The Heterogeneous Effects of Transportation Infrastructure: Evidence from Sub-Saharan Africa

Remi Jedwab
George Washington University

Adam Storeygard
Tufts University

Research Questions

- **How has intercity road upgrading affected local economic development in Sub-Saharan Africa?**
  - How do these effects differ by context?

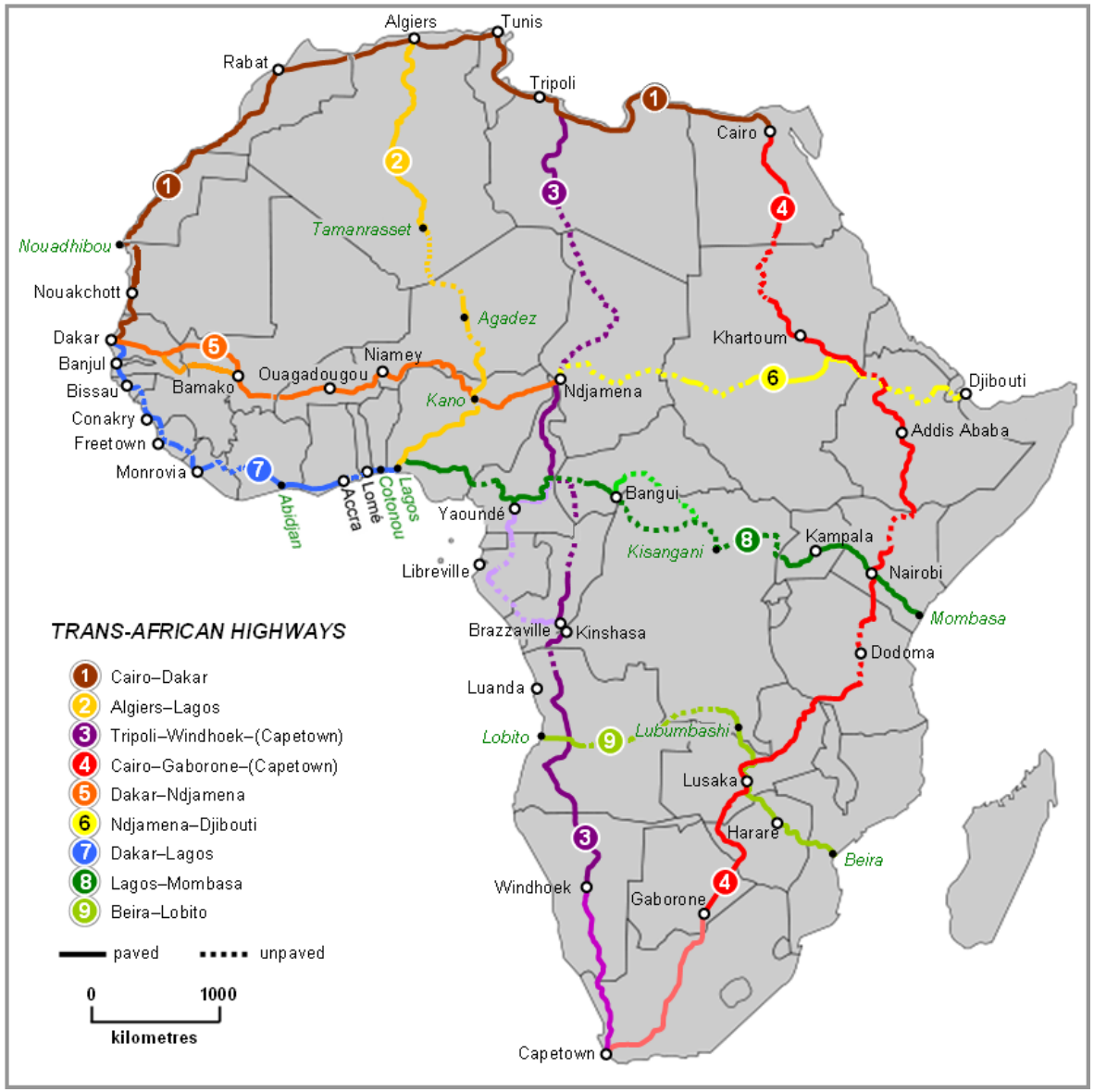
Research Questions

- **How has intercity road upgrading affected local economic development in Sub-Saharan Africa?**
  - How do these effects differ by context?
- **Implications for current/future road-building efforts?**
  - About 1/5 of World Bank lending on transport, 13% on roads
  - Large fraction of network still unpaved
  - Trans-African Highway network as coordinating mechanism: 55,000 km of planned highways (vs. 1,000 km of highways c. 2012)
    - Abidjan-Lagos Motorway: $8 billion
    - LAPSSET Project in Kenya-Ethiopia-South Sudan: $22 billion
    - Gauteng-Maputo Development Corridor: $5 billion

Research Questions

- **How has intercity road upgrading affected local economic development in Sub-Saharan Africa?**
  - How do these effects differ by context?
- **Implications for current/future road-building efforts?**
  - About 1/5 of World Bank lending on transport, 13% on roads
  - Large fraction of network still unpaved
  - Trans-African Highway network as coordinating mechanism: 55,000 km of planned highways (vs. 1,000 km of highways c. 2012)
    - Abidjan-Lagos Motorway: $8 billion
    - LAPSSET Project in Kenya-Ethiopia-South Sudan: $22 billion
    - Gauteng-Maputo Development Corridor: $5 billion
- **Implications for African urbanization?**
  - Expected increase 30% in 2010 to 50% in 2030: which cities?

Effects of Proposed Highway Networks?



What We Do

- Build **new panel data set** on road surface, city population and market access for 39 Sub-Saharan African countries 1960-2010
- Estimate the average effects of **market access** changes (as induced by road surface changes) on **city population growth**.
  - market access is trade-theory motivated measure of summarizing a location's access to all other locations.
  - find elasticity of 0.08–0.13 (vs. 0.04 OLS)
  - effect spread up to 30 years after road upgrading
- Investigate **heterogeneous effects** of road changes: remoteness, land suitability, ethnic homeland areas of head of state

Related work

- Market access and intercity transport costs:
  - Faber, 2014; Donaldson and Hornbeck, 2016; Donaldson, 2018; Jaworski and Kitchens, 2019; Baum-Snow et al., 2020 (reviewed in Redding and Turner 2015).
  - Market access more generally: Redding & Venables, 2004; Redding & Sturm, 2008
- Micro road surface/quality impacts
  - Casaburi et al., 2013; Gertler et al., 2015; Asher and Novosad, 2020; Gonzalez-Navarro and Quintana-Domeque 2016
- Transport and trade costs in Africa
  - rail (Jedwab & Moradi 2016; Jedwab, Kerby & Moradi 2017)
  - fuel prices (Storeygard 2016)
  - inferred from price changes (Atkin and Donaldson 2015)

Our contributions

- First systematic study of road-building across Africa
  - Scale: 140,000 km network, 39 countries, 6 time slices over 50 years: 1960, 1970, 1980, 1990, 2000 and 2010.
  - Timing of the effects
  - Heterogeneous effects
- Not just building highways: paving and improving (gravelling)
  - Marginal changes more likely in most contexts

Outline

- Data
- Estimation
- Results
- Conclusion

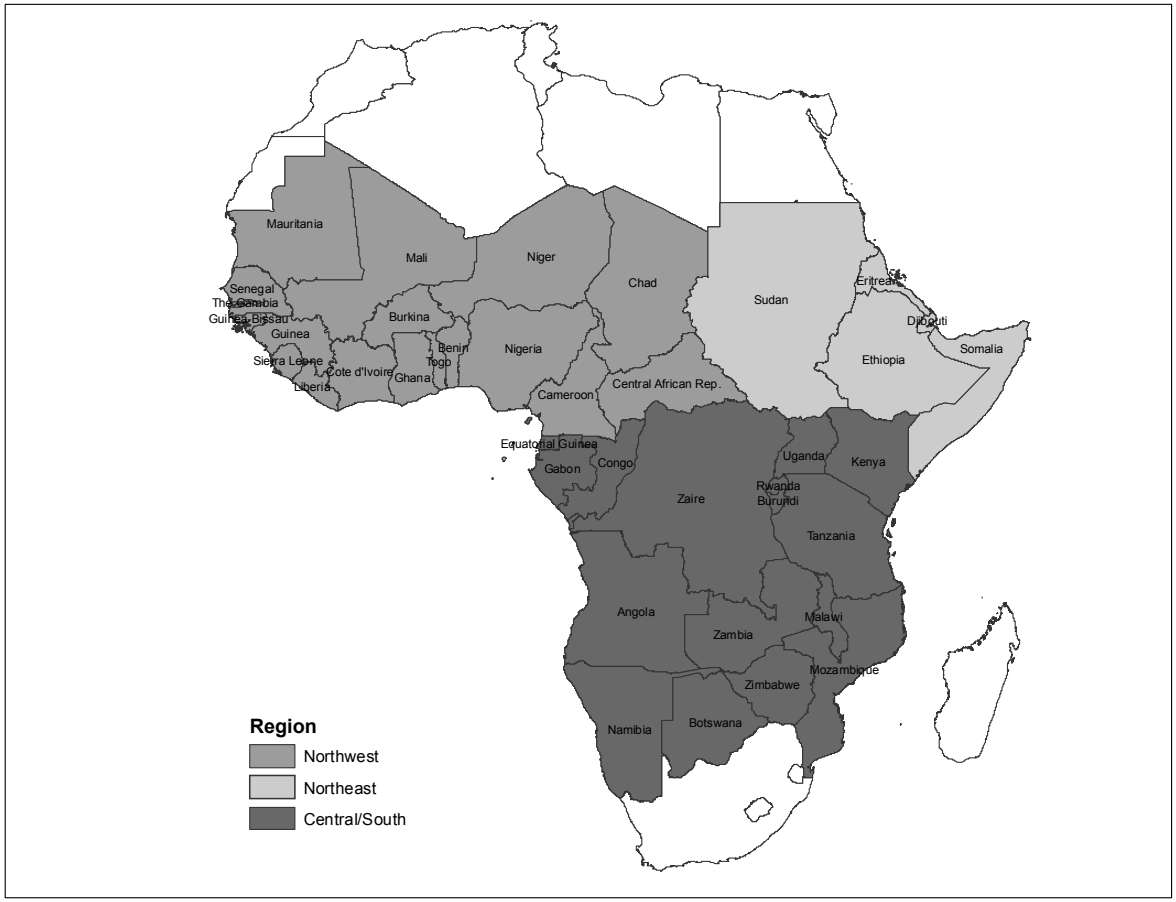
Outline

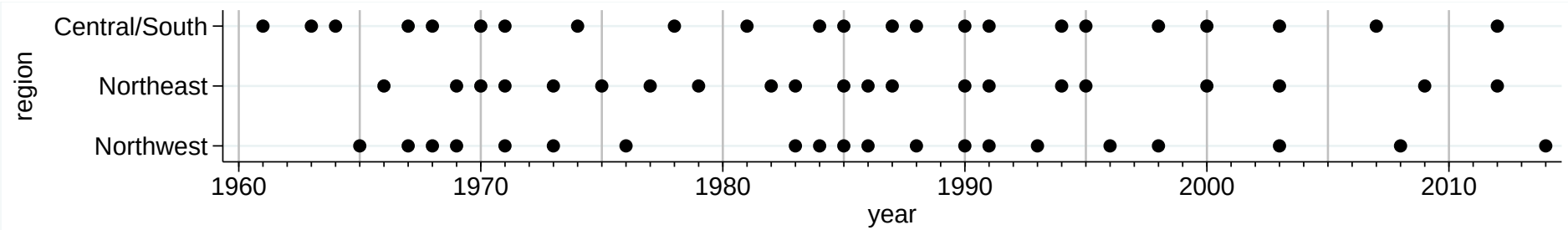
- **Data**
- Estimation
- Results
- Conclusion

GIS database of roads

- *Michelin* paper road maps for 39 Sub-Saharan African countries from the early 1960s to date. Sources:
  - Government maps
  - Feedback from customers (large network of tire distributors and correspondents)
- Map $\approx$ every 3 years, so 833 country-years
- Surface of each road: *Highway*, *Paved*, *Improved* and *Dirt* (vs. Primary, secondary, tertiary)
- *Dirt*: not systematically covered in Michelin maps. We use a separate basemap for the stock of all roads c. 2004, Michelin for surface conversions (which drive variation in the panel)
- No city streets

