely reflects the resourcing of the school more generally which is likely to influence education outcomes.

The completion rate for each year (2002 and 2016) is an average of the 3, 4, and 5 year completion rates up to the highest grade in a given school (as determined by its EMIS, the school identification number ((DBE, 2016)). To calculate the 3 year completion rate for 2016, the number of students who are registered in the highest grade of the school in 2016 is taken as a percent of the number of students in the grade 3 grades below the highest grade, 3 years prior (2013). This is repeated for the remaining completion rates for both 2002 and 2016. However, as there are no basic education graduation data in the data set, the completion rate is a misnomer as the students are only reaching the highest grade in the school rather than completing school itself.

These 3, 4, and 5 year completion rates are then averaged. If a school does not have a grade the required number of grades beneath its highest grade, that observation is dropped. 2016 is the highest year in the Snap data (see Basic Education (2016)) and the first year in the data is 1997, hence 2002 is the lowest year with a 5 year completion rate. The pseudo-code for creating this variable can be found in the Appendix (6.6.1).

Only schools with an EMIS number from 2014 onward are included in the 2002 sample.

This omission is due to the lack of accurate GPS coordinates in the 2002 sample. This should improve comparability as the schools in the 2002 sample but not in the 2014 sample are most likely to be the worst performing schools (as those are the schools most likely to have been closed). As the school completion rate has improved, the omission of the worst performing schools in 2002 will likely downward bias a 2002-2016 comparison.

Schools per square kilometre (after controlling for population density) is used as a measure of access to schooling. However, this comes with several important caveats. First, access to schooling is also a function of the transport infrastructure. As it is known that the homelands have poorer infrastructure, more schools per square kilometre may be needed in order to maintain similar travel times. Second, many schools with only a few grades may entail that for a given student in a given grade, the nearest school with that grade may remain distant even with a high density of schools. Finally, many small schools may be substantially less efficient due to high fixed costs. Thus, there may be a trade-off between the number of schools and their size. Consequently, the final variable of interest is the number of teachers per school which should roughly define the school's size.

4.4.4 Robustness

The regression discontinuity design was followed as per Chapter 3. However, a significant advantage of this chapter is the use of point data, a GPS location of every school in the country, and not imputed raster data. Consequently, the standard errors reflect the actual number of observations in the underlying data and remain robust to the numerous robustness checks employed.

To determine the schools per square kilometre, a 100km^2 sampling matrix was used, with sample size and thus standard errors reflecting the number of elements in the matrix (12,716 elements nationally, aggregating 25,312 school observations). The inclusion of a population covariate, where population is positively correlated with the homelands (see Section 4.2) and positively correlated with the completion rate, leads to a positive bias on the homelands coefficient. As I find that the effect of the homelands is negative, the absolute effect size is downward biased, hence the results found are conservative estimates.

Schools per square kilometre controls for spatial autocorrelation in the sampling matrix as per Chapter 3. However, in the data set, there are schools which share a single site location, entailing the duplication of GPS coordinates for multiple observations. Due to this, the `spmatrix create` function could not create a spatial weighting matrix. As such, spatial autocorrelation is controlled for with longitude and latitude and their squares in the remaining regressions, as per Dell (2010). Municipal level clustering was employed as there are no publicly available spatial files of school districts, a perhaps more suitable cluster level for this analysis. The robustness checks, clustering at different levels, can be found in the Appendix (6.7).

4.4.5 Results

4.4.5.1 Accessibility

Table 4.5: **Schools per Square Kilometre**

Specification/ Bandwidth:	50km	25km	10km
1) Linear polynomial in distance to boundary			
Homeland	0.0367***	0.0397***	0.0402***
	(0.00263)	(0.00198)	(0.00173)
2) Interacted linear polynomial in distance to boundary			
Homeland	0.0165***	0.0302***	0.0346***
	(0.00446)	(0.00278)	(0.00209)
3) Quadratic polynomial in distance to boundary			
Homeland	0.0364***	0.0396***	0.0403***
	(0.00261)	(0.00194)	(0.00168)
4) Ordinary least squares			
Homeland	0.0370***	0.0405***	0.0411***
	(0.00261)	(0.00192)	(0.00163)
5) Quadratic polynomials			
Homeland	0.0365***	0.0393***	0.0398***
	(0.00264)	(0.00202)	(0.00176)
6) Interacted quadratic polynomials in lat and long			
Homeland	0.0365***	0.0392***	0.0398***
	(0.00264)	(0.00202)	(0.00176)
7) Thiessen segment fixed effects (Benchmark)			
Homeland	0.0372***	0.0390***	**0.0401*** **
	(0.00262)	(0.00215)	**(0.00185)**
8) Log-linear with Thiessen segment fixed effects			
Homeland	0.869***	0.838***	**0.720*** **
	(0.031)	(0.033)	**(0.039)**
Observations	2741	2309	1543

Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Spatial autocorrelation robust.
Homelands is a dummy variable of the lands of the *former* homelands, within the given bandwidth.
Specifications can be found in Table 6.9.
These results are spatial autocorrelation robust as per Chapter 3.
The bandwidth is the sampling distance on either side of the former homeland border.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.
Regression 5-8 include location polynomials to further control for spatial autocorrelation.

In Table 4.5, taking regression 7 at a 10km bandwidth as the benchmark, this result shows that 0.0401 more schools per square kilometre can be attributed to the former homelands. Turning to regression 8 at 10km (0.720), after the appropriate transformation into percentages, approximately 105% more schools per square kilometre can be attributed to the former homelands. As population density is controlled for, this suggests significantly greater accessibility to schools in the former homelands. However, this comes with the important transport infrastructure and number of grades qualification. Nonetheless, this staggering density of schools, even after controlling for population density, can be clearly seen in Figure 4.3.

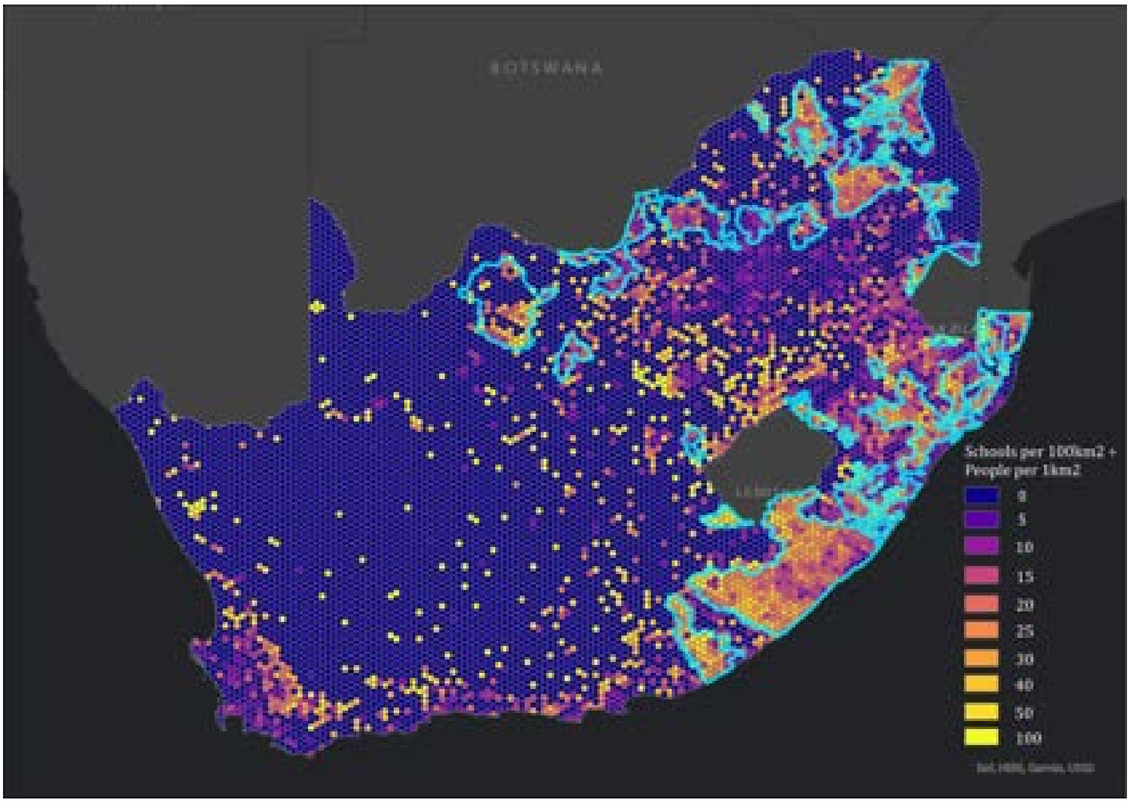

Figure 4.3: Hexagonal heat map of schools per 100km^2 as a fraction of the population density (people per square kilometre), 2016. Source: Author

Note the relative lack of schools in Gauteng due to the high population density, yet the homelands remain densely schooled.

Accessibility, although a vital education parameter, has an inherent trade-off with size, should the same number of teachers serve the same area. Many small schools likely creates great inefficiencies as the fixed costs of the schools, such as administration, would need to be duplicated for each school. To test whether the homeland schools are in fact smaller, I use the number of teachers as a proxy for the size of the school in Table 4.6.

Table 4.6: **2016 Teachers per School**

Specification/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homeland	-3.381***	-3.087***	-2.386***
	(0.497)	(0.575)	(0.573)
2) Interacted linear polynomial in distance to boundary			
Homeland	-2.324***	-1.112**	-1.363*
	(0.540)	(0.489)	(0.752)
3) Quadratic polynomial in distance to boundary			
Homeland	-3.490***	-3.176***	-2.534***
	(0.489)	(0.551)	(0.537)
4) Ordinary least squares			
Homeland	-3.311***	-3.141***	-2.489***
	(0.507)	(0.560)	(0.543)
5) Quadratic polynomials			
Homeland	-3.070***	-2.970***	-2.351***
	(0.560)	(0.613)	(0.582)
6) Interacted quadratic polynomials in lat and long			
Homeland	-3.061***	-2.968***	-2.353***
	(0.562)	(0.613)	(0.583)
7) Thiessen segment fixed effects (Benchmark)			
Homeland	-3.125***	-2.954***	-2.349***
	(0.505)	(0.602)	(0.606)
8) Log-linear Thiessen segment fixed effects			
	-0.121***	**-0.128*** **	-0.119***
	(0.0354)	**(0.0383)**	(0.0379)
Observations	15440	11287	7796

Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Municipality level clustering.
Specifications can be found in Table 6.9.
These results are spatial autocorrelation robust as per Chapter 3.
The bandwidth is the sampling distance on either side of the former homeland border.
Regression 7 and 8 include location polynomials to further control for spatial autocorrelation.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.
Homelands is a dummy variable of the lands of the *former* homelands, within the given bandwidth.

Table 4.6 clearly shows that the within homeland schools are significantly smaller than the schools immediately outside the homelands. As per specification 8 at 10km, there are approximately 13.66% fewer teachers per school in the homelands than immediately outside the homelands. This read in conjunction with the more than doubling of the density of schools per square kilometre (after controlling for population) shows that the trade-off described above, between accessibility and school size, likely pertains. This is primarily driven by the splintering of schools into the four DBE phases of education: Junior Primary, Senior Primary, Junior Secondary, Senior Secondary (Basic Education, 2005: 33). Moreover to duplicated fixed costs, this splintering reduces continuity of teaching as student must move to a new school after 3 years. Yet there are proposed benefits to small schools beyond accessibility: "some commentators argue that small schools encourage democratic participation in school affairs as the school community is well known to the broader community. Moreover, such schools are reported to have fewer disciplinary problems than larger comprehensive schools" (Basic Education, 2005).

