Migration Shocks and the Composition of African Migrants

Andre Gray [1]

[1]Approved by Sam, Sara

May 15, 2025

Motivation

What is the effect of internal migration on urban productivity in Africa?

Migration affects 2 factors:

1. Labor – The levels, number of people arriving (Castells-Quintana, 2017)
2. Composition – Second moment, mix of people arriving (Alesina and Ferrara, 2005)
   - Ethnicity, Religion, Language
   - Skill complementarity, labor market concentration

Ethnic diversity makes African migration different

- Linguistic barriers within country (Wang, 2024)
- Ethnic conflict (Arbatli et al., 2020; McGuirk and Nunn, 2024)

What We Do

Research Q: What is the effect of migration levels and composition on destination outcomes?

1. Build a panel that connects workers in destination d to homeland o using ethnicity/language/birthplace
   - Proxy of origin-destination panel of migration flows

What We Do

Research Q: What is the effect of migration levels and composition on destination outcomes?

1. Build a panel that connects workers in destination d to homeland o using ethnicity/language/birthplace
   - Proxy of origin-destination panel of migration flows

2. Use shift-share instrument to estimate impact of migration level and ethnic concentration on output, productivity, conflict.
   - Innovation: add nonlinearity to standard shift-share
   - Use push shocks to predict the shifts (outmigration rate)
     - ★ *Droughts, Commodity Prices*

Estimating Equation

People in destination d at time t are from origins o:

$$\Delta y_{d,t} = \alpha \Delta l_{d,t} + \gamma \Delta div_{d,t} + \upsilon_t + \epsilon_{d,t} \tag{1}$$

- $\Delta y_{d,t}$: Light density growth, Conflict events, Consumption, Ag-share
- $\Delta l_{d,t}$: Change in in-migration to d from all o as $\sum_o count_{odt}$
- Δdiv_{dt}: Change in ethnic concentration Herfindahl $HHI = \sum_o \pi_{od}^2$, where π_{od} is share of people in d from o $\frac{count_{od}}{count_d}$

Production Fucntion

Empirical Challenges

$$\Delta y_{d,t} = \alpha \Delta \widehat{l_{d,t}} + \gamma \Delta \widehat{div_{d,t}} + \upsilon_t + \epsilon_{d,t} \tag{2}$$

1. Data
   - No subnational migration flows to get $\Delta l_{d,t}, \Delta div_{dt}$
   - Limited time variation (sparse census waves)
2. Identification: Need 2 instruments
   - Unrelated to local productivity changes of the city $y_{d,t}$
   - Move migration levels (labor) and composition (diversity)

Empirics 1: General Shift-Share Design

Construct general bartik-shift share design for both flows and ethnic concentration.

$$\Delta y_{d,t} = \alpha \Delta \widehat{l_{d,t}} + \gamma \Delta \widehat{div_{d,t}} + \upsilon_t + \epsilon_{d,t} \tag{3}$$

Shift-Share Instrument For Linear Outcome

Each destination d has people coming from a set of origins o. Estimating labor in-migration is standard, baseline shares and exogenous shifts.

$$\Delta y_{d,t} = \alpha \Delta \widehat{l_{d,t}} + \gamma \Delta \widehat{div_{d,t}} + \upsilon_t + \epsilon_{d,t} \tag{4}$$

Each origin o has an out-of-homeland growth rate $g_{ot} = \frac{count_{ot} - count_{od,t-1}}{count_{od,t-1}}$.
Each destination d is exposed according to a baseline share of o.

$$\Delta \hat{l}_{dt} = \sum_o \frac{count_{od,t-1}}{\sum_d count_{od,t-1}} * g_{ot} * count_{o,t-1} \tag{5}$$

Shift Instrument For Nonlinear Outcome

The share of people in d that come from o is $\frac{count_{od}}{count_d} = \pi_{od}$

Change in diversity is change in ethnic HHI:

$$\Delta div_{dt} = \sum_o \pi^2_{odt} - \sum_o \pi^2_{od,t-1}$$

The initial HHI multiplied by the relative growth of each o to the overall destination growth rate $g_{dt} = \frac{l_{dt} - l_{d,t-1}}{l_{d,t-1}}$:

$$\Delta \hat{div}_{dt} = \sum_o \pi^2_{od,t-1} * \left(\frac{(1+g_{ot})^2}{(1+g_{dt})^2} - 1 \right)$$