Michelin Road Map Countries and Years



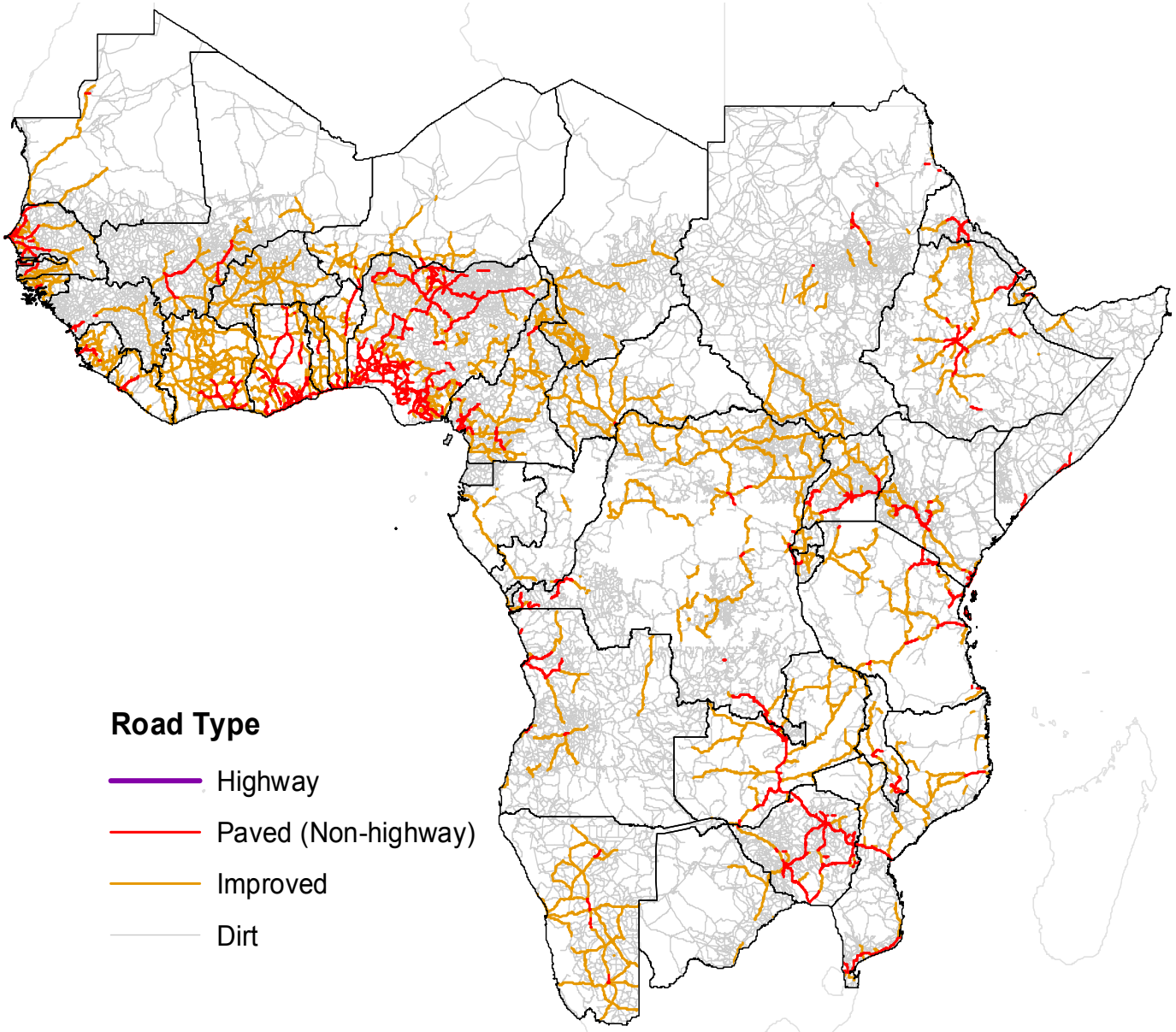


Michelin Road Map for Liberia in 1965

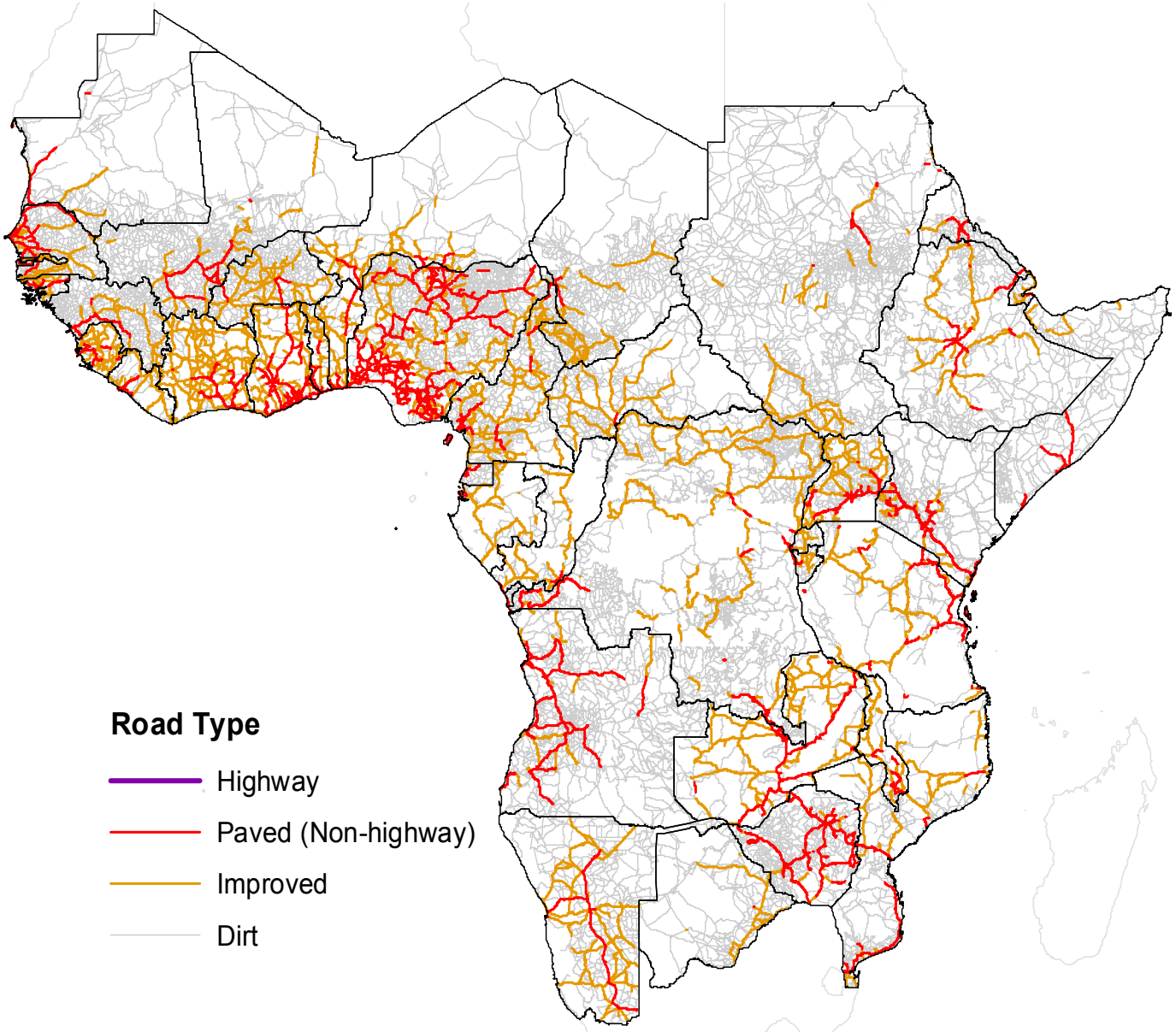
Surfaces aggregated into 4 categories: *Highway*, *Paved*, *Improved* and *Dirt*