4.4.5.2 Students per teacher

Table 4.7: **2016 Students per Teacher**

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homelands	1.498***	1.766***	1.642***
	(0.567)	(0.573)	(0.488)
2) Interacted linear polynomial in distance to boundary			
Homelands	1.880***	1.411**	0.432
	(0.714)	(0.637)	(0.803)
3) Quadratic polynomial in distance to boundary			
Homelands	1.612***	2.004***	1.897***
	(0.557)	(0.572)	(0.518)
4) Ordinary least squares			
Homelands	1.921***	2.117***	1.804***
	(0.575)	(0.632)	(0.523)
5) Quadratic polynomials			
Homelands	1.346**	1.754***	1.676***
	(0.576)	(0.590)	(0.525)
6) Interacted quadratic polynomials in lat and long			
Homelands	1.296**	1.749***	1.676***
	(0.588)	(0.594)	(0.531)
7) Thiessen segment fixed effects (Benchmark)			
Homelands	1.399***	**1.983*** **	1.727***
	(0.521)	**(0.520)**	(0.537)
8) Log-linear Thiessen segment fixed effects			
	0.0720***	**0.0738*** **	0.0522***
	(0.0189)	**(0.0188)**	(0.0188)
Observations	15921	11620	8000

Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Municipality level clustering.
Homelands is a dummy variable of the lands of the former homelands within the given bandwidth
Specifications can be found in Table 6.9.
The bandwidth is the sampling distance on either side of the former homeland border.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.

Moving to Table 4.7, taking regression 7 at 5km as the most complete benchmark specification, we can see that 1.983 more students per teacher at a given school can be attributed to the existence of the former homelands. After the appropriate transformation into percentages in specification 8, this amounts to approximately 7.66% increase in the number of students per teacher. This is a substantially larger teaching burden and robustly demonstrates the persistence of the Apartheid pattern of education subjugation into democracy. This pattern is clearly discernible on the map of students per teacher, Figure 4.4.

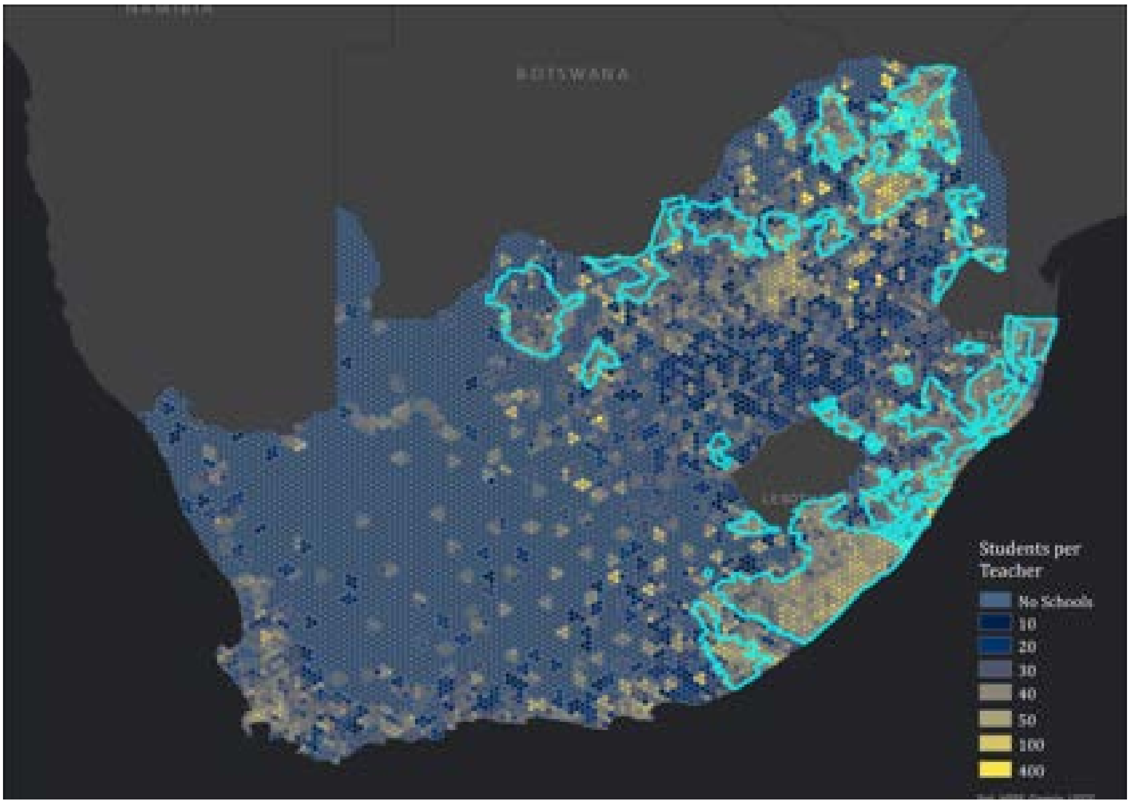

Figure 4.4: Hexagonal Heat map of students per teacher, 2016. Source: Author

Further, this pattern is visualised as a regression discontinuity plot (which does not account for covariates), in Figure 4.5.

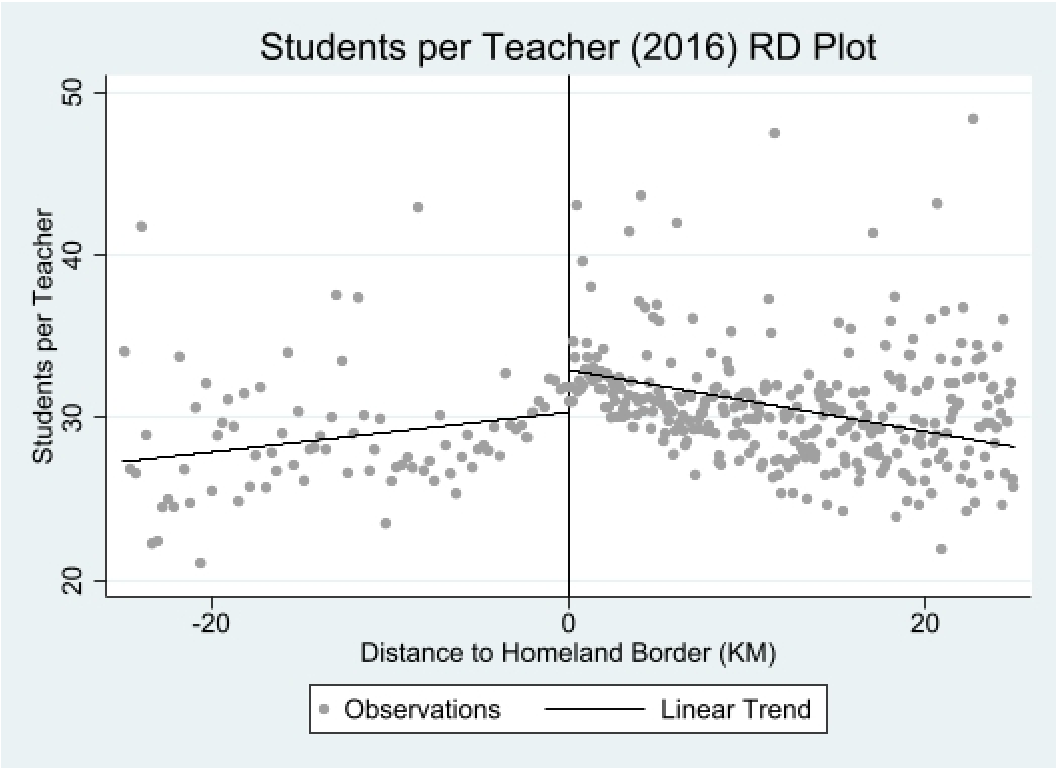

Figure 4.5: Students per teacher (2016) RD Plot. 0 is set to the homelands' borders. Distance is multiplied by -1 if the observation lies outside a homeland.

The 2002 regressions are not significant (see Appendix Table 6.10). This is likely in part due to the omission of schools not in the 2016 data set. Likewise, the student weighted PTR regressions are not significant.

Next, to see if these results are driven by particular homelands, I rerun the analysis for the TBVC[8] homelands and for KwaZulu.[9]

[8]The homelands which achieved nominal independence: Transkei, Bophuthatswana, Venda, and Ciskei.

[9]Included as this is the former homeland in my home province. All homelands were not run due to time constraints.

Table 4.8: **2016 Students per Teacher TBVCZ Decomposition**

	Transkei		Bophuthatswana		Venda		Ciskei		KwaZulu	
Bandwidth:	25km	10km	25km	10km	25km	10km	25km	10km	25km	10km
1) Linear polynomial in distance to boundary										
Homelands	0.503	2.669***	-0.360	0.547	0.755	0.771	4.559	6.776***	0.880	1.314
	(1.031)	(0.927)	(1.109)	(0.899)	(1.465)	(1.153)	(3.038)	(1.894)	(0.907)	(0.820)
2) Interacted linear polynomial in distance to boundary										
Homelands	4.699***	3.597*	-0.680	-2.402	-0.520	2.124*	10.56**	5.431**	1.646*	1.694**
	(1.572)	(1.932)	(1.388)	(1.571)	(1.664)	(1.106)	(4.402)	(2.316)	(0.953)	(0.669)
3) Quadratic polynomial in distance to boundary										
Homelands	1.994*	3.181***	0.0264	1.254	1.291	0.718	4.235	8.092***	0.631	1.031
	(1.148)	(1.077)	(1.212)	(1.034)	(1.352)	(1.179)	(4.700)	(2.566)	(0.993)	(0.928)
4) Ordinary least squares										
Homelands	1.922*	3.177***	1.157	1.187	1.622	0.723	3.346	7.679***	1.231	1.241
	(1.139)	(1.072)	(1.050)	(1.035)	(1.157)	(1.117)	(4.801)	(2.067)	(1.015)	(1.056)
5) Quadratic polynomials										
Homelands	1.597*	2.621***	-0.265	0.803	0.140	1.011	6.981***	6.976***	0.917	1.391*
	(0.904)	(0.977)	(1.196)	(0.958)	(1.674)	(1.578)	(2.080)	(1.923)	(0.886)	(0.811)
6) Interacted quadratic polynomials in lat and long										
Homelands	0.498	1.572	-0.478	0.703	0.594	1.155	11.81**	5.979***	0.995	1.449*
	(1.052)	(0.961)	(1.216)	(0.972)	(1.654)	(1.426)	(4.899)	(1.937)	(0.851)	(0.793)
7) Location polynomials with segmented Theissen fixed effects										
Homelands	0.214	1.431	-0.0290	0.760	0.885	1.319	9.102***	8.968***	1.288*	1.555**
Observations	2458	1121	1754	1268	736	484	953	603	5385	4441

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The homeland variable is a dummy of whether the school lies within the *former* homeland lands.
The bandwidth is the sampling distance on either side of the former homeland border
The segmented Theissen fixed effects segment the homeland borders into 50 segments

Table 4.8 shows that the results found for all homelands is broadly reflected at a statistically significant level for Transkei and Ciskei. The lack of statistical significance for the other homelands may simply reflect the reduced sample size at the level of individual homelands. Nonetheless, it is clear from these results that there is large variation between the homelands. Ciskei appears to have by far the worst PTR with at least 6 more students per teacher than lands immediately outside the homelands.

4.4.5.3 Completion rate

The initial results reflect the average completion rate *at the level of the school.* As such, a school with 500 students and a 50% completion rate is counted equivalently to a school with 10 students and a 50% completion rate. However, as 250 students failing is not equivalent to 5 students failing, a student weighted completion rate is also calculated. To do so, the average completion rate is multiplied by the number of students in that school as a fraction of the total number of students in the country. Counting at the level of students is evidently important, however, the coefficients do not have a straightforward interpretation. As such, Tables 4.9 and 4.11 report the more readily interpretable school level completion rates for the years 2016 and 2002 respectively. While Tables 4.10 and 6.7 report the student weighted results for 2016 and 2002 respectively.

