For Online Publication

Online Appendix to "Moved to Poverty? A Legacy of the Apartheid Experiment in South Africa"

Bladimir Carrillo Carlos Charris Wilman Iglesias

August 9, 2023

A Data Description and Summary Statistics

This appendix provides a description of all the data used in the paper.

1980 Census data: Our main analysis uses individual-level data from the 1980 census complete count file, obtained from Statistics South Africa.[1] The 1980 census is particularly well suited for the purposes of this study. It was conducted during the apartheid regime and when all legal restrictions on internal mobility were present. This implies that most migrants from White regions are likely to reside in the homelands where they were relocated. This census is the only one with detailed information on place of birth, including Magisterial District.[2] This information allows us to identify migrants of White-place origin and study possible mechanisms.

The analysis sample consists of individuals residing in a homeland at census time, all of whom we match to homeland formalization laws based on their place of residence. We then restrict the sample to individuals born in the 1937-1969 birth cohorts who were between 2 and 23 years old when their homeland of residence was established. The number of observations drops sharply when considering individuals who were under age 2 at homeland formalization because several homelands were established after 1970, and we observe their outcomes in 1980. This is why we exclude individuals who were under age 2 at homeland formalization.

The core census file provides information on all individuals in South Africa, except those in the homelands of Transkei, Venda, and Bophuthatswana. The South African Bureau of Statistics released a separate census data file for Bophuthatswana, but it does not include data on the province or District of birth for those born outside of the homelands. Although we exclude individuals in Bophuthatswana from the baseline analysis sample, we show that this omission does not affect our results in Appendix Table G.4. In the end, our main analysis draws records of individuals residing in one of seven homelands: Ciskei, KwaZulu, Gazankulu, Lebowa, QwaQwa, KaNgwane, and KwaNdebele.

1985 Census data: In 1985, the apartheid government carried out another population census. It covered all areas of South Africa, excluding the homelands of Bophuthatswana, Ciskei, Transkei, and Venda. This census includes a more restricted number of questions. Unlike the 1980 census, the information on an individual's place of birth is more limited: census enumerators asked respondents to indicate the homeland/province of birth, but provinces

[1] These data are publicly available at http://www.statssa.gov.za, last accessed on February 11, 2020.

[2] The provinces under apartheid include Cape, Natal, Transvaal, Oranje Free State, and the 10 homelands. In the census, each homeland is labeled as a different province. Magisterial District is a second-order administrative division, similar to a county in the United States. There are about 380 Magisterial Districts.

in White areas were recorded as "Republic of South Africa." More detailed geographical information at the Magisterial District of birth was not collected, limiting our ability to investigate mechanisms with these data. Despite these limitations, we are still able to identify migrants of White-place origin and examine the basic patterns.

1991 Census data: We also have access to the full census conducted in 1991.[3] As in the 1985 census, enumerators collected information on place of birth at the province/homeland level but not at the Magisterial District level. This census wave covered all the homelands, but individuals in the homeland of Transkei were separately enumerated, and such data are not readily available. Our analysis also excludes individuals residing in the Ciskei homeland because information on an individual's place of birth was not consistently recorded, making it impossible to identify migrants.[4] It is also important to mention that many particularly poor and remote areas within the homelands were not enumerated due to political violence and budget restrictions in 1991 (p. 20 Human Sciences Research Council, 2007). Therefore, some areas suffer from severe incomplete coverage, and the results from these data should be viewed with caution.

As in the 1980 census, our analysis sample consists of individuals residing in a homeland at census time. We focus on individuals aged 15 to 53 (inclusive) who were 23 years old or younger at homeland formalization. Since the birth cohorts of interest are now older, we can include those who were under age 2 when their homeland of residence was established. The basic analysis sample consists of individuals residing in one of eight homelands: Bophuhatswana, KwaZulu, Gazankulu, Lebowa, QwaQwa, KaNgwane, KwaNdebele, and Venda.

1996 Census data: We use the post-apartheid census conducted in 1996 in supplementary analyses. We use the 10 percent randomly drawn sample available from Statistics South Africa. Unlike previous census data, the 1996 census covers all the former homelands. Census enumerators did not collect information on an individual's place of birth in 1996, so we use the following questions to identify migrants of White-place origin: do you live now? Where did you live before this?." We define migrants from White areas as those who reside in a former homeland area at census time but whose Magisterial District of previous residence is in a White area. The rest of the individuals, which include those who have never moved

[3] This census was conducted in March 1991, three months before the apartheid regime was dismantled.

[4] To be more precise, the Ciskei datafile only provides information on the Magisterial District of birth for individuals born in Ciskei. For those born out of Ciskei, we have information only on the country of birth (South Africa or abroad). Therefore, it is impossible to know if individuals born in the rest of South Africa were born in a homeland or a White region.

and migrants from other homelands, are used as the comparison group. Approximately 75 percent of individuals in the comparison group have never moved across Magisterial District boundaries. This suggests that the vast majority of individuals in the control group were born and always lived in the current Magisterial District of residence. For those who never moved or have moved across Magisterial Districts only once, our assignment of childhood exposure will be relatively accurate. Data from the South African Internal Migration Survey 1999-2000 indicate that only 5 percent of individuals have moved more than once. Consistent with these figures, Dinkelman (2017) convincingly shows that these questions provide accurate information on the Magisterial District of birth for the most of individuals

Following the same idea as in the baseline analysis, we limit the sample to individuals who reside in a former homeland area at census time. Administrative areas in the 1996 census do not necessarily coincide with that in the apartheid era, as some Magisterial Districts were split into two or more Districts or aggregated to other Districts during the post-democratization period. We match each Magisterial District in the 1996 census to homeland boundaries using a high-resolution GIS map of South Africa during the apartheid era. We assign Magisterial Districts to the former homeland areas if more than 60 percent of their territory overlaps with a given homeland. The results are largely unchanged if we instead use either 50, 70, or 80 percent cutoffs for homeland assignment.

1994/1995 October Household Surveys: In Appendix Section G.5, we use these household surveys to look at fertility around homeland establishment. This survey provides basic demographic characteristics for about 300,000 individuals, including the Magisterial District of birth. For all women aged 12 to 54, the survey provides detailed information on fertility histories, including the timing of all births. We identify individuals residing in the 10 former homelands based on the Magisterial District of residence, which aligns closely with the White area/homeland divisions.

Date of homeland establishment: Data on the year of homeland establishment come from Dinkelman (2017). Table A.1 lists this information for each homeland.

Table A.1: Timing of Homeland Establishment and Coverage in Apartheid Censuses

		Coverage		
Homeland	Year of establishment	1980	1985	1991
Bophuthatswana	1961	✓		✓
Ciskei	1961	✓		
Gazankulu	1971	✓	✓	✓
Kangwane	1976	✓	✓	✓
KwaZulu	1970	✓	✓	✓
Kwandebele	1977	✓	✓	✓
Lebowa	1971	✓	✓	✓
Qwaqwa	1969	✓	✓	✓
Transkei	1959			
Venda	1962			✓

Notes: Data on homeland establishment come from Dinkelman (2017). The 1991 census did collect information on individuals in the Ciskei homeland, but we exclude this homeland in the analysis involving this census because information on an individual's place of birth was not consistently recorded.

Table A.2: Summary Statistics of Key Variables

	Mean	Standard deviation	Number of observations
Panel A: 1980 Census			
Years of education	4.737	2.395	2,489,596
Migrant	0.171	0.377	2,489,596
Age	20.449	6.589	2,489,596
Male	0.447	0.497	2,489,596
Panel B: 1985 Census			
Years of education	4.566	3.233	2,346,208
Migrant	0.22	0.414	2,346,208
Age	24.567	6.307	2,346,208
Male	0.409	0.492	2,346,208
Panel C: 1991 Census			
Years of education	6.396	2.112	2,139,840
Currently employed	0.296	0.246	2,179,334
Log total income	8.398	0.531	631,070
Migrant	0.231	0.421	2,139,840
Age	30.683	8.048	2,139,840
Male	0.392	0.488	2,139,840

Notes: Table reports summary statistics of main outcomes and basic demographic statistics of the samples used in the paper.

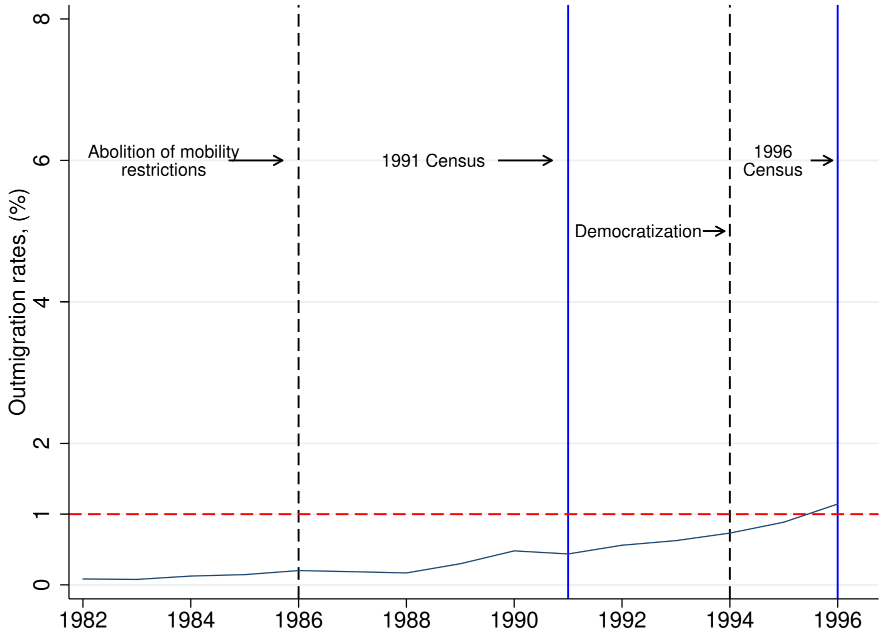
B Living Conditions in Homelands

It is widely acknowledged that the homelands were disadvantaged relative to the rest of South Africa. These areas were mainly rural, with high unemployment levels and pervasive poverty. Due to the central government's designation of the homelands as "self-sufficient entities", education, public health services, and other important welfare features suffered from severe underfunding (De Beer, 1986; Tanaka, 2014). The homelands were characterized by sanitation problems, including high rates of diseases, malnutrition, prevalence of respiratory infections, kwashiorkor and marasmus as well as measles and even gastroenteritis among children (Horrell, 1973; De Beer, 1986; Dinkelman, 2017). Educational attainment was low compared to residents in White South Africa, which is perhaps unsurprising given the underfunding of education in the homelands. Families depended on subsistence agriculture to survive. Arable land was scarce and relatively infertile, with droughts being a major source of adverse shocks. Although homeland residents could seek jobs as temporary migrants in White areas under special permissions, their access to these labor markets was limited.

