NOT FOR PUBLICATION

ONLINE APPENDICES

The Value of Democracy: Evidence from Road Building in Kenya

by Robin Burgess, Remi Jedwab, Edward Miguel, Ameet Morjaria and Gerard Padró i Miquel

December 2014

Appendix A: Data Sources

This appendix describes all of the data sources used in the paper. Summary statistics are reported in Appendix E: Table A2.

1. Road Expenditure Data

We construct an annual district-year road development expenditure (in constant 2000 US$) panel data set for the period 1963-2011. The *Development Estimates* of Kenya an official government publication, allows us to track road development expenditure at the district level.[1] The *Development Estimate*'s document programmatically road projects (for instance, project *Thika Main Road* is from *Thika* town to *Nyeri* town via *Limuru* town) and their actual cost.[2] When a road project spans more than one district, we use GIS tools to lay out the road segment in question and calculate the length of kilometers within each district. Hence, for projects which span multiple districts the expenditure share is weighted by distance. For the period 1963-1973, road development estimates do not document individual road projects, instead, only large nation-wide road programs are reported. We supplement our data with the government's four-year development plans and the World Bank's road project documentation to construct comparable, project level expenditure data for the 1963-1973 period.[3]

2. Road Construction Data

We create a district-map year paved road construction panel data set by constructing a GIS database of the Kenyan road network for the years we have road maps. To construct the GIS road network, we first use as a baseline the most recent GIS database that contains contemporary roads (*Global GIS*). We then proceed with our series of historical road maps to recreate the evolution of the road network in GIS.[4] Our maps

[1] Government of Kenya (a, 1963/64-2010/11): road expenditure is reported in East African pounds (1963-1966), Kenyan pounds (1967-1999), and Kenyan shillings (2000-present). We use Officer (2009) and IMF (2011), to convert these amounts to current US$ and use a US$ deflator series to convert to constant 2000 US$.

[2] We supplement our primary source of Government of Kenya (a, 1963/64-2010/11) with additional ministerial reports, Government of Kenya (2007/2008–2011/2012) to ensure we have accounted for all projects. Medium Term Expenditure Framework Reports available from `www.treasury.go.ke`, accessed on December 2012.

[3] We use the Government of Kenya's four year development plans for the following years: 1964-1966, 1966-1970, 1970-1974 and 1974-1978. Road construction programs in Africa during that period were primarily fully or partially funded under International Development Agency (IDA) financing program of the World Bank. We collate all the Road Program Operational Reports (available on `http://www.worldbank.org/projects`, accessed on November 2011) and these assist us to provide the sub-projects and their relative costs for the period 1963-1973.

[4] We use the road map series published by Michelin (1964-2002) for the years 1964, 1967, 1969, 1972, 1974, 1979, 1981, 1984, 1987, 1992 and 2002 (i.e. 11 maps). Two additional maps were published during our study period, in 1989 and 2010, unfortunately both of these maps are an exact re-print of the 1987 and 2002 editions, respectively.

limit us to consistently trace only the evolution of paved roads. The evolution of non-paved roads (classified as improved, laterite and dirt roads) provides a challenge due to inconsistent categorization and definitional changes in the legends of the Michelin maps. We use categories defined in the Michelin map as motorways and hard-surfaced roads as our measure of paved roads. Using GIS tools we splice the road network for the respective years with the 1964 district boundaries to create a paved road length (in kilometers) district-year panel data set of 451 observations (41 districts tracked for 11 years that we have maps). While these maps provide the stock of roads built by that particular map year, we are interested in measuring new road construction. We obtain this by taking the difference of the road length between two subsequent maps. Hence, the number of observations in our road building analysis is 410 (=451-41).

The details on how we construct the counterfactual road network data sets are described in Appendix B. Briefly, for both the paved road construction and the road development expenditure series, we use three different methodologies to create three different counterfactual datasets. The counterfactual datasets rely on two primary datasets: (i) the constructed GIS data of the paved and non-paved road network just at the turn of independence from the Michelin Map of 1964 and (ii) the population distribution of towns/cities in Kenya (42) and in the neighboring countries (7). We identify a town/city from the population census definition of urban settlements above and equal to 2000 inhabitants. For Kenya we use the 1962 population census (Government of Kenya 1965) and for the neighboring countries the nearest census year available to 1962 and interpolate.[5]

3. Ethnic Census

We use the population census of 1962 (Government of Kenya 1965) to obtain our district ethnic demographics. We scan, digitize and geo-reference a 1963 district map which allows us to construct for each district (41) its ethnic demographics by linking district names across the map and the census. The population census reports 41 ethnic classifications. In line with studies on the politics of Kenya, we aggregate the ethnic classifications into 13 groups.[6] **Coethnic District [d,t]** is a binary indicator equal to 1 if $\geq 50\%$ of district d's population is coethnic to the president in year t. The coethnicity of the president evolves as follows: between 1963-1978 the president is Kenyatta and the ethnic group is Kikuyu, between 1979-2002 the president is Moi and the ethnic group is Kalenjin and between 2003-2011 the president is Kibaki and the ethnic group is Kikuyu. **Democracy [t]** is a binary indicator equal to 1 if t is a democratic year. Democratic years are identified as those when the constitution of Kenya allows multiple parties to

[5]The border towns/cities in the neighboring countries are as follows: Yabelo (Ethiopia), for which we use Ethiopia's population censuses of 1956 and 1967; Afmadu (Somalia), for which we use Somalia's population censuses of 1953 and 1963; Kapoeta (Sudan, now in South Sudan) for which we use Sudan's population censuses of 1955 and 1966; Moshi, Arusha and Musoma (all in Tanzania) for which we use Tanzania's population censuses of 1957 and 1967; Tororo (Uganda) for which we use Uganda's population censuses of 1959 and 1969.

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

contest elections. The variable takes the value of 1 during the period 1963-1969 and 1993-2011 (both inclusive) and 0 in the interim periods.[7] **Kikuyu District [d,1962]** (**Kalenjin District [d,1962]**, **Kamba District [d,1962]**, **Luhya District [d,1962]** and **Luo District [d,1962]**) is a binary indicator equal to one if $\geq 50\%$ of district d's population is Kikuyu (Kalenjin or Kamba or Luhya or Luo, respectively) according to the 1962 population census. The **Coethnic Group [e,t]** is a binary indicator equal to one if the president belongs to ethnic group e in year t. The **VP-Coethnic District [d,t]** is a binary indicator equal to one if $\geq$50% of district d's population is from the ethnic group of the vice-president in year t. The **VP-Coethnic Group [e,t]** is a binary indicator equal to one if the vice-president belongs to ethnic group e in year t. **Non-Coethnic Majority <80% [d,1962]** is a binary indicator equal to one if the main non-coethnic group in district d accounts for <80% of its total population. Appendix Table A1 (Panel A) provides the national population share of the major ethnic groups across post-independence Kenya. The data tabulated on ethnic composition was obtained from all published population censuses (1962, 1969, 1979, 1989 and 2009). The 1999 population census did not disclose the ethnic demographics.

4. Demographic and Socioeconomic Variables

We use various archival sources to construct three sets of control variables at the district level: (i) *demography*: district population and urbanization rates are obtained from the population census (1962) and district area is estimated using GIS tools, (ii) *economic activity*: the Statistical Abstracts of Kenya (Government of Kenya 1963-66) are used to construct total formal district employment (1963) and total formal district earnings (1966) in constant 2000 US$, the value of cash crop exports is constructed using the Government of Kenya (1964) which provides reports of cash crop production for the year 1964/65[8], (iii) *economic geography*: GIS tools are used to create a binary variable which takes the value of one if the district is on the Mombasa-Nairobi-Kampala highway corridor, another binary variable is created which takes the value of one if any part of the district borders Tanzania or Uganda, the two main trading partners. Lastly the euclidean distance between the district centroid and the national capital, Nairobi, is calculated.

5. Cabinet Composition

We source archival data and construct a panel dataset of the ethnicity and position of all cabinet members between the years 1963 and 2011, after every general election (13

[7]Note Kenya's fiscal year is from July to June, *Development Estimates* for year t provide expenditure for the period July $t-1$ to June t. Moi takes presidency from 1979 (fiscal cycle July 1978-June 1979) and Kibaki takes presidency from 2003 (fiscal cycle July 2002-June 2003). Similarly, the transition to autocracy in November 1969 is considered from 1970 (fiscal cycle July 1969-June 1970) and the transition to democracy took place in December 1992 and is considered from 1993 (fiscal cycle July 1992-June 1993).

[8]The data is reported in Kenyan Shillings, using Officer (2009) and IMF (2011), we convert these amounts to current US$ and deflate the series to obtain figures in constant 2000 US$. The 1965 export price in constant 2000 US$ (FAO 2011) is used to calculate the district's total value of cash crop exports in 1965.

cabinets). This allows us to track the evolution of each ethnic group's representation in politics. We use two primary sources of data to compile this: the official listing *The National Assembly: List of Members*, *Organization of the Government of Kenya*, and Middleton (2007).[9] While the ethnicities of prominent cabinet members is well-known, information on other politicians is obtained by consulting several secondary sources and triangulating. We use: (i) the *Weekly Review* magazine, which would often discussed the ethnicity of cabinet members after each election, (ii) research done by political scientists on Kenya, especially Hornsby (1985) and Ahluwalia (1996), and (iii) direct assistance by several journalists from the top dailies in Kenya. Combining all these sources allows us to calculate the cabinet's ethnic representation in a particular year. Appendix Table A (Panel B) tabulates the evolution of the ethnic share across the political history of Kenya.