Building a Proxy $o - d$ panel

- African Census Districts *[1970 - 2019]*
  - Link individuals in district d and wave t to their homelands o (Müller-Crepon et al., 2022)
  - Unbalanced time spans between census waves (avg. 10 year gaps)
  - For a subset, we could use birthplace instead
- We can also do this with DHS/Afrobarometer
  - Aggregate points to a city level
  - Small samples, but more places and years

Tribes Limitations DHS Example Sanity Check

Census-Birthplace Example

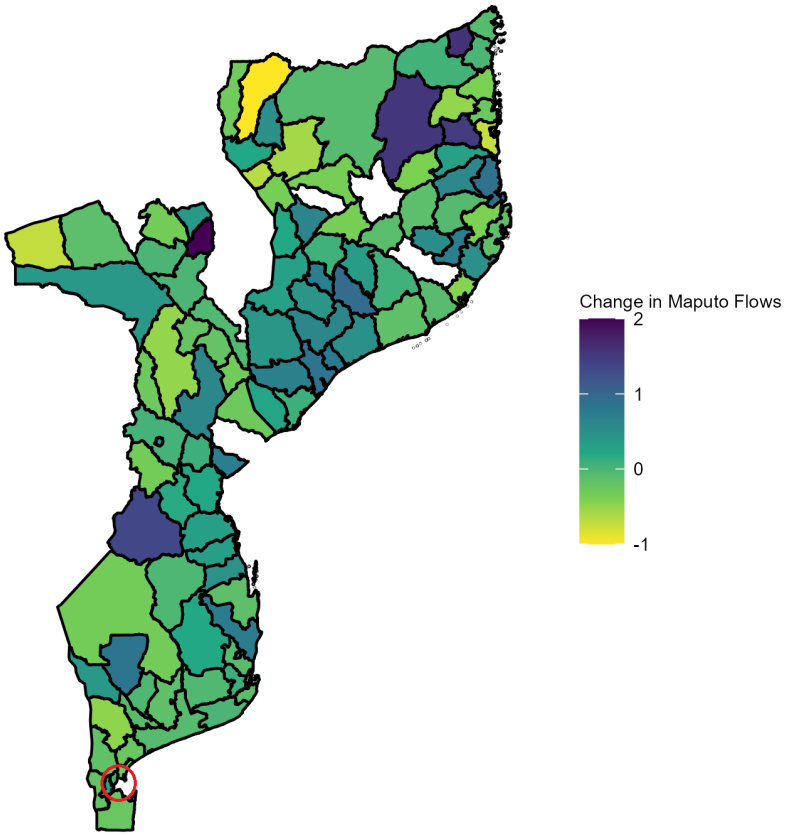


Figure: Mozambique Maputo Inflow, 1997-2007 Proportion Change

Shift-Share First Stage

	Δ l	Δ div	div
Predicted Δ l	0.481	0.029	0.108
	[0.044]***	[0.004]***	[0.022]***
Predicted Δ div	-0.002	0.006	0.105
	[0.030]	[0.003]**	[0.015]***
Mean Dep.	8.06	0.02	-1.88
Observations	444	444	444
Fstat	60.638	26.306	31.462

Note: $real_{d,t} = \alpha\widehat{\Delta l_{d,t}} + \gamma\widehat{\Delta div_{d,t}} + \upsilon_t + \mu_c$. Include country, year fixed effects. Geographic controls include malaria suitability.

Shift-Share Second Stage

$$\Delta y_{d,t} = \alpha \Delta \widehat{l_{d,t}} + \gamma \Delta \widehat{div_{d,t}} + \upsilon_t + \epsilon_{d,t} \tag{6}$$

Choices for y_{it}:

- Light Density – Harmonized VIIRS and DMSP 1992-2021
- Conflict Events – UCDP Geolocated Battles 1980 - 2023
- Other (to-do): Household Consumption, Port Efficiency

Shift-Share Second Stage

Table: OLS Results

	Log(lights)	Log(Δ lights)	Δ Lights	Δ Lights/Capita
Δ l	1.148	1.175	645.779	-0.099
	[0.293]***	[0.520]**	[505.060]	[0.217]
Δ div	8.107	11.407	3525.404	5.289
	[3.091]***	[5.490]**	[5335.864]	[2.215]**
Mean Dep.	3.35	-3.55	5116.83	-6.09
Observations	512	512	512	421

Note: $y_{d,t} = \alpha \Delta l_{d,t} + \gamma \Delta div_{d,t} + \upsilon_t + \mu_c$. Include country, year fixed effects. Geographic controls include malaria suitability. Lights are nighttime light density.