A variety of historical narratives are consistent with the notion that White-place migrants were worse off in the homelands in terms of opportunities. Between 1980 and 1983, a group of independent researchers sent by the Surplus People Project (1983, SPP) visited the homelands and collected information through interviews and direct observation of the living conditions of relocated people. Common opinions about opportunities collected by the SPP team include *"Not enough supply of work"*, *"in this place unemployment is very rife"*, *"our problem is that we have not got enough money as a consequence of scarcity of work"*, *"here we are not working because jobs are scarce"*, *"we could get jobs there [White areas], here there are no jobs"*, *"Work was not as scarce as it is here."*

C Out-Migration After 1986

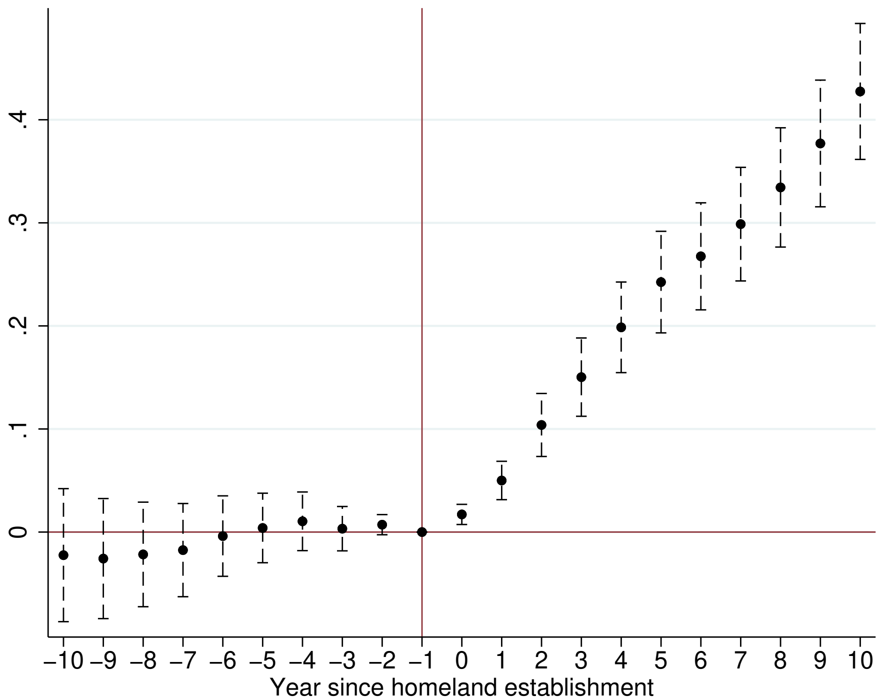
Figure C.1: No Significant Migration out of the Homelands After 1986



Notes: This figure plots the share of homeland residents who moved to White areas by year. Estimates based on the 1996 census.

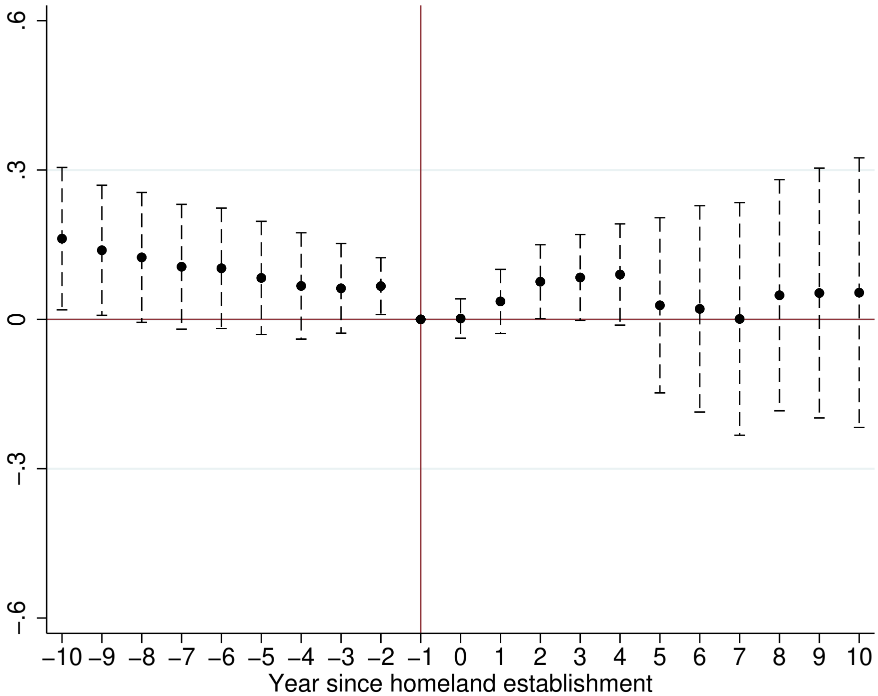
D Additional Results: Homeland Establishment and In-migration

Figure D.1: Estimated Effects on Probability of Living in a Homeland (Mobility from White Areas to Homelands - Unbalanced sample)



Notes: This figure plots β_z (and 95 confidence intervals) from estimating equation (1). Results based on the 1996 census. Using information on year of arrival, we define an event in individuals' residential history as a binary variable equal to one for the years in current homeland of residence. This generates a microdata panel at the individual-year level. The sample is limited to individuals whose previous place of residence is in a White area. There are no restrictions on whether or not individuals are observed during both the entire pre- and post-policy periods, so the panel is unbalanced. Controls include fixed effects for year and individual. Robust standard errors are clustered at the homeland $\times$ birth decade level.

Figure D.2: Falsification: Estimated Effects on Probability of Living in Actual Homeland (Mobility from Homeland to Homeland)



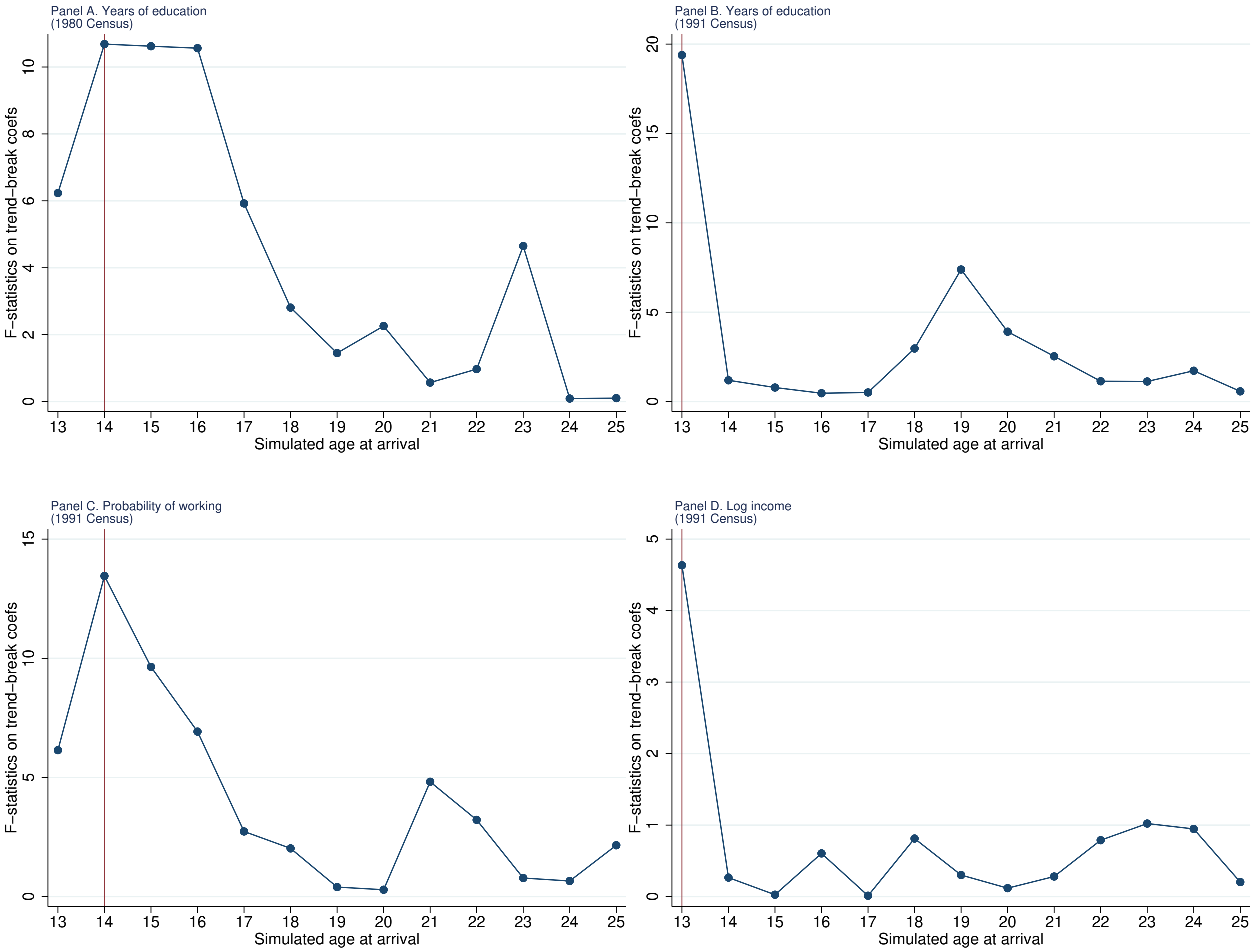
Notes: This figure plots β_z (and 95 confidence intervals) from estimating equation (1). Results based on the 1996 census. Using information on year of arrival, we define an event in individuals' residential history as a binary variable equal to one for the years in current homeland of residence. This generates a microdata panel at the individual-year level. The sample is limited to individuals whose previous place of residence is in a homeland. To estimate this event study on a fully balanced sample of individuals, the sample is limited to individuals who are observed during both the entire pre- and post-policy periods. Controls include fixed effects for year and individual. Robust standard errors are clustered at the homeland $\times$ birth decade level.