6. Electoral Data

Electoral data for the 1992 multiparty elections are obtained from the National Election Monitoring Unit (1993). Election results are tabulated at the constituency level (188), we overlay a digital geo-referenced map of constituencies (sourced from Morjaria 2014) on the geo-referenced district map (41) to allow aggregation of election results to the district level. We focus on the presidential elections. Electoral data are tabulated for each constituency and the number of votes won by each party that stood for elections is reported. We construct two variables: (i) **Margin of Victory [d,1992]**, the difference between the voting shares (%) of the winner and the runner up parties in district d and (ii) **Party Competition Herfindhal Index [d,1992]**, the Herfindahl index of voting shares of all the parties competing in district d.

7. Newspaper Articles

For the two main daily newspapers in Kenya (*The Daily Nation* and *The Standard*) which were in circulation both before and after the arrival of democracy in 1992 we employed a team of Kenyan journalists (supervised by one of the authors) to read 25 years worth of the daily editions of these two papers (i.e. close to 18,250 newspapers when we include both titles). These archives are not digitized and are in the form of microfiche and hard copies and so *LexisNexis* searches and the like were not an option. The task set for the team of journalist was to read through and catalogue whether or not a story pertaining to roads was in each of these daily editions across the 1985-2010 period. Note the journalists were not aware of our research hypotheses.

8. Growth, Ethnic Diversity and Democracy

Data on political regimes in Sub-Saharan Africa is obtained from Polity IV (2013). We use the variable **Combined Polity Score** which takes values from -10 (hereditary

[9] Government of Kenya (b, 1963/64-2010/11).

monarchy) to +10 (consolidated democracy). Polity IV categorizes regimes into *autocracies* (−10 to −6), *anocracies* (−5 to +5) and *democracies* (+6 to +10). In the analysis used in this paper we classify all regimes that are not autocracies as democracy, i.e. we add anocracies (imperfect democracies) and democracies (mature democracies). The average combined polity score for Sub-Saharan Africa is computed using the individual polity scores and weighted by the population of each country obtained from World Bank (2011). GDP per capita growth in Sub-Sahara Africa is obtained from World Bank (2011).

For Table 6, we obtain Easterly and Levine's data and append their decadel dataset with two additional decades, the 1990s and 2000s using identical sources as mentioned in Easterly and Levine (1997).[10] The variables updated are, *initial income* and *annual GDP per capita*, they are both obtained from *Penn World Tables 7.1*. Annual GDP per capita is used to calculate the growth of per capita real GDP. **Democracy [c,t]** is a binary indicator equal to one if country c is not an autocracy in decade t, specifically if the average combined polity score for the whole decade t is $\geq$-5. **Ethnic [c,1960]** is obtained from Easterly and Levine (1997) and is the ethnolingustic fractionalization of country c in 1960.

Additional References for Appendix A

Ahluwalia, P. (1996) *Post Colonialism and the Politics of Kenya*, New York: Nova Science.

Barro, R. and J. Lee (2010) "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*, 104: 184-198.

Easterly, W. and R. Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics*, 112(4): 1203-1250.

FAO. (2011) FAOSTAT, Rome: Food and Agricultural Organization.

Government of Kenya. (1963-66) *Statistical Abstracts of Kenya*, Nairobi: Government Printers.

Government of Kenya. (1964) *Development Plan of Kenya*, 1964-1970, Nairobi: Government Printers.

Government of Kenya. (a, 1963/64 - 2010/11) *Development Estimates for Year 19../..*, Nairobi: Government Printers.

Government of Kenya. (b, 1963/64 - 2010/11) *The National Assembly: List of Members, Organization of the Government of Kenya*, Nairobi: Government Printers.

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

Government of Kenya. (2007/08-2011/12) *Physical Infrastructure Sector: Medium Term Expenditure Framework*, Nairobi: Government Printers.

[10] Available on `http://go.worldbank.org/K7WYOCA8T0`, accessed on December 2012.

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.

Michelin. (1964-2002) *Michelin Map Africa Central & Southern Africa: 746 Series*, Paris: Michelin.

Middleton, J. (2007) *Encyclopedia of Africa South of the Sahara*, New York: C. Scribner's Sons.

Morjaria, A. (2014) "Is Democracy Detrimental for the Environment in Developing Countries? Evidence from Kenya.", Working Paper, Harvard University.

National Election Monitoring Unit. (1993) *The Multi-Party General Elections in Kenya 29 December 1992: The Report of the National Election Monitoring Unit*, Nairobi: NEMU.

Ng'weno, H. (1979-1999) *Weekly Review*, Nairobi.

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

Polity IV (2013) *Polity IV Project, Political Regime Characteristics and Transitions, 1800-2013*, Vienna VA: Center for Systemic Peace.

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

Appendix B: Constructing Counterfactual Road Networks

This appendix describes the steps we undertake to construct the counterfactual panel datasets that are used in **Table 3** and **Figure 3** (paved road construction) and **Appendix Table A4** and **Appendix Figure A2** (road development expenditure). We set out the details first for paved road construction and then discuss the series for road development expenditure.

Step #1: Calculating the kilometers of paved road that were constructed. In our counterfactual exercise we take as given the total length of roads constructed in each between-maps period. Availability of maps restricts how many years of paved road data we have available, in particular recall that we have maps for the years: 1964, 1967, 1969, 1972, 1974, 1979, 1981, 1984, 1987, 1992 and 2002. Digitizing the maps allows us to compute the number of kilometers of paved roads that were built between 1964 and 2002 (total paved road constructed in length is 5286 kilometers). Appendix Table A9 tabulates the kilometers of paved road that were built in each between-maps period. For instance, between 1964-1967, 212 kilometers of roads were paved, between 1967-1969, 386 kilometers of roads were paved, etc. This exercise gives us two pieces of information (i) the total number of paved roads constructed for the period 1962-2002 and (ii) the number of kilometers that are paved between available map years. The counterfactuals take these kilometers of paved road as given (this can be thought of as our *budget constraint*) and proceeds to allocate them according to efficiency criteria that we spell out below.

Step #2: Creating the set of potential road segments. Appendix Figure A6 illustrates the distribution of the paved and unpaved road networks in Kenya at independence (1964). Since all the paved roads constructed during the period 1964-2002 were initially unpaved and were already in existence in 1964, the unpaved network in 1964 indicates all the road segments that could potentially be paved. Our counterfactual simulation exercise sequentially paves these unpaved segments depending on the ranking criteria we outline below. The key advantage of using the entire road network in 1964 is that it allows us to take account of the physical geography of the country. For an illustration, the unpaved roads unambiguously circumvent Mount Kenya, Mount Elgon, Lake Victoria and Lake Naivasha, as well as the national forest reserves.

Step #3: Data to generate efficiency criteria to allocate paved roads. We posit that a social planner would be interested in connecting pairs of towns/cities that are already economically active or have the potential to be active. Data on the aggregate incomes of the 42 towns/cities in Kenya and the 7 border towns/cities in 1964 are not available. We instead use the population of each town/city as a proxy for economic activity.[11] We use Kenya's population census (Government of Kenya 1965) and

[11]This is a common approach when limited economic activity measures exists, see for instance, De Long and Shleifer (1993) and Acemoglu, Johnson and Robinson (2002), among others.

Google Earth to construct a GIS database of towns/cities as defined by the census of localities with inhabitants of equal and above 2000 people.[12] To locate where these towns/cities are spatially in relation to the road network, we use our road map for 1964 (the available map closest in time to Kenya's independence in 1963) and the population censuses of the neighboring countries for towns/cities in neighboring countries (using the same 2000 inhabitants definition). We identify that there are 7 border towns/cities that we should account for in assessing which pairs of settlements should be connected. Appendix Figure A6 further illustrates the border towns/cities in the neighboring countries.

Step #4: Generating efficiency criteria using the information available in 1964. We keep using our towns/cities as proxies for local economic development. We construct various criteria that allow us to obtain values in connecting town/city pairs. The procedure is as follows, there are 42 towns/cities, which imply 861 possible town/city pairs within Kenya to connect and 7 border towns/cities thus 294 possible pairs of town/city between Kenya and its neighbors, this adds up to a total of 1155 possible pairs.[13] We rank these 1155 town/city pairs using three different criteria:

(i) The first criterion creates a measure using only the population of the two settlements and sums the two, i.e., maximizes the population sum of the town/city pair, *max* $(P_i + P_j)$, where town/city are denoted by i and j. By construction, this criterion gives precedence to segments that connect populated towns/cities.

(ii) The second criterion creates a measure using only the distance between two settlements and minimizes the sum of the two, i.e. minimizes the euclidean distance between the town/city pair, *min* (D_{ij}). By construction, this criterion prioritizes shorter roads which can be seen as a shorthand for minimizing costs.

(iii) The third criterion creates a measure using both the population and distance between two settlements, known as *market potential*. This criterion maximizes $(P_i + P_j)/D_{ij}$. By construction, those cities that are close to each other and have a large number of inhabitants have a connection with higher market potential. Appendix Table A3 displays the top 20 and bottom 20 potential bilateral connections in terms of their market potential, as well as when they become paved in the counterfactual simulation.

Step #5: Ranking all potential road segments according to the three efficiency criteria to create the three counterfactual. The three different criteria allow us to rank the 1155 potential pairs in their order of importance. The ranking order depends on the counterfactual criteria used: population alone, distance alone, and market potential. Appendix Table A3 provides an illustration of how we go about creating

[12] We will restrict our criteria to towns/cities in 1962 to abstract away from concerns of town/city growth due to political factors and leadership changes.