Four Road Surface Categories

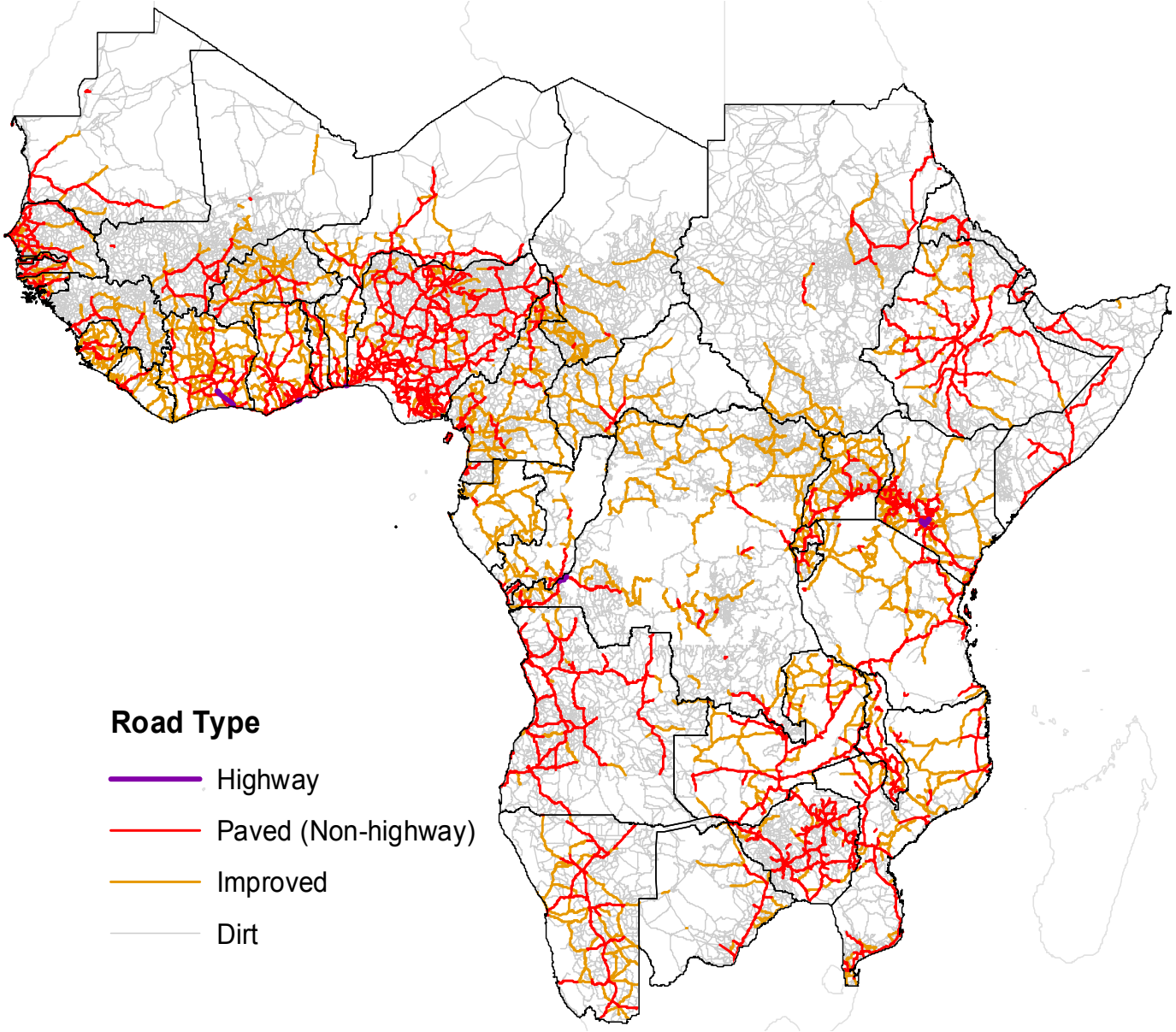
Roads in 1960



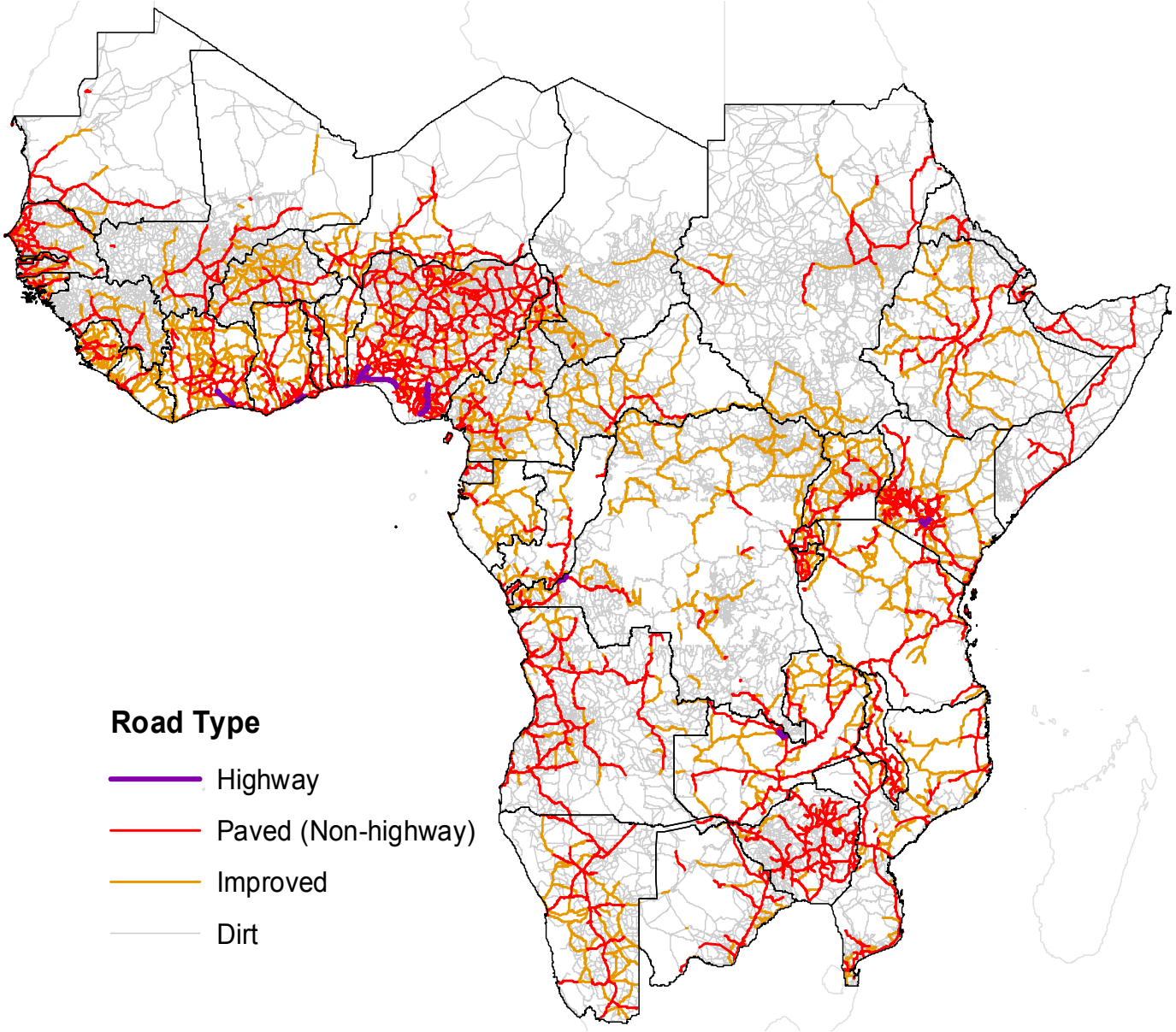
Roads in 1970



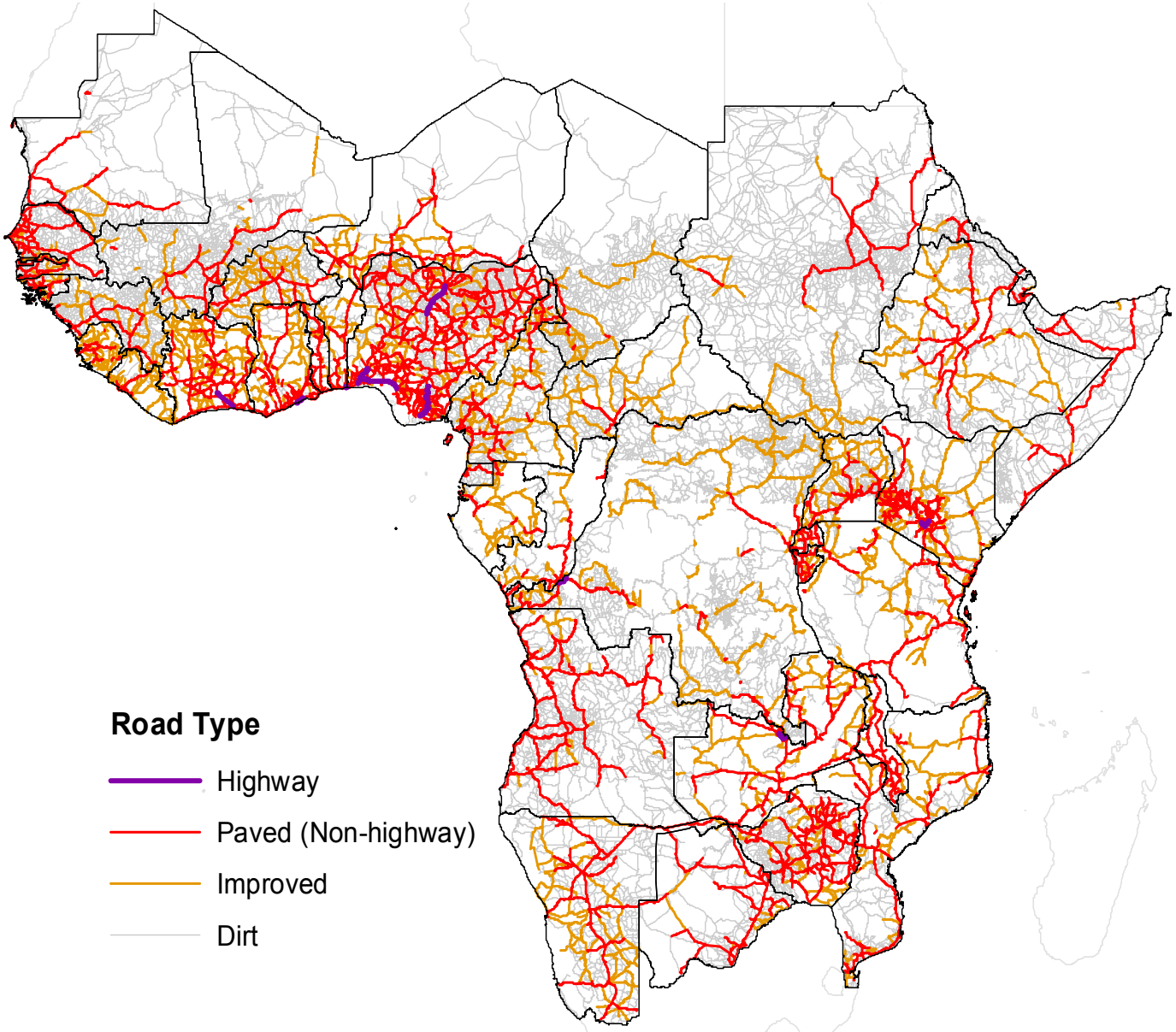
Roads in 1980



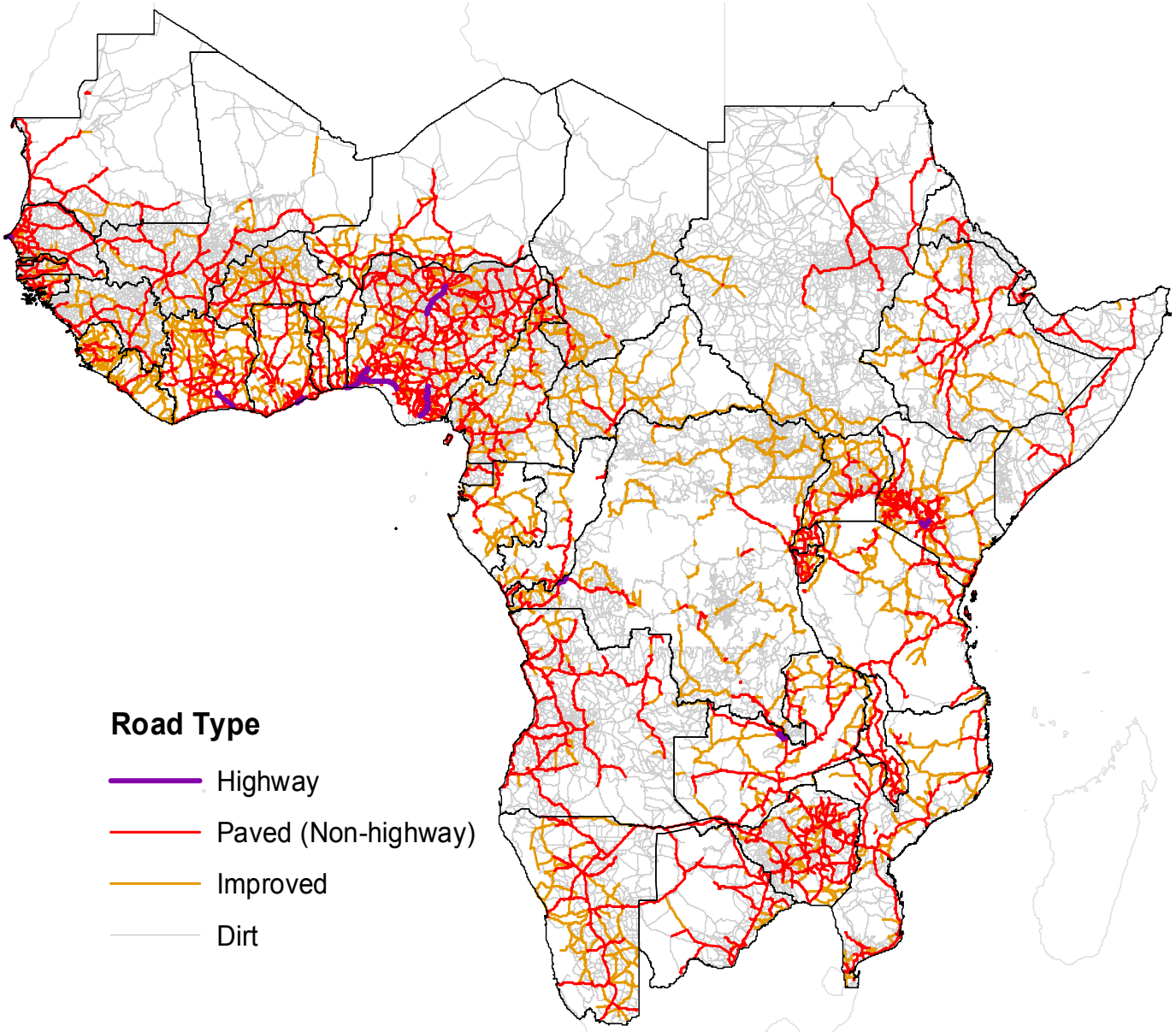
Roads in 1990



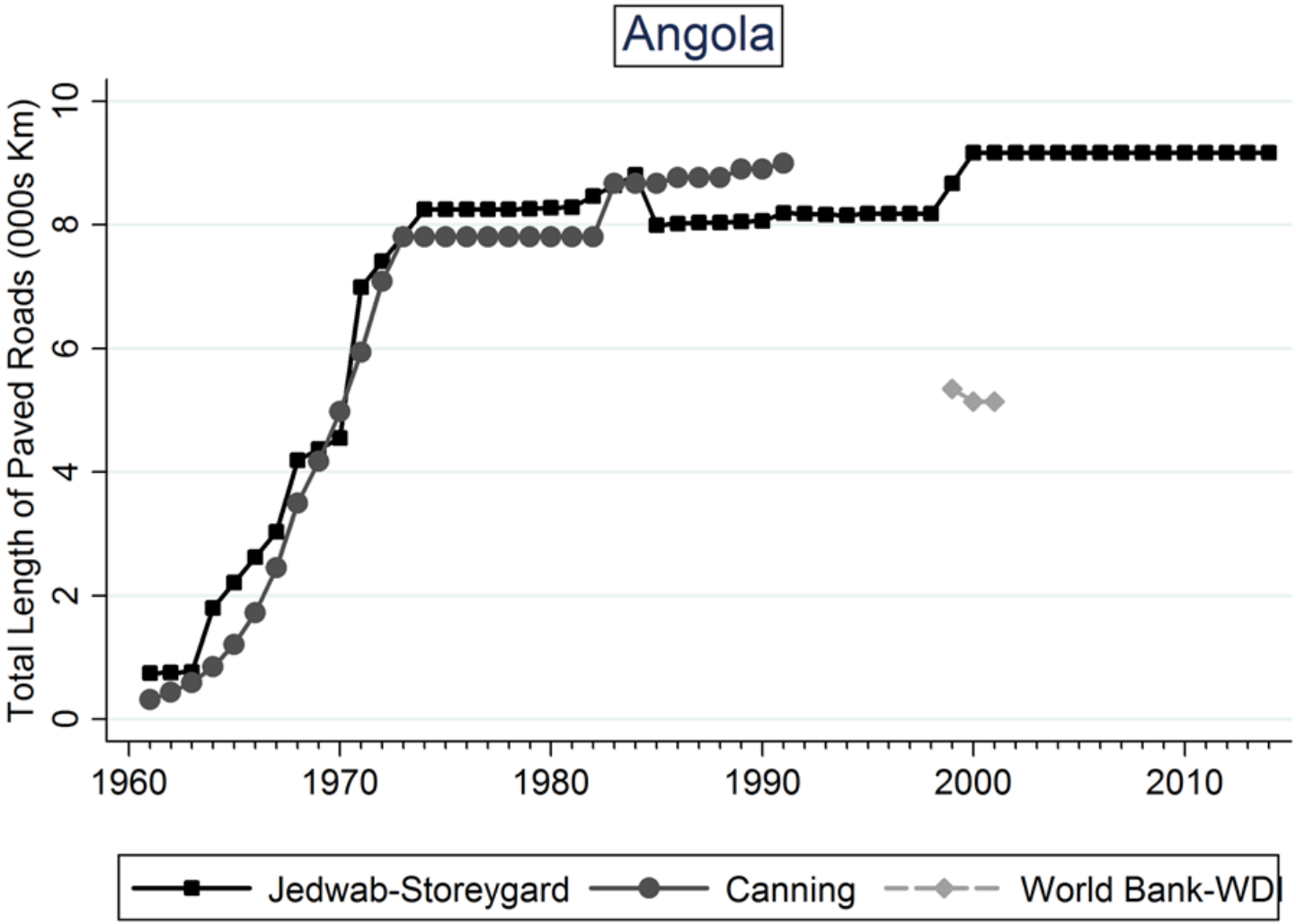
Roads in 2000



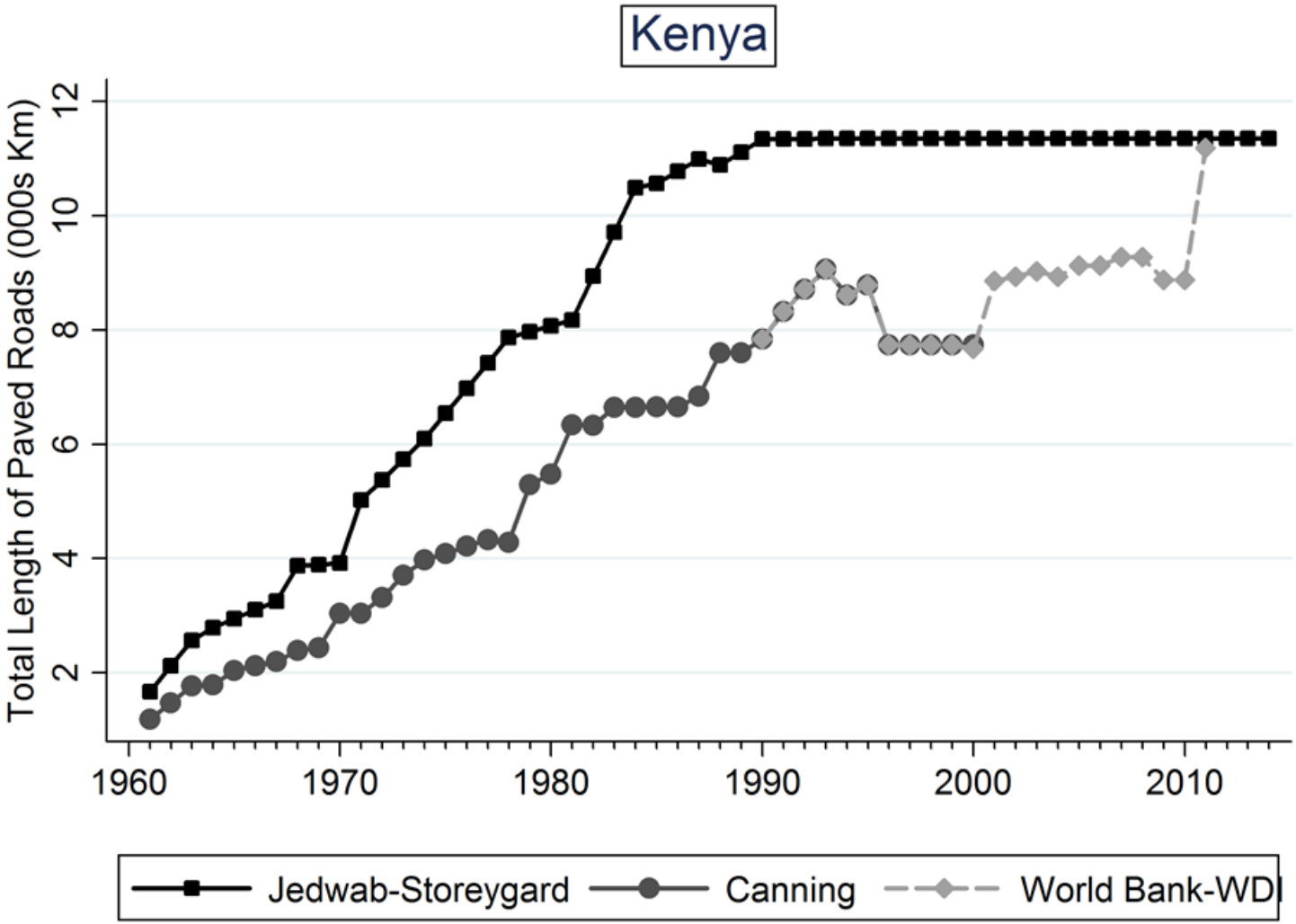
Roads in 2010



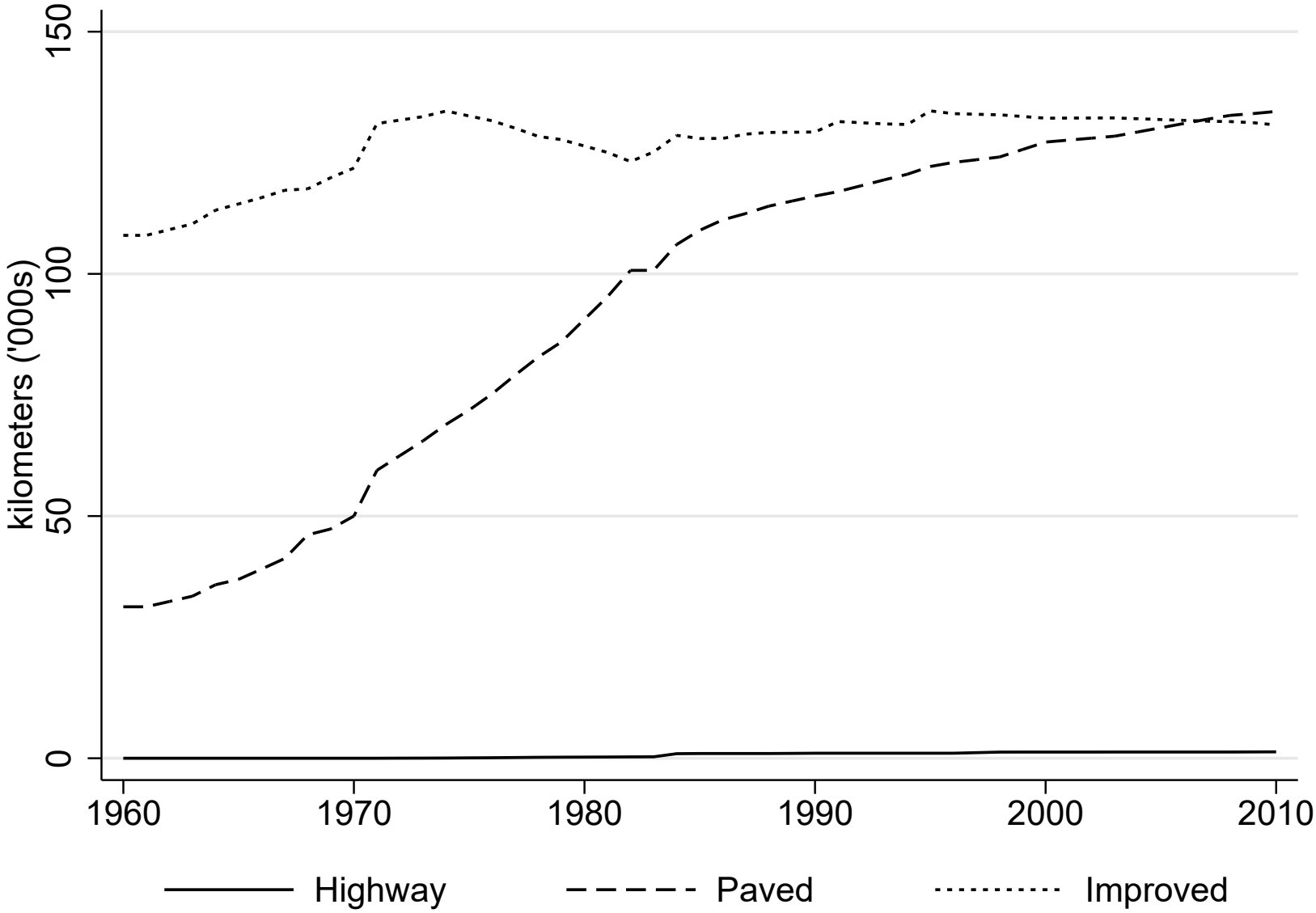
Road Length: Michelin vs. Canning (2008) vs. World Bank



Road Length: Michelin vs. Canning (2008) vs. World Bank



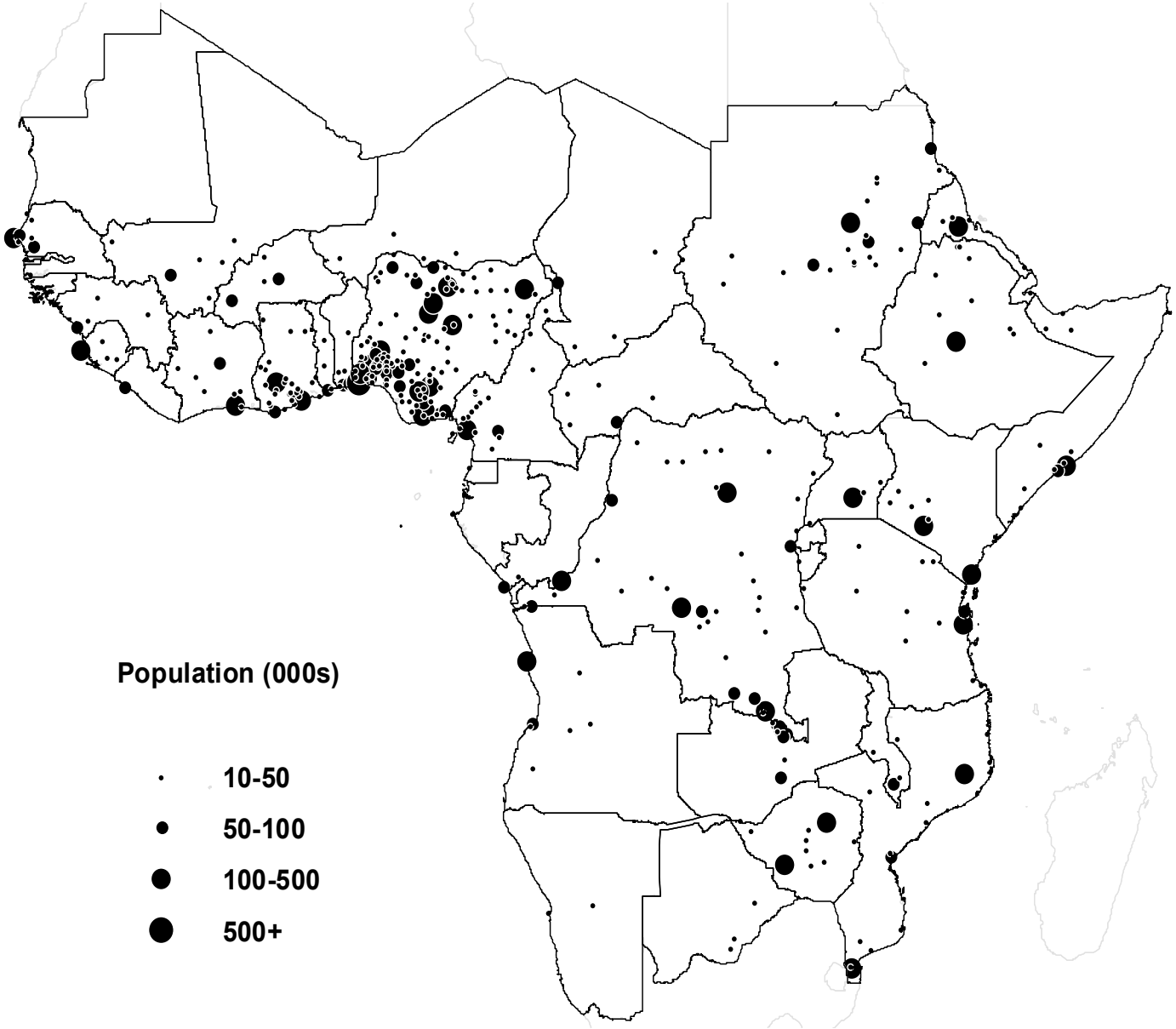