E Defining Childhood Exposure

Our measure of exposure considers age 15 as the upper age limit of exposure. This age cutoff is not chosen at random. To guide our choice of the relevant years of childhood, we conduct a series of trend-break tests on the age-at-shock coefficients in the spirit of Greenstone and Hanna (2014). Specifically, we estimate regressions where the dependent variable are the simulated age-at-arrival coefficients estimated from equation (5) on a linear trend, an indicator for a possible break age and the interaction between both variables. This specification allows for possible breaks in levels and trends. We estimate this regression for each possible age break within a window of the middle 50 percent of the possible break ages. We then calculate the corresponding F-statistics that tests the hypothesis that the coefficients on the break indicator and its interaction with the linear trend are both equal to zero. This F-statistics formally tests the hypothesis of a structural break in the given age of exposure. The maximum of the F-statistics is used to identify the best possible break age of exposure. We conduct this test for all educational and labor market outcomes used in the main analysis to have a broad picture of the test.

Figure E.1 presents these results. They confirm the visual inspection of the non-parametric estimates of exposure at different ages of exposure. The test suggests a structural break between ages 13 and 15. Based on this evidence, we consider age 15 as the upper age limit of exposure to summarize our main results.

Figure E.1: Trend-Break F-Statistics



Notes: F-statistics are from the joint test of the equality of the age-at-formalization variable, its interaction with a dummy for age-at-formalization lower than or equal to x (where x is given by the x-axis in the figure).

F Robust Inference

In our main estimates, we cluster standard error at the homeland-birth decade level because the variation in our key independent variable of interest occurs largely at this level and because we have a relatively small number of homelands (only seven). In Appendix Table F.1, we explore the robustness of our results to alternative inference approaches. A possible concern with our baseline approach is that error terms might be correlated across individuals from the same homeland but in different cohort groups. To investigate this issue, we cluster standard errors at the homeland only and correct for the small number of clusters using wild bootstrap (Cameron et al., 2008). We also present the corresponding p-values from standard errors clustered at the homeland-birth decade level. We also examine if the results from the flexible specification (5) are affected when changing the level of clustering. Appendix Figure F.1 shows the corresponding p-values for each age-exposure coefficient. In general, there are no meaningful differences between these alternative inference approaches. In fact, our baseline approach yields p-values that are larger in some cases (see Panel A of Figure F.1).

We also conduct a non-parametric permutation test following the same idea as Chetty et al. (2009). Specifically, we randomly assign placebo formalization dates and reconstruct our measure of childhood exposure. We then rerun the baseline equation (4) using the placebo measure of treatment and repeat this procedure 500 times. The share of the absolute placebo coefficients that are larger in magnitude than the "true" coefficient can be taken as a measure of how likely are our results to arise by chance. The key advantage of this inference method is that it does not rely on parametric assumptions on the distribution of the error term. Figure F.2 plots the distribution of the placebo coefficients, with the red line denoting the actual coefficient. As shown in the figure, the actual coefficient falls far in the left tail of the distribution. The implied p-values are always below 0.01, a result that is consistent with that computed using conventional inference and it suggests that our results are very unlikely to be an artifact of the data.

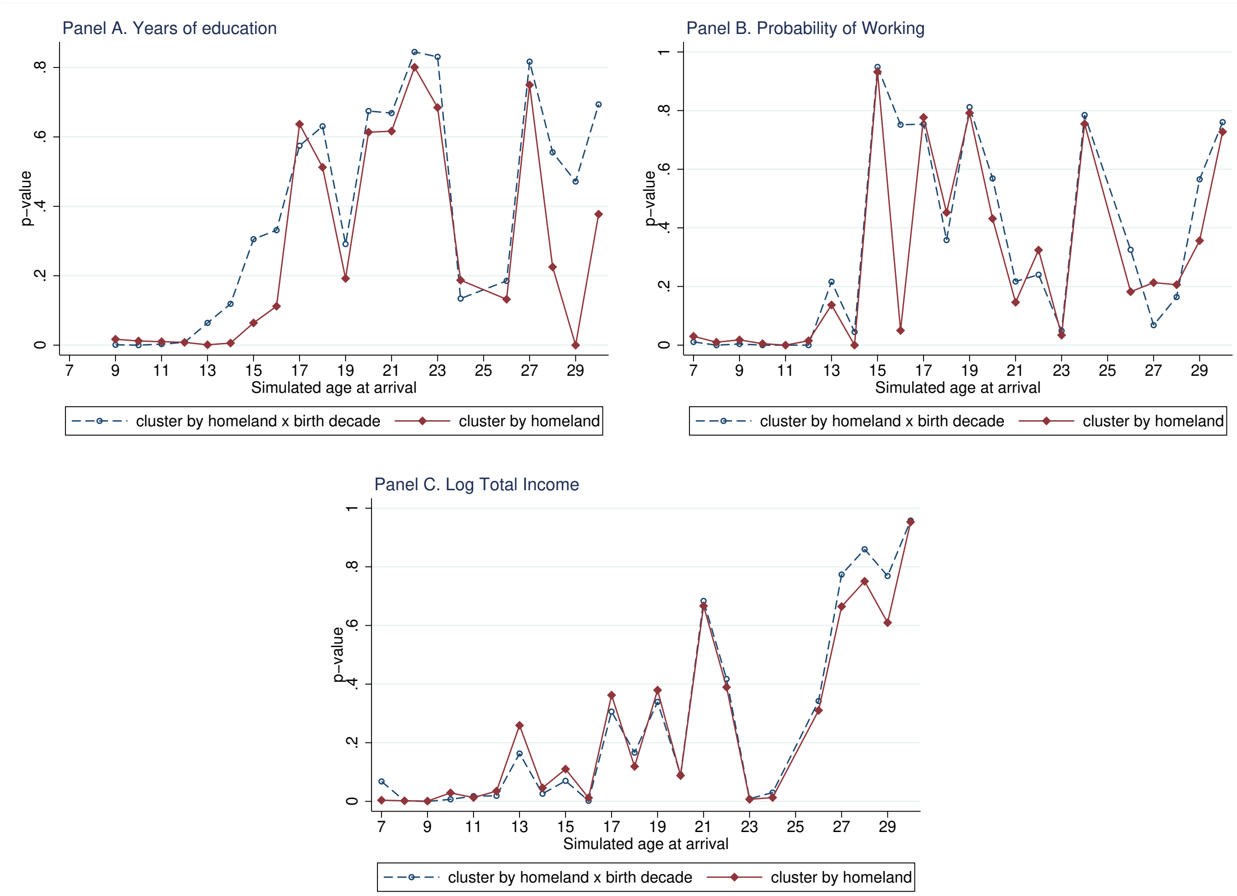
Given that our results are not appreciably affected when considering more demanding inference approaches, we focus on our computationally simple baseline inference method throughout the paper.

Table F.1: Alternative Approaches to Inference

	Dependent variable:		
	(1)	(2)	(3)
	years of education	Probability of working	Log total income
Simulated Exposure × Migrant	-0.116	-0.0086	-0.0077
	[0.0198]	[0.0014]	[0.0021]
Wild Bootstrap p-value:			
cluster by homeland - birth decade	0.001	0.001	0.01
cluster by homeland	0.001	0.017	0.03
Homeland × birth-year FE	✓	✓	✓
Province-of-birth FE	✓	✓	✓
Magisterial District FE	✓	✓	✓
Demographic controls	✓	✓	✓
Observations	2,489,596	2,179,334	631,070