Table 4.9: **2016 School Completion Rate**

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homeland	-2.910*	-4.733**	-4.163*
	(1.624)	(1.862)	(2.155)
2) Quadratic polynomial in distance to boundary			
Homeland	-2.913*	-5.218***	-4.060*
	(1.638)	(1.987)	(2.171)
3) Ordinary least squares			
Homeland	-4.253**	-5.307***	-4.143**
	(1.733)	(1.905)	(2.031)
4) Quadratic polynomials			
Homeland	-3.081*	-4.966**	-3.986*
	(1.612)	(1.933)	(2.197)
5) Interacted quadratic polynomials in lat and long			
Homeland	-2.968*	-4.958**	-3.985*
	(1.607)	(1.928)	(2.198)
6) Thiessen segment fixed effects (Benchmark)			
Homeland	-1.987	**-3.862****	-2.741
	(1.625)	**(1.955)**	(2.323)
Observations	20630	14614	4287

Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
The completion rate is a 3,4,5 year average completion rate, dropping schools without the required number of years below the highest grade.
Coefficients reflect the school level completion rate, weighting schools of different sizes equally.
Municipality level clustering.
Homelands is a dummy variable of the lands of the former homelands within the given bandwidth
Specifications can be found in Table 6.9.
The bandwidth is the sampling distance on either side of the former homeland border.
Regressions 4,5,6 include location polynomials to control for spatial autocorrelation.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.

Table 4.9, specification 6 at a 10km bandwidth, shows that a 3,86% lower school level completion rate can be attributed to the homelands. This is significant at the 5% level, with all but two specifications of similar magnitude and significant at least at the 10% level. This robustly demonstrates the lower educational quality within the homelands, as supported by Figure 4.6.

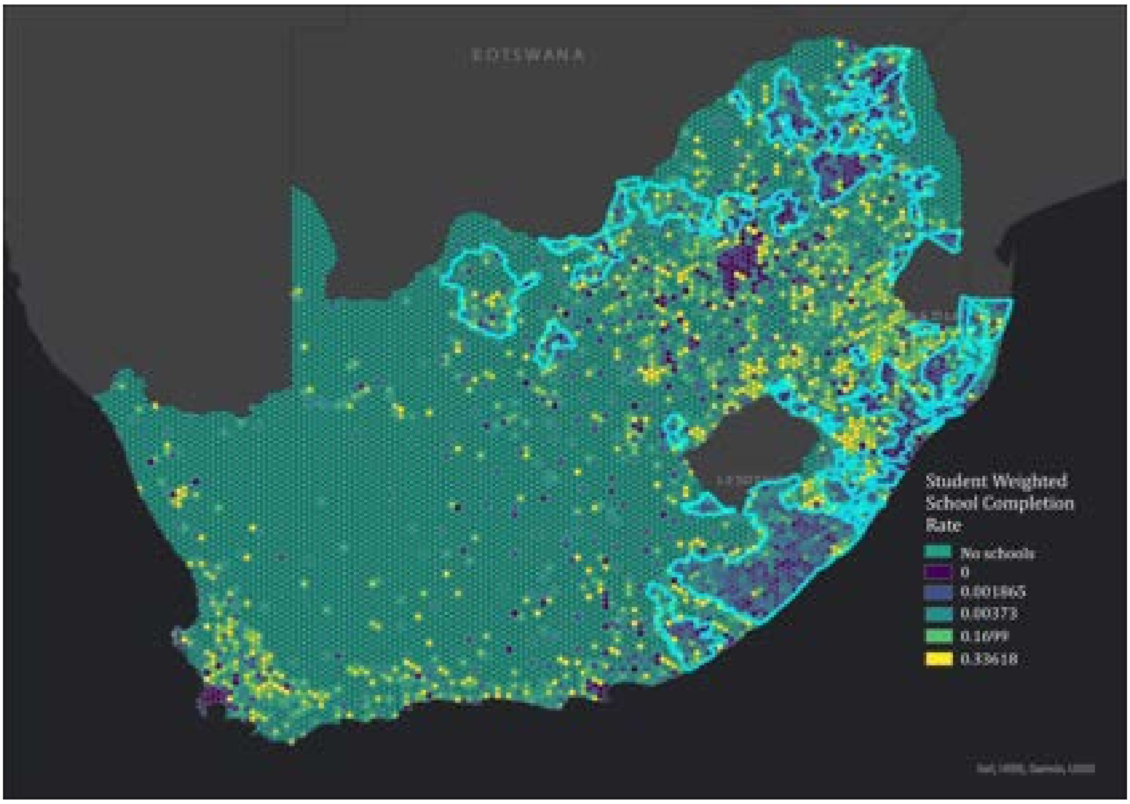

Figure 4.6: Hexagonal heat map of the student weighted school completion rate. Source: Author

The sign and significance of these results is supported even further, typically to the 1% level, by the student weighted 2016 school completion rate, as per Table 4.10.

Table 4.10: **2016 School Completion Rate (Student Weighted)**

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homelands	-0.000412***	-0.000535***	-0.000483***
	(0.000146)	(0.000143)	(0.000138)
2) Interacted linear polynomial in distance to boundary			
Homelands	-0.000536***	-0.000408**	-0.000507***
	(0.000197)	(0.000163)	(0.000176)
2) Quadratic polynomial in distance to boundary			
Homelands	-0.000431***	-0.000567***	-0.000512***
	(0.000148)	(0.000149)	(0.000140)
3) Ordinary least squares			
Homelands	-0.000428***	-0.000541***	-0.000469***
	(0.000146)	(0.000153)	(0.000136)
4) Quadratic polynomials			
Homelands	-0.000393***	-0.000501***	-0.000395***
	(0.000141)	(0.000146)	(0.000143)
5) Interacted quadratic polynomials in lat and long			
Homelands	-0.000394***	-0.000501***	-0.000391***
	(0.000141)	(0.000146)	(0.000145)
6) Thiessen segment fixed effects (Benchmark)			
Homelands	-0.000314**	-0.000396***	-0.000335***
	(0.000134)	(0.000124)	(0.000120)
7) Log-linear with Thiessen segment fixed effects			
Homelands	0.0966**	0.0544	0.0188
	(0.0417)	(0.0448)	(0.0452)
Observations	7303	10630	14614

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The homeland variable is a dummy of whether the school lies within the *former* homeland lands.
The bandwidth is the sampling distance on either side of the former homeland border
The segmented Theissen fixed effects segment the homeland borders into 50 segments.

Finally, this result is visualised in a regression discontinuity (RD) plot (which does not account for the covariates), Figure 4.7.

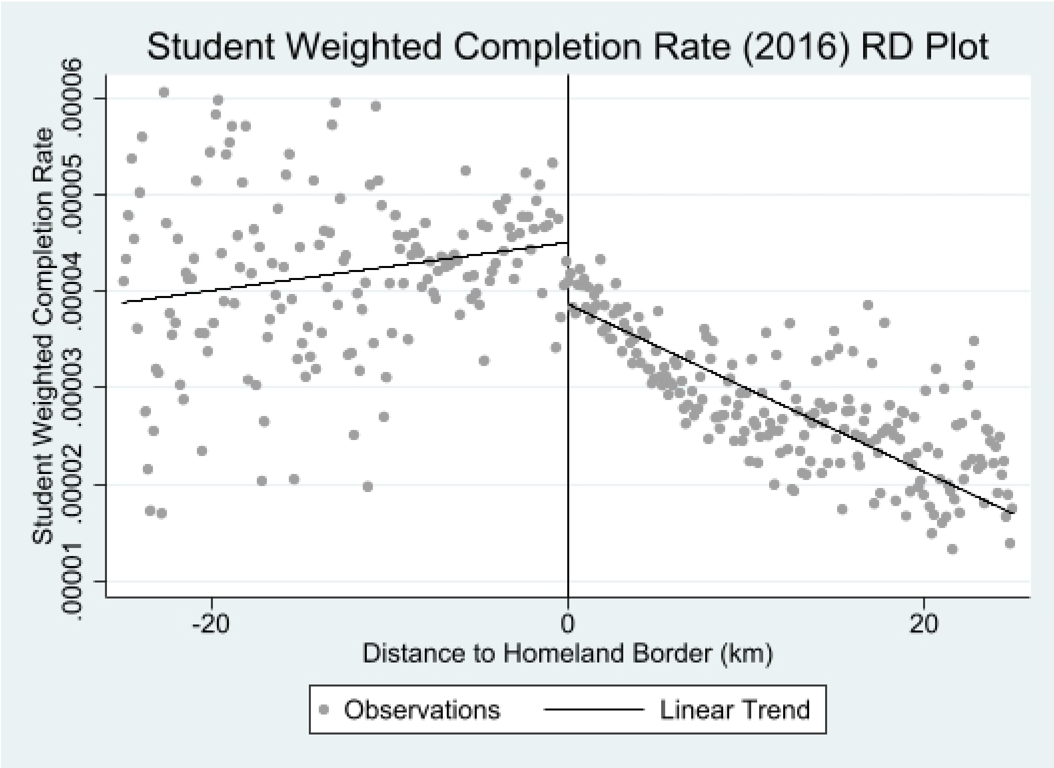

Figure 4.7: Student weighted completion rate (2016) RD Plot. 0 is set to the homelands' borders. Distance is multiplied by -1 if the observation lies outside a homeland.

Turning to the 2002 results (Table 4.11), specification 6 at a 10km bandwidth, shows that a 4.112% lower school level completion rate can be attributed to the homelands, but only at the 10% level. Further, results are not significant at the 5km bandwidth. This is perhaps due to the lack of schools immediately outside the homelands in the smaller 2002 sample.

Table 4.11: **2002 Average School Completion Rate**

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homeland	-7.429***	-4.898**	-3.070
	(2.434)	(1.940)	(2.365)
2) Quadratic polynomial in distance to boundary			
Homeland	-7.135***	-4.563**	-2.794
	(2.290)	(2.018)	(2.439)
3) Ordinary least squares			
Homeland	-7.373***	-4.527**	-2.866
	(2.550)	(2.010)	(2.367)
4) Quadratic polynomials			
Homeland	-6.498***	-4.527**	-2.689
	(2.215)	(1.923)	(2.404)
5) Interacted quadratic polynomials in lat and long			
Homeland	-6.572***	-4.553**	-2.732
	(2.238)	(1.922)	(2.391)
6) Thiessen segment fixed effects (Benchmark			
Homeland	-5.181**	**-4.112***	-1.445
	(2.423)	**(2.136)**	(2.607)
Observations	11863	8505	5834

Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
The completion rate is a 3,4,5 year average completion rate, dropping schools without the required number of years below the highest grade.
Coefficients reflect the school level completion rate, weighting schools of different sizes equally.
Municipality level clustering.
Homelands is a dummy variable of the lands of the former homelands within the given bandwidth
Specifications can be found in Table 6.9.
The bandwidth is the sampling distance on either side of the former homeland border.
Regressions 4,5,6 include location polynomials to control for spatial autocorrelation.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.
Only schools with an EMIS in 2016 are included

Turning to the student weighted 2002 results (in the Appendix, Table 6.7), significance is lost for most specifications. At the 5km bandwidth, several specifications are significant at the 5% level. However, as there is no family-wise error rate correction (such as a Bonferroni correction), the robustness of these tables is supported by the predominance of significance, rather its occurrence. As such, these results provide little support for the school level completion rate results for 2002.

The lack of significance for the 2002 results with high significance for the 2016 results may be driven by the dropped observations (those in the 2002 sample but not in the 2016 results). Otherwise, this may be occurring due to post-Apartheid structures which correspond to the homelands, namely traditional leadership structures, such as the Ingonyama Trust in KwaZulu-Natal. These structures may have further eroded the educational attainment in these areas, as exampled by the village education project in Maputaland closed by the traditional authority, as per Chapter 2. However, the descriptive statistics do not support this latter explanation. Finally, turning to the TBVC (plus KwaZulu) decomposition, the only statistically significant results are for KwaZulu. This is perhaps driven by the KwaZulu homeland being the closest homeland to urban areas (namely Durban). However, as rurality is controlled for, the weight of a rural school and an urban school are not equal. Again, here the influence of the Ingonyama Trust must be considered.