Shift-Share Second Stage

Table: 2SLS Results

	Log(lights)	Log(Δ lights)	Δ Lights	Δ Lights/Capita
Δ l	5.228	6.650	-4101.383	1.618
	[1.778]***	[3.087]**	[2424.976]*	[1.088]
Δ div	8.619	0.906	92569.826	12.351
	[13.597]	[23.599]	[18541.000]***	[8.024]
Mean Dep.	2.42	-4.98	4747.00	-6.57
Observations	289	289	289	245
Cragg-Donald Fstat	10.284	10.284	10.284	10.719

Note: $y_{d,t} = \alpha\widehat{\Delta l_{d,t}} + \gamma\widehat{\Delta div_{d,t}} + \upsilon_t + \mu_c$. Include country, year fixed effects. Geographic controls include malaria suitability. Lights are nighttime light density.

Shift-Share Second Stage

Table: Conflict Outcomes

	OLS Conflict	2SLS Conflict	OLS Δ Conf	2SLS Δ Conf
Δ l	0.009	-0.001	0.004	0.001
	[0.007]	[0.037]	[0.007]	[0.040]
Δ div	-0.024	-0.159	-0.048	-0.452
	[0.068]	[0.277]	[0.066]	[0.324]
Mean Dep.	0.02	0.02	0.02	0.03
Observations	544	310	432	236
Cragg-Donald Fstat		8.625		8.318

Note: Include country, year fixed effects. Geographic controls include malaria suitability. "Conflict" is number of battle events in a year.

Empirics 2: Push-Shock Design

Predict migration shifts and concentration using push-shocks at origin over time.

- Drought
- Agricultural and Mineral Commodities
- Effects on labor and labor-efficiency units (skill)
- First stage is non-obvious (limited prior work)

Instrument Intuition: From Shocks to Diversity

Previously, g_o is a growth rate from aggregate shifts of migrants from o.
Now:

$$g_{ot} = \omega ShockExposure_{ot} + \upsilon_o + \beta_t + \epsilon_{ot} \tag{7}$$

Plug $\hat{g_{ot}}$ into shift-share.

$$\Delta \hat{l}_{dt} = \sum_o \frac{count_{od}}{\sum_d count_{od}} * \hat{g_o} * count_{o,t-1}$$

$$\Delta \hat{div}_{dt} = \sum_o \pi^2_{od,t-1} * \left(\frac{(1+\hat{g_o})^2}{(1+g_d)^2} - 1 \right)$$

Shock 1: Droughts in Homelands

Claim: Droughts move people out of origins

- Not ex-ante obvious the direction of the effect on migration flows
- Only limited research on climate-urbanization connection in Africa (Henderson et al., 2017; Kamuikeni and Naito, 2024)
  - Rainfall predicts urbanization in certain types of cities (industrialized)

Example

Shock 1: Data on Droughts

Standardised Precipitation-Evapotranspiration Index (SPEI) :

- Combines temperature and precipitation data to measure intensity and duration of drought conditions
- Monthly data at a 0.5x0.5 degree cell resolution (1900 - 2022)
- Aggregate to our ethnic origins
  - Code each month as being in drought (0/1) if SPEI score is in the bottom decile
  - $DroughtExposure_{ot}$: Aggregate to yearly drought intensity (share of year in drought)