[13] The 1155 connections are obtained as follows, within Kenya 861 (42x41)/2 pairs and between Kenyan towns/cities and border towns/cities 294 (42x7) pairs.

the counterfactual. Upon ranking all the bilateral pairs (see *Rank* column in Appendix Table A3) we first systematically eliminate all bilateral pairs that are already connected by a paved road link in 1964, using the 1964 map as guidance. This then gives us a new ranking (see *Conditional Ranking*, in Appendix Table A3). For each ranked pair we ask the question: are the pairs already connected via the existing paved road network? If yes, we omit this connection and repeat the same procedure for the next connection in the rank. If the ranked pair is ***not*** connected, we pave this connection, using the shortest route via both the paved and unpaved road network using the 1964 map. We proceed in this manner, until we have allocated all of paved road kilometers that are available between available maps years, as discussed in **Step #1** (e.g., 212 km between 1964-1967, 386 km between 1967-1969, until we exhaust the total 5286 km between 1964-2002). In the second last column of Appendix Table A3, we illustrate when the pair gets paved in the counterfactual simulation, and the last column illustrates the number of kilometers allocated to that link.[14]

Step #6: How do we now create the counterfactual data series on paved road construction at the district level between 1964-2002? Due to the geo-spatial nature of our data, we know the length and location (and thus districts) of the road linking each town/city pair. We are able to splice these segments into kilometers of paved roads within the respective districts. This allows us to construct a data series very similar in structure to the actual paved road length data, namely, the change in the total length of paved road for each district d over time. Note that we are able to repeat **Step #5** for the other two counterfactual criteria (population alone and distance alone). This allows us to create three different counterfactual data series and hence the same dependent variable used in our main analysis (Table 2) can now we be re-computed and re-analyzed using these data. The summary statistics for these counterfactual paved road construction outcomes are reported in Appendix Table A2 (Panel D).

Step #7: How do we create the counterfactual data series for the annual development expenditure series using the information we have from paved roads constructed between 1964-2002? For the analysis on road development expenditure, we use our original road development expenditure which allow us to obtain yearly amounts of road investment. Since our counterfactual exercise is based on spatially connecting settlements and hence about reallocating paved roads between two geo-spatial points using the criteria outlined in **Step #4**, we can only construct the counterfactual series for the period 1964 to 2002 and not from 1963 to 2011, which is the time period we have available for our actual road expenditure data series. From **Step #1**, we know that we have to reallocate 5286 km of paved roads during our study period. Since we know the budget contribution for each year t to the total road development expenditure budget for the nation across the whole period 1964-2002, we can

[14]For the cases when the establishment of a connection spans across two periods (for instance the 1964-1967 and 1967-1969), we allocate the segments that are closer to the largest of the two cities to the first period (e.g., 1964-1967) and the segments that are farther to the second period (e.g., 1967-1969).

compute the km of paved roads that are to be allocated every year between 1964 and 2002, using the average cost to construct 1 km of paved road. This methodology thus assumes that the cost of constructing 1 km of paved road remains constant throughout the study period.[15] Using the expenditure series, we can thus estimate how many km of paved roads can be constructed in each year. This relies on another assumption, namely that the whole roads budget is being allocated to construction of paved roads. The World Bank Operational Reports on roads indicate that the bulk of the road budget is dedicated to constructing paved roads, but once again, the need for this reasonable but strong assumption leads us to prefer the paved road length counterfactual (described above).[16] See Appendix Table A10 for these figures.

We use our estimate of the counterfactual paved road length built in each district-map year (based on the underlying road expenditures in that year), as well as the total national km of paved roads in that year (using the same cost per 1 km constructed approximation discussed above), to compute the district's share of national road expenditures in that year. We once again apply the three ranking criteria used above (see **Step #4**) to construct three counterfactual data series on district-year road expenditures, similar to the three counterfactual series created in the paved road length counterfactual above. The summary statistics for the three counterfactual road expenditure outcomes are reported in Appendix Table A2 (Panel D).

Note: The town/city pair rankings for all six counterfactual series (three each for the road length counterfactuals and for the road expenditure counterfactuals) are available from the authors (in *MS-Excel* spreadsheet format).

Additional References for Appendix B

Acemoglu, D., S. Johnson, and J. A. Robinson. (2002) "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *The Quarterly Journal of Economics*, 117(4): 1231-1294.

De Long, J. Bradford, and A. Shleifer. (1993) "Princes and Merchants: European City Growth before the Industrial Revolution." *Journal of Law and Economics*, 36(2): 671-702.

[15] Note that the paved road length counterfactual described above does not rely on this sort of "scaling" assumption, and thus is arguably more attractive, hence our primary focus on the paved road length counterfactual in the main text.

[16] See discussion on World Bank Operational Reports in Appendix A.

Appendix C: Theory

We start by stating more formally a few features of the model.

We assume that $R(\eta)$ is strictly increasing and concave and satisfies $R'(0) = \infty$ and $R'(\infty) = 0$.

Denote by $\omega_t \in \{A, B\}$ a state variable that captures the ethnic type of the president at time t.

Formally stated, the timing of the game, given ω_t, is as follows:

1. The president announces the policy vector $P_t = (\tau^{\omega_t}, \eta^{A\omega_t}, \eta^{B\omega_t})$

2. The citizens of group ω_t decide whether to support the leader, $s_t = 1$ or not $s_t = 0$

3. If $s_t = 1$, P_t is implemented and payoffs are realized. Next period starts with $\omega_{t+1} = \omega_t$ with probability $\bar{\gamma}$. With probability $1 - \bar{\gamma}$ the president loses power and the next president is from the other group.

4. If $s_t = 0$, the leader is immediately ousted and the transition policy vector $P = (0, 0, 0)$ is implemented. After the transition, with probability $\underline{\gamma}$ the new ruler belongs to the same group as the ousted ruler and hence $\omega_{t+1} = \omega_t$. With probability $1 - \underline{\gamma}$ the new president belongs to the other group.

We now proceed to the proof of Proposition 1.

We search for the Markov Perfect Equilibrium (MPE) of the game. Strategies can therefore only be conditioned on the payoff relevant state variables and past play within the stage game. Note that the only payoff-relevant state variable is ω_t.

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.

We 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 (λ and μ 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 π^i or π^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

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

So we can now set up the value functions

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

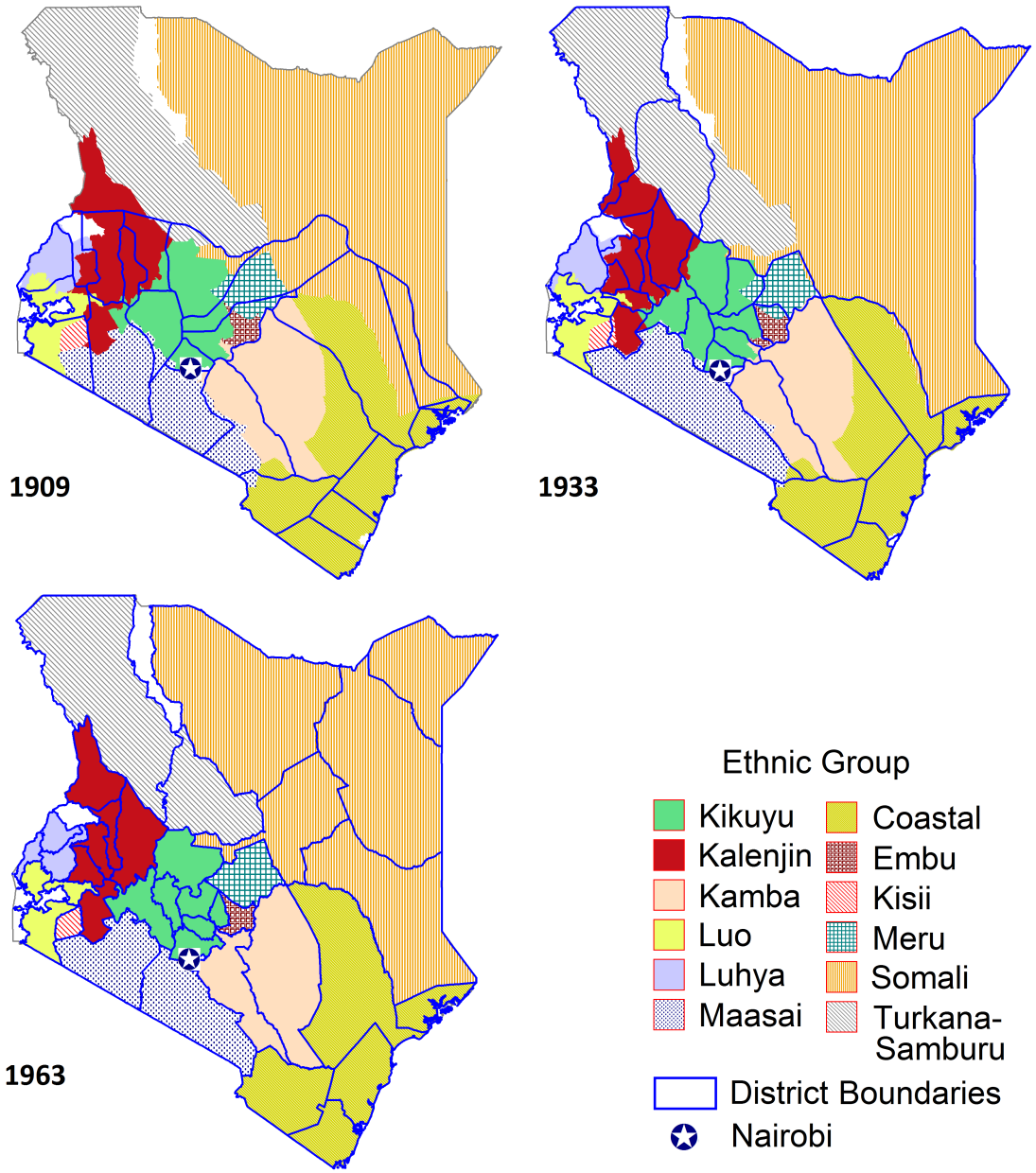
and in addition we know that the two versions of (1) are binding

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

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.