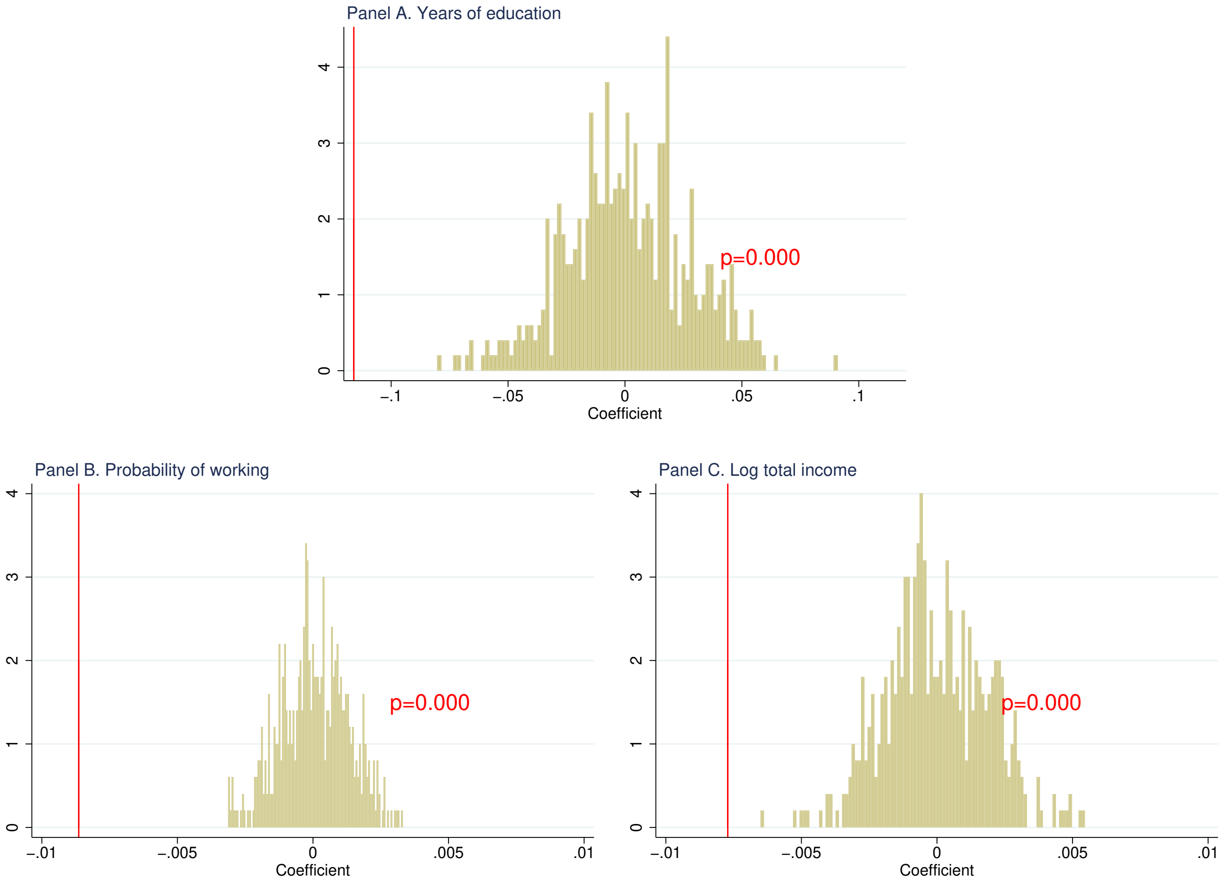
Notes: This table reports wild bootstrap p-values from error terms clustered at the homeland-birth decade level and at the homeland only. Results in column 1 are based on the 1980 census, while those in columns (2) and (3) are based on the 1991 census. Demographic controls include dummies for gender and ethnicity, except in columns (2) and (3) where information on ethnicity is not available.

Figure F.1: Wild-bootstrap p-values
(alternative clustering)



Notes: This figure reestimates the non-parametric equation 5 and plots wild bootstrap p-values based on clustering at the homeland-birth decade level (baseline) and based on clustering at the homeland level.

Figure F.2: Permutation Test

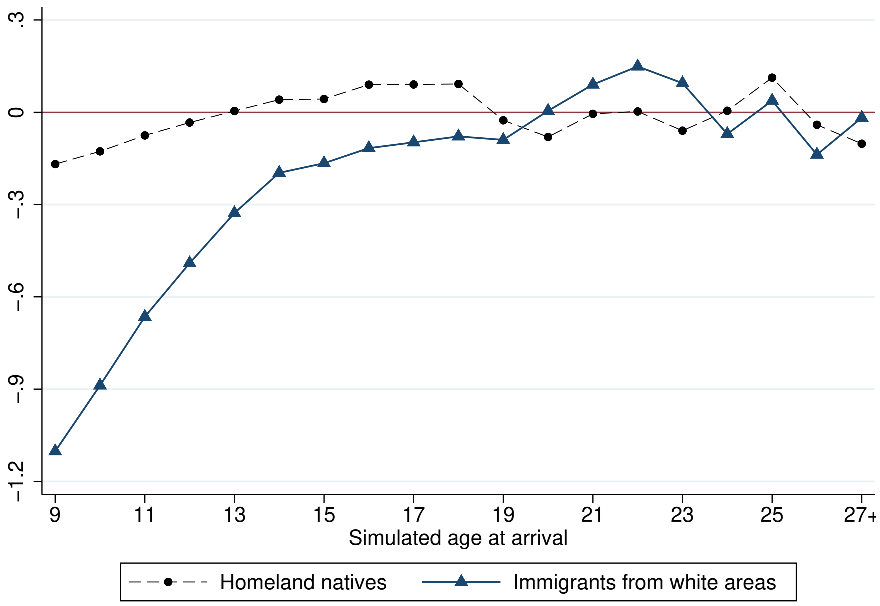


Notes: This figure plots the empirical distribution of placebo effects for years of education. This distribution is constructed from 500 estimates of β using the specification in column (1) of Table 1. The vertical line shows the estimated effect reported in Table 1.

G Additional Results and Robustness Checks

G.1 Flexible Estimates for Migrants and Natives Separately

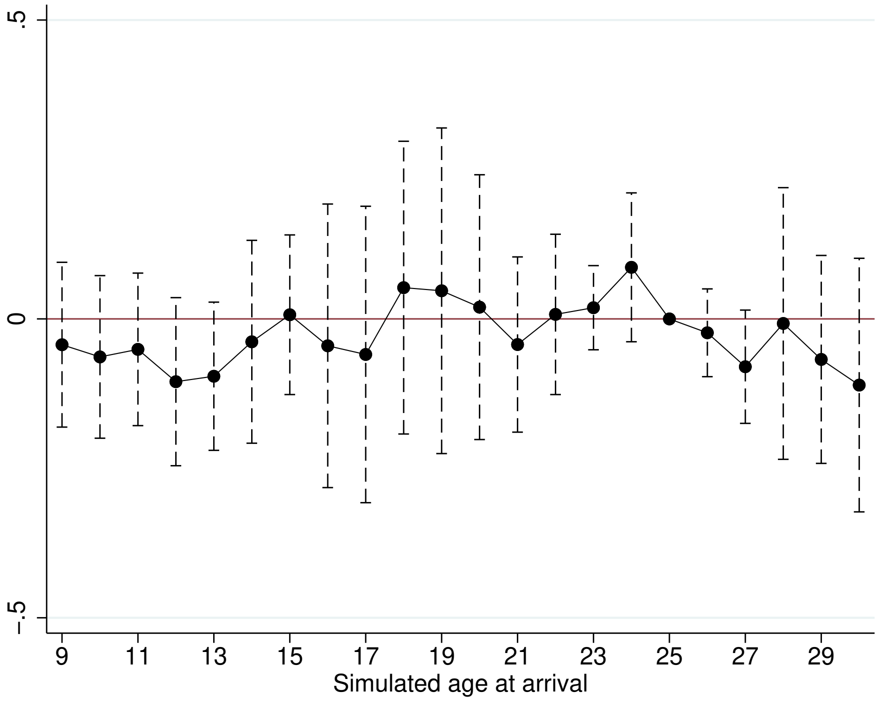
Figure G.1: Estimated Effects on Years of Education
(Migranst and Locals)



Notes: This figure plots β_τ from estimating the analogous version of equation (5) for migrants and locals separately. Controls include Magisterial District fixed effects, province-of-birth fixed effects, birth-year fixed effects, indicators for gender and ethnicity, and homeland-specific linear cohort pre-trends. The homeland-specific linear pre-trends are interactions between a linear cohort trend and homeland dummies, which are set to zero for simulated ages at arrival below 18. Robust standard errors are clustered at the homeland-birth decade level.

G.2 Falsification

Figure G.2: Falsification: Migrants from other Homelands
(No Effect on Education)



Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating equation (5). The treatment group is replaced by the placebo treatment of migrants from other homelands. The sample is based on the 1980 Census. Controls include homeland-of-residence × birth-year fixed effects, Magisterial District fixed effects, province-of-birth fixed effects, and indicators for gender and ethnicity. Robust standard errors are clustered at the homeland- birth decade level.

G.3 Evidence from the 1996 Post-Apartheid Census

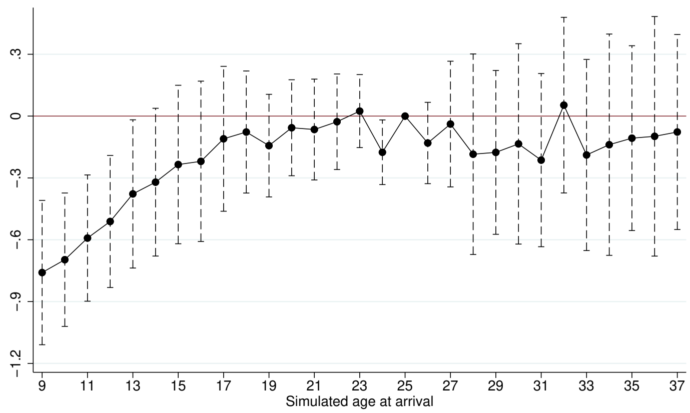
Table G.1: Evidence from the 1996 Post-Apartheid Census

	Dependent variable: years of education	
	(1)	(2)
	1996 Census (10 percent sample)	
	Baseline sample	Extended sample
Simulated Exposure × *Migrant*	-0.0683	-0.0439
	[0.0270]	[0.0190]
Homeland × birth-year FE	✓	✓
Place-of-origin FE	✓	✓
Magisterial District FE	✓	✓
Observations	119,762	206,941

Notes: All estimates are based on the 1996 census. Baseline sample limits the sample to the same birth cohorts and 7 homelands used in the 1980 estimation sample. Extended sample includes all the 10 homelands available in the 1996 census. Place-of-origin is defined as prior province of residence. Since provinces in the 1996 census do not correspond exactly with the provinces in the apartheid-censuses, we distinguish between provinces × homeland-status groups to account for the fact that some Districts were part of the former homelands within each province. The regressions include a dummy for gender and use population weights. Robust standard errors (reported in brackets) are clustered at the homeland-birth decade level.