Raw SPEI Data

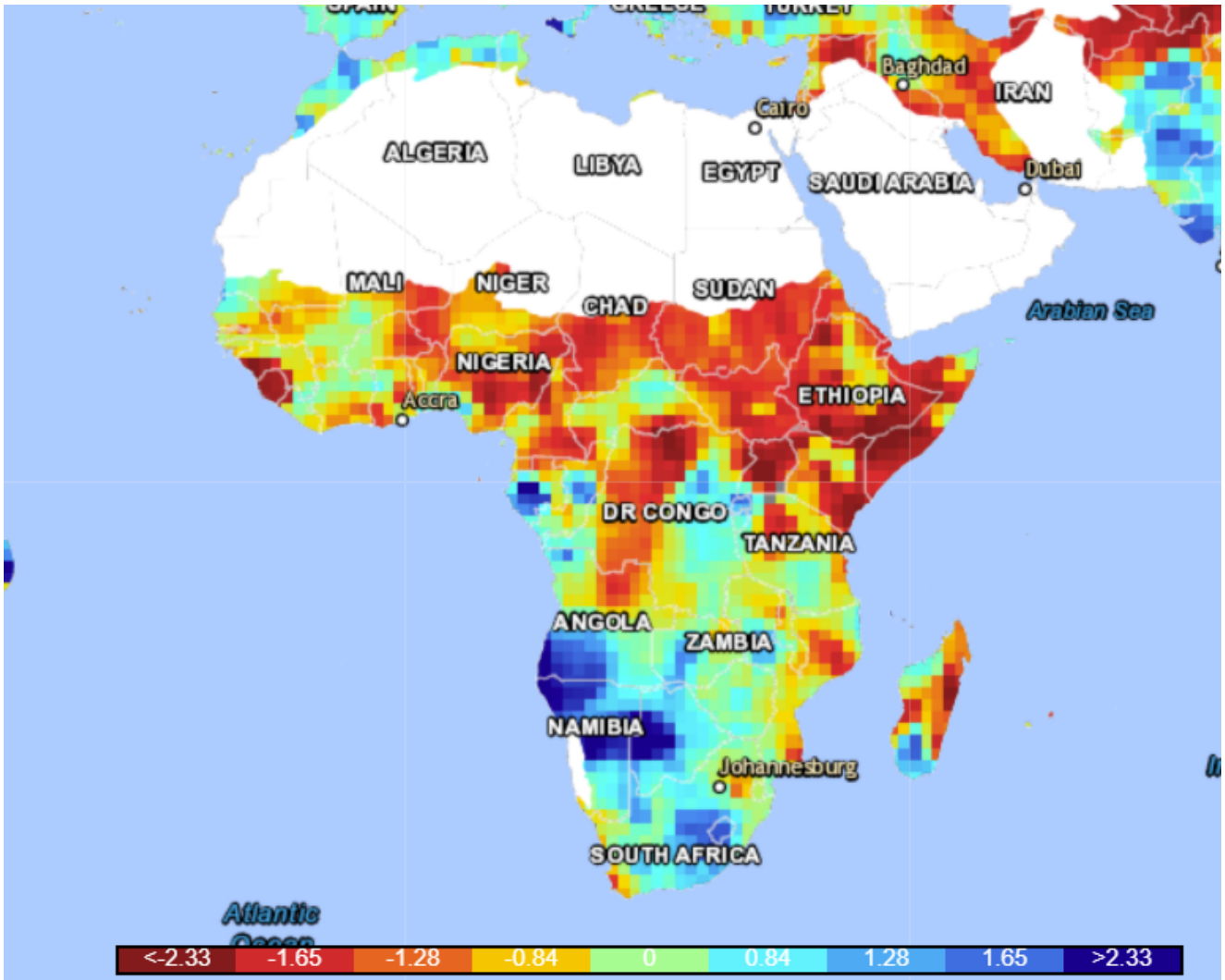


Figure: SPEI Example, 2011

Impact of Drought on Outmigration

- Count is number of people from tribal homeland o in year t living outside of o:

$$log(count)_{ot} = \omega DroughtExposure_{ot} + \upsilon_o + \beta_t + \epsilon_{ot} \tag{8}$$

Drought Shocks and Homeland Composition

	Log(Count)	Log(Count)	Log(Count)	Log(Count)
Drought Contemporary	-0.012			-0.011
	[0.005]**			[0.005]**
Drought 1 year		0.013		0.013
		[0.006]**		[0.007]*
Drought 10 year			0.005	-0.002
			[0.006]	[0.008]
Mean Dep.	7.14	7.14	7.14	7.14
Observations	878	878	878	878
Fstat	5.304	4.954	0.603	3.171

Note: $log(count)_{ot} = \omega DroughtExposure_{ot} + \upsilon_o + \beta_t + \epsilon_{ot}$. Origin fixed effects, year fixed effects. Outcome is defined as log of people from homeland o not in o

