Identifying the Rich: Registration, Taxation, and Access to the State in Tanzania
Jeremy Bowles

Outcome datasets

census.dta

This file contains outcomes as observed in the 2012 Population and Housing Census. District of birth, as observed in this file, is merged into the covariate datasets (treatment_map.dta, schools.dta, taxation.dta) to define treatment assignment (see make_tables_census.do). Source: Tanzania National Bureau of Statistics.

b_region_txt	: Region of birth (2012 administrative units).
b_district_txt	: District of birth (2012 administrative units).
b_district	: District of birth code (2012 administrative units).
b_region	: Region of birth code (2012 administrative units).
b_cert	: Respondent possesses a birth certificate.
yob	: Year of birth of respondent.
age	: Age of respondent.
pl_male	: Respondent is male.
pl_tanzanian	: Respondent is Tanzanian national.
pl_alive_father	: Respondent's father is alive.
pl_alive_mother	: Respondent's mother is alive.
lit_any	: Respondent is literate in any language.
lit_kiswa	: Respondent is literate in Kiswahili.
lit_eng	: Respondent is literate in English.
ed_primary	: Respondent has any primary education (above P1).
ed_secondary	: Respondent has any secondary education (above S1).
ed_uni	: Respondent has any university education.
ed_attain_1	: Respondent has at least Primary standard 1 education.
ed_attain_2	: Respondent has at least Primary standard 2 education.
ed_attain_3	: Respondent has at least Primary standard 3 education.
ed_attain_4	: Respondent has at least Primary standard 4 education.
ed_attain_5	: Respondent has at least Primary standard 5 education.
ed_attain_6	: Respondent has at least Primary standard 6 education.
ed_attain_7	: Respondent has at least Primary standard 7 education or Post-primary training.
ed_attain_8	: Respondent has at least Secondary form 1 education.
ed_attain_9	: Respondent has at least Secondary form 2 education.

ed_attain_10 : Respondent has at least Secondary form 3 education.
ed_attain_11 : Respondent has at least Secondary form 4 education.
ed_attain_12 : Respondent has at least Secondary form 5 education.
ed_attain_13 : Respondent has at least Secondary form 6 education or Post-secondary training.
ed_attain_14 : Respondent has University and other related education.
i_hhid : Household identifier.
z_asset_pca : Standardized first principal component from full set of indicators for asset ownership recorded in census (radio; telephone; mobile phone; bike; motor vehicle; motorcycle; guta; bajaj; television; electric iron; charcoal iron; cooker; fridge; computer; internet; plough; power tiller; hand hoe; wheelbarrow; oxen; donkey/camel; house; land).
s_any : Respondent is in a household with access to any social security fund.
s_nssf : Respondent is in a household with access to National Social Security Fund (NSSF).
s_pension_parastatal : Respondent is in a household with access to Parastatal Pension Fund (PPF).
s_pension_pubservice : Respondent is in a household with access to Public Service Pension Fund (PSPF).
s_pension_govemp : Respondent is in a household with access to Government Employee Provident Fund (GEPF).
s_pension_locauth : Respondent is in a household with access to Local Authority Pension Fund (LAPF).
s_pens_priv : Respondent is in a household with access to any social security fund indicating prior private sector employment (NSSF or PPF).
s_pens_pub : Respondent is in a household with access to any social security fund indicating prior public sector employment (PSPF, GEPF or LAPF).
s_nih : Respondent is in a household with access to National Health Insurance Fund (NHIF).
s_other : Respondent is in a household with access to any other social security fund.

nps.dta

This file contains outcomes as observed in the R2 and R4 of the National Panel Survey. District of birth, as observed in this file, is merged into the covariate datasets (treatment_map.dta, schools.dta, taxation.dta) to define treatment assignment (see make_tables_nps.do). Source: Tanzania National Bureau of Statistics.

b_district : District of birth code (2012 administrative units).
yob : Year of birth of respondent.
age : Age of respondent.
b_cert : Respondent has a birth certificate.
ed_parent_pri : Respondent has at least one parent with any primary education (above P1).
ed_parent_sec : Respondent has at least one parent with any secondary education (above S1).
ed_parent_uni : Respondent has at least one parent with university education.
pl_male : Respondent is male.

tax_all : Household paid any tax to government in prior years.
tax_fee : Household paid any government fees or fines in prior years.
tax_local : Household paid any local government taxes or council rates in prior years.
tax_central : Household paid any central government taxes on income or property in prior years.
z_asset_pca : Standardized first principal component from full set of indicators for asset ownership recorded in NPS (see NPS Section N documentation for full list of assets recorded).

Data/District-level/treatment_map.dta
This file maps districts as observed in the outcome data (both census and NPS) to administrative units as existing during the reform period, including information on their assignment to the reform during that period.

census_2012 : Text for district name provided in 2012 census file (Swahili).
district_2012 : Text for district name provided in 2012 census file (English).
region_1967 : Region (as defined in 1967 census).
district_1967 : District (as defined in 1967 census).
year_reform_parent : Year compulsory registration reform was first applied to any district sharing the same parent district.
year_reform : Year compulsory registration reform was applied to district (See Appendix A1 for additional information).
district_urban : District is urban (as defined in 1967 census) (indicator).
dcode : District code (as defined in 2012 census).
g_parent : District is either treated with compulsory registration in 1966 (g_parent=1) or shares pre-reform parent district with a district treated in 1966 based on year_reform_parent=1966 (g_parent=0) (indicator). Used for baseline definition of analysis sample.
g_urban : District is either treated with compulsory registration in 1966 (g_urban=1) or was an urban locality as defined in 1967 but not treated with reform (g_urban=0) (indicator). Used for first stage robustness in Table A4.
wave : Reform wave during which district had compulsory registration reform applied (1966, 1980s, 2009) [see Table A11].
district_1967_g : District (as defined in 1967 census) plus reform status for districts containing a treated town. Used for baseline fixed effects in regressions.

Data/District-level/taxation.dta
This file contains district-level covariates from Jensen & Mkama (1968) and Lee (1965). Used primarily in Table 4.

district_1967 : District (as defined in 1967 census).
jm_pop_dens : Population density (population per square mile).
jm_gdp_cap : GDP per capita (‘000s of Tanzanian shillings).
jm_rate_payers_cap : Number of tax payers per capita.
jm_tax_cap : Total amount of tax collected per assessed taxpayer (‘000s of Tanzanian shillings).

lee_diff_tax	: Absolute difference in tax rate faced by highest bracket relative to tax rate faced by lowest bracket (Tanzanian shillings).
jm_s_enrollment_cap	: School enrollment per 1000 population.
jm_h_beds_cap	: Hospital beds per 1000 population.
jm_hh_size	: Average household size.
jm_pop_growth	: Population growth rate (1967 relative to 1957).
jm_emp_ag_share	: Employment share of agriculture.

Data/District-level/schools.dta

This file contains district-level information on the presence of primary/secondary schools existing in a given district prior to the initial compulsory registration reform (based on administrative data).

dcode	: District code (as defined in 2012 census).
sch_primary_n_66	: Number of primary schools existing in district in 1966.
sch_secondary_any_66	: Secondary school existed in district in 1966.