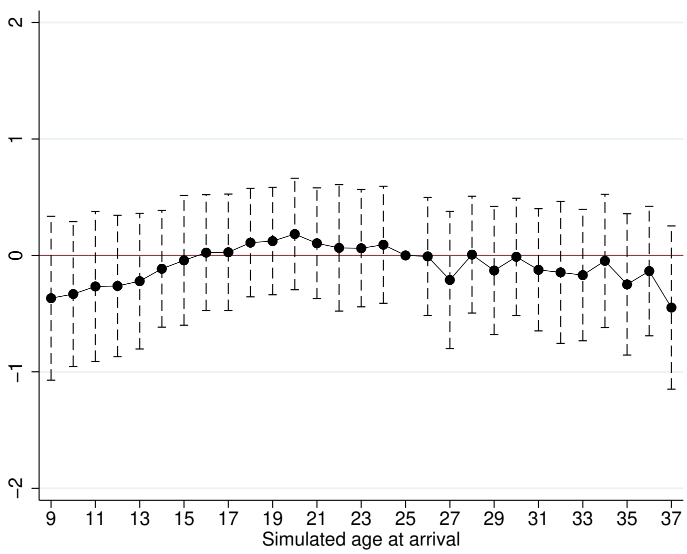
G.4 Extended Sample and Cohort Sizes

Figure G.3: Estimated Effects on Years of Education (Extended Sample)



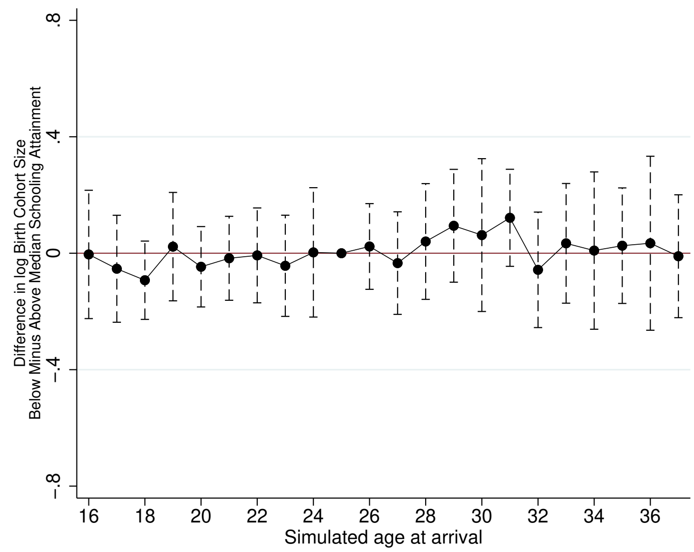
Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating the equation (5) using the 1980 census. Robust standard errors are clustered at the homeland-birth decade level.

Figure G.4: Estimated Effects on Log Cohort Sizes



Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating the equation (5) on log cohort size. Cohort size is calculated at the province-of-birth, homeland, Magisterial District and birth-year level. Sample is based on the 1980 census. Controls include homeland $\times$ birth-year fixed effects, Magisterial District fixed effects and province-of-birth fixed effects. Robust standard errors are clustered at the homeland $\times$birth decade level.

Figure G.5: Estimated Effects on Differences in Birth Cohort Sizes between High- and Low-Education Cohorts



Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating the equation (5) on differences in birth cohort size between high- and low-education cohorts. Cohort size is calculated at the province-of-birth, homeland, Magisterial District and birth-year level. Sample is based on the 1980 census. Controls include homeland × birth-year fixed effects, Magisterial District fixed effects and province-of-birth fixed effects. Robust standard errors are clustered at the homeland ×birth decade level.

Table G.2: Estimated Effects on Cohort Sizes

	Dependent variable: log cohort size
	(1)
Simulated Exposure × Migrant	-0.075
	[0.057]
Observations	10,668

Notes: Estimates are based on the 1980 census. Cohort size is calculated at the province-of-birth, homeland, Magisterial District and birth-year level. Controls include homeland × birth-year fixed effects, Magisterial District fixed effects, and province-of-birth fixed effects. Robust standard errors are clustered at the homeland× birth decade level.

G.5 Fertility around Homeland Establishment

As mentioned above, one may be worried that families anticipated homeland establishment and responded by changing fertility decisions. While we do not observe significant changes in cohort sizes correlated with homeland establishment, we further explore this issue using the October Household Survey (OHS, 1994-95). The OHS provides retrospective information on fertility for all women aged 12 to 54, including the timing of all births. Importantly, these data contain information on an individual place of birth and thus it is possible to identify migrants from White areas.

Using these data, we construct a pseudo panel of women-by-year and estimate a dynamic model of the timing of birth. The initial year for each woman is the one in which they were 12, whereas the final year is the same as that of the survey. We then create a dummy indicator equal to one if a woman gave birth in a given year and zero otherwise. Appendix Figure G.6, Panel A informally shows that the levels and trends in birth rates around homeland establishment are strikingly similar for migrants and locals. Importantly, it does not appear to have an abnormal, abrupt change in the years prior to formalization for migrants.

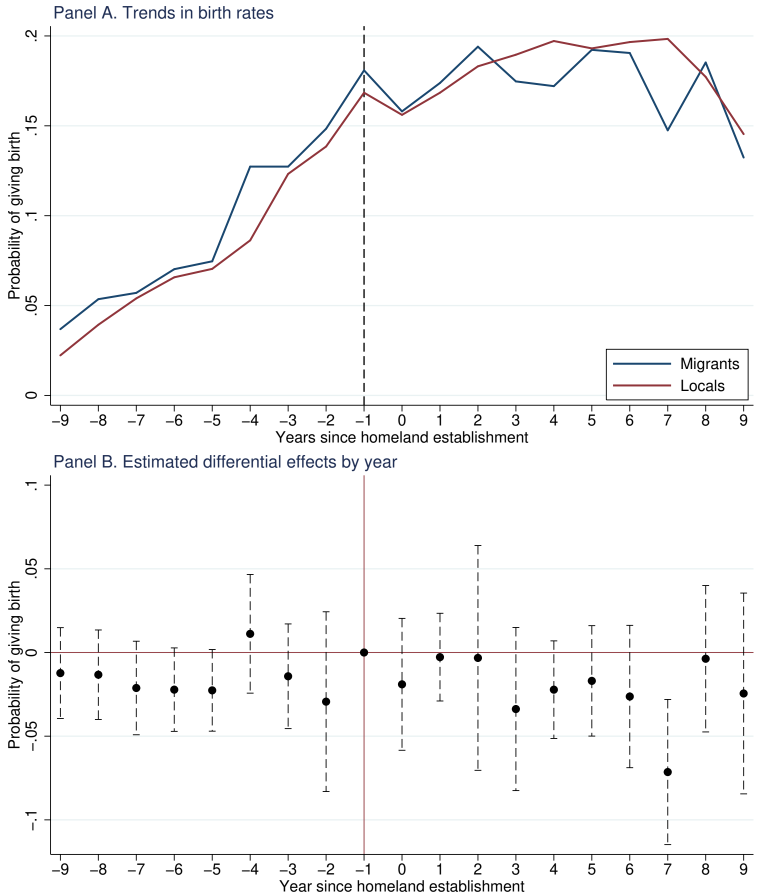
We then formally test whether there were differential trends by estimating the following model of the timing of birth:

$$
\begin{aligned}
Pr(birth=1)_{ijt} = \alpha \ + \ & \underbrace{\sum_{z=0}^{T} \beta_z \left(Establishment{=}z\right)_{jt} \times \left(Migrant\right)_i}_{\text{post-policy period}} + \\
& \underbrace{\sum_{z=-T}^{-2} \beta_z \left(Establishment{=}z\right)_{jt} \times \left(Migrant\right)_i}_{\text{pre-policy period}} + \lambda_i \ + \ \gamma_{jt} \ + \ \xi_{ijt}
\end{aligned}
\tag{G.7}
$$

where $Pr(birth = 1)_{ijt}$ is the birth indicator for woman i in homeland j and year t. $Establishment$ is an indicator for z years between homeland formalization and year t for the individual belonging to the homeland j. The omitted category is -1. We include individual fixed effects (λ_i) and homeland-by-year fixed effects (γ_{jt}). We allow the idiosyncratic error term to be correlated across individuals within a homeland-birth decade cohort group. To estimate this event study on a fully balanced sample of individuals, we exclude individuals who are not observed during the entire pre- and post-policy periods. Balancing the panel implies a smaller sample but alleviates selection concerns due to different birth cohorts being observed in different moments in time. In practice, this restriction does not materially affect the results.

