# FOR ONLINE PUBLICATION

## APPENDIX 1: DATA SOURCES

This appendix describes all of the data sources used in the paper. Summary statistics for the variables we construct are reported in Appendix Table 1.

### 1. Road Expenditure Data

We construct a district-year road expenditure (in 2000 USD) panel data set annually for the period 1963 to 2011. Total road expenditure is the sum of development expenditure [new investments] and recurrent expenditure [maintenance]. The *Annual Development Estimates of Kenya*, our main data source, allows us to track only road development expenditure at the district level.[25] In particular, Annual Development Estimate list programmatically individual road projects (e.g. project *Thika Main Road* from *Thika* town to *Nyeri* town via *Limuru* town) and their related costs.[26] When a road project spans more than one district, we use GIS tools to deconstruct the road network in question and calculate the length of kilometers within each district. Hence, for projects which span across more than one district the expenditure share is distance-weighted. For the period 1963-1973, development estimates for road expenditure are not documented as individual road projects, instead, only large nation-wide road programs are reported (and their costs). To breakdown these aggregate road programs into individual projects and to obtain the district level expenditure, we supplement our data with the Development Plans of Kenya (usually a four year plan) and World Bank project reports.[27] This exercise allows us to construct a road expenditure district-year panel data set of 2009 observations (41 districts tracked for 49 years).

### 2. Road Construction Data

We create a district-year road construction (unbalanced) panel data set using a novel GIS database of the Kenyan road network on regular intervals for 1963-2011. In particular we have road maps for the years: 1964, 1967, 1969, 1972, 1974, 1979, 1981, 1987, 1989, 1992 and 2002. To construct the road network, we use a baseline the latest GIS database containing contemporary roads from *Global GIS*. We use a series of historical maps to recreate the evolution of the road network in GIS.[28] We are only able to consistently track the evolution of the *paved* roads

[25] Road expenditure is reported in East African Pounds 1963-1966, Kenyan Pounds 1967-1999, and Kenya Shillings 2000 onwards. For consistency we use Officer (2009) and IMF (2011), to convert these amounts to current US$ and further using a US$ deflator series these amounts are constructed in constant 2000 USD.

[26] These reports are the Development Estimates of Kenya 1963-2011, Physical Infrastructure Sector MTEF Report of Kenya 2007/2008-2009/2010 and 2009/2010-2011/2012. We further use Recurrent Estimates of Kenya 1963-2011 to make comparisons with aggregate spending on road expenditure to: (i) the total budget, (ii) expenditure in other public goods (education, health and water development) and (iii) the road maintenance budget.

[27] The Development Plan of Kenya: 1964-1966, 1966-1970, 1970-1974 and 1974-1978. Most road programs during this era were either fully or partially funded under the IDA program of the World Bank, we collate all the Road Program Operational Reports [available on `http://www.worldbank.org/projects`, accessed on November 2011]. These Operational Reports contain for each program the individual projects and their respective costs.

[28] We use the road map series *Michelin National Map for Central and South Africa* for 1964, 1967,

network. Tracking of non-paved roads (improved, laterite, dirt roads) provides a challenge due to definitional changes in the legends of the maps as well as omissions on several occasions. We use categories in the Michelin map *motorways* and *hard-surfaced roads* to define our measure of paved roads. The few *Survey of Kenya* maps available during the period 1967-1991 are further used as a robustness check to the Michelin series. Two limitations on recent map availability are worth mentioning. Firstly, we use the 2002 *Survey of Kenya* map to recreate the paved network in 2002. Secondly, no updated road maps are available for the post-2002 period. The latest government survey of the road network was undertaken in 2002 and a more updated inventory was planned in 2010 but this has not surfaced yet in the public domain. Further, commercial mapping agencies have not been able to update their mapping series post-2002 as these agencies use the official survey maps as a baseline.[29] Using GIS we splice the road networks for the respective years with district boundaries to create a paved road construction (km) district-year panel data set of 451 observations (41 districts tracked for 11 years).