- *Number of people from homeland o living outside increases with homeland drought intensity*

Shock 2: Prices in Homelands

Claim: International cash crop prices change relative gains of rural-urban move

- Limited macro evidence connecting commodity prices and push-urbanization in Africa (Gollin et al., 2016)
  - Measuring subnational price exposure in Africa – ideas from conflict literature (McGuirk and Burke, 2020)

Shock 2: Data on Price Exposure in Homelands

- *Production Shares by Product:*
  - FAO GAEZ hectare area for major cash crops at cell level
  - Static, estimated in 2000
  - Examples – maize, wheat, sorghum
- *International Prices*
  - IMF International Financial Statistics
  - Yearly international prices for major cash crops to 2023 (Bazzi and Blattman, 2014)
- Exclude products where African homeland may not be price taker (ex. cassava)
- Yearly exposure defined as sum of international prices weighted by product hectare area

FAO Production Data

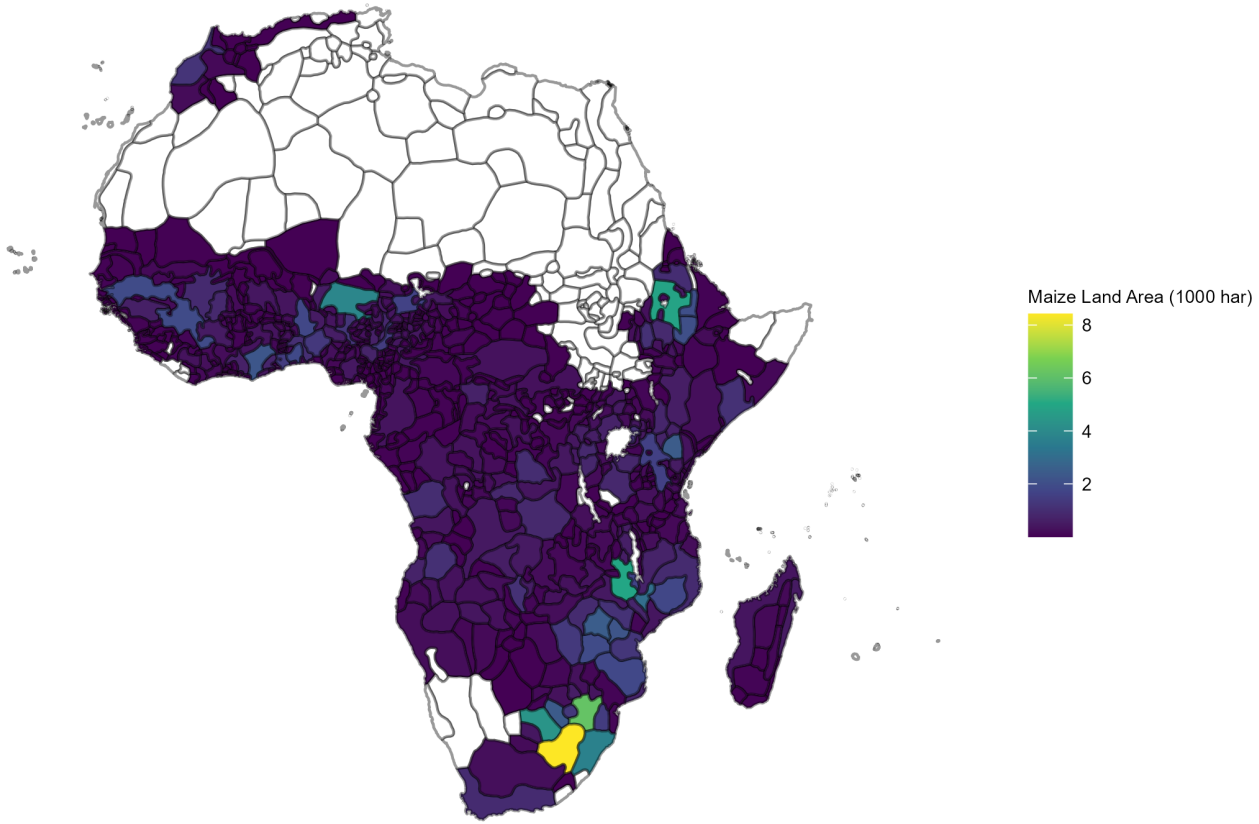


Figure: Maize Production 2000 from FAO (1000 ha)

Shock 2: Building Price Exposure in Homelands

Price Exposure in a homeland o is a composite of the static measure of products produced in that homeland and the time variation in their prices.