Road Length in Sub-Saharan Africa (39 Countries)



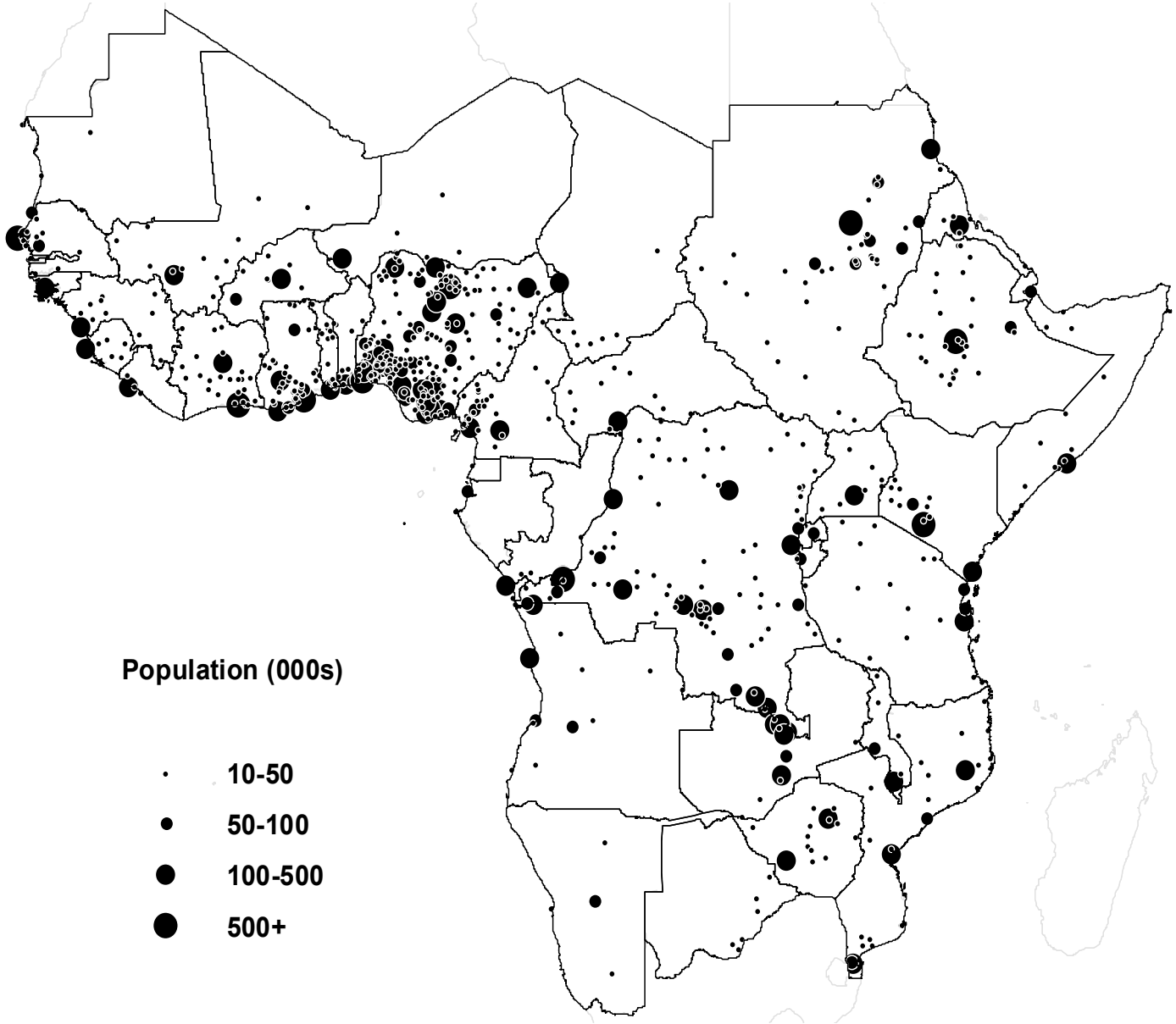
Data: GIS database of cities

- Population of localities ever above 10,000 inh. for the same 39 countries in 1960, 1970, 1980, 1990, 2000 and 2010
- Proxy for local economic development in the absence of other data (no land prices, no systematic rural populations before c. 1990, no night lights before 1992).
- Sources: *Africapolis I* and *II* for 33 countries + *Population Census* data for 6 countries (similar methodology)
- We have population estimates for many city-year observations when below 10,000 (but not for all of them).