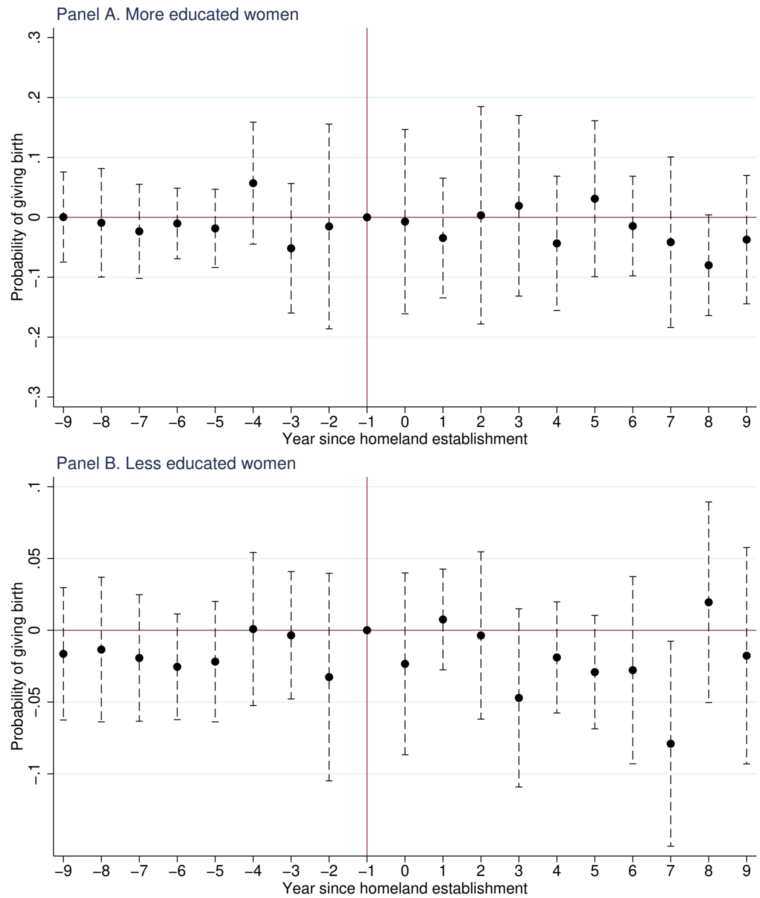
Appendix Figure G.6, Panel B plots estimates of β_z and respective confidence intervals. As can be seen from the figure, there are no differential trends in the years around homeland establishment. Furthermore, one can see that there is not any clear tendency toward improving or deteriorating birth rates in the years just before formalization. One may still be worried about the possibility that these patterns are different for more and less advantaged women. Appendix Figure G.7 repeats the same exercise for more and less educated women. Since education is itself endogenous to the policy shock, we limit the sample to women who had certainly completed their schooling decisions at homeland establishment (i.e., after age 16 at homeland establishment). This restriction only excludes 1 percent of the baseline sample. Reassuringly, there are no meaningful differences in birth trends for low and education mothers. Summarizing, there is no strong indication that sample selection due to changes in fertility is a major issue.

Figure G.6: Probability of Giving Birth Relative to Homeland Establishment



Notes: Panel A plots trends in birth rates for the years around homeland establishment. Panel B plots estimates of β_z and respective 95 percent confidence intervals based on equation (G.7). Robust standard errors are clustered at the homeland-birth decade cohort level.

Figure G.7: Probability of Giving Birth Relative to Homeland Establishment (More and Less Educated Mothers)



Notes: This figure plots estimates of β_z and respective 95 percent confidence intervals based on equation (G.7) separately for more and less educated mothers. More-educated mothers represents those mothers who completed at least primary education. Less-educated mothers are those who completed less than primary education. Robust standard errors are clustered at the homeland-birth decade cohort level.

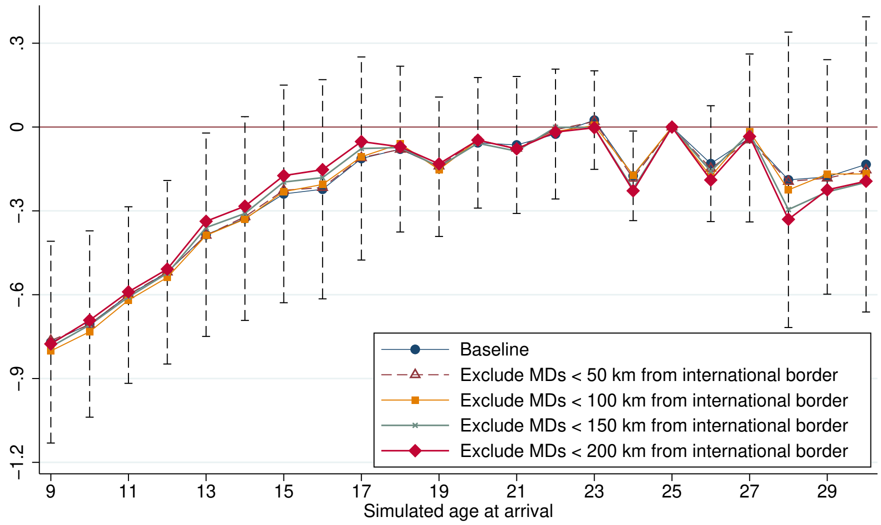
G.6 Selective Outmigration from South Africa

Table G.3: Excluding Immigrants from Areas Close to the International Border

	Dependent variable: years of education				
	(1)	(2)	(3)	(4)	(5)
		Exclude immigrants from Magisterial Districts with Distance from Neighboring Countries:			
	Baseline estimate	< 50 km	<100 km	< 150 km	< 200 km
Simulated Exposure × Migrant	-0.116	-0.1156	-0.1211	-0.1169	-0.1147
	[0.0198]	[0.0201]	[0.0223]	[0.0225]	[0.0230]
Homeland × birth-year FE	✓	✓	✓	✓	✓
Province-of-birth FE	✓	✓	✓	✓	✓
Magisterial District FE	✓	✓	✓	✓	✓
Observations	2,489,596	2,482,181	2,451,655	2,410,457	2,398,251

Notes: Estimates are based on individuals residing in a homeland at the time they are observed in the 1980 census. Simulated exposure refers to the predicted number of childhood years in the homeland of residence. Migrant is an indicator for migrants from White areas. All regressions include controls for gender and ethnicity. Each column reports results that exclude migrants from Magisterial Districts in White areas depending on the distance to the international border. Robust standard errors (reported in brackets) are clustered at the homeland × birth decade level.

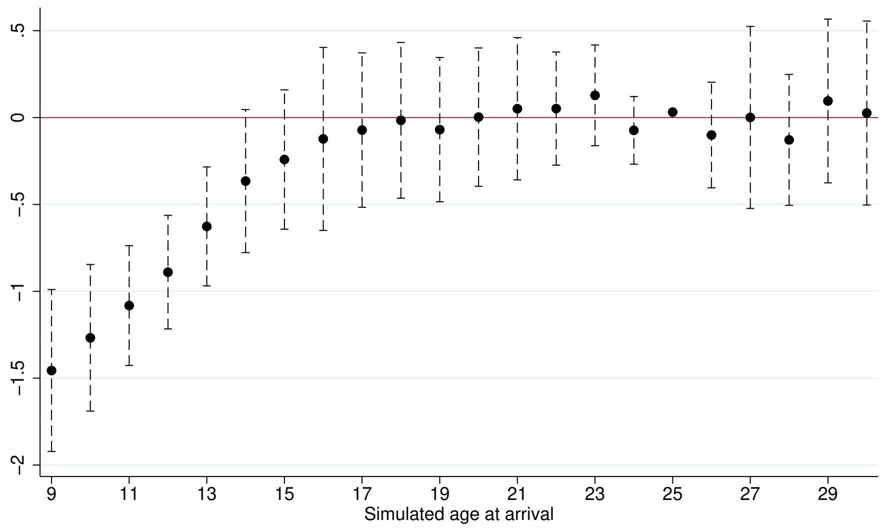
Figure G.8: Estimated Effects on Years of Education
(Exclude Magisterial Districts close to the international border)



Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating equation (5). Figure also shows results from alternative specifications. The sample is based on the 1980 census. Controls include homeland-of-residence × birth-year fixed effects, Magisterial District fixed effects, province-of-birth fixed effects, and indicator for gender and ethnicity. Robust standard errors are clustered at the homeland × birth decade level.

G.7 Movers-versus-Stayers Design

Figure G.9: Estimated Effects on Years of Education
(Stayers as comparison group)



Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating the equation (5) using stayers as comparison group. Robust standard errors are clustered at the homeland-birth decade level.

G.8 Alternative Specifications

In this section, we examine the robustness of our key findings that the homeland policy reduced educational attainment to a variety of alternative specifications. Appendix Tables G.4 through G.5 present these robustness checks, with column (1) of each table repeating our baseline.

Columns (2) and (3) of Appendix Table G.4 examine the sensitivity of our results to alternative controls. Column (2) removes the demographic controls for gender and ethnicity and the fixed effects for Magisterial District of residence. The omission of these controls reduces slightly the magnitude of the relationship, and standard error increases by about 50 percent, but the coefficient remains highly significant at the conventional levels of significance. Column (3) reincorporates these controls and also adds fixed effects for the Magisterial Districts of birth. The inclusion of Magisterial District of birth fixed effects identifies the parameter of interest from within comparisons between cohorts from the same granular place of origin and residence but with different ages of exposure. If anything, we find that the relationship between the homeland policy and educational attainment tends to be more negative and estimated with greater precision in this alternative specification.

Our main results include homeland-to-homeland movers in the comparison group. Column (4) limits the comparison group to homeland natives who were residing in their homeland of birth. The point estimate is virtually unchanged and remains highly significant.

In the rest of columns, we investigate how sensitive results are to different geographical restrictions. The baseline results exclude the Bophuthatswana homeland because there is no information on the province of birth for individuals residing in this homeland but born in White South Africa. Column (5) explores the robustness of the results to this omission. Rather than including the robust set of fixed effects for province of birth, we use an indicator for being born in a White area. In this way, we can estimate our baseline model including Bophuthatswana in our sample. The estimated coefficient is slightly smaller in magnitude compared to the baseline and estimated with less precision, but it remains highly significant. Reduced precision can come about because including this homeland increases sampling variance or because we are unable to include the robust set of province-of-birth fixed effects. For completeness, column (6) repeats our baseline specification in the main analysis sample that excludes Bophuthatswana but removes province-of-birth fixed effects and includes instead an indicator for being born in a White area. The estimated coefficient and standard error are very similar to that derived from the extended sample that includes Bophuthatswana. This suggests that it is the omission of the province-of-birth fixed effects that drives the reduction in precision in column (5) and not the inclusion of the Bophuthatswana homeland. In sum, the omission of Bophuthatswana is unlikely to affect quantitatively and qualitatively our main findings.

Column (7) explores how results vary when excluding the largest homeland in our sample, the Gazankulu Bantustan, which accounts for about 40 percent of all observations. In addition, Appendix Figure G.12, Panel A reports estimates when each homeland is excluded one by one. Column (8) excludes individuals residing in the largest Magisterial Districts. Columns (9) and (10) drop individuals born in the largest provinces/homelands and Magisterial Districts respectively. These restrictions hardly affect our coefficient of interest and its significance. This suggests that our findings are unlikely to be driven by a particularly influential place or cohort.