Product prices are normalized to 100 in 2000, and summed with weights by the hectare area.

Exposure is weighted by share hectares dedicated to that crop.

$$PriceExposure_{ot} = \sum_{p} Price_t * HectareShare_{op} \tag{9}$$

Shock 2: Impact of Price Exposure on Flows to City

- Count is number of people from tribal homeland o in year t living outside of o:

$$log(count)_{ot} = \omega PricetExposure_{ot} + \upsilon_o + \beta_t + \epsilon_{ot} \tag{10}$$

Price Shocks and Homeland Composition

	Log(Count)	Log(Count)	Log(Count)	Log(Count)
Price Contemporary	0.250			-0.541
	[0.500]			[0.739]
Price 1 year		0.381		-0.355
		[0.413]		[0.658]
Price 10 year			2.524	3.432
			[0.814]***	[1.045]***
Mean Dep.	7.14	7.14	7.14	7.14
Observations	817	817	817	817
Fstat	0.251	0.853	9.620	3.906

Note: $log(count)_{ot} = \omega PricetExposure_{ot} + \upsilon_o + \beta_t + \epsilon_{ot}$. Origin fixed effects, year fixed effects. Outcome is defined as log of people from homeland o not in o

- *Lag of price exposure in homeland o increases number of out-migrants*

Effect of Shocks on Migrant Skill Composition

Define skill in as average human capital school years of migrants from o

	Skill	Skill	Skill	Skill	Skill	Skill
Log(DroughtShock)	0.002	0.003			0.025	0.001
	[0.002]	[0.003]			[0.003]***	[0.003]
Log(PriceShock)			-2.719	-4.433	-1.471	-2.328
			[0.656]***	[1.045]***	[0.667]**	[1.404]*
Mean Dep.	4.631	5.695	4.222	5.391	4.222	4.222
Observations	26,814	11,814	12,536	4,872	12,536	12,536
UrbanOnly	No	Yes	No	Yes	No	No
YearFE	No	No	No	No	No	Yes

Note: Origin-destination fixed effects, year fixed effects. Outcome is defined as average school year of people from homeland o for census

Next Steps

Conclusions:

- Shift-share can be leveraged to predict concentration and flows
- Evidence that migrants boost productivity in a shift-share design.
- Push-shocks predict migration rates and skill composition.

Next Steps:

- Build second stage from shock-predicted migration
- Expand list of second-stage outcomes

References I

Alesina, A. and Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3):762–800.

Arbatli, C. E., Ashraf, Q. H., Galor, O., and Klemp, M. (2020). Diversity and Conflict. *Econometrica*, 88(2):727–797.

Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, 9(2):219–255.

Bazzi, S. and Blattman, C. (2014). Economic shocks and conflict: Evidence from commodity prices. *American Economic Journal: Macroeconomics*, 6(4):1–38.

Castells-Quintana, D. (2017). Malthus living in a slum: Urban concentration, infrastructure and economic growth. *Journal of Urban Economics*, 98:158–173.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *American Economic Review*, 106(3):479–524.

Gollin, D., Jedwab, R., and Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth*, 21:35–70.

Henderson, J. V., Storeygard, A., and Deichmann, U. (2017). Has climate change driven urbanization in africa? *Journal of development economics*, 124:60–82.

Kamuikeni, C. and Naito, H. (2024). The effect of climate change on internal migration: Evidence from micro census data of 16 sub-saharan african countries.

McGuirk, E. and Burke, M. (2020). The economic origins of conflict in africa. *Journal of Political Economy*, 128(10):3940–3997.

References II

McGuirk, E. F. and Nunn, N. (2024). Transhumant pastoralism, climate change and conflict in africa. *Review of Economic Studies*, page rdae027.

Müller-Crepon, C., Pengl, Y., and Bormann, N.-C. (2022). Linking ethnic data from africa (leda). *Journal of Peace Research*, 59(3):425–435.