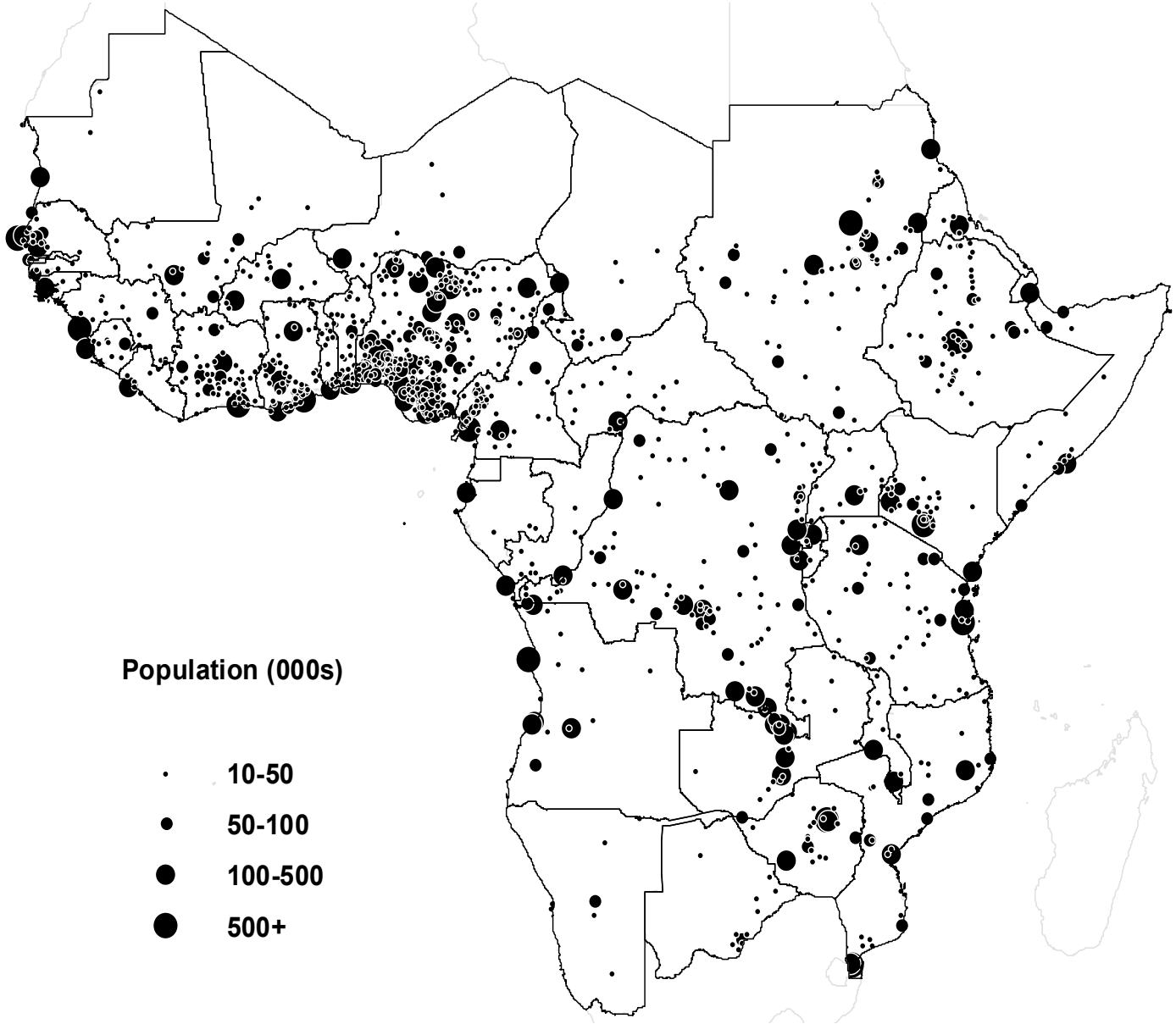
Cities ($\geq$ 10,000 Inh.) in 1960



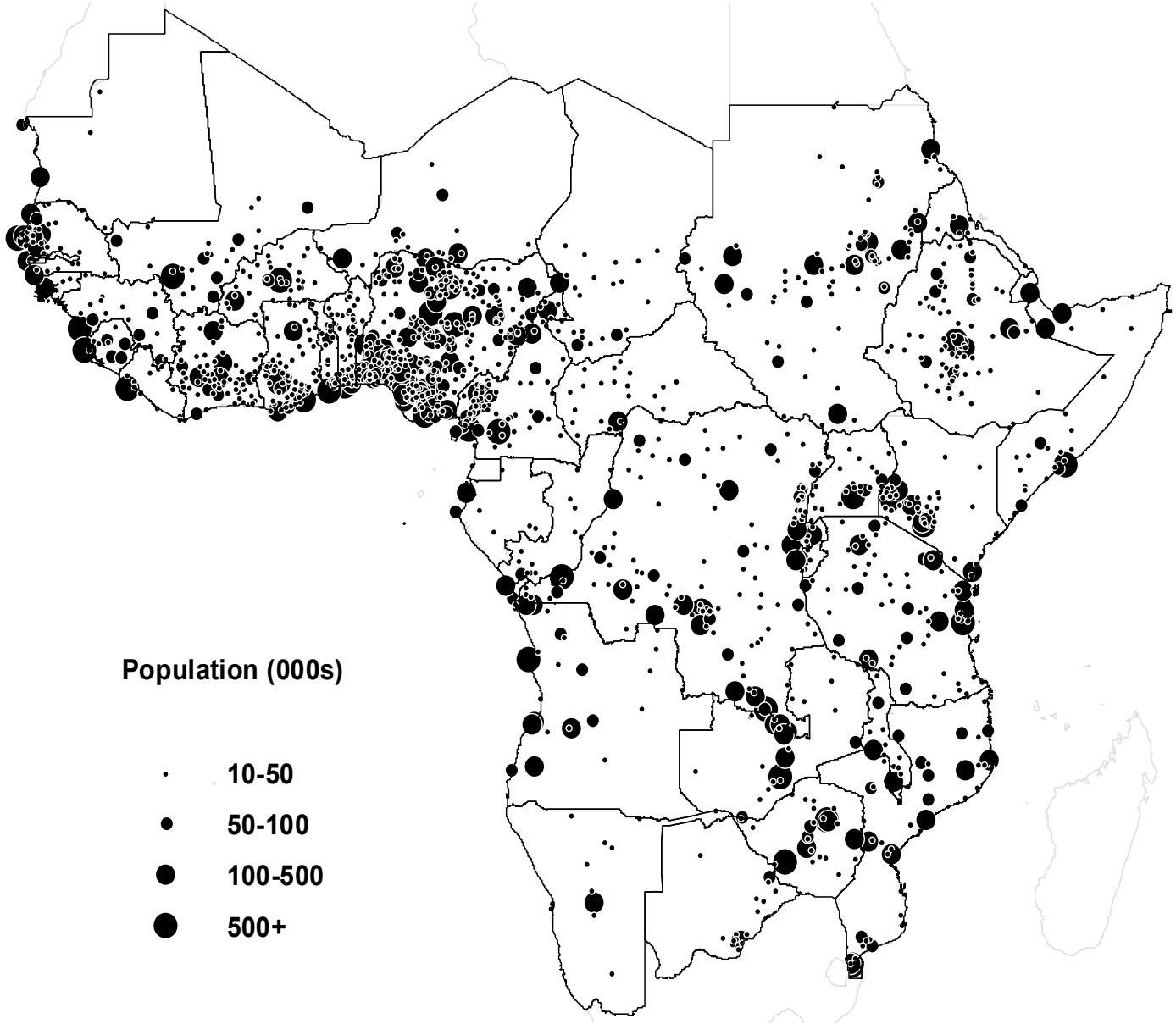
Cities ($\geq$ 10,000 Inh.) in 1970



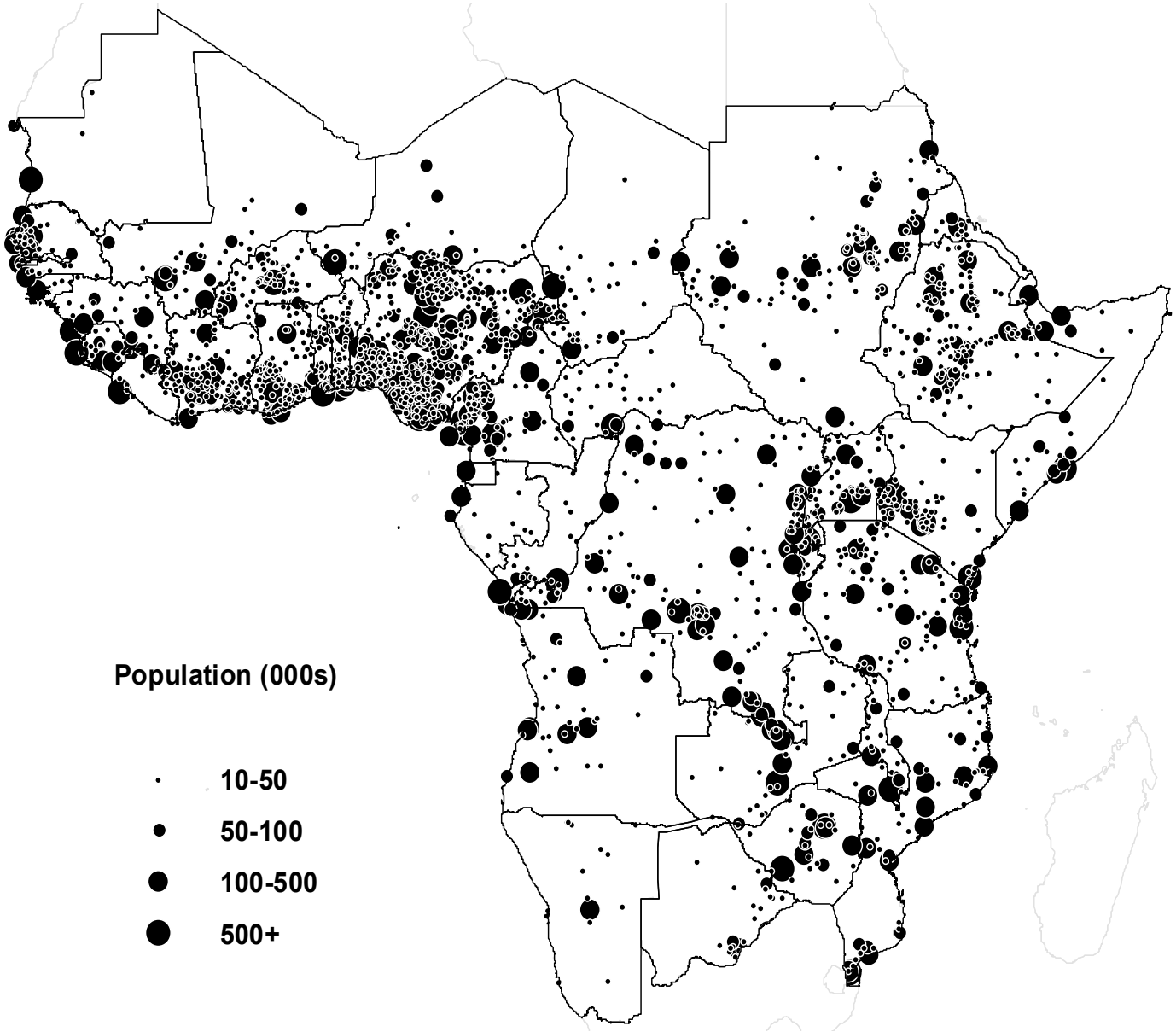
Cities ($\geq$ 10,000 Inh.) in 1980



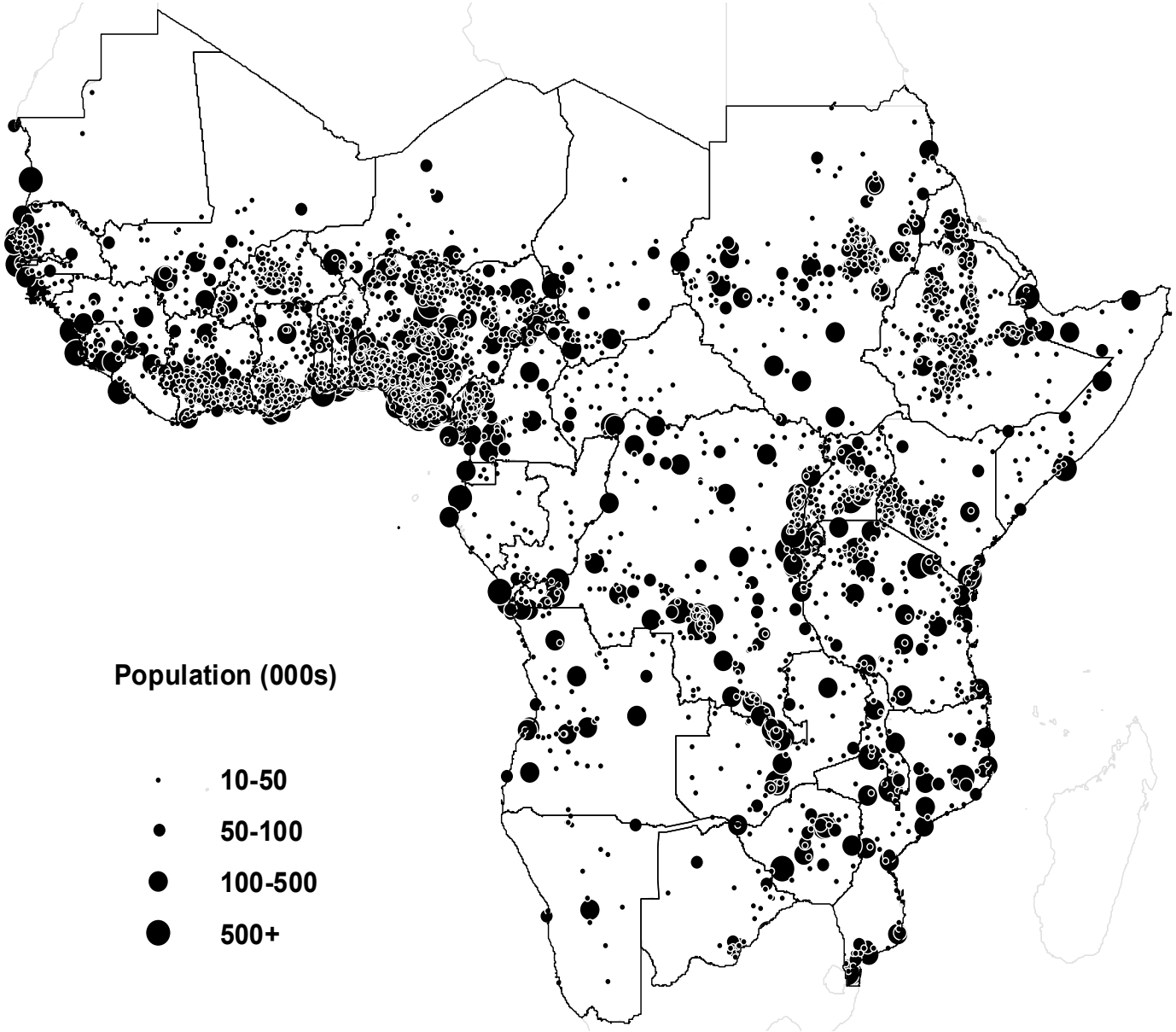
Cities ($\geq$ 10,000 Inh.) in 1990



Cities ($\geq$ 10,000 Inh.) in 2000



Cities ($\geq$ 10,000 Inh.) in 2010



Outline

- Data
- **Estimation**
- Results
- Conclusion

Unit of Analysis: Grid squares

- 0.1x0.1 degree (~11x11 km; due to computational constraints)
- Potentially multiple roads and cities per cell
  - Select best road in the cell:

    Highway $>$ paved $>$ improved $>$ dirt
  - Use the sum of city populations within cell (fortunately, only 3% of populated cells have multiple cities)

Sample

- Full sample: 5,906 city-years for 2,127 cities (population > 10,000 in at least two consecutive years)
  - 2010: 2,119
  - 2000: 1,514
  - 1990: 1,094
  - 1980: 746
  - 1970: 433
  - 4,725 city-years for 2,126 cities when including two lags

Market Access

- The market access is the travel cost-discounted sum of the population of all other cities:
  - $M_o = \Sigma_{d \neq o} P_d \tau_{od}^{-\theta}$, where
  - P_d = Population of city d
  - τ_{od} is the cost of travel from o to d, and
  - θ is the trade elasticity, baseline = 3.8
- Approximation to a recursive formulation that arises from Eaton-Kortum-type models.
- Focus on changes in market access due to changes in roads, not due to changes in population:

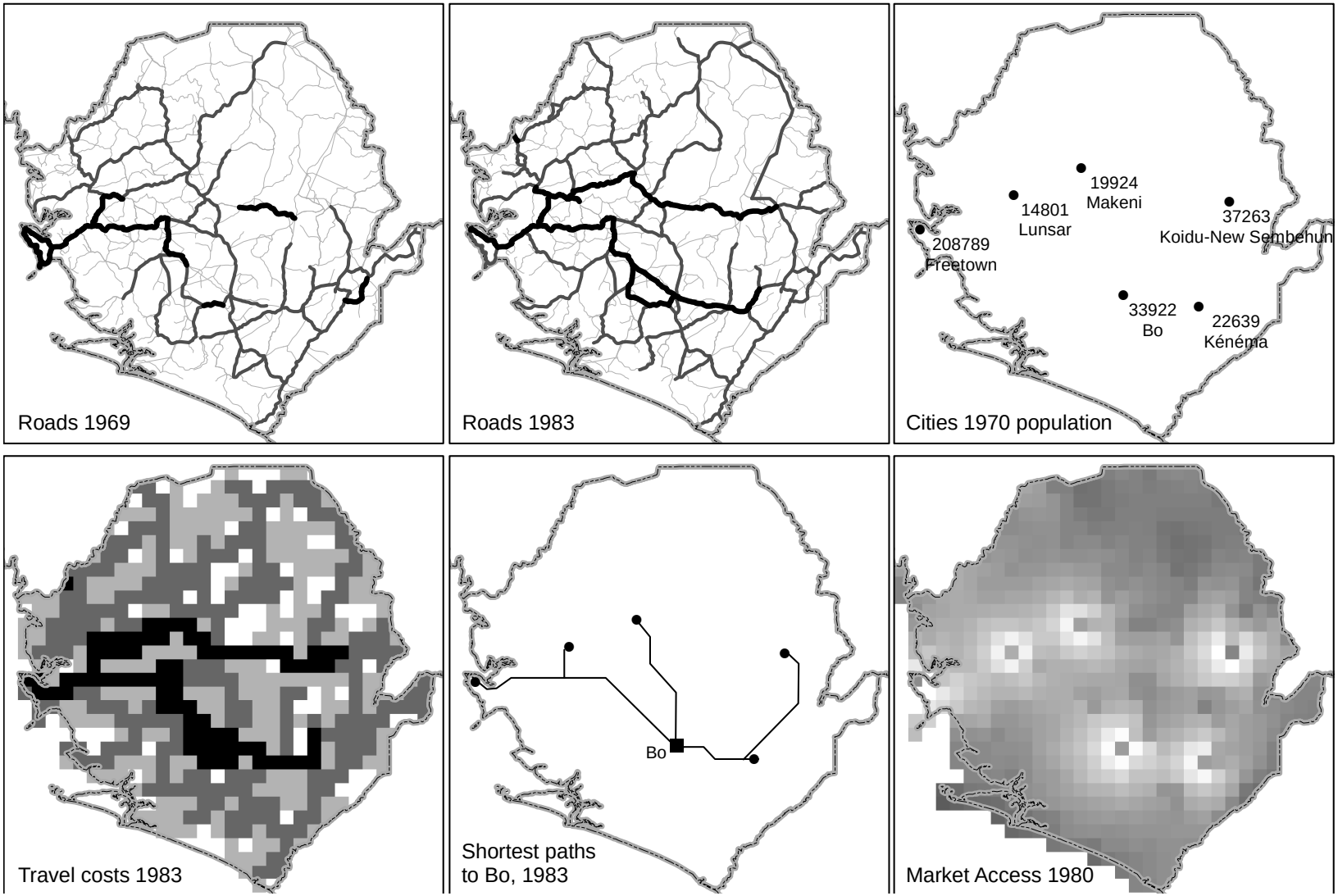
$$\Delta \ln M_{ot} = \ln \Sigma_{d \neq o} P_{d,t-10} \tau_{odt}^{-\theta} - \ln \Sigma_{d \neq o} P_{d,t-10} \tau_{od,t-10}^{-\theta}$$

Travel Speeds Used to Compute the Driving Times

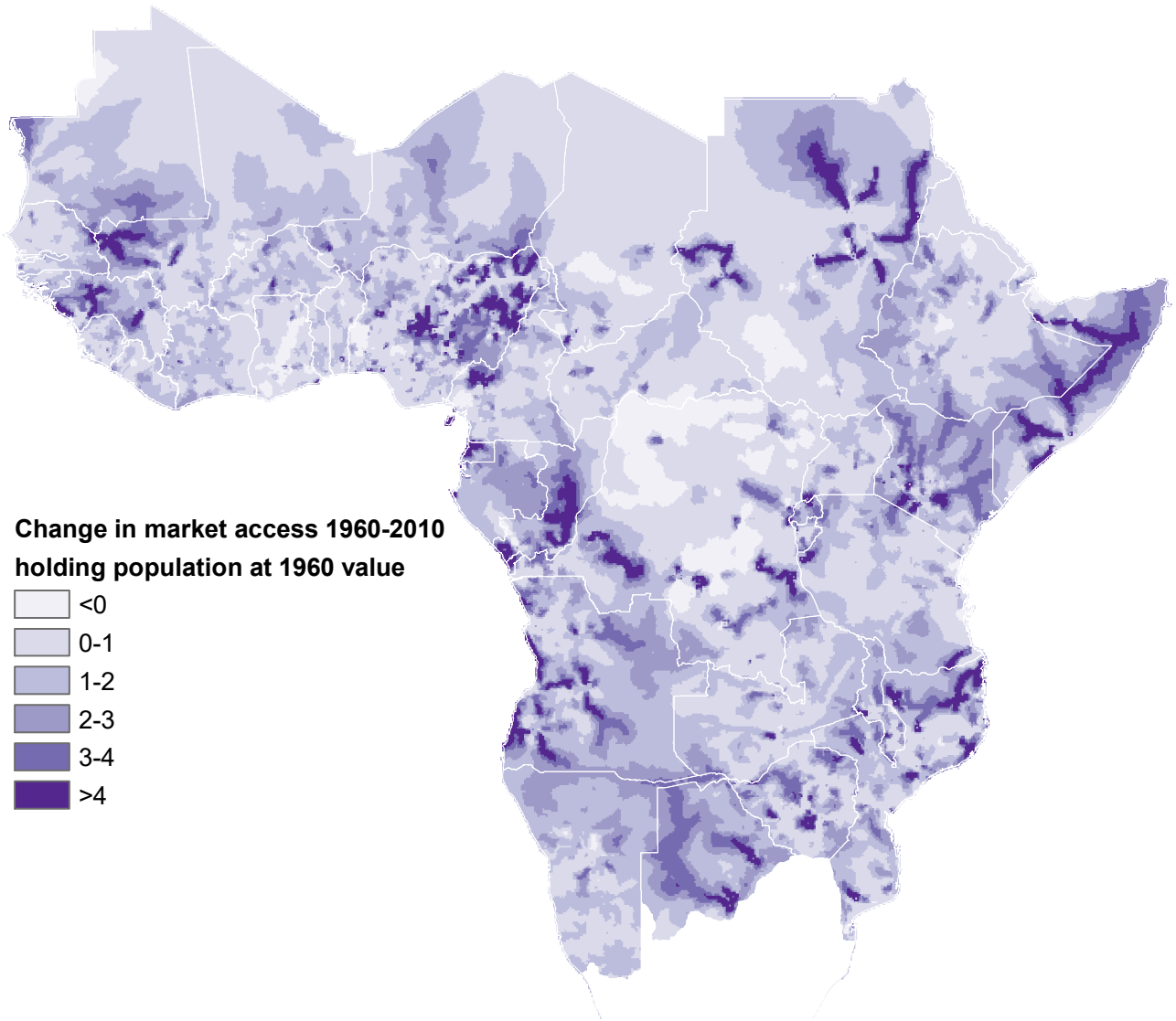
Category	This Paper	India (Alder, 2015)		Ethiopia (Shiferaw et al., 2013)	
Highway	80	Golden Quadrilateral	75	Post-Rehab Asphalt	70
Paved Road	60	Conventional Highways	35	Pre-Rehab Asphalt	50
Improved Road	40	Roads of Lower Quality	25	Pre-Rehab Federal Gravel	35
Dirt Road	12	Unpaved/No Roads	10	Pre-Rehab Regional Gravel	25
No Road	6	Unpaved/No Roads	10	Pre-Rehab Earth	20

We will show that results do not depend on the speeds used.

Example for Sierra Leone, 1970-1980



Change in market access due to road changes, 1960-2010



Specification

- Estimating equation:

$$\Delta \ln P_{ot} = \beta \Delta \ln MA_{ot} + lags + controls + u_{ot} \qquad (1)$$

P = urban population

MA = market access

- Donaldson and Hornbeck (2016) derive this specification from an Eaton-Kortum style Ricardian model of trade across US counties.

Summary stats (baseline specification, 2 lags, N=4,725)

Main Variable:	Mean	Std. Dev.	Min	Max
Δ_{t-10}^{t} ln urban pop	0.318	0.209	-1.533	2.343
Δ_{t-10}^{t} ln MA	0.655	0.892	-8.236	10.618
Δ_{t-20}^{t-10} ln MA	0.901	1.099	-8.236	11.537
Δ_{t-30}^{t-20} ln MA	1.161	1.288	-8.236	13.291
ln urban pop$_{t-10}$	10.247	0.990	9.210	15.902

Identification concerns

- **Omitted variables:** unobserved productivity shocks that drive road building and city growth.
- **Reverse causality:** roads built to cities expected to grow (or expected to lag).
- **Measurement error:** speed assumptions are rough proxies, road quality, etc.

Identification

- **Baseline specification:**
  - First-difference estimator to control for fixed productivity/amenities
  - Additional controls:
    - Initial log urban population (mean reversion)
    - Country-year fixed effects.
    - Decade-specific third-order polynomial in longitude/latitude.

Identification

- **Baseline specification:**
  - First-difference estimator to control for fixed productivity/amenities
  - Additional controls:
    - Initial log urban population (mean reversion)
    - Country-year fixed effects.
    - Decade-specific third-order polynomial in longitude/latitude.
- **Instrumental variables:**
  - Exploit **non-local changes in road market access** as they are more likely to be exogenous to local changes in city population:
    - Hold population fixed at $t - 10$ (1960 in robustness checks)
    - Road changes 5, 10, 15 cells away ($\approx$55, 110, 165 km).

Instrument

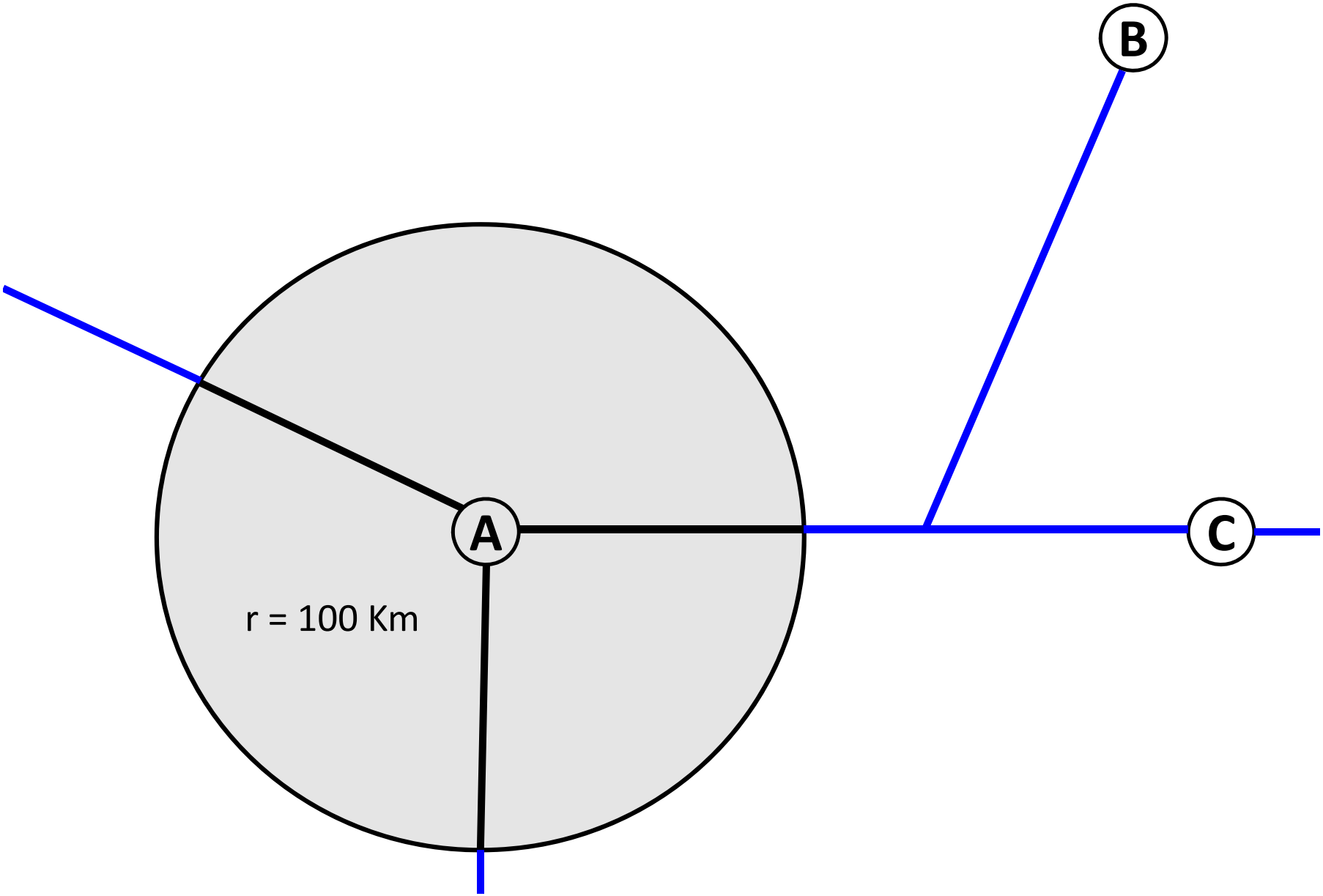
For distance $j \in \{5, 10, 15\}$ cells,

$$\Delta_R^{out,j} \ln MA_{ot} = \ln\left(\sum_{d:\delta(d,o)\geq j} P_{d,t-10}\tau_{od,t}^{-\theta} + \sum_{d:0<\delta(d,o)<j} P_{d,t-10}\tau_{od,t-10}^{-\theta}\right) - \ln\left(\sum_{d\neq o} P_{d,t-10}\tau_{od,t-10}^{-\theta}\right)$$

where

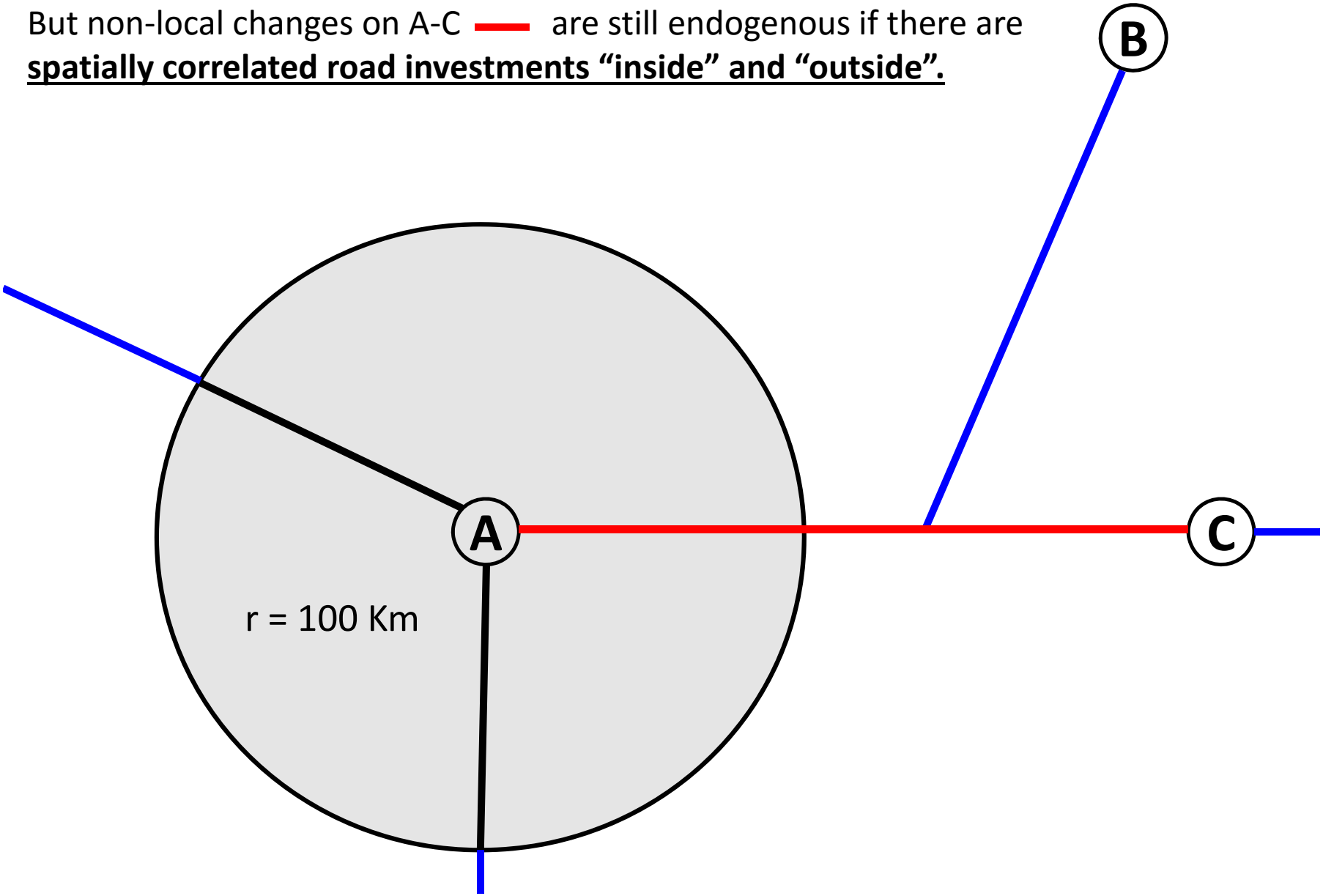
- P_d = Population of city d
- τ_{od} is the cost of travel from o to d
- θ is the trade elasticity
- $\delta(d, o)$ is the Euclidean distance between d and o.