Table 4.12: **2016 School Completion Rate TBVCZ Decomposition**

	Transkei		Bophuthatswana		Venda		Ciskei		KwaZulu	
Bandwidth:	25km	10km	25km	10km	25km	10km	25km	10km	25km	10km
1) Linear polynomial in distance to boundary										
Homelands	5.138	3.850	1.629	0.972	-13.77	-20.60	3.229	0.201	-5.357***	-7.021**
	(4.706)	(6.091)	(4.269)	(4.859)	(9.718)	(13.71)	(3.592)	(2.622)	(2.048)	(2.814)
2) Interacted linear polynomial in distance to boundary										
Homelands	-0.170	6.504	10.87	-7.195	-24.67	-53.72	-1.777	-3.758	-7.466**	-3.343
	(7.487)	(10.93)	(7.866)	(5.981)	(16.31)	(35.58)	(2.788)	(2.328)	(3.722)	(3.254)
3) Quadratic polynomial in distance to boundary										
Homelands	5.279	1.287	3.004	0.238	-11.21	-13.85	4.034	0.633	-5.897***	-7.439**
	(4.599)	(5.732)	(4.406)	(4.867)	(9.933)	(14.93)	(2.860)	(2.661)	(2.161)	(2.973)
4) Ordinary least squares										
Homelands	5.264	1.307	-2.966	0.444	-8.789	-16.32	2.434	0.453	-6.881***	-7.544***
	(4.604)	(5.671)	(4.497)	(4.846)	(8.028)	(11.69)	(3.386)	(2.557)	(1.969)	(2.880)
5) Quadratic polynomials										
Homelands	5.944	3.455	3.814	0.807	-15.36	-20.70	1.735	-0.197	-4.563**	-6.207**
	(5.003)	(6.228)	(4.697)	(4.953)	(9.745)	(13.55)	(3.485)	(1.833)	(2.023)	(2.770)
6) Interacted quadratic polynomials in lat and long										
Homelands	7.135	4.545	3.748	0.123	-14.94	-23.99*	3.017	1.194	-4.143**	-5.729**
	(5.649)	(6.593)	(4.887)	(5.027)	(12.71)	(14.26)	(2.789)	(1.795)	(1.983)	(2.719)
7) Location polynomials with Thiessen segment fixed effects										
Homelands	-0.415	-1.252	4.842	2.198	-16.30	-22.35	1.696	0.358	-3.977*	-4.846*
	(5.740)	(6.549)	(4.566)	(4.760)	(10.38)	(14.90)	(3.589)	(1.986)	(2.271)	(2.737)
Observations	2458	1121	1754	1268	736	484	953	603	5385	4441

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The homeland variable is a dummy of whether the school lies within the *former* homeland lands.
The bandwidth is the sampling distance on either side of the former homeland border
The segmented Theissen fixed effects segment the homeland borders into 50 segments

4.4.6 Conclusion

These results and visualisations robustly demonstrate how the former homelands continue to define a pattern of education inequality in South Africa. However, this deficit approach must be paired with embracing the capacities of these communities, as suggested by the epigraph to this section. There are four primary areas in need of further research on this topic. First, a difference-in-difference analysis of the change since 2002. Second, a gender decomposition. Third, an accessibility analysis by grade. Finally a robustness check with more robust educational attainment data (not publicly available).

Although I find that accessibility may be improved, with more schools per square kilometre, this comes with important caveats including the number of grades per school, transport infrastructure, and efficiencies of scale. The average classroom size, or students per teacher, is significantly larger in the former homelands. Likewise, the school completion rate is significantly lower in the former homelands. 2016 is 22 years after these borders ceased to exist. That being 5km on the wrong side of a nonexistent border can impact whether your child will pass or fail is a tragedy with which democratic South Africa must contend.

CHAPTER 5

CONCLUSION

This thesis has robustly shown that a spatial pattern of deprivation, determined by the Apartheid regime, has persisted into democracy. Those who live in the densely populated rural former homelands continue to endure worse welfare conditions than those immediately outside the former homelands. With 29.5% of South African's currently living in the former homelands, the implications of this for the welfare of the nation as a whole are substantial.

This research has shown that rural livelihoods are greatly affected by living just on the wrong side of a now non-existent border, a former homeland border. Likely a major cause of the relative deprivation of these areas is their extraordinarily high rural population density, which has persisted since Apartheid. Here, I find that the homelands have induced a doubling of the population density in contemporary South Africa compared to the areas immediately outside the homelands. Yet, the subsequent results found preclude population density as the cause of the agricultural and educational deprivations estimated, as population density is controlled for. It is the cluster of oppressive homeland legislation and administration that bears the blame for the magnitude of the harms identified in this thesis. As illustrated in Chapter 2, the causes of these harms include the limited size of the homelands, denaturalization, the migrant labour system, parent absenteeism, Apartheid rural policy (including 'influx control' and 'Betterment'), 'Native Law', Bantu education, and property dispossession.

Subsistence agriculture is vital for the welfare of many rural South Africans. Yet, the viability of farming in the homelands has been curtailed by high population density and the reduction in topsoil nutrients caused by the homelands. Here, I find that the homelands have caused at least a 2.7% reduction in topsoil nitrogen levels. This combined with the topsoil in the homelands containing 17.87% less organic carbon than the country as a whole, shows that at the margin, there are many people for whom farming cannot be a viable source of welfare.

Some of the most poignant results have been found and visualised for education. Education attainment is greatly curtailed by the homelands—the homeland schools experience a 2.19% reduction in the school completion rate. Likewise, education inputs are significantly reduced by the homelands, with 7.66% more students per teacher. Finally, the homelands have caused a doubling of the number of schools per square kilometre. However, the latter result does not necessarily entail greater school access, for reasons explored in Chapter 4.

The education section provides multiple visualisations for the education patterns identified. The maps provide an opportunity to perceive the pattern of deprivation caused by the homelands without any econometric knowledge. It is hoped that through publication, these results can raise awareness for the particular plight of the residents of the former homelands.[1]

Backward looking considerations are most important when they tell us something about the state of livelihoods today. This research has explicitly drawn the link between Apartheid oppression and current welfare. Yet, this research is also fundamentally forward looking. The aspects of Apartheid induced deprivation that have not persisted do not affect the estimated results. As such, the link with the past is illustrative only insofar as it describes the nature of current deprivation. It is current deprivation that must be the focus of Apartheid redress. Most current Apartheid redress measures are based on race, not welfare.[2] As such, more than two decades after the end of apartheid, the welfare of contemporary South Africans remains significantly reduced by living just 5km on the wrong side of a now non-existent homeland border. Former homeland specific policy is required to contend with this injustice.

[1]To briefly editorialise, I have shown these maps to many South Africans, from all walks of life. My childhood domestic worker became emotional upon seeing these maps and seeing so graphically for the first time why her children dropped out of school, and why it was not their fault. I have shown it to friends from Gazankulu, who greatly appreciated seeing why their mother worked away from home, in Johannesburg, and how universal that experience is for South Africans. Although the costs of Apartheid are understood by most, it is novel for many to see the geographic pattern and degree of persistence of these harms.

[2]For example, Broad-Based Black Economic Empowerment, the government's cornerstone Apartheid redress policy.

Bibliography

Abadie, Alberto et al. (2017). *When should you adjust standard errors for clustering?* Tech. rep. National Bureau of Economic Research.

Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). "The colonial origins of comparative development: An empirical investigation". In: *American economic review* 91.5, pp. 1369–1401.

– (2002). "Reversal of fortune: Geography and institutions in the making of the modern world income distribution". In: *The Quarterly journal of economics* 117.4, pp. 1231–1294.

Alix-Garcia, Jennifer et al. (2013). "The ecological footprint of poverty alleviation: evidence from Mexico's Oportunidades program". In: *Review of Economics and Statistics* 95.2, pp. 417–435.

Ally, Shireen and Arianna Lissoni (2017). *New Histories of South Africa's Apartheid-era Bantustans*. Taylor & Francis.

Ardington, Cally, Anne Case, and Victoria Hosegood (2009). "Labor supply responses to large social transfers: Longitudinal evidence from South Africa". In: *American economic journal: Applied economics* 1.1, pp. 22–48.

Asher, Sam, Teevrat Garg, and Paul Novosad (2020). "The ecological impact of transportation infrastructure". In: *The Economic Journal* 130.629, pp. 1173–1199.

Basic Education, Department of (2005). *Report of the ministerial committee, on rural education: A new vision for rural schooling.*

– (2016). *Snap Survey of Ordinary Schools 1997-2016 [dataset]*. Version 1.6. DOI: `https://doi.org/10.25828/2e3e9x02`. URL: `https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/689/get-microdata`.

Batjes, Niels H et al. (2017). "WoSIS: providing standardised soil profile data for the world". In: *Earth System Science Data* 9.1, pp. 1–14.

Beinart, W. and C. Bundy (1980). "State intervention and rural resistance: the Transkei, 1900-1965". In: *Peasants in Africa: Historical and Contemporary Perspective*. Ed. by M Klein. Sage, pp. 271–315.

Biko, Steve (1978). *I write what I like: selected writings*. University of Chicago Press.

Bundy, Colin (1979). *The rise and fall of the South African peasantry*. Vol. 28. Univ of California Press.

Butler, Jeffrey, Robert I Rotberg, and John Adams (1978). *The black homelands of South Africa: The political and economic development of Bophuthatswana and Kwa-Zulu*. Vol. 396. Univ of California Press.

Cagé, Julia and Valeria Rueda (2016). “The long-term effects of the printing press in sub-Saharan Africa”. In: *American Economic Journal: Applied Economics* 8.3, pp. 69–99.

Chimere-Dan, Orieji (1992). “Apartheid and demography in South Africa”. In: *African Population Studies* 1992.7, pp. 28–38.

Chong, Alberto and Cesar Calderon (2000). “Institutional quality and poverty measures in a cross-section of countries”. In: *Economics of Governance* 1.2, pp. 123–135.

Collinson, M and Pn Kok (2006). *Migration and urbanization in South Africa. Report 03-04-02.*

Conley, Timothy G (1999). “GMM estimation with cross sectional dependence”. In: *Journal of econometrics* 92.1, pp. 1–45.

Cooper, Carole et al. (1994). *Race Relations Survey.* Johannesburg: South African Institute of Race Relations.

Crush, J, A Jeeves, and D Yudelman (1991). *South Africa's Labour Empire:A History of Black Migrancy to the Gold Mines.*

DBE (2016). *National Oridnary Schools EMIS data.* URL: `https://www.education.gov.za/Programmes/EMIS/EMISDownloads.aspx`.

De Wet, Christopher J (1995). *Moving together, drifting apart: Betterment planning and villagisation in a South African homeland.* Witwatersrand University Press.

Dell, Melissa (2010). “The Persistent Effects of Peru's Mining Mita”. In: *Econometrica* 78.6, pp. 1863–1903.

Dell, Melissa and Benjamin A Olken (2020). “The development effects of the extractive colonial economy: The dutch cultivation system in java”. In: *The Review of Economic Studies* 87.1, pp. 164–203.

Donaldson, Dave (2018). “Railroads of the Raj: Estimating the impact of transportation infrastructure”. In: *American Economic Review* 108.4-5, pp. 899–934.

Elder, Glen S (2003). “Malevolent traditions: Hostel violence and the procreational geography of apartheid”. In: *Journal of Southern African Studies* 29.4, pp. 921–935.

Fick, S.E. and R.J. Hijmans (2017). “WorldClim 2: new 1km spatial resolution climate surfaces for global land areas”. In: *International Journal of Climatology* 37.12, pp. 4302–4315.

Firth, Adrian (2013). URL: `https://adrian.frith.dev/images/dotmaps/Durban-2011.png`.

Geldenhuys, Deon (1981). “South Africa's black homelands: past objectives, present realities and future developments”. In: South African Institute of International Affairs (SAIIA).