## 3. Ethnic Census

Our primary ethnic demographic data is obtained from the housing and population census of 1962 (Government of Kenya (1965)). The population census of 1962 collected ethnic demographics at the disaggregated level of the sub-location (168), a unit of administration below the district. Using GIS tools to create a post-independence 1963 digital district map we aggregate from the sub-location to district level construct of each district (41) its ethnic demographics. The population census reports 41 ethnic classifications. In line with studies on the political economy of Kenya we aggregate the classifications into 13 ethnic groups.[30] For each district we create a set of 13 dummies equal to 1 if more than 50% of district population is from a certain ethnic group $X$ and 0 otherwise. These district dummies are used to construct the other key variables in the analysis. *Coethnic District Indicator [d,t]* is a dummy equal to 1 if more than 50% of the district $d$ population is from the ethnic group of the president in year $t$. The evolution of this variable is as follows, the ethnic group of the president is Kikuyu during Kenyatta's era 1963-1978, Kalenjin during Moi's era 1979-2002 and Kikuyu again during Kibaki's tenure 2003-2011. The *Democracy Indicator [t]* is a dummy equal to 1 if the year $t$ is a democratic year and zero otherwise. The dummy variable takes the value 1 during the period 1963-1969 and 1993-2011 and 0 in the interim periods.[31] The *Kikuyu District Indicator [d,1962]* (resp. *Kalenjin District Indicator [d,1962]*) is a dummy equal to 1 if more than 50% of the district $d$'s

1969, 1972, 1974, 1979, 1981, 1984, 1987, 1989 and 1992; and *Survey of Kenya* maps for 1967, 1972, 1991 and 2002.

[29]This was reveled in several discussions with archivists at the British Library (London) and Michelin (Paris).

[30]Kikuyu, Kalenjin, Kamba, Luo, Luhya, Maasai, Coastal, Embu, Kisii, Meru, Somali, Turkana-Samburu and Other (which are Other Africans, Arabs, Asians, Non-Africans).

[31]Note Kenya's Development Estimates publications provide expenditures for the year $t$ for the period from July $t-1$ to June t. Moi has an influence from 1979 (July 1978-June 1979) and Kibaki has an influence from 2003 (July 2002-June 2003). Similarly, the transition to autocracy in November 1969 is considered from 1970 (July 1969-June 1970) and the transition to democracy took place in December 1992 and is in considered from 1993 (July 1992-June 1993).

population is Kikuyu (resp. Kalenjin) in the 1962 population and housing census. The *Coethnic Group Indicator [e,t]* is an indicator variable whose value is 1 if the president at time $t$ belongs to ethnic group $e$ and 0 otherwise. The *VP-Coethnic District Indicator [d,t]* is an indicator variable whose value is 1 if more than 50% of the population of district $d$ is from the ethnic group of the vice-president at time $t$. The *VP-Group Indicator [e,t]* is an indicator variable whose value is 1 if the vice-president at time $t$ belongs to ethnic group $e$. In Appendix Table 1 (Panel A) we provide the national population share of the main ethnic groups across modern Kenya. The data tabulated on ethnic composition was collected from housing and population census reports for the years: 1962, 1969, 1979, 1989 and 2009. The 1999 census though collected ethnic demographics however these were not released in the public domain. We supplement our sources with Kenya's DHS for the year 2003. The 2009 census released ethnic shares at the national level.

## 4. Control Variables: Demographic and Socioeconomic

We use supplementary datasets to construct three sets of control variables at the district level. Firstly, demographic variables: district population and urbanization rates are obtained from the population census of 1962. District area (in kms) is estimated using GIS area calculation of polygons on the digital district boundary created (Survey of Kenya, 1963). Secondly, initial economic activity variables are obtained using the *Statistical Abstracts of Kenya* to construct total district employment (1963) and total district earnings (1966) in 2000$ in the formal sector.[32] Further, the value of district cash crop exports is constructed at the district level using the *Development Plan of Kenya* 1964-1970 which provides reports of cash crop production (coffee, tea and sisal) at the district level for the year 1964/65.[33] Thirdly, economic geography variables: GIS tools are used to create a indicator variable which takes the value 1 if the district is on the main highway corridor between Mombasa-Nairobi-Kampala and 0 otherwise, another indicator variable is created which takes on the value 1 if any part of the district borders Tanzania or Uganda, the main two trading partners for the country and a district measure is created of the Euclidean distance between the centroid of the district and Nairobi (the capital).