We only use **non-local** road changes



We only use **non-local** road changes

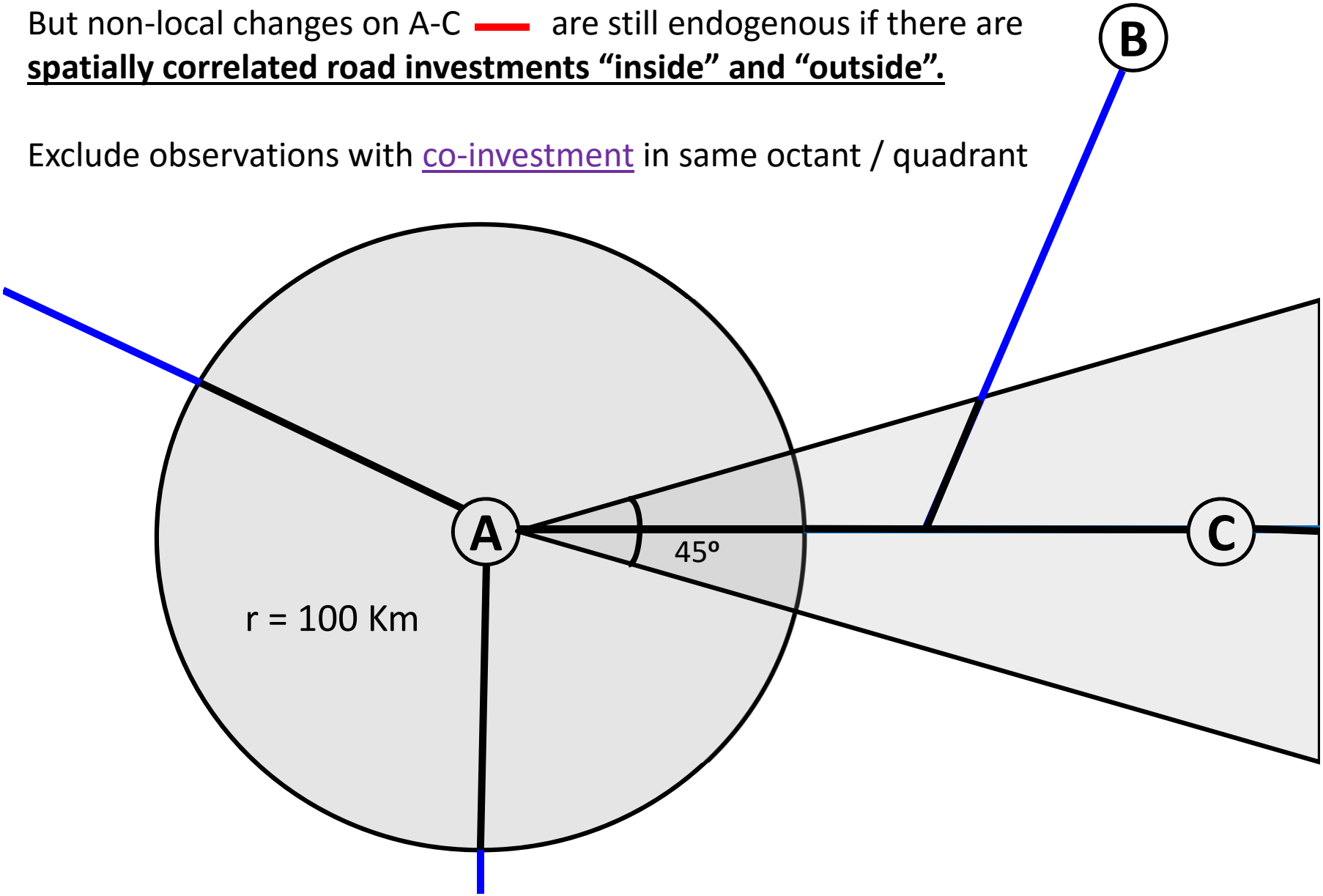
But non-local changes on A-C are still endogenous if there are **spatially correlated road investments "inside" and "outside".**



We only use **non-local** road changes

But non-local changes on A-C are still endogenous if there are **spatially correlated road investments "inside" and "outside".**

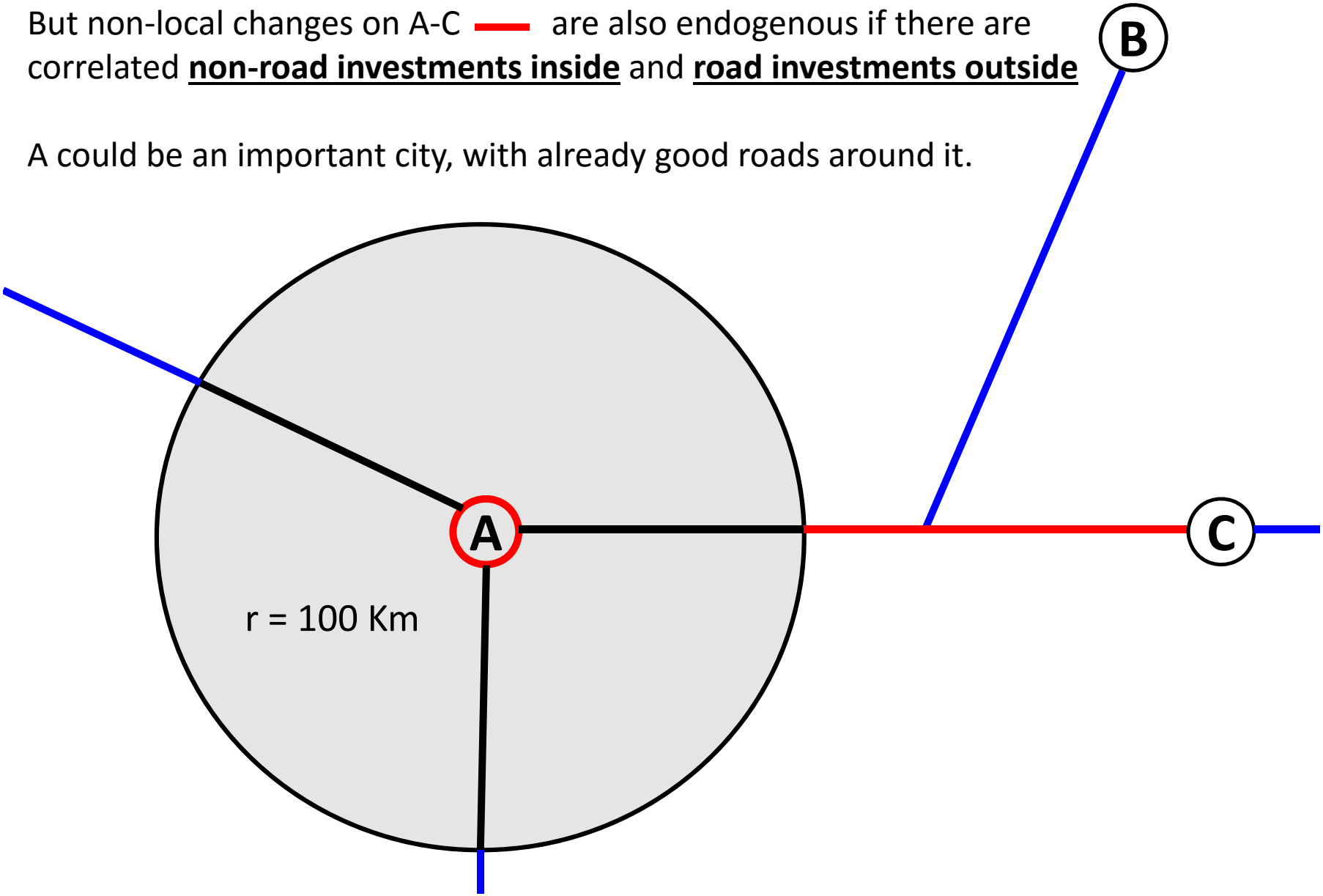
Exclude observations with co-investment in same octant / quadrant



We only use **non-local** road changes

But non-local changes on A-C are also endogenous if there are correlated **non-road investments inside** and **road investments outside**

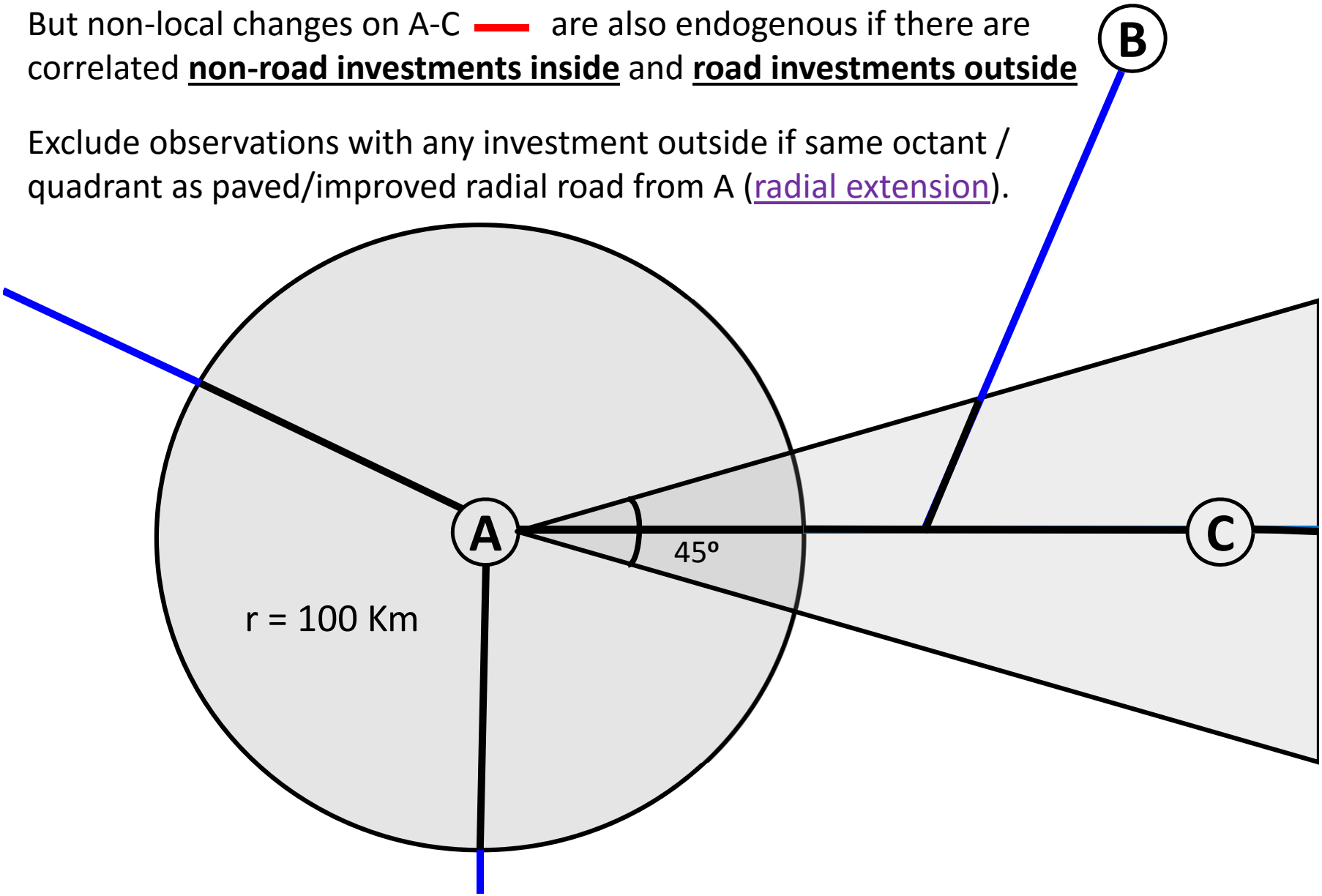
A could be an important city, with already good roads around it.