Our main measure of exposure given by equation (3) relies on the presumption that earlier exposure matters more, based on recent work that circumstances and events experienced at earlier ages are more important for skill formation and the subsequent acquisition of human capital (Heckman, 2007). This specification also receives support in the non-parametric figures. In Appendix Table G.5, we explore alternative exposure definitions. In column (2), we use an indicator for exposure at age 15 or younger rather than the baseline measure of the total number of childhood exposure. Column (3) uses linear spline terms for ages 9 to 15, 15 to 23 and 23 or older at homeland establishment. The slope of the segment for ages 16 and older allows for a test to detect pre-trends, while the slopes between the other segments provide a trend-break of the effects. The results from these alternative specifications are broadly consistent with the basic pictures shown in the Table 1 and Figure 5.

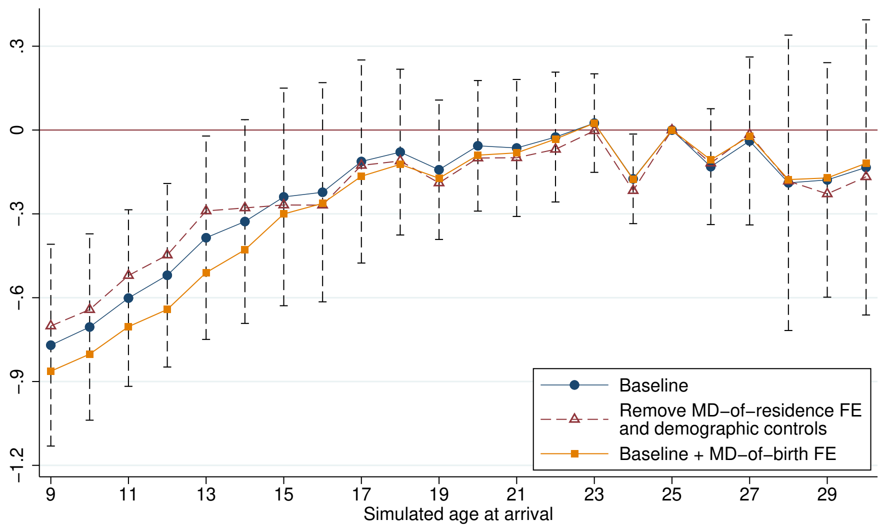
In Appendix Figure G.13, we explore wider age ranges to define exposure. When we widen the set of childhood ages, the estimated relationship becomes smaller and the pattern of coefficients becomes flatter, which is consistent with the shape shown in the non-parametric figures. This provides further support to our baseline measure of childhood exposure.

Table G.4: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: years of education									
Simulated Exposure × *Migrant*	-0.116	-0.0956	-0.1279	-0.1356	-0.1098	-0.1107	-0.1034	-0.0937	-0.1131	-0.1193
	[0.0198]	[0.0319]	[0.0198]	[0.0231]	[0.0271]	[0.0269]	[0.0285]	[0.0203]	[0.0232]	[0.0208]
Homeland × birth-year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Province-of-birth FE	✓	✓	✓	✓			✓	✓	✓	✓
Magisterial District-of-residence FE	✓		✓	✓	✓	✓	✓	✓	✓	✓
Magisterial District-of-birth FE			✓							
Demographic controls	✓		✓	✓	✓	✓	✓	✓	✓	✓
Migrant indicator					✓	✓				
Different geographic samples:										
Exclude homeland-to-homeland movers				✓						
Include Bophutatswana					✓					
Exclude largest homeland							✓			
Exclude largest Maisterial Districts								✓		
Exclude largest province of birth									✓	
Exclude largest Magisterial Districts of birth										✓
Observations	2,489,596	2,489,596	2,489,596	1,954,089	2,869,867	2,489,596	1,175,846	2,072,674	1,674,370	2,438,802

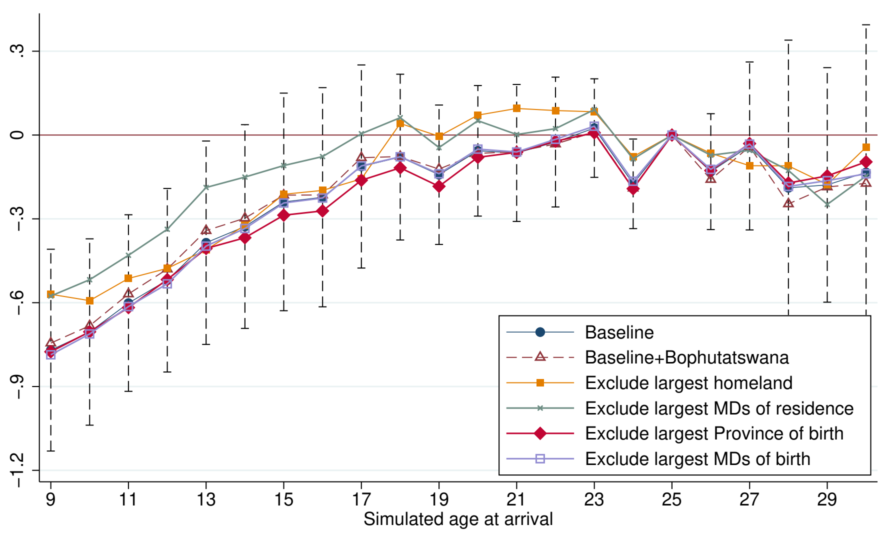
Notes: Estimates are based on individuals residing in a homeland at the time they are observed in the 1980 census. Simulated exposure refers to the predicted number of childhood years in the homeland of residence. Migrant is an indicator for migrants from White areas. All regressions include controls for gender and ethnicity. Each column reports results from alternative specifications.

Figure G.10: Estimated Effects on Years of Education
(Alternative controls)



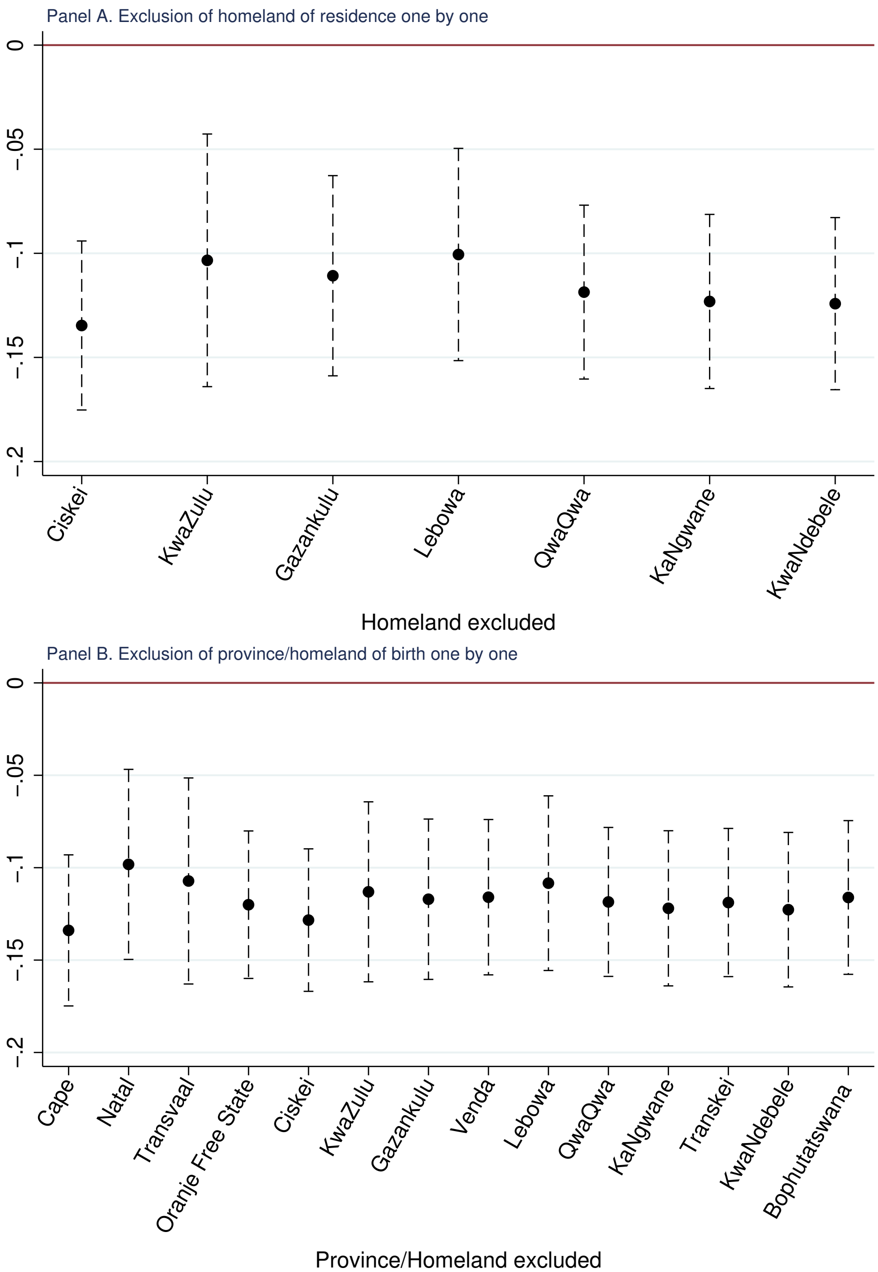
Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating equation (5). Figure also shows results from alternative specifications. The sample is based on the 1980 census. Controls include homeland-of-residence × birth-year fixed effects, Magisterial District fixed effects, province-of-birth fixed effects, and indicator for gender and ethnicity.

