For Online Publication

Appendix 1: Data

1. Rainfall Data and Drought

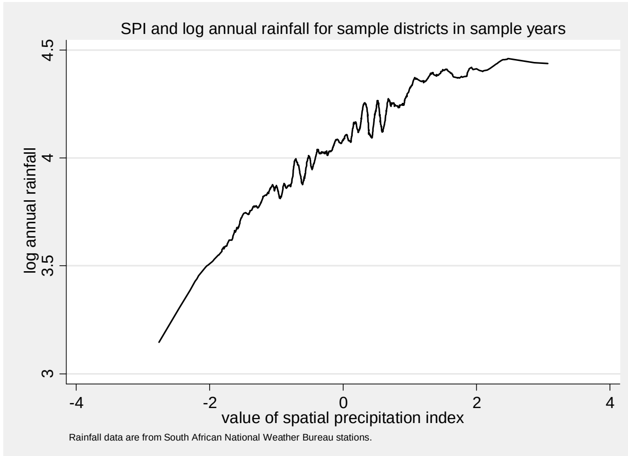
The South African Weather Service http://www.weathersa.co.za/web/ provided the raw historical rainfall data. These data contain monthly rainfall measures at the weather station level for over 1,600 weather stations across South Africa from 1920 to 2009. I spatially match the GIS locations of rainfall stations to corresponding district boundaries and aggregate rainfall totals to the district-year level.

To create a measure of drought, I construct the Standardized Precipitation Index (SPI) at the district and year level (McKee, Doesken and Kleist 1993). The SPI measures the probability of observing a recent rainfall event based on the distribution of all rainfall events for a given time scale and place. Since rainfall is not normally distributed, the SPI procedure calls for a gamma distribution to be fit to the empirical data distributions. I fit a gamma distribution to the annual total rainfall of each district. and generate estimates of the scale and location parameters for district-specific rainfall patterns. For each year in the data, and the district-specific gamma distribution, I compute the probability of observing the total rainfall that was measured in each year and translate this into a normally distributed random variable using the normal CDF. This number is the district-year-specific SPI, where positive numbers reflect above-average rainfall and negative values reflect below-average rainfall. The positive relationship between log rainfall and the SPI measure across all districts is shown in Appendix Figure 1.

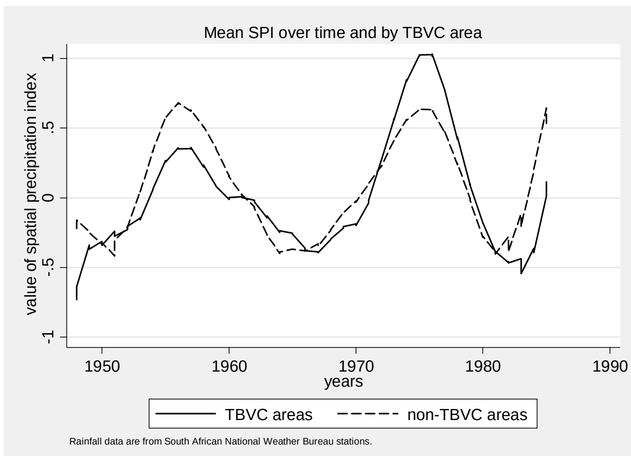
Following the climatological literature (e.g. McKee et al 1993) I define an indicator $DROUGHT_{dt}$ for each district (*d*) and year (*t*) that takes a value of 1 when the SPI is less than -1.5, and 0 otherwise. Appendix Figure 2 shows lowess-smoothed graphs of the district level mean SPI values across TBVC and non-TBVC areas for the years 1948 to 1986.

There is a tight link between the SPI measure and South African maize production. Using province-level data from the South African Maize Board for the period 1964 to 1984 and for the commercial maize-growing provinces (Transvaal and the Orange Free State), I estimate the relationship between the SPI measure and maize yields. Appendix Figure 3 shows the lowess-smoothed relationship between the log of South Africa's annual maize output (in tons) against the Spatial Precipitation Index using an Epanechnikov kernel with a 0.5 bandwidth. This positive relationship is asymmetric. Output appears more sensitive to low values of the SPI than it is to higher, positive values of the SPI. Figure 3 suggests that drought in particular captures an important negative productivity shock in agriculture.

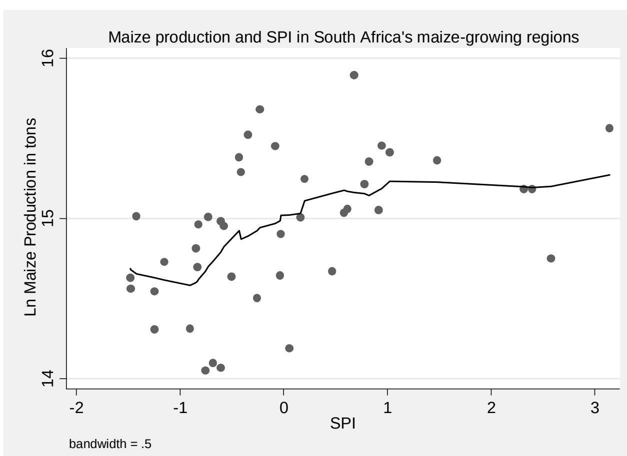
Appendix Figure 1



Appendix Figure 2



Appendix Figure 3



2. Homeland Boundary Data and the TBVC assignment

GIS data on sub-national boundaries for the 1996 and 2001 Census were obtained from Statistics South Africa (www.statssa.gov.za). I use the 2001 district council Census boundaries as the main geographic unit of observation since these areas are large enough to treat as distinct local labor markets and contain sufficient population in each year to make aggregation feasible.[1]