Appendix D: Additional Figures

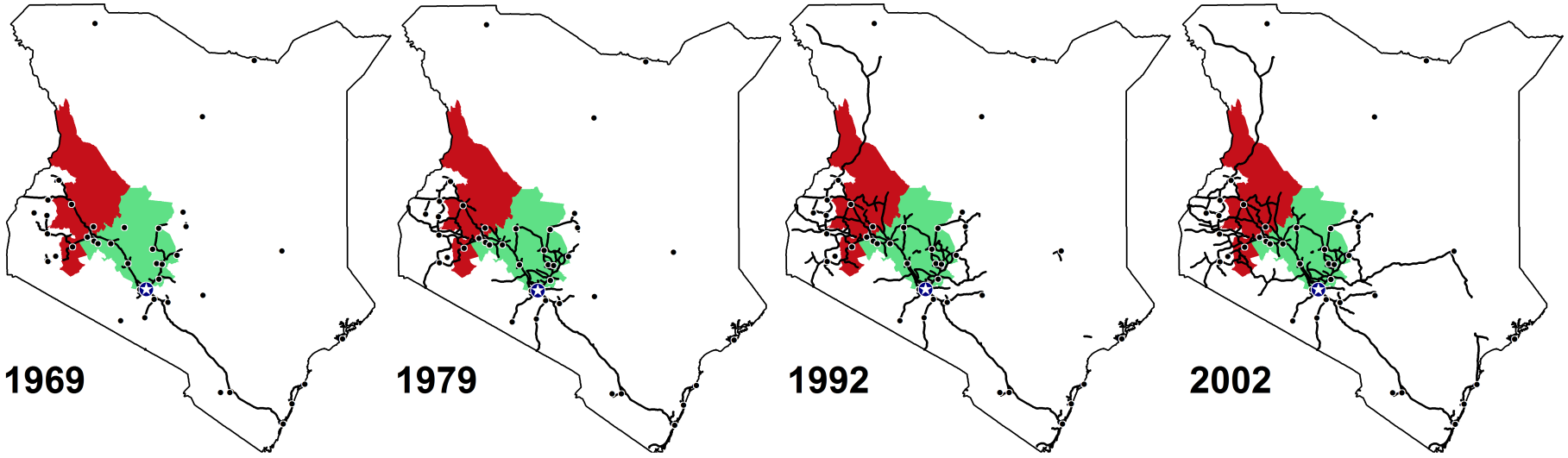
Appendix Figure A1: Evolving District Boundaries in Colonial Kenya and Ethnic Composition



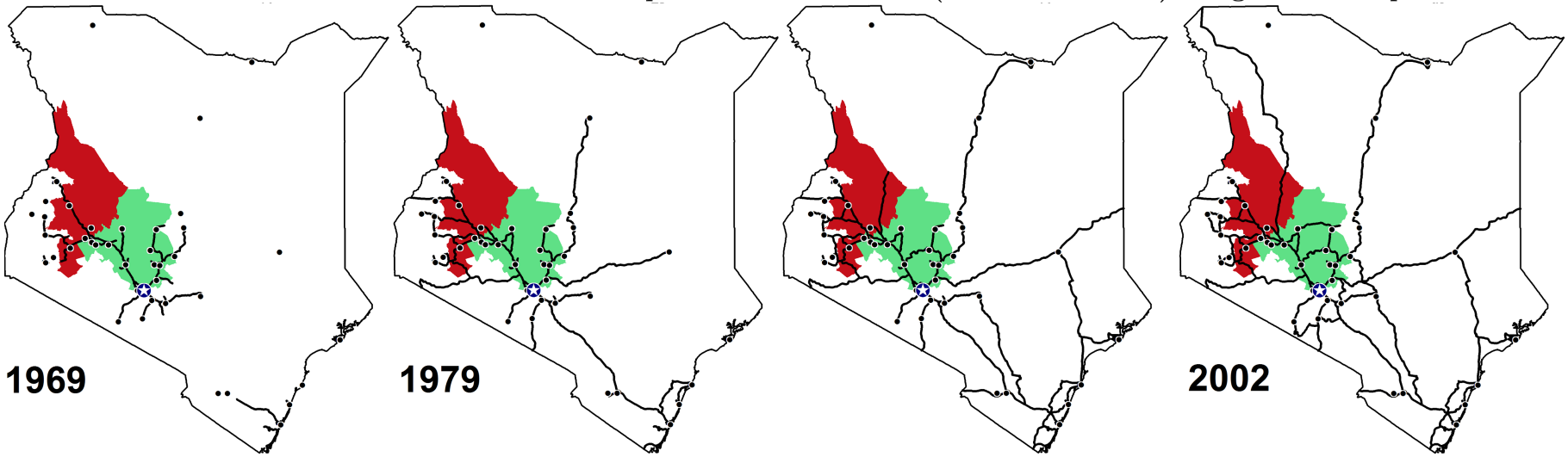
Notes: These figures illustrate the district ethnic composition, using the 1962 population census, and the evolution of district boundaries for selected years (1909, 1933, 1963) in Colonial Kenya. A district d is defined to be ethnic group e if $\geq 50\%$ of the district's population is ethnic group e. Only three districts are without a single ethnic majority group: Nairobi, Mombasa and Trans Nzoia. The 41 districts of the 1963 delineation of boundaries is used in all our analysis. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Figure A2: Evolution of Kenya's Paved Road Network

Panel A: Actual Network

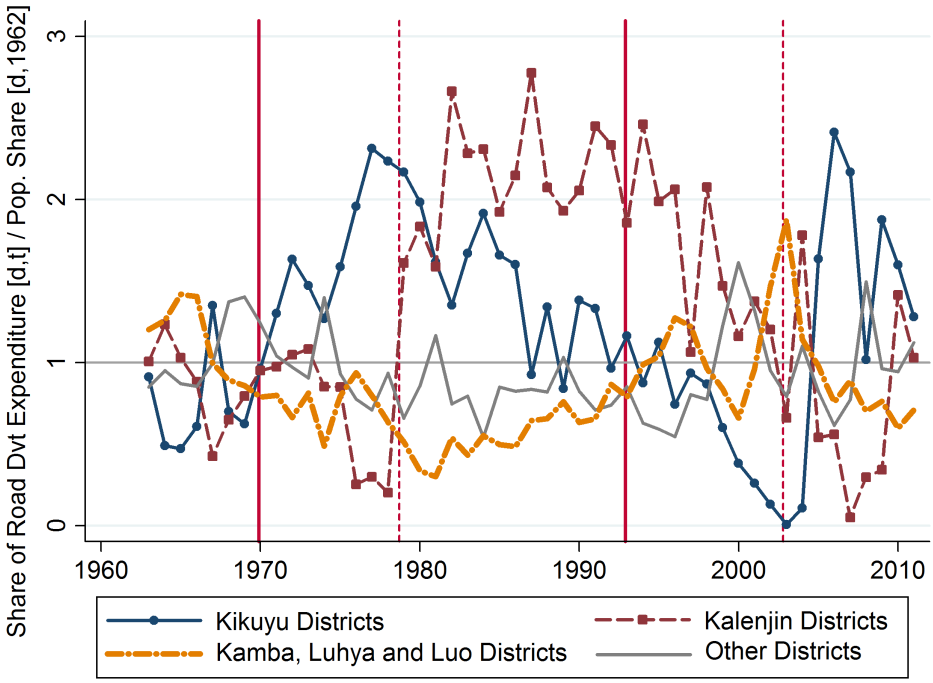


Panel B: Counterfactual Network based on Population and Distance (Market Potential) using Annual Expenditure