Figure G.11: Estimated Effects on Years of Education
(Different Geographical Samples)



Notes: This figure plots β_τ (and 95 percent confidence intervals) from estimating equation (5). Figure also shows results from alternative specifications. The sample is based on the 1980 census. Controls include homeland-of-residence × birth-year fixed effects, Magisterial District fixed effects, province-of-birth fixed effects, and indicator for gender and ethnicity.

Figure G.12: Estimated Effects on Years of Education (Exclusion of Different Areas One by One)



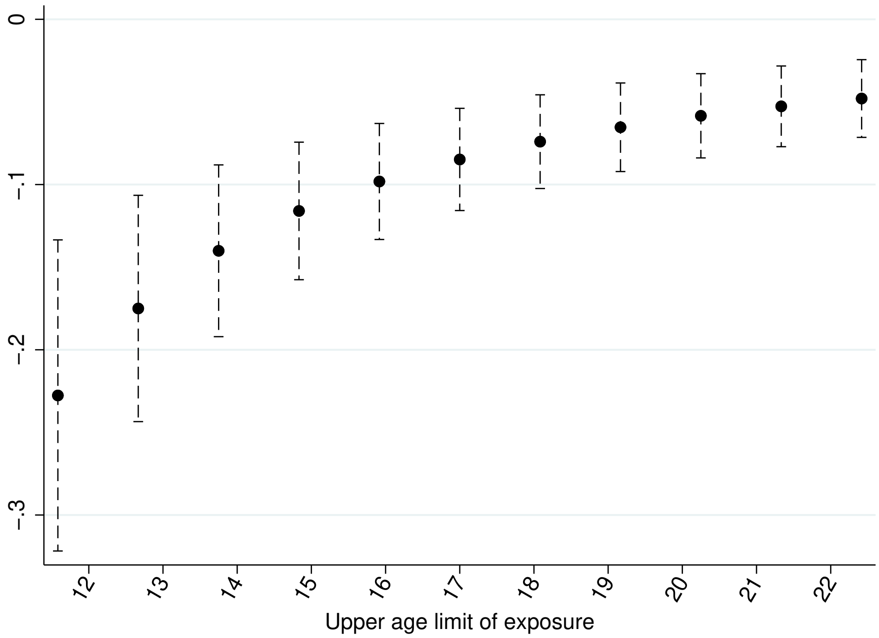
Notes: This figure plots β (and 95 percent confidence intervals) from estimating the equation (5). Panel A shows results from excluding homelands one by one. Panel B shows results from excluding province/homelands of birth one by one. Sample is based on the 1980 census. Controls include homeland $\times$ birth-year fixed effects, Magisterial District fixed effects, province-of-birth fixed effects, and demographic controls for gender and ethnicity. Robust standard errors are clustered at the homeland-birth decade level.

Table G.5: Alternative Exposure Definitions

	Dependent variable years of education		
	(1)	(2)	(3)
Simulated Exposure × Migrant	-0.116		
	[0.0198]		
Simulated Exposure Indicator × Migrant		-0.4356	
		[0.0819]	
Spline Coefficients:			
Simulated age at arrival <15			0.1022
			[0.0329]
15< Simulated age at arrival <23			0.0357
			[0.0255]
Simulated age at arrival >23			-0.0225
			[0.0305]
Homeland × birth-year FE	✓	✓	✓
Province-of-birth FE	✓	✓	✓
Magisterial District FE	✓	✓	✓
Demographic controls	✓	✓	✓
Observations	2,489,596	2,489,596	2,489,596

Notes: Estimates are based on individuals residing in a homeland at the time they are observed in the 1980 census. Simulated exposure refers to the predicted number of childhood years in the homeland of residence. Migrant is an indicator for migrants from White areas. All regressions include controls for gender and ethnicity. Each column reports results from alternative exposure definitions.

Figure G.13: Estimated Effects on Years of Education
(Upper Age Limit of Exposure)



Notes: This figure plots β (and 95 percent confidence intervals) from estimating the equation (5), but considering alternative age ranges to define childhood exposure. Sample is based on the 1980 census. Controls include homeland $\times$ birth-year fixed effects, Magisterial District fixed effects, province-of-birth fixed effects, and demographic controls for gender and ethnicity. Robust standard errors are clustered at the homeland-birth decade level.

G.9 Addressing Selection in the Income Sample

As discussed above, one complication with interpreting the results on the intensive margin of income is that the homeland policy led to a decline in employment rates, potentially changing the sample of individuals with positive income in our data. To address this issue, we impute information for missing observations as result of the homeland policy under best- and worst-scenarios. The lower bound assumes that the "extra" unemployed individuals as a consequence of the policy change would be located at the top of the income distribution. The upper bound assumes that "extra" unemployed individuals would be located in the bottom of the income distribution.

To implement this method, we first rerun the flexible specification (5) for employment and saved the coefficients for exposure between ages 7 to 14. We then condition the sample to migrants from White areas and define interactions between place of birth, homeland, Magisterial District, age, and sex, resulting in 1,008 mutually exclusive groups. For each subsample N_g and for each exposure age $\tau = 7, 8, .., 14$, we extend the sample by creating $N_{g,\tau} = N_g \times \beta_\tau$ new observations. This procedure therefore adds the "extra" number of observations of White-place migrants that are missing in the income sample as result of the homeland policy. To calculate lower bounds, we assign the seventy-fifth percentile of income distribution for all extra $N_{g,\tau}$ observations. Upper bounds assign the twenty-fifth percentile of income distribution for all extra $N_{g,\tau}$ observations. The twenty-fifth and seventy-fifth percentiles of the income distribution are calculated separately for each subgroup.

This procedure follows the idea underlying the approach proposed by Lee (2009), but instead of trimming observations in the comparison group we add the policy-induced missing observations in the treatment group. Our procedure is easier and more natural to implement in our context given the nature of our treatment variable.

G.10 Contribution of Education to the Effect on Income

In this section, we investigate to what extent the large and negative effect of the homeland policy on income is mediated by the negative effect on education. Appendix Table G.6 addresses this question by directly incorporating education in the income regression. Column (1) shows the baseline coefficient of -0.008. Column (2) includes completed years of education as a control variable. Column (3) controls more flexibly for education by incorporating dummies for each year of schooling. Once one accounts for the relationship between income and education, the estimated effect of the homeland policy on income falls by as much as 37 percent and is marginally significant at the 10 percent.

Of course, this approach may under- or over-estimate the role of education because education is an endogenous variable. As an alternative approach, we restrict the relationship between education and income by using estimates of the economic returns to schooling from well-identified studies in the literature. In the context of a developing country, Duflo (2001) finds estimates ranging from 6 to 14 percent. Columns (4) through (7) partial out the effect of schooling using these suggested rates of returns to schooling. We obtain coefficients that go from -0.06 to -0.05. Reduced educational attainment as a result of the homeland policy accounts for about 25-37 percent of the actual impacts on income. The remaining fraction may be due to changes in school quality or in the supply of other inputs that are important in the production function of child quality, but that are not completely captured by the number of years of schooling.

Table G.6: Effects on Income, as Mediated by Years of Schooling

	Dependent variable: log total income						
	Baseline estimate	Returns to schooling estimated		Returns to schooling constrained to a particular value			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Simulated Exposure × *Migrant*	-0.008	-0.005	-0.005	-0.006	-0.006	-0.005	-0.005
	[0.002]	[0.003]	[0.003]	[0.002]	[0.003]	[0.003]	[0.003]
Years of schooling		0.129		0.08	0.1	0.12	0.14
		[0.005]					
Dummies for years of schooling?			✓				
Homeland × birth-year FE	✓	✓	✓	✓	✓	✓	✓
Province-of-birth FE	✓	✓	✓	✓	✓	✓	✓
Magisterial District FE	✓	✓	✓	✓	✓	✓	✓
Observations	631,070	621,039	621,039	621,039	621,039	621,039	621,039

Notes: Estimates are based on individuals residing in a homeland at the time they are observed in the 1991 census. Simulated exposure refers to the predicted number of childhood years in the homeland of residence. Migrant is an indicator for migrants from White areas. All regressions includes a dummy for gender. Robust standard errors are clustered at the homeland-birth decade level. Columns (4)-(7) constraint the coefficient on years of schooling to the reported value. Robust standard errors are clustered at the homeland × birth decade level.

G.11 Returns to Schooling for Migrants and Locals

Table G.7: No Differential Returns to Schooling for Migrants

	(1)	(2)
	Dependent variable:	
	Probability of working	Log total income
Years of schooling	0.020	0.122
	[0.001]	[0.003]
Years of schooling × *Migrant*	-0.000	-0.007
	[0.002]	[0.006]
Observations	4,110,892	1,018,426

Notes: This table estimates whether there are differences in the marginal effect of schooling for migrants from White areas. Controls include an indicator for migrants, age, gender, ethnicity, and Magisterial District fixed effects. Robust standard errors are clustered at the homeland × birth decade level.

References

Beer, Cedric De, *The South African disease: Apartheid health and health services*, Africa World Press, 1986.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American economic review*, 2009, *99* (4), 1145–77.

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

Duflo, Esther, "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," *American economic review*, 2001, *91* (4), 795–813.

Greenstone, Michael and Rema Hanna, "Environmental regulations, air and water pollution, and infant mortality in India," *American Economic Review*, 2014, *104* (10), 3038–72.

Heckman, James J, "The economics, technology, and neuroscience of human capability formation," *Proceedings of the national Academy of Sciences*, 2007, *104* (33), 13250–13255.

Horrell, Muriel, *A survey of race relations in South Africa*, Vol. 26, Univ of California Press, 1973.

Lee, David S, "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Surplus People Project, "Forced removals in South Africa : Volume 1-5 of the Surplus People Project report," Technical Report, Cape Town: Surplus People Project 1983.

Tanaka, Shinsuke, "Does abolishing user fees lead to improved health status? Evidence from post-Apartheid South Africa," *American economic Journal: economic policy*, 2014, *6* (3), 282–312.