## 5. Economic Growth, Ethnic Diversity and Democracy

Data on political regimes in Sub-Saharan African countries and Kenya is obtained from the *Polity IV Project*. We use the variable "Combined Polity Score" which takes values from -10 (hereditary monarchy) to +10 (consolidated democracy). Polity IV categories regimes into *autocracies* (-10 to -6), *anocracies* (-5 to +5) and *democracies* (+6 to +10). The average combined policy score for Sub-Saharan Africa is computed using the individual polity scores and population of each country as weights (obtained from World Bank 2011). GDP per capita growth in Sub-Saharan Africa is obtained using the World Bank (2011). In Table 6 we conduct a similar exercise to

[32]The data is reported in Kenyan Shillings, using Officer (2009) and IMF (2011), we convert these amounts to current US$ and deflate them to get constant 2000$.

[33]Using the 1965 export price in 2000$ FAO (2011) to calculate the district total value of cash crop exports in 1965.

Easterly and Levine (1997). We obtain Easterly and Levine's dataset and append their decadel dataset with two additional decades (1990s and 2000s).[34] We source the same data to expand the variables for the additional two decades. Namely initial income and GDP is obtained from *Penn World Tables 7.1*. Annual GDP per capita is used to compute the growth of per capita real GDP. The indicator variable *democracy [c,d]* is a variable whose value is one if country $c$ is not an autocracy in decade $d$, specifcally if the average combined polity score for decade $d$ is strictly less than -5, the threshold defined by *Polity IV Project*. The variable *ethnic [c,1960]* is obtained from Easterly and Levine (1997) and is the ethnolingustic fractionalization of country $c$.

## 6. Cabinet Composition

We construct data on the ethnicity and position of all cabinet members between 1963 and 2011 (13 cabinets). This allows us to track the evolution of each ethnic group's representation in the politics of Kenya. Several data sources are used to compile the data which contains the name and position of each cabinet member (president, vice-president, prime minister from 2008, and ministers) between 1963 and 2011. These publications are: *The National Assembly: List of Members*, *Organization of the Government of Kenya*, and *Encyclopedia of Sub-Saharan Africa: Kenya*. While the ethnicity of prominent cabinet members is well-known, the information for the less prominent politicians is obtained in several ways. We use secondary sources: (i) the *Weekly Review* magazine in the Moi period would often list out the cabinet and ethnicity of cabinet members, (ii) the descriptive work done by various political scientists on Kenyan politicians, especially Hornsby (1985) and Ahluwalia (1996), and (iii) the direct help of several journalists from top dailies in Kenya. Collating these sources allows us to calculate the share of each ethnic group in the cabinet. Appendix Table 1 (Panel B) tabulates the evolution of the ethnic share across the political history of Kenya.

## Additional References


Ahluwalia, P. (1996): *Post Colonialism and the Politics of Kenya*, New York: Nova Science.
FAO (2011): *FAOSTAT*, Rome: Food and Agricultural Organization.

Government of Kenya (1965): *Housing and Population Census 1962*, Nairobi: Government Printers.

Heston, A., R.Summers, and B. Aten (2012). *Penn World Tables Version 7.1*, Centre for International Comparisons of Production, Income and Prices, Philadelphia: University of Philadelphia.

Hornsby, C. (1985). The Member of Parliament in Kenya, 1969-1983: the Election, Background and Position of the Representative and the Implications for his Role in the One-Party State, Unpublished thesis, Oxford University.

IMF (2011): *International Financial Statistics*, Washington DC: International Monetary Fund.


[34] Available on `http://williameasterly.org/academic-work`

Officer, L. (2009): *Exchange Rates between the United States Dollar and Forty-one Currencies*, Measuring Worth Project.

World Bank (2011): *World Development Indicators*, Washington DC: World Bank.

# APPENDIX 2: THEORY

Assume that $\theta < \max\{\frac{1}{\pi^A}, \frac{1}{\pi^B}\}$.