Gelman, Andrew and Guido Imbens (2019). “Why high-order polynomials should not be used in regression discontinuity designs”. In: *Journal of Business & Economic Statistics* 37.3, pp. 447–456.

Gilbert, Marius et al. (2018). "Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010". In: *Scientific data* 5.1, pp. 1–11.

Giliomee, Hermann (2009). "A note on Bantu education, 1953 to 1970". In: *South African Journal of Economics* 77.1, pp. 190–198.

Government, Union (1951). *Report of the Commission on Native Education (Eislesn Commission).*

Hall, Katharine and Winnie Sambu (2019). "Demography of South Africa's children". In: *South African child gauge*, pp. 216–220.

Hendricks, Fred T et al. (1990). *The pillars of apartheid: land tenure, rural planning and the chieftancy.* 32.

Hengl, Tomislav et al. (2017). "Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning". In: *Nutrient Cycling in Agroecosystems* 109.1, pp. 77–102.

Hill, Christopher R (1964). "Bantustans: the fragmentation of South Africa". In:

History, SA (2011). URL: `https://www.sahistory.org.za/article/pass-laws-south-africa-1800-1994`.

Horrell, Muriel (1969). *Bantu education to 1968*. South African Institute of Race Relations.

Houghton, D Hobart (1987). "The Land Question in South Africa: A preliminary assessment". In: *The South African Journal of Economics* 55.1.

Huillery, Elise (2009). "History matters: The long-term impact of colonial public investments in French West Africa". In: *American economic journal: applied economics* 1.2, pp. 176–215.

Jedwab, Remi, Edward Kerby, and Alexander Moradi (2017). "History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya". In: *The Economic Journal* 127.603, pp. 1467–1494.

Keele, Luke J and Rocio Titiunik (2015). "Geographic boundaries as regression discontinuities". In: *Political Analysis* 23.1, pp. 127–155.

Kenney, Henry (2016). *Verwoerd: Architect of apartheid.* Jonathan Ball Publishers.

Kepe, Thembela and Ruth Hall (2018). "Land redistribution in South Africa: Towards decolonisation or recolonisation?" In: *Politikon* 45.1, pp. 128–137.

Lahiff, Edward P (1997). *Agriculture and rural livelihoods in a South African 'homeland': A case study from Venda (PhD Thesis).* University of London, School of Oriental and African Studies (United Kingdom).

Letsoalo, EM and Christian M Rogerson (1982). "Rural "development" planning under apartheid: Betterment planning in Lebowa, South Africa". In: *Geoforum* 13.4, pp. 301–314.

Levin, Richard and Daniel Weiner (1991). "The Agrarian Question and the Emergence of Conflicting Agricultural Strategies in South Africa". In: *Towards a New Agrarian Democratic Order*. Ed. by S. Matlhape and A. Munz. SAERT.

– (1997). *No more tears: Struggles for land in Mpumalanga, South Africa.* Africa World Press.

Malinda, Khuthadzo (2015). URL: `http://daffarcgis.nda.agric.za/portal/home/item.html?id=a1f33777bfca47d8a2424456e566a269`.

Mamdani, Mahmood (1996). *Citizen and subject: contemporary Africa and the legacy of late colonialism.* Princeton University Press.

Mazibuko, Sibonginkosi (2014). "Rural governance in South Africa: Is there a place for neo-feudalism in a democracy?" In: *Mediterranean Journal of Social Sciences* 5.20, pp. 2455–2455.

Mbeki, Govan (1964). *South Africa: the peasants' revolt.* Penguin Books.

Molteno, Frank (1977). "The historical significance of the Bantustan strategy". In: *Social Dynamics* 3.2, pp. 15–33.

Nattrass, Jill (1976). "Migrant labour and South African economic development". In: *South African Journal of Economics* 44.1, pp. 65–83.

Nattrass, Nicoli and Jill Nattrass (1990). "South Africa, the homelands and rural development". In: *Development Southern Africa* 7.1, pp. 517–534.

Nelson Mandela Foundation, the (2005). *Emerging voices: A report on education in South African rural communities.* HSRC Press.

Norman, Nick (2006). *Geological journeys : a traveller's guide to South Africa's rocks and landforms.* Cape Town London: Struik New Holland distributor. ISBN: 9781770070622.

Pareto_Software (2021). URL: `https://simplemaps.com/data/za-cities`.

Platzky, Laurine and Cherryl Walker (1985). *The surplus people: Forced removals in South Africa.* Ravan Pressof South Africa.

RCMRD (2015). URL: `http://geoportal.rcmrd.org/layers/servir%5C%3Asouth_africa_srmt30meters`.

Robinson, James A and Daron Acemoglu (2012). *Why nations fail: The origins of power, prosperity and poverty.* Profile London.

Rogan, Michael, Likani Lebani, and Nompumelelo Nzimande (2009). "Internal migration and poverty in KwaZulu-Natal: Findings from censuses, labour force surveys and panel data". In:

ROSEA, UN OCHA (2018). URL: `https://data.world/ocha-rosea/061d4492-56e8-458c-a3fb-e7950991adf0`.

SAIRR (1942). *South African Institute of Race Relations' Fourteenth annual report.*

Sepuru, Terrence Koena and Timothy Dube (2018). "Understanding the spatial distribution of eroded areas in the former rural homelands of South Africa: Comparative evidence from two new non-commercial multispectral sensors". In: *International journal of applied earth observation and geoinformation* 69, pp. 119–132.

Seutloali, Khoboso E, Timothy Dube, and Onisimo Mutanga (2017). "Assessing and mapping the severity of soil erosion using the 30-m Landsat multispectral satellite data in the former South African homelands of Transkei". In: *Physics and Chemistry of the Earth, Parts A/B/C* 100, pp. 296–304.

South Africa, The Government of (1908). *The Framework of Union. A Comparison of the Constitutions of the United States, Canada, Australia, Switzerland and Germany.* Central News Agency.

South Africa, Union of (1922). *Report of the Interdepartmental Committee on the Native Pass Laws, 1920.*

Southern Africa, Development Bank of (1987). *Statistical abstracts on self-governing territories.*

Spaull, Nic (2019). "Equity: A Price Too High to Pay?" In: *South African Schooling: The Enigma of Inequality.* Ed. by Nic Spaull and Jonathan Jansen. Springer Nature Switzerland. Chap. 1, pp. 1–24.

Stuart, James (2013). *A History of the Zulu Rebellion 1906: And of Dinuzulu's Arrest, Trial and Expatriation.* Cambridge University Press.

Tapson, D (1985). "The agricultural development of the homelands: Problems and prospects". In: *Up against the fences: poverty, passes and privilege in South Africa*. Ed. by L Giliomee F & Schlemmer. St. Martin's Press.

Tatem, Andy (2015). *South Africa - Population density (2015).* URL: `https://energydata.info/dataset/south-africa-population-density-2015`.

Thompson, Leonard (2008). *A history of South Africa.* Yale University Press.

Tobler, Waldo R (1970). "A computer movie simulating urban growth in the Detroit region". In: *Economic geography* 46.sup1, pp. 234–240.

Valencia Caicedo, Felipe (2019). "The mission: Human capital transmission, economic persistence, and culture in South America". In: *The Quarterly Journal of Economics* 134.1, pp. 507–556.

Van der Horst, Sheila T (1971). *Native Labour in South Africa.* 4. London: Cass.

van Zyl, Johan and Johan van Rooyen (1991). "Agricultural Production in South Africa". In: *A Harvest of Discontent: The Land Question in South Africa.* Ed. by M. de Klerk. Institute for a Democratic Alternative for South Africa.

Vosloo, Christo (2020). “Extreme apartheid: the South African system of migrant labour and its hostels”. In: *Image & Text* 34, pp. 1–33.

Wellings, Paul and Anthony Black (1986). “Industrial decentralization under apartheid: The relocation of industry to the South African periphery”. In: *World Development* 14.1, pp. 1–38.

Welsh, David (1973). “The Roots of Segregation: Native Policy in Colonial Natal”. In: *The annals of the American Academy of Political and Social Science* 409.1, pp. 180–181.

Wessels, Konrad J et al. (2004). “Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1 km AVHRR NDVI time-series”. In: *Remote sensing of environment* 91.1, pp. 47–67.

Wolpe, Harold (1972). “Capitalism and cheap labour-power in South Africa: from segregation to apartheid”. In: *Economy and society* 1.4, pp. 425–456.

Worden, Nigel (2011). *The making of modern South Africa: Conquest, apartheid, democracy.* John Wiley & Sons.

Yawitch, Joanne (1982). *Betterment: The myth of homeland agriculture.* Vol. 852. 37. South African Institute of Race Relations.

Zmora-Nahum, Sharon et al. (2005). “Dissolved organic carbon (DOC) as a parameter of compost maturity”. In: *Soil Biology and Biochemistry* 37.11, pp. 2109–2116.

CHAPTER 6

APPENDIX

6.1 ONLINE APPENDIX

https://drive.google.com/drive/folders/17bv3bGrrrDtO3WageJmWWeqFx5Z5ZBkc?usp=sharing