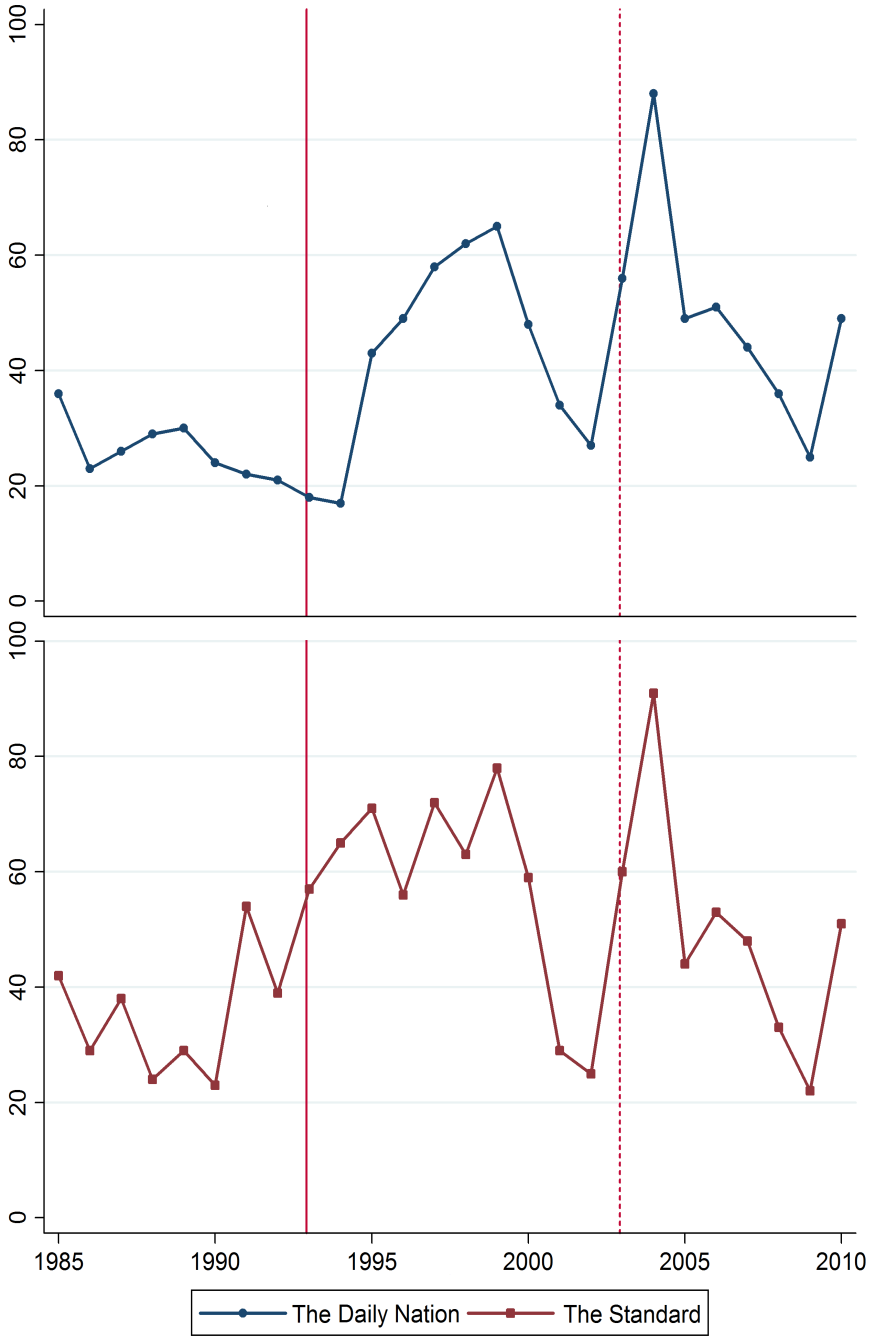
Notes: These figures illustrate the evolution of Kenya's actual and counterfactual paved road network for the key political and leadership transition years, 1969 (transition from democracy to autocracy), 1979 (from Kenyatta [Kikuyu] to Moi [Kalenjin] in 1978), 1992 (return to democracy) and 2002 (from Moi [Kalenjin] to Kibaki [Kikuyu]). The counterfactual network sequentially paves the unpaved bilateral connections with the highest market potential (based on population and distance). Border towns are not illustrated due to space constraint. The length of paved roads to be re-allocated every year is constrained by the amount of the total budget available for a year t as a share of the total road expenditure budget across the whole period of 1963-2011. Road maps are overlaid on ethnic demographics to illustrate the two presidential coethnic districts of Kikuyus and Kalenjins. **Coethnicity [d,t]** is defined if $\geq 50\%$ of district d's population is coethnic to the president in year t. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Figure A3: Road Expenditure in Kenyan Districts for the Largest Ethnic Groups and Other Groups, 1963-2011



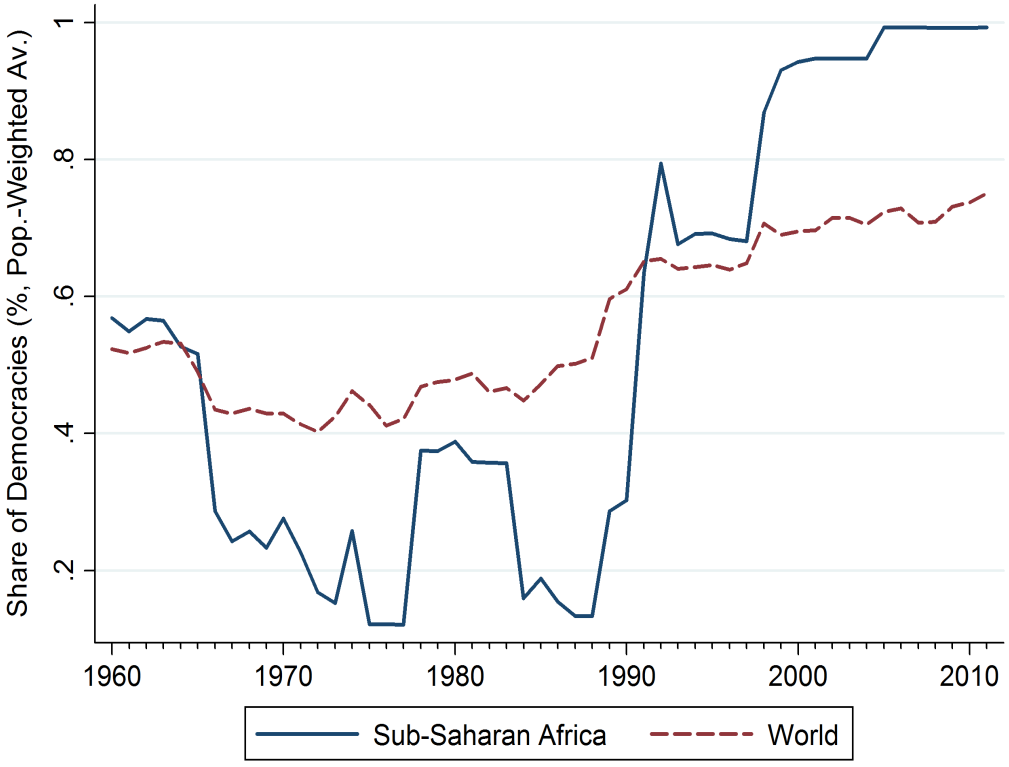
Notes: This figure plots the ratio between the share of road development expenditure in district d and year t to the share of population in 1962 for the main ethnic groups of Kenya (Kikuyu, Kalenjin and a single category for Kamba-Luhya-Luo), and the rest of the ethnic groups categorized as Other Ethnic districts. The main ethnic groups (Kikuyu, Kalenjin and Kamba-Luhya-Luo) and the Other Ethnic districts are defined as these types of district if $\geq$ 50% of the district's population is dominated by the main ethnic group or if it falls under other ethnic groups. There are 7 Kikuyu and 6 Kalenjin districts. The Kikuyu and Kalenjin districts are as defined in Figure 5. A Kamba-Luhya-Luo District is a district d if $\geq$ 50% of its population is either Kamba (2 districts), Luhya (3 districts) or Luo (3 districts) according to the 1962 population census. The vertical red lines represent political transitions, while the red vertical dotted lines represent leadership transitions as detailed in Figure 1. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Figure A4: Number of Road Articles in *The Daily Nation* and *The Standard* Newspapers, 1985-2010



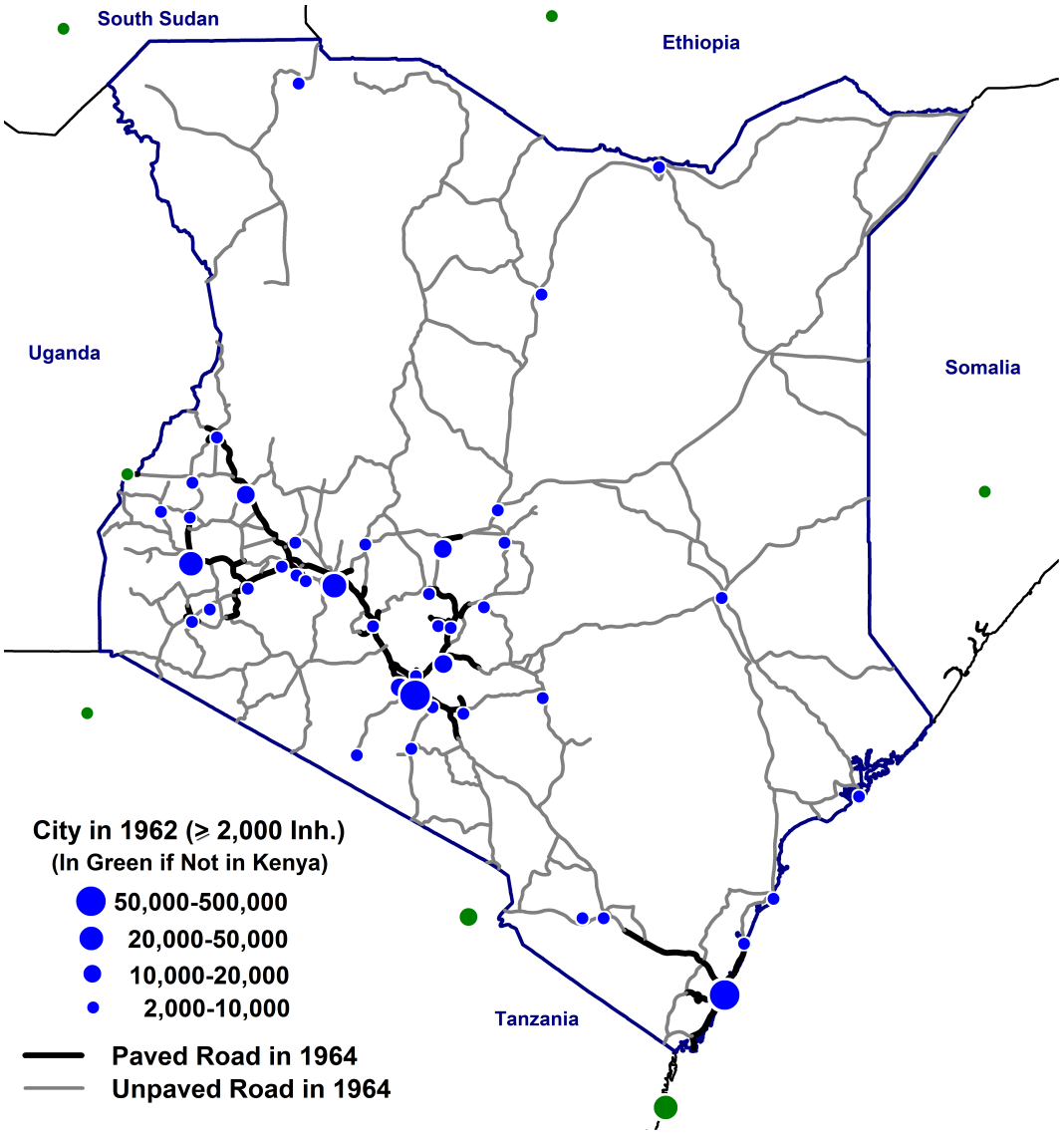
Notes: This figure illustrates the evolution of the number of articles which pertain to roads in Kenya's two leading independent dailies: *The Daily Nation* and *The Standard*. Data sources are described in Appendix A.