To define which of these districts belong to former homeland areas, I obtained online maps of the ten homelands with the predominant map dated 1986 (see Figure 1 in the main text). I overlaid these homeland maps onto Census boundaries and, where there was overlap, assigned districts to homelands. I created an indicator $TBVC_d$ that takes a value of 1 if a district overlapped with any of the TBVC homelands, and is 0 for those districts overlapping the remaining six homelands.[2] Of the 53 district councils in South Africa, 16 of them (30%) fall into prior rural homeland areas. Of these 16 areas, six fall in the former TBVC areas and the remaining 10 fall in the non-TBVC areas.

3. 1996 Census Data

The 10% sample of individual records from the 1996 South African Census was obtained from Statistics South Africa (www.statssa.gov.za).

i. Migration variables

Migration is difficult to measure well using household survey data. The benefit of using Census data is that it provides comprehensive coverage of migrant groups across the country, unlike general household surveys that draw from a subset of districts. The coverage of the Census also allows aggregation of migration data to broader geographic units (for our purposes, the district level).

I use information from Census questions: "Where do you live now? Where did you live before this? What year did you move here?" to construct migration variables. Specifically, I define:

- The district of current residence
- The district of former residence: this is the same as district of current residence for individuals who report never having moved, for individuals who have moved since the end of apartheid (1994-1996), for individuals who reported moving during childhood to their current residence, and for those who report moving within a district
  An indicator for whether a person moved before 1994 (the end of apartheid) and during their adult lives. This indicator is 1 if a person's former residence differs from their current residence and if they report the year they moved to their current residence.

[1] Magisterial districts are too small to contain sufficient population and rainfall measurements for my analysis.
[2] TBVC stands for Transkei, Boputhatswana, Venda and the Ciskei.

The sample is restricted to African adults aged 18 and older in 1996 who report a current (for never movers) or prior residence (for movers) in South Africa that is predominantly rural and located in one of the former homeland areas. I eliminate individuals reporting a current residence (for never movers) or a prior residence (for movers) in a district outside of any of the homelands, and those who report living in (for never movers) or formerly living in (for movers) districts outside of South Africa. Less than 1% of the sample has a usual residence outside South Africa and less than 5% have a prior residence outside of South Africa. Of the remaining sample of adults who report a former residence (for movers) or current residence (for never movers) located in rural South Africa, 97% have complete information on current and former district of residence and the year of moving to current residence. For the 3% who report a current residence and no information about year of moving, I assign them to be non-movers.

Using the migration variables, I construct a pseudo-panel dataset of individual-year observations capturing where each person lived in each year. This dataset indicates whether a person moved out of a given district in any given year, based on their "last move" information and allows me to observe the total number of adults (ages 18 and over) in each district in each year between 1948 and 1986 and the number who move away from each district in each year. I collapse the data to district-year level and generate the percent of adults living in each district that migrated away in each year. I use this district-year level dataset to describe historical outmigration from each district in each year in Tables 1, 2 and Figure 2.

ii. Health, fertility and population outcomes

For the disability analysis, and for the analysis of cohort size and sex composition, I use the sample of African adults who lived in any of the former homeland areas between 1948 and 1986. I match the cross-sectional data on outcomes at the individual level to the drought data on year of birth and prior district.

Note that the Census does not capture place of birth information, so I assume that a person's prior residence is their birth district. This means that birth district is potentially misclassified for people who move multiple times.[3]

For the analysis of fertility and child mortality outcomes, I restrict the sample to African women aged 40 to 60 in 1996 who have completed childbearing. I create a variable that represents the fraction of their childbearing years (ages 15-40) that they experienced drought. I assign drought exposure at the district level using the prior district reported by these women.

[3] Data from the 2007 South African Community Survey and the 2007 Cape Area Panel Study indicate that the fraction of Africans who move more than once in the past five years is between 0.01 and 0.13 respectively. The 2007 Community Survey collects data on more than 300,000 African adults including their province of current and prior residences. The 2007 Cape Area Panel Survey is a sample of young adults (ages 24-33) drawn from a province with a highly mobile population, hence the higher rates of misclassification. Older data from the 1997 October Household Survey corroborates these numbers: among African adults aged 15 and older in 1986, only about 4% of them report moving at all across magisterial districts in any year prior to 1986. And, under 1% report multiple cross-district moves during the *apartheid* period.

4. Baseline District-level control variables

To control for key district level variables that could contribute to differential responses of disability rates to drought events, I create:

- District-level population density recorded in the 1946 South African Census. These data were digitized from hardcopies of Census aggregate reports and matched to later district boundaries
- The straight-line distance (in kilometers) from the midpoint of each district polygon to the nearest large city. Distances were calculated using 1996 Census maps in ArcGIS version 10
- A district's median score on the FAO's maize suitability index. This index captures how suitability land is for maize production at a fine grid level. I use these values to create the median value of the suitability index across all points in given district in ArcGIS. Low values of this number represent greater suitability of the soil for maize production

References

McKee, T. B., N. Doesken and J. Kleist 1993 "The relationship of drought frequency and duration to time scales", Preprints, 8th Conference on Applied Climatology, pp. 179–184. January 17–22, Anaheim, California