6.2 LaTeXFILE

https://www.overleaf.com/read/fqwrqxjjrvqw

6.3 CHAPTER 1 APPENDIX

page left intentionally blank

6.4 CHAPTER 2 APPENDIX

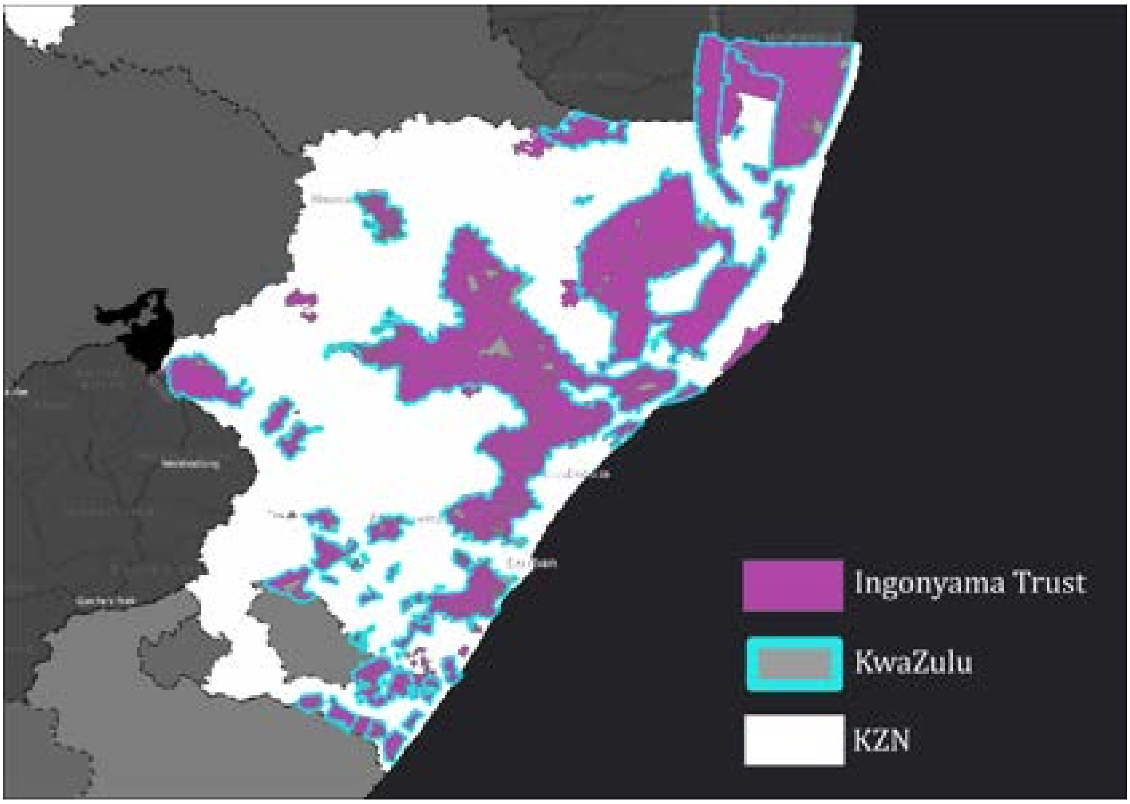

Figure 6.1: The current extent of the Ingonyama Trust

6.5 CHAPTER 3 APPENDIX

Table 6.1: Regression discontinuity variable descriptions

Variable name	Description	Unit	Source	Specification
OID	Object Identification Number	n/a	GIS calculation	n/a
km^2	Area of the observation	sqkm	GIS calculation	n/a
dhome	Nearest distance to a homeland border	km	GIS calculation	1,2,3,4,5,6,7,8
dhome2	dhome squared	n/a	Auxiliary calculation	3
dhomexhome	dhome interacted with home	n/a	Auxiliary calculation	2,
dhomexhome2	dhome2 interacted with home	n/a	Auxiliary calculation	n/a
dcity	Nearest distance to a city (in the top 20 most populated cities)	km	(Pareto_Software, 2021)	1,2,3,4,5,6,7,8
home	Homeland Dummy: 1= homeland 0=outside homeland	binary	(Malinda, 2015)	1,2,3,4,5,6,7,8
area	Categorical variable for each homeland and province	factor: 1-19	(ROSEA, 2018)	n/a
elev	Elevation	m	(RCMRD, 2015)	1,2,3,4,5,6,7,8
slope	Slope	degrees	GIS calculation	1,2,3,4,5,6,7,8
pop	Sum of population in observation	integer	(Tatem, 2015)	n/a
rain	Average Rainfall	mm	(Fick and Hijmans, 2017)	1,2,3,4,5,6,7,8
nitr	Average Topsoil Nitrogen Content	cg/kg	(Hengl et al., 2017)	Table 4.2
muni	Categorical variable for municipalities	Factor: 1-212	(ROSEA, 2018)	Clustering
pop_dens	Population density (pop/sqkm)	integer	Auxiliary calculation	1,2,3,4,5,6,7,8
lon	Longitude	decimal degrees	GIS calculation	1,5,6,7,8
lat	latitude	decimal degrees	GIS calculation	1,5,6,7,8
lon2	Longitude squared	n/a	Auxiliary calculation	1,6,7
lat2	Latitude squared	n/a	Auxiliary calculation	1,6,7
lonxlat	longitude and latitude interaction	n/a	Auxiliary calculation	7
lon2xlat2	lonigitude suqared and latitude squared interaction	n/a	Auxiliary calculation	7
seg	Categorical variable for each Thiessen polygon	factor:1-50	GIS Calculation	Clustering
seg_home	Homeland assigned Thiessen polygons	factor: 10-19	GIS calculation	Clustering
c	Running variable (dhome*-1 if outside homelands, dhome if within)	n/a	Auxiliary calculation	Clustering
homexseg	seg interacted with home	n/a	Auxiliary calculation	Clustering
homexseg_home	home interacted with seg_home	n/a	Auxiliary calculation	Clustering

1882 MELISSA DELL

TABLE III

SPECIFICATION TESTS[a]

	Dependent Variable						
	Log Equiv. Household Consumption (2001)			Stunted Growth, Children 6–9 (2005)			
Sample Within:	< 100 km of Bound. (1)	< 75 km of Bound. (2)	< 50 km of Bound. (3)	< 100 km of Bound. (4)	< 75 km of Bound. (5)	< 50 km of Bound. (6)	Border District (7)
Alternative Functional Forms for RD Polynomial: Baseline I							
Linear polynomial in latitude and longitude							
Mita	−0.294***	−0.199	−0.143	0.064***	0.054**	0.062**	0.068**
	(0.092)	(0.126)	(0.128)	(0.021)	(0.022)	(0.026)	(0.031)
Quadratic polynomial in latitude and longitude							
Mita	−0.151	−0.247	−0.361	0.073*	0.091**	0.106**	0.087**
	(0.189)	(0.209)	(0.216)	(0.040)	(0.043)	(0.047)	(0.041)
Quartic polynomial in latitude and longitude							
Mita	−0.392*	−0.324	−0.342	0.073	0.072	0.057	0.104**
	(0.225)	(0.231)	(0.260)	(0.054)	(0.050)	(0.048)	(0.042)
Alternative Functional Forms for RD Polynomial: Baseline II							
Linear polynomial in distance to Potosí							
Mita	−0.297***	−0.273***	−0.220**	0.050**	0.048**	0.049**	0.071**
	(0.079)	(0.093)	(0.092)	(0.022)	(0.022)	(0.024)	(0.031)
Quadratic polynomial in distance to Potosí							
Mita	−0.345***	−0.262***	−0.309***	0.072***	0.064***	0.072***	0.060*
	(0.086)	(0.095)	(0.108)	(0.023)	(0.022)	(0.023)	(0.032)
Quartic polynomial in distance to Potosí							
Mita	−0.331***	−0.310***	−0.330***	0.078***	0.075***	0.071***	0.053*
	(0.086)	(0.100)	(0.097)	(0.021)	(0.020)	(0.021)	(0.031)
Interacted linear polynomial in distance to Potosí							
Mita	−0.307***	−0.280***	−0.227**	0.051**	0.048**	0.043*	0.076***
	(0.092)	(0.094)	(0.095)	(0.022)	(0.021)	(0.022)	(0.029)
Interacted quadratic polynomial in distance to Potosí							
Mita	−0.264***	−0.177*	−0.285**	0.033	0.027	0.039*	0.036
	(0.087)	(0.096)	(0.111)	(0.024)	(0.023)	(0.023)	(0.024)
Alternative Functional Forms for RD Polynomial: Baseline III							
Linear polynomial in distance to mita boundary							
Mita	−0.299***	−0.227**	−0.223**	0.072***	0.060***	0.058**	0.056*
	(0.082)	(0.089)	(0.091)	(0.024)	(0.022)	(0.023)	(0.032)
Quadratic polynomial in distance to mita boundary							
Mita	−0.277***	−0.227**	−0.224**	0.072***	0.060***	0.061***	0.056*
	(0.078)	(0.089)	(0.092)	(0.023)	(0.022)	(0.023)	(0.030)
Quartic polynomial in distance to mita boundary							
Mita	−0.251***	−0.229**	−0.246***	0.073***	0.064***	0.063***	0.055*
	(0.078)	(0.089)	(0.088)	(0.023)	(0.022)	(0.023)	(0.030)
Interacted linear polynomial in distance to mita boundary							
Mita	−0.301*	−0.277	−0.385*	0.082	0.087	0.095	0.132**
	(0.174)	(0.190)	(0.210)	(0.054)	(0.055)	(0.065)	(0.053)
Interacted quadratic polynomial in distance to mita boundary							
Mita	−0.351	−0.505	−0.295	0.140*	0.132	0.136	0.121*
	(0.260)	(0.319)	(0.366)	(0.082)	(0.084)	(0.086)	(0.064)
Ordinary Least Squares							
Mita	−0.294***	−0.288***	−0.227**	0.057**	0.048*	0.049*	0.055*
	(0.083)	(0.089)	(0.090)	(0.025)	(0.024)	(0.026)	(0.031)
Geo. controls	yes	yes	yes	yes	yes	yes	yes
Boundary F.E.s	yes	yes	yes	yes	yes	yes	yes
Clusters	71	60	52	289	239	185	63
Observations	1478	1161	1013	158,848	115,761	100,446	37,421

(Continues)

TABLE III—Continued

[a] Robust standard errors, adjusted for clustering by district, are in parentheses. All regressions include geographic controls and boundary segment fixed effects (F.E.s). Columns 1–3 include demographic controls for the number of infants, children, and adults in the household. Coefficients significantly different from zero are denoted by the following system: *10%, **5%, and ***1%.

Figure 6.2: (Dell, 2010) regression discontinuity specification table

6.6 CHAPTER 4 APPENDIX

Table 6.2: Topsoil Nitrogen Homeland Segments Fixed Effects

VARIABLES	Homeland Segements Fes
home	-4.137***
	(0.707)
dhome	-0.176
	(0.192)
Bophutatswana	base
Ciskei	-15.43
	(9.690)
Gazunkulu	-12.60**
	(5.342)
KaNgwane	-45.21***
	(8.537)
KwaNdebele	-2.035
	(3.483)
KwaZulu	-38.56***
	(7.469)
Lebowa	-7.098*
	(4.090)
Qwa Qwa	-47.49***
	(8.367)
Transkei	-32.15***
	(8.263)
Venda	-7.658
	(5.654)
Constant	-939.1***
	(55.25)
Observations	2,202

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.9.

Table 6.3: Topsoil Nitrogen RDD with homeland centroid clustered SEs

Specification / Bandwidth:	10km	25km	50km
1) linear polynomial in distance to boundary			
Homelands	-3.234***	-5.185***	-7.624***
	(1.209)	(0.754)	(0.736)
2) Interacted linear polynomial in distance to boundary			
Homelands	-2.042***	-0.473**	-0.290***
	(0.366)	(0.238)	(0.0793)
3) Quadratic polynomial in distance to boundary			
Homelands	-5.102**	-7.197***	-7.339***
	(2.092)	(1.497)	(1.365)
4) Ordinary Least Squares			
Homelands	-4.912**	-7.387***	-7.836***
	(2.122)	(1.517)	(1.320)
5) Linear polynomial in lat and long			
Homelands	-3.626**	-5.503***	-7.673***
	(1.394)	(0.922)	(0.751)
6) Quadratic polynomial in lat and long			
Homelands	-3.234***	-5.185***	-7.624***
	(1.209)	(0.754)	(0.736)
7)Interacted quadratic polynomials in lat and long			
Homelands	-3.219***	-5.640***	-8.076***
	(1.079)	(0.692)	(0.704)
8) Linear with segmented fe			
Homelands	-4.585***	-6.381***	-8.594***
	(0.914)	(0.644)	(0.471)
9) Linear with segmented homeland proxy fixed effects			
Homelands	-4.548***	-7.546***	-9.765***
	(1.306)	(0.816)	(0.649)
10) Linear with interacted homelands segments			
Homelands	6.745	9.466***	9.879***
	(4.156)	(2.851)	(2.625)
Observations	5112	3734	2202

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.4: Topsoil Nitrogen RDD with municipality clustered SEs

Specification / Bandwidth:	10km	25km	50km
1) linear polynomial in distance to boundary			
Homelands	-2.529**	-4.616***	-7.075***
	(1.239)	(1.293)	(1.431)
2) Interacted linear polynomial in distance to boundary			
Homelands	-1.635***	-0.457*	-0.290**
	(0.585)	(0.255)	(0.117)
3) Quadratic polynomial in distance to boundary			
Homelands	-4.625**	-6.592***	-6.790***
	(2.187)	(2.283)	(2.250)
4) Ordinary Least Squares			
Homelands	-4.449**	-6.880***	-7.360***
	(2.178)	(2.248)	(2.312)
5) Linear polynomial in lat and long			
Homelands	-2.879**	-4.989***	-7.133***
	(1.364)	(1.355)	(1.506)
6) Quadratic polynomial in lat and long			
Homelands	-2.529**	-4.616***	-7.075***
	(1.239)	(1.293)	(1.431)
7)Interacted quadratic polynomials in lat and long			
Homelands	-2.672**	-5.187***	-7.634***
	(1.195)	(1.273)	(1.420)
8) Linear with segmented fe			
Homelands	-4.617***	-6.204***	-8.376***
	(1.063)	(1.053)	(1.141)
9) Linear with segmented homeland proxy fixed effects			
Homelands	-3.683***	-6.664***	-8.953***
	(1.246)	(1.315)	(1.417)
10) Linear with interacted homelands segments			
Homelands	7.220	9.754*	10.04**
	(5.116)	(4.972)	(4.832)
Observations	5112	3734	2202