Appendix Figure A5: Democratic Change in the World and Sub-Saharan Africa, 1960-2011



Notes: This figure illustrates the annual share of democracies (%, pop.-weighted averages) for the world and for Sub-Saharan Africa for the period 1960-2011. A country is democratic if it is not autocratic in the Polity IV data set (combined polity score of ≥-5). Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Figure A6: Road Network and Urban Settlements in Kenya, 1962-1964



Notes: This figure illustrates the paved and unpaved road network at independence (1964) and the spatial distribution of urban settlements on the eve of independence (1962). Urban settlements are those towns/cities that have ≥2000 inhabitants. Kenya at independence has 42 towns/cities (Nairobi being the largest and the capital, followed by the port city of Mombasa). The map also depicts in green the 7 border towns/cities in neighboring countries: Ethiopia (1), Somalia (1), Sudan (1), Tanzania (3) and Uganda (1). Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix E: Additional Tables

Appendix Table A1: Ethnic Group representation in Population and Cabinet, 1962-2011

***Panel A*: Population Share (%) of Main Ethnic Groups**							
Census	Kikuyu	Kalenjin	Luo	Luhya	Kamba	Other	Pop. (Millions)
1962	18.8	10.8	13.4	12.7	10.5	33.8	8.6
1969	20.1	10.9	13.9	13.3	11.0	30.8	11.0
1979	20.9	10.8	13.2	13.8	11.3	30.0	15.3
1989	20.8	11.5	12.4	14.4	11.4	29.5	21.4
2009	17.2	12.9	10.8	13.8	10.1	35.2	38.6
***Panel B*: Cabinet Share (%) of Main Ethnic Groups**							
Cabinet	Kikuyu	Kalenjin	Luo	Luhya	Kamba	Other	Cabinet Size
1963	35.3	0.0	0.0	5.9	0.0	35.3	17
1964	31.6	5.3	5.3	5.3	9.1	27.6	19
1966	27.3	4.6	4.6	9.1	9.1	36.3	22
1969	31.8	9.1	9.1	9.1	18.2	22.7	22
1974	31.8	9.1	9.1	9.1	9.1	31.8	22
1979	29.6	14.8	7.4	11.1	9.1	28.0	27
1983	20.8	16.7	12.5	12.5	0.0	37.5	24
1988	25.0	11.8	14.7	11.8	0.0	36.7	34
1993	6.0	20.0	4.0	16.0	9.1	44.9	25
1998	5.4	25.0	0.0	17.9	9.1	42.6	28
2003	21.2	7.7	15.4	19.2	18.2	18.3	26
2005	22.8	6.1	3.0	24.2	18.2	25.7	33
2008	17.4	13.9	11.6	18.6	9.1	29.4	43

Notes: Panel A tabulates the national share of the main ethnic groups for each population census. The 1999 population census did not disclose the national ethnic demographics, the national population was at 28.7 million. *Panel B* tabulates the ethnic profile of the appointed cabinet post-general elections. The cabinet includes the president, vice-president, ministers with portfolios and two other ex-officios. The solid lines in *Panel B* denote leadership transitions: from Kenyatta (Kikuyu) to Moi (Kalenjin) in August 1978, and from Moi (Kalenjin) to Kibaki (Kikuyu) in December 2002. The dashed lines in *Panel B* denote democratic regime changes in Kenya: December 1969 is the transition from democracy to autocracy, while December 1992 is the return of democracy. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A2: Summary Statistics and Data Sources

(Details provided in **Appendix A**)

	Mean	SD	Obs	Source
Panel A: Main Dependent Variables				
(a) Share of Road Expenditure [d,t]/Population Share [d,1962]	1.25	2.80	2009	*Expenditure*: Government of Kenya (a, 1963/64-2010/11); *Population*: Government of Kenya (1965).
(b) Share of Road Expenditure [d,t]/Area Share [d]	3.62	7.81	2009	*Expenditure*: sources in (a); *Area*: GIS tools on map of districts.
(c) Share of Paved Road Construction [d,t]/Pop. Share [d,1962]	1.37	7.29	410	*Paved Roads*: Michelin (1964-2002); *Population*: sources in (a).
(d) Share of Paved Road Construction [d,t] / Area Share [d]	2.90	8.30	410	*Paved Roads*: sources in (c); *Population*: sources in (a).
Panel B: Main Regressors				
(a) Coethnic District [d,t]	0.16	0.37	2009	*Coethnic*: Political & Leadership changes (see Figure 1); *Ethnic population*: Government of Kenya (1965).
(b) Democracy [t]	0.53	0.50	49	*Democratic years*: Political & Leadership changes (see Figure 1).
(c) Coethnic Share [d,t]	0.12	0.29	2009	*Coethnic*: Political & Leadership changes (see Figure 1); *Ethnic population*: sources in (a).
(d) Kikuyu District [d,1962]	0.17	0.38	41	*Ethnic Population*: sources in (a).
(e) Kalenjin District [d,1962]	0.15	0.36	41	*Ethnic Population*: sources in (a).
Panel C: Control Variables				
(a) Population ('000) [d,1962]	211	164	41	*Population*: Government of Kenya (1965).
(b) Area ('000 Sqr km) [d]	13.9	17.4	41	*Area*: GIS tools on map of districts.
(c) Urbanization (%) [d,1962]	7.4	20.0	41	*Ethnic Population*: sources in (a).
(d) Total Earnings in the Formal Sector (2000$) [d,1966]	10.5	21.0	41	*Earnings:* Government of Kenya (1963-66).
(e) Total Employment in the Formal Sector ('000) [d,1963]	42.6	77.2	41	*Employment*: Government of Kenya (1963-66).
(f) Total Value of Cash Crop Production (2000$) [d,1963]	8.3	20.1	41	*Cash Crop*: Government of Kenya (1964); *Value*: Officer (2009) and IMF (2011).
(g) Mombasa-Nairobi-Kampala Corridor [d]	0.37	0.48	41	*Corridor*: GIS tools on map of districts.
(h) Border district [d]	0.27	0.44	41	*Border*: GIS tools on map of districts.
(i) Euclidean Distance (km) to Nairobi [d]	268.3	146.1	41	*Distance*: GIS tools on map of districts.

Appendix Table A2: Summary Statistics and Data Sources

(continued, details provided in **Appendix A**)

	Mean	SD	Obs	Source
Panel D: Counterfactual Dependent Variables				
(a) Share of Paved Road Construction [d,t]/Population Share [d,1962]				
Counterfactual criteria based on:				
(i) population	1.86	7.95	410	*Paved Roads*: Michelin (1964-2002); *Population*: Government of Kenya (1965).
(ii) distance	1.69	8.01	410	*Paved Roads*: Michelin (1964-2002); *Location of settlements*: GIS map of towns/cities from Government of Kenya (1965).
(iii) population and distance (market potential)	2.25	11.50	410	*Paved Roads*: Michelin (1964-2002); *Population and location of settlements*: Government of Kenya (1965).
(b) Share of Road Expenditure [d,t]/Population Share [d,1962]				
Counterfactual criteria based on:				
(i) population	2.01	13.20	1599	*Paved Roads*: Michelin (1964-2002); *Population*: Government of Kenya (1965).
(ii) distance	1.76	10.67	1599	*Paved Roads*: Michelin (1964-2002); *Population and location of settlements*: Government of Kenya (1965).
Panel E: Other Main Variables				
(a) Ethnic share of cabinet [e,t]/Pop. share [e,1962]	1.03	0.80	169	*Cabinet*: Government of Kenya (b, 1963-2011), Ahluwalia (1996), Hornsby (1985), Ng'weno (1979-1999) and Middleton (2007); *Ethnic Population*: Government of Kenya (1965).
(b) Coethnic Group [e,t]	0.08	0.27	169	*Coethnic*: Political and Leadership changes (see Figure 1); *Ethnic Population*: sources in (a).
(c) VP-Coethnic Group [e,t]	0.09	0.29	169	*VP & Ethnic Population*: see sources in (a).
(d) VP-Coethnic District [d,t]	0.15	0.36	2009	*VP & Ethnic Population*: see sources in (a).
(e) Kamba-Luhya-Luo District [d,1962]	0.20	0.40	2009	*Ethnic Population*: sources in (a).
(f) Non-Coethnic Majority $< 80\%$ [d,1962]	0.37	0.48	2009	*Ethnic Population*: sources in (a).