Denote by $V^i(j)$ a MPE utility for a citizen of type $i$ starting in a subgame with a president of type $j$.

Proceed by backwards induction. Assume a president of type $i$ announces $P^i = (\tau^i, \eta^{Ai}, \eta^{Bi})$.

For group $i$ to support the policy it must be that

$$\begin{aligned} R(\eta^{ii}) - \tau^i + \bar{\gamma} V^i(i) + (1-\bar{\gamma}) V^i(j) &\geq \underline{\gamma} V^i(i) + \left(1-\underline{\gamma}\right) V^i(j) \\ R(\eta^{ii}) - \tau^i + \left(\bar{\gamma} - \underline{\gamma}\right)\left(V^i(i) - V^i(j)\right) &\geq 0 \end{aligned} \tag{2}$$

The President thus maximizes his instanteneous utility subject to (2) and (1).

$$\begin{gathered} \max_{\tau^i, \eta^{ii}, \eta^{ij}} \pi^i \left(\tau - \eta^{ii}\right) + \pi^j \left(\tau - \eta^{ij}\right) \\ R(\eta^{ii}) - \tau^i + \left(\bar{\gamma} - \underline{\gamma}\right)\left(V^i(i) - V^i(j)\right) \geq 0 \\ \eta^{ii} \leq \theta \left(\pi^i \eta^{ii} + \pi^j \eta^{ij}\right) \\ \eta^{ji} \geq 0 \end{gathered}$$

Note that the last constraint cannot bind: if $\eta^{ji} = 0$ then due to (1) we would have $\eta^{ii} \leq \theta \pi^i \eta^{ii}$ which directly contradicts $\theta < \max\{\frac{1}{\pi^A}, \frac{1}{\pi^A}\}$.

The first order conditions of the problem yield ($\lambda$ and $\mu$ as multipliers)

$$\begin{aligned} \pi^i + \pi^j - \lambda &= 0 \\ -\pi^i + \lambda R'(\eta^{ii}) + \mu\left(\theta \pi^i - 1\right) &= 0 \\ -\pi^j + \mu \theta \pi^j &= 0 \end{aligned}$$

This solves to

$$\begin{aligned} \lambda &= 1 \\ R'(\eta^{ii}) &= \frac{1}{\theta} \\ \mu &= \frac{1}{\theta} \end{aligned}$$

which means that both constraints are binding. Since this does not depend on $\pi^i$ or $\pi^j$ (the only differences across groups), we have that $R'(\eta^*) \equiv R'(\eta^{ii}) = R'(\eta^{jj}) = \frac{1}{\theta}$.

Also, since (2) is binding, we have

$$\eta^{ji} = \eta^* \frac{1-\theta\pi^i}{\theta\pi^j}$$
$$\eta^{ij} = \eta^* \frac{1-\theta\pi^j}{\theta\pi^i}$$

So we can now set up the value functions

$$V^i(i) = R(\eta^*) - \tau^i + \bar{\gamma}V^i(i) + (1-\bar{\gamma})\,V^i(j)$$
$$V^i(j) = R(\eta^{ij}) - \tau^i + \bar{\gamma}V^i(i) + (1-\bar{\gamma})\,V^i(j)$$
$$V^j(j) = R(\eta^*) - \tau^j + \bar{\gamma}V^j(j) + (1-\bar{\gamma})\,V^j(i)$$
$$V^j(i) = R(\eta^{ji}) - \tau^j + \bar{\gamma}V^j(j) + (1-\bar{\gamma})\,V^j(i)$$

and in addition we know that the two versions of (1) are binding

$$R(\eta^*) - \tau^i + \left(\bar{\gamma} - \underline{\gamma}\right)\left(V^i(i) - V^i(j)\right) = 0$$
$$R(\eta^*) - \tau^j + \left(\bar{\gamma} - \underline{\gamma}\right)\left(V^j(j) - V^j(i)\right) = 0.$$

This gives us a linear system of six equations in six unknowns $\left(V^i(i), V^i(j), V^j(j), V^j(i), \tau^i, \tau^j\right)$. This has a unique solution, and hence uniqueness of MPE is proven.