We only use **non-local** road changes

But non-local changes on A-C are also endogenous if there are correlated **non-road investments inside** and **road investments outside**

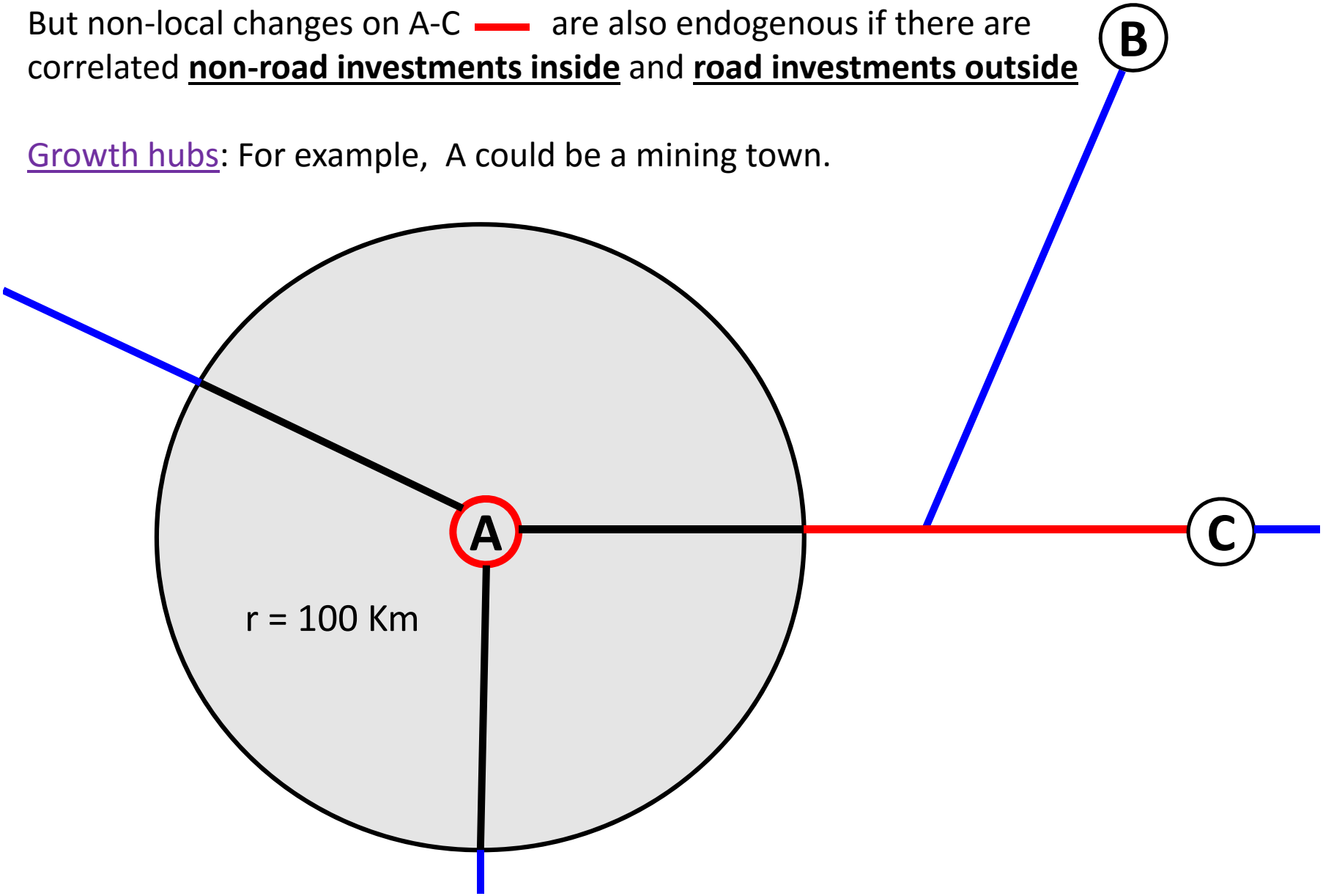
Exclude observations with any investment outside if same octant / quadrant as paved/improved radial road from A (radial extension).



We only use **non-local** road changes

But non-local changes on A-C are also endogenous if there are correlated **non-road investments inside** and **road investments outside**

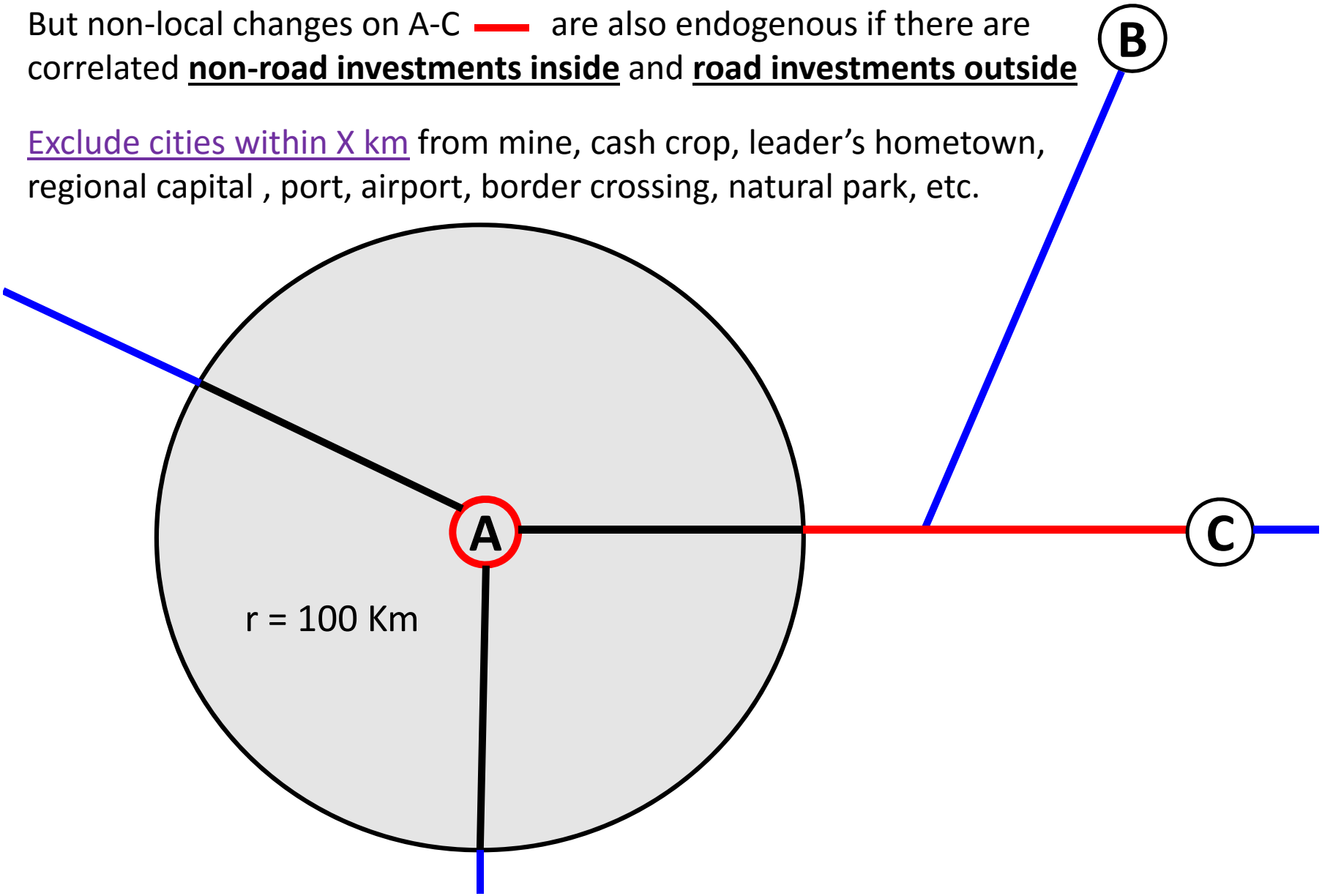
Growth hubs: For example, A could be a mining town.



We only use **non-local** road changes

But non-local changes on A-C are also endogenous if there are correlated **non-road investments inside** and **road investments outside**

Exclude cities within X km from mine, cash crop, leader's hometown, regional capital , port, airport, border crossing, natural park, etc.



Outline

- Data
- Estimation
- **Results**
- Conclusion

Table 1: Average Effect of Market Access on Urban Population

Panel A: OLS	(1)	(2)	(3)	(4)	(5)
Δ^{t}_{t-10} ln Market Access	1.33***	1.33***	1.56***	1.57***	1.52***
	[0.38]	[0.42]	[0.37]	[0.46]	[0.41]
Δ^{t-10}_{t-20} ln Market Access		0.98***	1.18***	1.49***	1.10***
		[0.27]	[0.31]	[0.36]	[0.34]
Δ^{t-20}_{t-30} ln Market Access			0.73***	0.80**	0.76**
			[0.24]	[0.32]	[0.30]
Δ^{t-30}_{t-40} ln Market Access				0.31	
				[0.25]	
Δ^{t+10}_{t} ln Market Access					0.72
					[0.56]
Overall Effect	1.33***	2.31***	3.47***	4.18***	3.39***
($t-40$ to t)	[0.38]	[0.58]	[0.63]	[0.83]	[0.71]
Observations	5,906	5,472	4,725	3,630	2,607
Adj. R-squared	0.23	0.19	0.17	0.16	0.19

Panel B: IV	IV: Exclude 5 (1)	IV: Exclude 10 (2)	IV: Exclude 15 (3)
Δ_{t-10}^{t} ln Market Access	3.09***	4.45**	5.55*
	[1.10]	[1.82]	[2.99]
Δ_{t-20}^{t-10} ln Market Access	3.04***	5.56***	6.68**
	[0.87]	[1.50]	[2.62]
Δ_{t-30}^{t-20} ln Market Access	2.23**	2.88**	4.27**
	[0.88]	[1.38]	[1.94]
Overall Effect	8.35***	12.89***	16.49***
($t-30$ to t)	[2.16]	[3.23]	[4.60]
Observations	4,725	4,725	4,725
1st stage Kleibergen-Paap F	98.56	42.94	11.90

Robustness Checks 1/2

Results are robust to:

- Calculating MA based on fixed 1960 Pop or Δ^{t}_{t-10}Pop or using alternate speeds, contiguity, closing borders
- Changing the trade elasticity (1–13; magnitudes change much, effect of 1 SD change little when $\theta \geq 3$)
- Removing country-year FEs or lagged population
- Controlling for dist. to coast/rivers, rain, soils, altitude, 1960 MA by year, railroads, or railroad MA, past pop. growth
- Balancing panel with population estimates

Robustness Checks 2/2

Results are also robust to:

- Separately dropping: largest few cities & capital in each country; regional capitals; cities w/**co-investment/radial extension**; cities w/in 100km of mines, ports, airports, head of state's hometown, cash crop areas; South Africa neighbors; Maghreb neighbors; 2000s; country-decades with wars, multi-year droughts or poor population data; outliers

Table 3: Effect of Market Access on Night Lights

	(1) OLS	(2) IV: Excl. 5	(3) IV: Excl. 10	(4) IV: Excl. 15
Δ^{t}_{t-10} ln MA	0.39	22.29**	43.98***	69.37***
	[3.06]	[9.97]	[11.98]	[18.56]
Δ^{t-10}_{t-20} ln MA	1.70	12.52	8.82	5.84
	[2.90]	[8.12]	[12.50]	[16.91]
Δ^{t-20}_{t-30} ln MA	0.84	3.90	0.71	-3.01
	[2.11]	[4.32]	[7.33]	[10.21]
Overall Effect	2.93	38.70***	53.51***	72.19***
	[5.22]	[11.39]	[18.31]	[26.49]
1st stage Kleibergen-Paap F		53.24	29.74	9.837

Notes: See Table 1. Outcome variable is $100\Delta^{t}_{t-10}$ ln (Light Intensity). N = 3,591. Robust SEs, clustered by 1960 province, are in brackets. *, **, *** = 10, 5, 1% significance.

Roads increase economic activity quickly. The population effects take longer to evolve. Suggests **migration flows** take substantial time to develop.

Induced Urbanization vs. Reallocation, Natural Increase

- Cannot directly test whether increases are due to
  - *Induced rural-urban migration*
  - *Reallocation across cities*
  - *Natural increase* (births and deaths)

Induced Urbanization vs. Reallocation, Natural Increase

- Cannot directly test whether increases are due to
  - *Induced rural-urban migration*
  - *Reallocation across cities*
  - *Natural increase* (births and deaths)
- Evidence consistent with primary role for induced urbanization
  - Restrict sample to initially **less urbanized** country-periods ($< 10\%$ or 7%; less scope for urban reallocation): similar results
  - Use **mega-cells** of 3–9 cells (33–99 km) (and drop those with national/regional/largest cities, since reallocation less likely between non-neighboring local cities). Noisy but still sizable effect
  - Test for direct effect on **natural increase** (Demographic and Health Survey data): suggests no more than half of overall effect

Summary of Average Effects

- Naive effect of a 10% change in market access: $\approx$ 0.1%–0.15% per decade for three decades (total 30-year effect: 0.3–0.4%).
- **IV: $\approx$ 0.8–1.3% over 30 years.**
- Concentrated in first two decades (i.e. decade of construction and following decade)
- No measurable effect in fourth decade.
- Source of growth: rural areas, other cities, or natural increase?
  - Natural increase and urban reallocation unlikely to be large share of effect. Rural-urban migration must be large share.

Comparison to literature

- Somewhat smaller than railroads in the 19th century US using similar method (Donaldson & Hornbeck 2016): ≈20–35%.
- More comparable to other contexts.
- **Contextual differences:**
  - Not a transportation revolution like in the 19th century US. Railroads already existed in Africa before roads (and poor roads existed before good roads).
  - Migration costs likely higher at least for large distances.
  - Lower economic growth.

Heterogeneous Effects

- We classify the cities into two groups depending on:
  - High vs. low initial **market access**.
  - High vs. low **land suitability** for crops.
  - Ethnic homeland areas of **head of state** vs. rest (we use newly collected data on place of origin and ethnicity of 189 heads of state 1960–2010).
- See if the overall effect of a same change in road market access varies across the two groups.
- Possibly important for policy.

Table 5: Heterogeneous Effects of Market Access on Urban Population

	OLS	Col. (2)–(4): IV5			IV10	IV15
	Diff.	0	1	Diff.	Diff.	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)
(1) > Med. Dist. Top Cities	7.54***	2.16	9.21***	7.05***	13.41***	21.13***
(F: _; 41.3; 10.5; 2.1)	[1.30]	[2.65]	[2.20]	[2.66]	[3.74]	[5.76]
(2) Crop Suitability < 25%	-0.92	6.95***	11.75***	4.80	11.28*	20.54***
(F: _; 17.0; 8.6; 3.9. Sh: 0.16)	[1.34]	[2.12]	[4.32]	[4.42]	[6.03]	[7.87]
(3) Leader's Origin 150km	-2.20*	9.68***	1.68	-8.00**	-7.35	-9.88
(F:_; 12.6; 5.7; 5.2, Sh: 0.24)	[1.26]	[2.20]	[3.94]	[4.00]	[5.21]	[6.29]

Larger effects for smaller and more remote places (decentralization).

Bigger effects for areas with worse agricultural land (trade specialization?)

Smaller in ethnic homeland areas of head of state ("roads to nowhere")

Conclusion

- Study the effects of road construction and market access on city population growth in Sub-Saharan Africa in 1960-2010.
- New panel data set on road surface and city population for 39 African countries every ten years in 1960-2010.
- Average effect of a 100% change in market access $\approx$ 8–13%.
- Heterogeneity in the effects. Need to understand local context when evaluating the impact of transport investment.