Appendix Table A3: The Top and Bottom 20 Bilateral Connections for Construction of Paved Road Counterfactual based on Population and Distance (Market Potential)

Rank	Origin (i)	Destination (j)	Market Potential$_{ij}$ $(P_i + P_j)/D_{i,j}$	Conditional Ranking	Construction Year (in the counterfactual)	Kilometers Constructed
1	NAIROBI	KIAMBU	26310.44	$*$	–	–
2	KIKUYU	NAIROBI	22121.72	1	1967	17.5
3	NAIROBI	ATHI RIVER	13483.00	$*$	–	–
4	NAIROBI	THIKA	8810.95	$*$	–	–
5	NAIROBI	MACHAKOS	6240.04	$*$	–	–
6	NAIROBI	KAJIADO	5651.97	2	1967	51.5
7	KARURI	NAIROBI	5033.10	$*$	–	–
8	MURANGA	NAIROBI	4814.91	$*$	–	–
9	NAIVASHA	NAIROBI	4522.46	$*$	–	–
10	NAIROBI	MAGADI	3843.05	3	1967	103.1
11	KILIFI	MOMBASA	3623.13	$*$	–	–
12	NAIROBI	NYERI	3581.14	$*$	–	–
13	EMBU	NAIROBI	3188.80	4	1967	15.9
14	NAKURU	NAIROBI	2775.15	$*$	–	–
15	NAIROBI	KITUI	2643.00	5	1967 & 1969	89.1
16	NAIROBI	NANYUKI	2386.33	6	1969	50.2
17	NYAHURURU	NAIROBI	2235.42	7	1969	50.7
18	NAIROBI	ELBURGON	2207.71	8	1969	31.9
19	NAIROBI	MOLO	2067.63	$*$	–	–
20	MERU	NAIROBI	1978.64	9	1969	78.8
...	...	...	...	...	...	...
...	...	...	...	...	...	...
1136	MOYALE	NYAMIRA	6.21	*695*	NA	NA
1137	VOI	YABELO [ETHIOPIA]	6.00	*696*	NA	NA
1138	KAJIADO	LOKITAUNG	5.92	*697*	NA	NA
1139	VOI	MOYALE	5.90	*698*	NA	NA
1140	LOKITAUNG	GARISSA	5.88	*699*	NA	NA
1141	NYAMIRA	AFMADU [SOMALIA]	5.74	*700*	NA	NA
1142	LOKITAUNG	AFMADU [SOMALIA]	5.64	*701*	NA	NA
1143	GARISSA	KAPOETA [SUDAN]	5.62	*702*	NA	NA
1144	MUMIAS	KILIFI	5.50	*703*	NA	NA
1145	MUMIAS	AFMADU [SOMALIA]	5.46	*704*	NA	NA
1146	WUNDANYI	YABELO [ETHIOPIA]	5.42	*705*	NA	NA
1147	KILIFI	YABELO [ETHIOPIA]	5.25	*706*	NA	NA
1148	VOI	KAPOETA [SUDAN]	5.20	*707*	NA	NA
1149	MOYALE	WUNDANYI	5.19	*708*	NA	NA
1150	KILIFI	MOYALE	5.11	*709*	NA	NA
1151	LOKITAUNG	VOI	4.99	*710*	NA	NA
1152	WUNDANYI	KAPOETA [SUDAN]	4.75	*711*	NA	NA
1153	WUNDANYI	LOKITAUNG	4.44	*712*	NA	NA
1154	KILIFI	KAPOETA [SUDAN]	4.36	*713*	NA	NA
1155	KILIFI	LOKITAUNG	4.12	*714*	NA	NA

Notes: The above tabulation displays the top and bottom 20 bilateral connections for the construction of paved roads based on maximizing market potential (population and distance). We use data for 42 towns/cities in Kenya and 7 towns/cities in bordering countries in 1962. There are 1155 bilateral connections to consider (42*41/2 = 861 pairs within Kenya and 42*7 = 294 pairs between Kenyan towns/cities and bordering countries). Market potential for a town/city i and town/city j is defined as the sum of its population's P divided by the Euclidean distance D (km) between the pair: $(P_i + P_j)/D_{ij}$. We use the initial road network at independence (1964) to establish roads that are paved and unpaved. If a town/city pair is already connected in 1964 we ignore the pair (denoted above as $*$) and re-rank the initial ranked list, hence the conditional ranking column. If the pair has not been already paved, this pair than features in the ranking and is in line for potential paving, note this is conditional on an existence of an unpaved road connection between the town/city pair. We proceed similarly until we exhaust the amount of paved roads available (5286 km of paved roads is the total paved road constructed between 1964 and 2002, 51 town/city pairs are connected and the ranking of the last pair connected is ranked at 91 hence ranked pairs in *italics* are those that never get paved in the counterfactual as there is not enough paved roads available to allocate across the remaining years). Sudan refers to the now South Sudan. Construction of the counterfactual is described in Appendix B.

Appendix Table A4: Counterfactual Road Expenditure, Ethnicity and Democratic Changes, 1963-2011

Dependent Variable	Share of road development expenditure [d,t] / Population share [d,1962]		
Counterfactual Ranking	Population	Distance	Population and Distance (Market Potential)
	(1)	(2)	(3)
Panel A			
Coethnic District [d,t]	0.41	0.11	0.40
	(0.70)	(0.82)	(0.49)
Panel B			
Coethnic District [d,t]	0.59	-0.16	0.11
	(0.97)	(0.95)	(0.66)
Coethnic District [d,t] * Democracy [t]	-0.52	0.79	0.85
	(1.35)	(1.31)	(1.50)
F-test [*p-value*]	0.01	0.30	0.67
H_0: Coethnic + (Coethnic*Democracy) = 0	*[0.94]*	*[0.59]*	*[0.42]*
Observations	1599	1599	1599
Year and district fixed effects	Y	Y	Y
Controls*trend	Y	Y	Y

Notes: OLS regressions on annual district-year counterfactual expenditure panel dataset of 41 districts for the period of 1963-2011. The amount of paved roads to be constructed every year is determined by the contribution of road development expenditure t to the national total road development expenditure for the whole period of 1964-2002. The counterfactual dataset paves sequentially the network starting with the unpaved bilateral connections of cities i and j with the highest value of the ranking metric. We use 42 towns/cities in Kenya and 7 border towns/cities in neighboring countries classified in 1962 with population of $\geq$ 2000 habitants. This results in 1155 (=42*41/2 + 42*7) bilateral connections across these towns/cities. In Column (1), the counterfactual is based on maximizing population (P_{ij}) between two bilateral pairs ($P_i + P_j$) to obtain the ranking of the connections. In Column (2), the counterfactual is based on minimizing distance between two bilateral pairs (D_{ij}) to obtain the ranking of the connections. In Column (3), the counterfactual is based on maximizing market potential, the population and distance between two city pairs, $(P_i + P_j)/D_{ij}$. **Coethnic District [d,t]** is a binary indicator equal to one if $\geq$ 50% of district's (d) population is coethnic to the president in year t. **Democracy [t]** is a binary indicator equal to one if year t is a democratic year. Democratic years are identified as those when the constitution of Kenya allows multiple parties to contest elections. The *F-test* is for the hypothesis that coethnic and non-coethnic districts have equal outcomes under democracy. Columns (1)-(3) include initial controls interacted with a time trend (1964-2002). Refer to Table 1 notes for description of controls. Robust standard errors clustered at district level are reported in the parentheses with stars indicating *** p <0.01, ** p <0.05, * p <0.1. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A5: Road Expenditure, Ethnicity and Democratic Changes: Robustness to Excluding Selected Districts, 1963-2011

Dependent Variable	Share of road development expenditure [d,t] / Population share [d,1962]				
Sample	EXCLUDING SELECTED DISTRICTS				
	White Highlands	Nairobi and Adjacent	Mombasa-Kampala	Nairobi-Kampala	Five Richest
	(1)	(2)	(3)	(4)	(5)
Panel A					
Coethnic District [d,t]	1.29***	1.08***	0.87***	0.86***	0.84**
	(0.43)	(0.36)	(0.31)	(0.30)	(0.33)
Panel B					
Coethnic District [d,t]	2.33***	1.74***	1.69**	1.97***	1.78***
	(0.79)	(0.52)	(0.63)	(0.63)	(0.59)
Coethnic District [d,t] * Democracy [t]	-2.10**	-1.23*	-1.46*	-1.98**	-1.75**
	(0.91)	(0.67)	(0.74)	(0.77)	(0.65)
F-test [*p-value*]	0.49	1.29	0.53	0.00	0.01
H_0: Coethnic + (Coethnic*Democracy) = 0	*[0.49]*	*[0.26]*	*[0.47]*	*[0.99]*	*[0.91]*
Observations	1568	1813	1274	1568	1764
No. of districts	32	37	26	32	36
Year and district fixed effects	Y	Y	Y	Y	Y
Controls*trend	Y	Y	Y	Y	Y

Notes: OLS regressions on annual district-year expenditure panel dataset of 41 districts for the period of 1963-2011. **Coethnic District [d,t]** is a binary indicator equal to one if $\geq$ 50% of district's (d) population is coethnic to the president in year t. **Democracy [t]** is a binary indicator equal to one if year t is a democratic year. Democratic years are identified as those when the constitution of Kenya allows multiple parties to contest elections. Column (1): excludes the 9 former White Highland districts (which are defined as districts which had $\geq$ 50% of it's area under the former White Highland area); column (2) excludes Nairobi and its 3 adjacent districts; Column (3) excludes 15 districts which are on the Mombasa-Nairobi-Kampala corridor; column (4) excludes the North-Western corridor connecting Nairobi to Kampala and column (5) excludes the 5 richest districts just prior to independence in 1962 (the criteria being highest African formal sector earnings at the district level). The *F-test* is for the hypothesis that coethnic and non-coethnic districts have equal outcomes under democracy. Columns (1)-(5) include the same controls as in Table 1, column 4 interacted with a time trend. Refer to Table 1 notes for description of controls. Robust standard errors clustered at district level are reported in the parentheses with stars indicating *** p <0.01, ** p <0.05, * p <0.1. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A6: Robustness Checks with Different Dependent Variables