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.5: Topsoil nitrogen 10km RDD quadratic polynomial in distance to boundary full output

nitr	Coef.	Std. Err.	z	P
home	-3.455737	0.7069364	-4.89	0
dhome	-0.2226467	0.1271546	-1.75	0.08
dcity	-0.1626291	0.0320452	-5.07	0
elev	0.0256529	0.0043747	5.86	0
slope	6.18443	0.2601832	23.77	0
pop_dens	0.0075736	0.0024768	3.06	0.002
rain	0.4784244	0.0540275	8.86	0
lon	-235.0598	24.31797	-9.67	0
lat	183.1831	18.18616	10.07	0
lon2	4.413729	0.4405652	10.02	0
cons	5526.865	585.5716	9.44	0

Table 6.6: Schools per square kilometre with municiplaity clustered standard errors

Specification/ Bandwidth	50km	25km	10km
1) Linear polynomial in distance to boundary			
Home	0.0356***	0.0403***	0.0422***
	(0.00258)	(0.00240)	(0.00247)
3) Quadratic polynomial in distance to boundary			
Home	0.0349***	0.0398***	0.0425***
	(0.00258)	(0.00243)	(0.00254)
4) Ordinary least squares			
Home	0.0356***	0.0405***	0.0430***
	(0.00261)	(0.00242)	(0.00249)
6) Quadratic polynomial in lat and long			
Home	0.0359***	0.0405***	0.0420***
	(0.00255)	(0.00242)	(0.00251)
7)Interacted quadratic polynomials in lat and long			
Home	0.0359***	0.0404***	0.0415***
	(0.00255)	(0.00247)	(0.00262)
8) Linear with segmented fe			
Home	0.0373***	0.0413***	0.0416***
	(0.00278)	(0.00265)	(0.00271)
9) Linear with segmented homeland proxy fixed effects			
Home	0.0368***	0.0411***	0.0422***
	(0.00263)	(0.00248)	(0.00262)
10) Linear with interacted homelands segments			
Home	0.0546***	0.0683***	0.0697***
	(0.0113)	(0.0153)	(0.0153)
Observations	4287	3130	1843

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.7: 2002 School Completion Rate (Student Weighted)

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homeland	-0.000190	-0.000249	-0.000439**
	(0.000209)	(0.000183)	(0.000208)
2) Interacted linear polynomial in distance to boundary			
Homeland	-2.18e-05	2.59e-05	0.000182
	(0.000281)	(0.000253)	(0.000330)
3) Quadratic polynomial in distance to boundary			
Homeland	-0.000219	-0.000276*	-0.000457**
	(0.000203)	(0.000163)	(0.000188)
4) Ordinary least squares			
Homeland	-0.000214	-0.000262	-0.000415*
	(0.000201)	(0.000165)	(0.000216)
5) Quadratic polynomials			
Homeland	-0.000136	-0.000177	-0.000307
	(0.000194)	(0.000176)	(0.000193)
6) Interacted quadratic polynomials in lat and long			
Homeland	-0.000140	-0.000178	-0.000311
	(0.000193)	(0.000174)	(0.000195)
7) Location polynomials with Thiessen segment fixed effects			
Homeland	-4.38e-05	-0.000138	-0.000255
	(0.000147)	(0.000156)	(0.000185)
8) Log-linear location polynomials with Thiessen segment fixed effects			
Homelands	.0544347	0.0188	0.0966**
	(0.0448)	(0.0453)	(.0417)
Observations	5834	8505	11863

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The homeland variable is a dummy of whether the school lies within the *former* homeland lands.
The bandwidth is the sampling distance on either side of the former homeland border
The segmented Theissen fixed effects segment the homeland borders into 50 segments
Only schools with an EMIS in 2016 are included

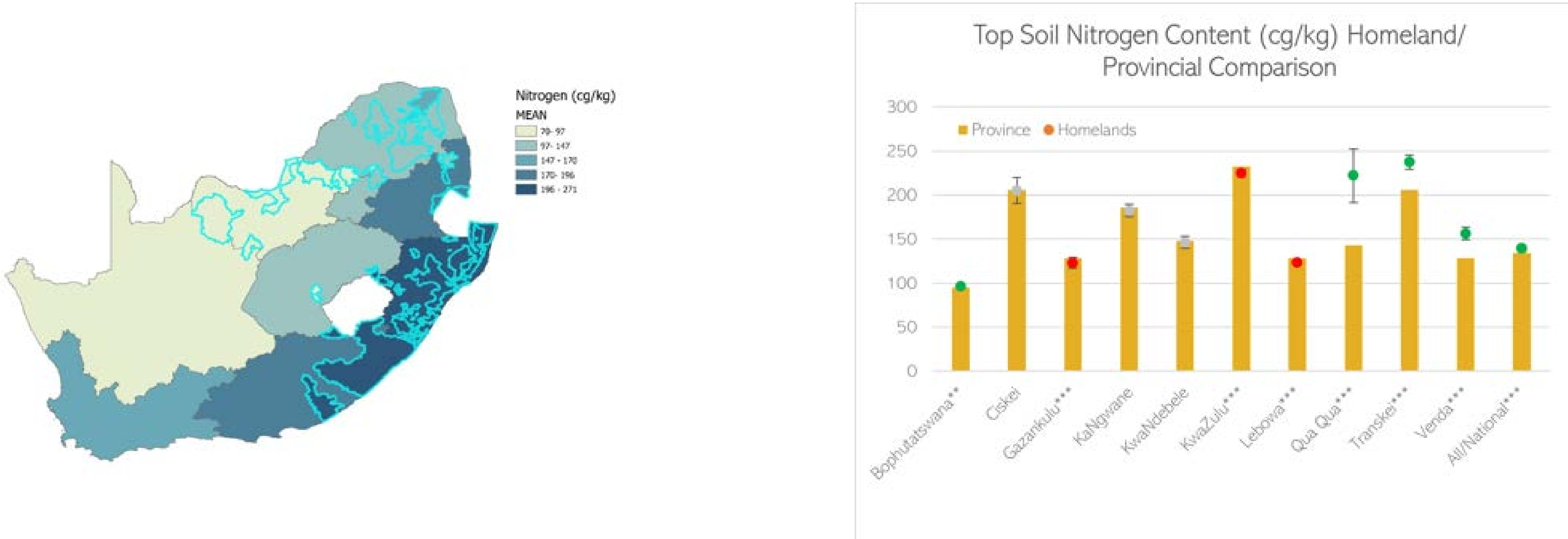

Figure 6.3: Map and histogram of homeland & province topsoil nitrogen (cg/kg) comparison

6.6.1 School completion rate pseudo code

1. Create the primary data set in ArcGIS Pro
   1.1. Import the cleaned EMIS data (DBE, 2016) (located missing GPS coordinates from subsequent datasets).
   1.2. Create points from the GPS coordinates.
   `XY Table to Point`
   1.3. Calculate distance variables.
   `Near`
   1.4. Buffer points to 5km radius (5km chosen as a feasible catchment distance for students to walk to school)
   `Buffer`
   1.5. Erase the areas of the buffers that extend beyond the borders of SA
   `Clip`
   1.6. Use zonal statistics to determine the mean of the raster covariates beneath the buffer (summed in the case of the population point data, divided by buffer area to determine population density (Tatem, 2015).)
   `Zonal Statistics as a Table`
   1.7. Export to Excel and calculate the remaining variables (such as PTR)
   `Table to Excel`
2. Determine the number of students in the highest grade in 2016 (snaps-2016-learner-enrolment-v1.1)
   2.1. Create a new variable of the highest grade per school in 2016
   `bysort natemis(gradecd) : gen grade_max = gradecd[_N]`
   emphnatemis = the school identification number
   2.2. Drop all observations that are not in the highest grade
   `drop if grade_max != gradecd[_N]`
   2.3. Sum the male and female pupils in the highest grade (only remaining variables per natemis)
   `egen sum = total(answer), by natemis`
   2.4. Drop duplicates
   ```
   sort natemis
   quietly by natemis : gen dup = cond(_N==1,0,_n)
   drop if dup > 1
   ```
   2.5. save summed data

`save ‘...highest_grade’`

2.6. Open original data and merge the highest grade `grade_max` and `sum` to the cleaned GIS (primary) data

`merge 1:1 emis ’...highest_grade’`

3. Determine the number of students who were enrolled in the `grade_max`−5 in 2011 (snaps-1997-2013-learner-enrolment-v1.4)

3.1. Drop all variables not in 2011

`drop if datayear != 2011`

3.2. In the primary data set, create a new variable of the grade 5 years below the `grade_max`

gen grade_max−5 = grade_max −5

3.3. Attach the `grade_max`−5 to each emis number along all grades and genders in 2011

`merge 1:1 emis "...highest grade"`

3.4. Drop if the grade is not the grade_max−5

`drop if grade != grade_max−5`

3.5. Sum the male and females in the grade_max−5

`egen sum2 = total(quantity), by natemis`

3.6. Drop duplicates

```
sort natemis
quietly by natemis :  gen dup = cond(_N==1,0,_n)
drop if dup > 1
```

3.7. Assign the primary data set the number of people in grade_max−5 in 2011

`merge 1:1 emis ‘highest_grade’`

4. Calculate the 5 year completion rate

$(sum \div sum2) \times 100$

5. Calculate the remaining completion rates for 2016 and all rates for 2002

6. Create an average completion rate not counting missing observations for each year

```
egen mean_2002_compl_rte = rmean(compl_rte_9702 compl_rte_9802
compl_rte_9902)
egen mean_2016_compl_rte = rmean(compl_rte_1116 compl_rte_1216
compl_rte_1316)
```

Table 6.8: Table of Variables

Variable Name	Description	Unit/ Data Type	Source	Specifications
ptr2016	Students per teacher in 2016	Integer	Auxiliary calculation	Table 4.7
ptr2002	Students per teacher in 2002	Integer	Auxiliary calculation	Table 6.10
mean2002complrte	3,4,&5 year average completion rate in 2002	Decimal	Auxiliary calculation	Table 4.11
mean2016complrte	3,4,&5 year average completion rate in 2016	Decimal	Auxiliary calculation	Table 4.9
mean2002complrte_stdweight	mean2002complrte*stdnt_weight_2002	Decimal	Auxiliary calculation	Table 6.7
mean2016complrte_stdweight	mean2016complrte*stdnt_weight_2016	Decimal	Auxiliary calculation	Table 4.10
lnptr2016	Natural log of ptr2016	Decimal	Auxiliary calculation	Table 4.7
lnptr2002	Natural log of ptr2002	Decimal	Auxiliary calculation	Table 6.10
emis	School identification number	Integer	(DBE, 2016)	n/a
lon	Longitude of school	Decimal degrees	(DBE, 2016)	1,5,6,7
lat	Latitude of school	Decimal degrees	(DBE, 2016)	1,5,6,7
lon2	lon^2	Decimal	Auxiliary calculation	5,6,7
lat2	lat^2	Decimal	Auxiliary calculation	5,6,7
lonxlat	lon*lat	Decimal	Auxiliary calculation	6
lon2xlat2	$lon^2 * lat^2$	Decimal	Auxiliary calculation	6
rural	Rural school = 1; urban/ suburban = 0	D: (0,1)	(DBE, 2016)	1,2,3,4,5,6,7
dcity	Distance to closest populous city	km	GIS calculation	1,2,3,4,5,6,7
dhome_nb	Distance to the nearest homelands borders (excluding SA border)	km	GIS calculation	1,2,3,4,5,6,7
dborder	Distance to the SA border	km	GIS calculation	Insignificant
elev	Elevation	m	(RCMRD, 2015)	Insignificant
rain	Rainfall	mm	(Fick and Hijmans, 2017)	1,2,3,4,5,6,7
pop_dens	Population density (people per square kilometre)	Decimal	(Tatem, 2015)	1,2,3,4,5,6,7

Table 6.9: Table of Variables Cont.