Dependent Variable	Share of road dvt exp.[d,t] / Pop.sh.[d,1962]		Share of road dvt exp.[d,t] / Area sh.[d]		Share of paved road con.[d,t] / Pop.sh.[d,1962]		Share of paved road con.[d,t] / Area sh.[d]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Coethnic District [d,t]	1.00***	1.72***	1.78*	2.98***	3.71**	4.26**	4.54***	5.67***
	(0.35)	(0.49)	(0.87)	(0.99)	(1.69)	(1.74)	(1.54)	(1.68)
Coethnic District [d,t] * Democracy [t]		-1.32**		-2.22*		-2.38*		-4.82
		(0.62)		(1.29)		(1.36)		(2.89)
F-test [*p-value*]		0.88		0.46		0.98		0.11
H_0: Coethnic + (Coethnic*Democracy) = 0		*[0.36]*		*[0.50]*		*[0.33]*		*[0.75]*
Panel B								
Coethnic Share [d,t]	1.23***	2.28***	2.28**	4.36***	4.20**	5.03**	4.22**	5.76***
	(0.38)	(0.56)	(1.00)	(1.14)	(1.88)	(1.98)	(1.93)	(1.76)
Coethnic Share [d,t] * Democracy [t]		-1.88***		-3.74**		-2.85*		-5.35
		(0.66)		(1.45)		(1.66)		(4.03)
F-test [*p-value*]		0.94		0.25		1.07		0.01
H_0: Coethnic + (Coethnic*Democracy) = 0		*[0.34]*		*[0.62]*		*[0.31]*		*[0.92]*
Observations	2009	2009	2009	2009	410	410	410	410

Notes: Columns (1)-(4): OLS regressions on annual district-year expenditure panel dataset of 41 districts for the period of 1963-2011. Columns (5)-(8): OLS regressions on paved road construction panel dataset of 41 districts for the period of 1963-2002. In columns (1)-(2) and (5)-(6) the dependent variables' numerator is normalized by the population share of the district in 1962. In columns (3)-(4) and (7)-(8), the numerator is normalized by the share of the district area. All regressions include district and year fixed effects. **Coethnic District [d,t]** is a binary indicator equal to one if $\geq$ 50% of district's (d) population is coethnic to the president in year (t). **Democracy [t]** is a binary indicator equal to one if year t is a democratic year. Democratic years are identified as those when the constitution of Kenya allows multiple parties to contest elections. **Coethnic Share [d,t]** is the population share of the ethnic group of the president in district d at time t. The *F-test* is for the hypothesis that coethnic and non-coethnic districts have equal outcomes under democracy. Columns (1)-(8) include the same controls interacted with a time trend as in Table 1, column (4). Refer to Table 1 notes for description of controls. Robust standard errors clustered at district level are reported in the parentheses with stars indicating *** p <0.01, ** p <0.05, * p <0.1. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A7: Robustness Checks with Different Specifications

Dependent Variable	Share of road development expenditure [d,t] / Population share [d,1962]				
	Table 1, column (4)				
	(1)	(2)	(3)	(4)	(5)
Panel A					
Coethnic District [d,t]	1.00***	1.25**	1.63	1.00***	1.00***
	(0.35)	(0.51)	(0.98)	(0.33)	(0.28)
Panel B					
Coethnic District [d,t]	1.72***	2.17***	1.84*	1.72***	1.72***
	(0.49)	(0.70)	(0.96)	(0.48)	(0.38)
Coethnic District [d,t] * Democracy [t]	-1.32**	-1.81	-1.26**	-1.32**	-1.32**
	(0.62)	(1.09)	(0.50)	(0.61)	(0.48)
F-test [*p-value*]	0.88	0.22	0.41	0.95	1.05
H_0: Coethnic + (Coethnic*Democracy) = 0	*[0.36]*	*[0.64]*	*[0.52]*	*[0.34]*	*[0.30]*
Observations	2009	2009	2009	2009	2009
No. of districts	41	41	41	41	41
Year and district fixed effects	Y	Y	Y	Y	Y
Controls*trend	Y	N	Y	Y	Y
Controls*year fixed effects	N	Y	N	N	N
Number of years coethnic district	N	N	Y	N	N
Clustering / Conley standard errors	District	District	District	200 km	400 km

Notes: OLS regressions on annual district-year expenditure panel dataset of 41 districts for the period of 1963-2011. **Coethnic District [d,t]** is a binary indicator equal to one if $\geq$ 50% of district's (d) population is coethnic to the president in year (t). **Democracy [t]** is a binary indicator equal to one if year t is a democratic year. Democratic years are identified as those when the constitution of Kenya allows multiple parties to contest elections. Column (1): replication of Table 1, column 4. Column (2): We interact our controls with year fixed effects. Column (3): We include the number of years a district has been a coethnic district. Column (4): Standard errors corrected for spatial clustering using a 200 km threshold. Column (5) Standard errors corrected for spatial clustering using a 400 km threshold. The *F-test* is for the hypothesis that coethnic and non-coethnic districts have equal outcomes under democracy. Columns (1)-(5) include the same controls as in Table 1, column 4 interacted with a time trend. Refer to Table 1 notes for description of controls. Robust standard errors clustered at district level are reported [for columns (1)-(3)] in the parentheses with stars indicating *** p <0.01, ** p <0.05, * p <0.1. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A8: Road Expenditure, Democratic Changes and Coalition Politics, 1963-2011

Dependent Variable	Share of road development expenditure [d,t] / Population share [d,1962]					
	Table 1, column (4)	ASSESSING POLITICAL COMPETITION EFFECTS				
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B						
Coethnic District [d,t]	1.72***	1.70***	1.73***	1.71***	1.69***	1.60***
	(0.49)	(0.49)	(0.49)	(0.49)	(0.47)	(0.46)
Coethnic District [d,t] * Democracy [t]	-1.32**	-1.29**	-1.34**	-1.30**	-1.27**	-1.14*
	(0.62)	(0.63)	(0.63)	(0.64)	(0.60)	(0.58)
Kamba District [d,1962] * Democracy [t]		0.68				
		(0.57)				
Luhya District [d,1962] * Democracy [t]			-0.25			
			(0.53)			
Luo District [d,1962] * Democracy [t]				0.22		
				(0.37)		
Margin of Victory [d,1992] * Democracy [t]					-0.26	
					(0.67)	
Party Competition Herfindhal Index [d,1992] * Democracy [t]						-1.17
						(1.02)
F-test [*p-value*]	0.88	0.94	0.80	0.89	0.94	1.16
H_0: Coethnic + (Coethnic*Democracy) = 0	*[0.36]*	*[0.34]*	*[0.38]*	*[0.35]*	*[0.34]*	*[0.29]*
Observations	2009	2009	2009	2009	2009	2009
No. of districts	41	41	41	41	41	41
Year and district fixed effects	Y	Y	Y	Y	Y	Y
Controls*trend	Y	Y	Y	Y	Y	Y

Notes: OLS regressions on annual district-year expenditure panel dataset of 41 districts for the period of 1963-2011. **Coethnic District [d,t]** is a binary indicator equal to one if $\geq$ 50% of district's (d) population is coethnic to the president in year (t). **Democracy [t]** is a binary indicator equal to one if year t is a democratic year. Democratic years are identified as those when the constitution of Kenya allows multiple parties to contest elections. **Kamba (*or* Luhya *or* Luo) District [d,1962]** is a binary indicator equal to one if $\geq$ 50% of district's (d) population is Kikuyu (Kamba *or* Luhya *or* Luo) according to the 1962 population census. There are 2 Kamba, 3 Luhya and 3 Luo districts. **Margin of Victory [d,1992]** is the difference between the voting shares (%) of the winner and the runner up parties at the constituency level for the presidential elections of 1992 aggregated to the district level. **Party Competition Herfindhal Index [d,1992]** is the Herfindahl index of the voting shares of all parties at the constituency level aggregated to the district level from the presidential elections of 1992. Columns (1)-(6) include the same controls as in Table 1, column 4 interacted with a time trend. Refer to Table 1 notes for description of controls. Robust standard errors clustered at district level are reported in the parentheses with stars indicating *** p <0.01, ** p <0.05, * p <0.1. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A9: Kilometers of Paved Roads Constructed between 1964-2002

Map Year	Next Map Year	Length (km)
1964	1967	212
1967	1969	386
1969	1972	590
1972	1974	504
1974	1979	896
1979	1981	151
1981	1984	1,149
1984	1987	220
1987	1992	209
1992	2002	969
1964	2002	5286

Notes: The table shows kilometers of paved roads constructed between every map year for the period 1964-2002. Data sources and construction are described in Appendix A and Appendix E: Table A2.

Appendix Table A10: Kilometers of Paved Roads Constructed between 1964-2002 (Road Expenditure Counterfactuals)

Year	Length (km)
1964	43
1965	42
1966	91
1967	66
1968	113
1969	134
1970	146
1971	221
1972	234
1973	236
1974	258
1975	192
1976	120
1977	146
1978	114
1979	159
1980	161
1981	145
1982	204
1983	199
1984	203
1985	159
1986	137
1987	135
1988	108
1989	164
1990	149
1991	124
1992	106
1993	60
1994	94
1995	99
1996	134
1997	115
1998	118
1999	75
2000	102
2001	69
2002	115
1964-2002	*5286*

Notes: This table shows how many kilometers of paved roads must be constructed every year for the period 1964-2002 for use in the counterfactual road expenditure exercise. In total, 5286 km of paved roads must be constructed between 1964-2002. Since we know the contribution of each year t to the total amount of road development expenditure in 1964-2002, we can back out the amount of paved roads (km) that needs to be constructed every year between 1964 and 2002 assuming constant cost per km of paved road constructed. Data sources and construction are described in Appendix A and Appendix E: Table A2.