Variable Name	Description	Unit/ Data Type	Source	Specifications
slope	Slope	degrees	GIS calculation from (RCMRD, 2015)	1,2,3,4,5,6,7
muni	Municipality (1-212)	F: (1-212)	(ROSEA, 2018)	1,2,3,4,5,6,7
cen_seg	Homeland centroid Thiessen segments	F: (1-467)	GIS operation	
tcity	Travel time to nearest city	min	**weissmapping**	1,2,3,4,5,6
educators2002	Number of teachers per school in 2002	Integer	DBE (2016)	
educators2016	Number of teachers per school in 2016	Integer	DBE (2016)	Table 4.6
students2002	Number of students per school in 2002	Integer	DBE (2016)	n/a
students2016	Number of students per school in 2016	Integer	DBE (2016)	n/a
seg	Thiessen 200km border segments (1-50)	F:(1-50)	GIS operation	7
loc	Homelands (10-19); Provinces sans homelands (1-9)	F:(1-9)(10-19)	GIS operation	n/a
home	Within homeland = 1; Outside homeland = 0	D: (0,1)	Malinda (2015)	1,2,3,4,5,6,7
dhome_nbxhome	dhome_nb*home	Decimal	Auxiliary calculation	2
dhome_nb2	$dhome\_nb^2$	Decimal	Auxiliary calculation	3,5,6,7
dkwazulu	Distance to KwaZulu border	km	GIS calculation	Table 4.12 & 4.8
dciskei	Distance to Ciskei border	km	GIS calculation	Table 4.12 & 4.8
dtranskei	Distance to Transkei border	km	GIS calculation	Table 4.12 & 4.8
dbop	Distance to Bophuthatswana border	km	GIS calculation	Table 4.12 & 4.8
dvenda	Distance to Venda border	km	GIS calculation	Table 4.12 & 4.8
stdnt_weigth_2016	students2016 ÷ Σstudents2016	Decimal	Auxiliary calculation	n/a
stdnt_weight_2002	students2002 ÷ Σstudents2002	Decimal	Auxiliary calculation	n/a

Table 6.10: 2002 Students per teacher

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
10 5 Homeland	-1.656	-0.505	-2.872
	(3.280)	(2.972)	(3.321)
2) Interacted linear polynomial in distance to boundary			
Homeland	2.863	-4.902	-3.433
	(2.939)	(3.756)	(3.182)
3) Quadratic polynomial in distance to boundary			
Homeland	-0.758	0.740	-1.930
	(3.222)	(3.002)	(3.566)
4) Ordinary least squares			
Homeland	-0.730	0.835	-1.812
	(3.433)	(3.071)	(3.590)
5) Quadratic polynomials			
Homeland	-0.478	0.0514	-2.824
	(3.254)	(2.924)	(3.232)
6) Interacted quadratic polynomials in lat and long			
Homeland	-0.517	0.146	-2.673
	(3.055)	(2.811)	(3.074)
7) Location polynomials with segmented Thiessen fixed effects			
Homeland	-0.457	0.0882	-1.115
	(1.875)	(1.873)	(2.125)
8) Log-linear location polynomials with Thiessen segment fixed effects			
	0.0483*	0.0565*	0.0342
	(0.0268)	(0.0296)	(0.0314)
Observations	12664	9135	6254

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Municipality level clustering.
Homelands is a dummy variable of the lands of the former homelands within the given bandwidth
Specifications can be found in Table 6.9.
The bandwidth is the sampling distance on either side of the former homeland border.
Regressions 4,5,6,7 include location polynomials to control for spatial autocorrelation.
The Thiessen segment fixed effects segment the homeland borders into 50 segments.

Table 6.11: Schools per square kilometre with centroid clustered standard errors

Specification/ Bandwidth	50km	25km	10km
1) Linear polynomial in distance to boundary			
Homelands	0.0428***	0.0410***	0.0359***
	(0.00160)	(0.00191)	(0.00286)
2) Quadratic polynomial in distance to boundary			
Homelands	0.0430***	0.0404***	0.0353***
	(0.00179)	(0.00204)	(0.00295)
3) Ordinary least squares			
Homelands	0.0435***	0.0412***	0.0362***
	(0.00168)	(0.00185)	(0.00274)
4) Quadratic polynomial in lat and long			
Homelands	0.0426***	0.0411***	0.0362***
	(0.00164)	(0.00195)	(0.00297)
5)Interacted quadratic polynomials in lat and long			
Homelands 0.0421***	0.0410***	0.0362***	
	(0.00159)	(0.00191)	(0.00301)
6) Thiessen segmented fixed effects			
Homelands	0.0425***	0.0422***	0.0382***
	(0.00177)	(0.00189)	(0.00264)
7) Linear with segmented homeland proxy fixed effects			
Homelands	0.0431***	0.0421***	0.0376***
	(0.00163)	(0.00189)	(0.00280)
Observations	4287	3130	1843

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.12: Schools per square kilometre with municiplaity clustered standard errors

Specification/ Bandwidth	50km	25km	10km
1) Linear polynomial in distance to boundary			
Home	0.0356***	0.0403***	0.0422***
	(0.00258)	(0.00240)	(0.00247)
3) Quadratic polynomial in distance to boundary			
Home	0.0349***	0.0398***	0.0425***
	(0.00258)	(0.00243)	(0.00254)
4) Ordinary least squares			
Home	0.0356***	0.0405***	0.0430***
	(0.00261)	(0.00242)	(0.00249)
6) Quadratic polynomial in lat and long			
Home	0.0359***	0.0405***	0.0420***
	(0.00255)	(0.00242)	(0.00251)
7)Interacted quadratic polynomials in lat and long			
Home	0.0359***	0.0404***	0.0415***
	(0.00255)	(0.00247)	(0.00262)
8) Linear with segmented fe			
Home	0.0373***	0.0413***	0.0416***
	(0.00278)	(0.00265)	(0.00271)
9) Linear with segmented homeland proxy fixed effects			
Home	0.0368***	0.0411***	0.0422***
	(0.00263)	(0.00248)	(0.00262)
10) Linear with interacted homelands segments			
Home	0.0546***	0.0683***	0.0697***
	(0.0113)	(0.0153)	(0.0153)
Observations	4287	3130	1843

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.13: Students per teacher with centroid clustered standard errors

Specification/ Bandwidth	50km	25km	10km
1) Linear polynomial in distance to boundary			
Home	1.399***	1.355***	1.486***
	(0.299)	(0.324)	(0.420)
2) Interacted linear polynomial in distance to boundary			
Home	1.736***	1.316*	-0.353
	(0.396)	(0.682)	(0.525)
3) Quadratic polynomial in distance to boundary			
Home	1.361***	1.517***	1.552***
	(0.313)	(0.349)	(0.448)
4) Ordinary least squares			
Home	1.521***	1.606***	1.454***
	(0.300)	(0.357)	(0.433)
5) Linear polynomial in lat and long			
Home	1.524***	1.482***	1.386***
	(0.301)	(0.331)	(0.411)
6) Quadratic polynomial in lat and long			
Home	1.253***	1.252***	1.462***
	(0.307)	(0.329)	(0.429)
7)Interacted quadratic polynomials in lat and long			
Home	1.266***	1.265***	1.459***
	(0.307)	(0.330)	(0.438)
8) Linear with segmented fe			
Home	1.088***	1.206***	1.445***
	(0.346)	(0.374)	(0.423)
9) Linear with segmented homeland proxy fixed effects			
Home	1.095***	1.199***	1.470***
	(0.315)	(0.333)	(0.416)
10) Linear with interacted homelands segments			
Home	2.272***	2.808***	3.662***
	(0.818)	(0.871)	(0.929)
Observations	2741	2309	1543

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.14: Schools per square kilometre with municipality clustered standard errors

Specification/ Bandwidth	50km	25km	10km
1) Linear polynomial in distance to boundary			
Home	1.480***	1.381***	1.413***
	(0.322)	(0.328)	(0.309)
2) Interacted linear polynomial in distance to boundary			
Home	0.448***	1.287***	1.764***
	(0.144)	(0.442)	(0.403)
3) Quadratic polynomial in distance to boundary			
Home	1.589***	1.564***	1.423***
	(0.338)	(0.330)	(0.310)
4) Ordinary least squares			
Home	1.495***	1.643***	1.581***
	(0.341)	(0.343)	(0.313)
5) Linear polynomial in lat and long			
Home	1.387***	1.494***	1.549***
	(0.327)	(0.340)	(0.312)
6) Quadratic polynomial in lat and long			
Home	1.457***	1.278***	1.281***
	(0.318)	(0.327)	(0.311)
7)Interacted quadratic polynomials in lat and long			
Home	1.455***	1.282***	1.284***
	(0.318)	(0.327)	(0.314)
8) Linear with segmented fe			
Home	1.433***	1.220***	1.151***
	(0.323)	(0.336)	(0.315)
9) Linear with segmented homeland proxy fixed effects			
Home	1.468***	1.222***	1.136***
	(0.320)	(0.343)	(0.327)
10) Linear with interacted homelands segments			
Home	4.010***	3.178***	2.786***
	(1.138)	(0.800)	(0.829)
Observations	2741	2309	1543

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Homelands is a dummy variable of the lands of the *former* homelands.
Standard errors assume each raster pixel is an individual observation. Specifications can be found in Table 6.1

Table 6.15: 2002 average school completion rate with centroid segment clusters

Variable/ Bandwidth:	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homelands	-7.429***	-4.898**	-3.070
	(2.434)	(1.940)	(2.365)
2) Interacted linear polynomial in distance to boundary			
Homelands	1.000	2.002	2.748
	(3.583)	(2.466)	(3.771)
3) Quadratic polynomial in distance to boundary			
Homelands	-7.135***	-4.563**	-2.794
	(2.290)	(2.018)	(2.439)
4) Ordinary least squares			
Homelands	-7.373***	-4.527**	-2.866
	(2.550)	(2.010)	(2.367)
5) Quadratic polynomials			
Homelands	-6.498***	-4.527**	-2.689
	(2.215)	(1.923)	(2.404)
6) Interacted quadratic polynomials in lat and long			
Homelands	-6.572***	-4.553**	-2.732
	(2.238)	(1.922)	(2.391)
7) Location polynomials with Thiessen segment fixed effects			
Homelands	-5.181**	-4.112*	-1.445
	(2.423)	(2.136)	(2.607)
Observations	11863	8505	5834

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Homelands is a dummy variable of the lands of the *former* homelands. Specifications can be found in Table 6.9.

Table 6.16: 2016 average school completion rate with centroid segment clusters

Variable/ Bandwidth	25km	10km	5km
1) Linear polynomial in distance to boundary			
Homeland	-2.910*	-4.733***	-4.163**
	(1.756)	(1.806)	(1.621)
2) Interacted linear polynomial in distance to boundary			
Homeland	-2.694	-4.737**	-3.051
	(3.614)	(2.121)	(3.308)
3) Quadratic polynomial in distance to boundary			
Homeland	-2.913*	-5.218***	-4.060**
	(1.748)	(1.896)	(1.643)
4) Ordinary least squares			
Homeland	-4.253**	-5.307***	-4.143***
	(1.959)	(1.826)	(1.553)
5) Quadratic polynomials			
Homeland	-3.081*	-4.966***	-3.986**
	(1.766)	(1.844)	(1.668)
6) Interacted quadratic polynomials in lat and long			
Homeland	-2.968*	-4.958***	-3.985**
	(1.773)	(1.846)	(1.659)
7) Location polynomials with Thiessen segment fixed effects			
Homeland	-1.987	-3.862**	-2.741
	(1.725)	(1.901)	(1.780)
Observations	20630	14614	4287

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The homeland variable is a dummy of whether the school lies within the *former* homeland lands.
The bandwidth is the sampling distance on either side of the former homeland border
The segmented Theissen fixed effects segment the homeland borders into 50 segments
Only schools with an emis in 2016 are included