Porteous, O. (2019). High trade costs and their consequences: An estimated dynamic model of african agricultural storage and trade. *American Economic Journal: Applied Economics*, 11(4):327–366.

Wang, Y. (2024). Linguistic distance, internal migration and welfare: Evidence from indonesia. *Internal Migration and Welfare: Evidence from Indonesia*(February 01, 2024)*.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, 128(4):1727–1785.

Urban Food Prices

Do our shocks move urban food prices?

- Market level panel data for a subset of products from Famine Early Warning (FEWS NET) and World Food Program (VAM) (Porteous, 2019)
- Run same estimation at the product level with same shocks:

For a tribal homeland o in year t at destination i and product p:

$$log(price)_{poit} = \omega Exposure_{ot} + \upsilon_{poi} + \beta_o + \alpha_{op} + \gamma_p + \epsilon_{poit} \tag{11}$$

Market-level Food Prices and Shocks

Table: Drought and International Prices on Local Food Prices

	Log(price)	Log(price)	Log(price)	Log(price)
Drought Exposure	-0.228	-0.154		
	[0.106]**	[0.241]		
Price Exposure			0.006	0.006
			[0.001]***	[0.001]***
Mean Dep.	6.100	6.234	6.068	5.913
Observations	44,329	18,004	5,247	2,403
YearFE	No	No	No	No
50k+ Ony	No	Yes	No	Yes

Note: Origin-destination-product fixed effects, origin fixed effects, product fixed effects. Outcome is defined as log of product price in local currency.

Selection Problems in a Roy Framework

- Potential migrants i born in homeland o receive a productivity draw z for moving to urban center.
- Move if urban wage w minus a cost of moving c_o greater than agricultural income, which is a function of local cash crop prices p_o and rainfall σ_o

$$z_i w - c_o > \sigma_o p_o \tag{12}$$

- Potentially permanent drops in prices or rainfall increase flows of migrants (at lower skill-level)
- Rainfall and prices might affect urban real wage $\frac{w}{p}$ (urban food prices, market towns, agglomeration)
- Credit constraints make flow direction and skill-level uncertain (Bazzi, 2017)

return

Population Data: Africapolis 1950-2000

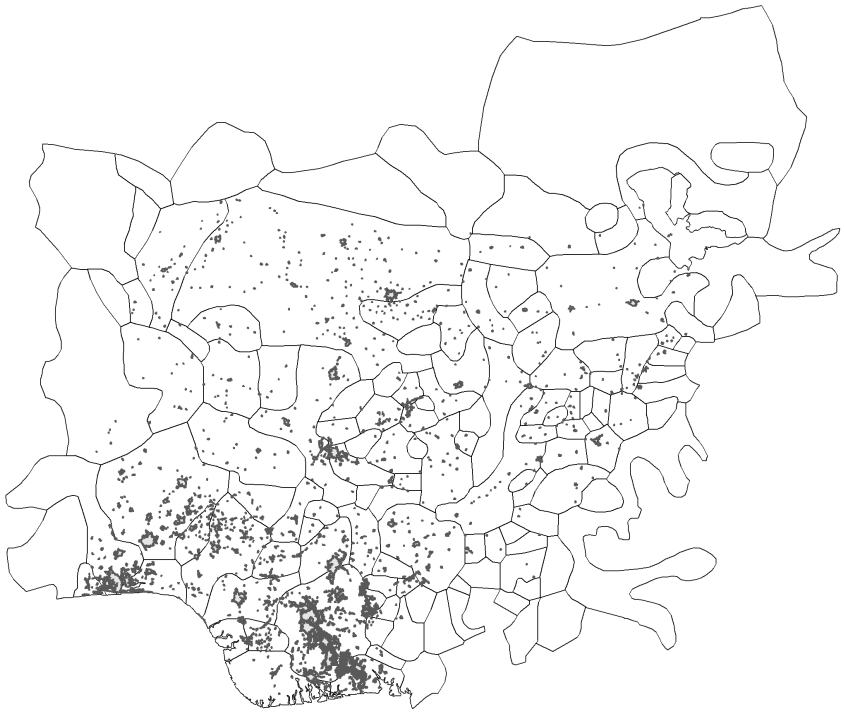


Figure: Towns/Settlements > 10k, Africapolis

return

Aggregating Afro/DHS Points From Neighborhoods to the City Level

return

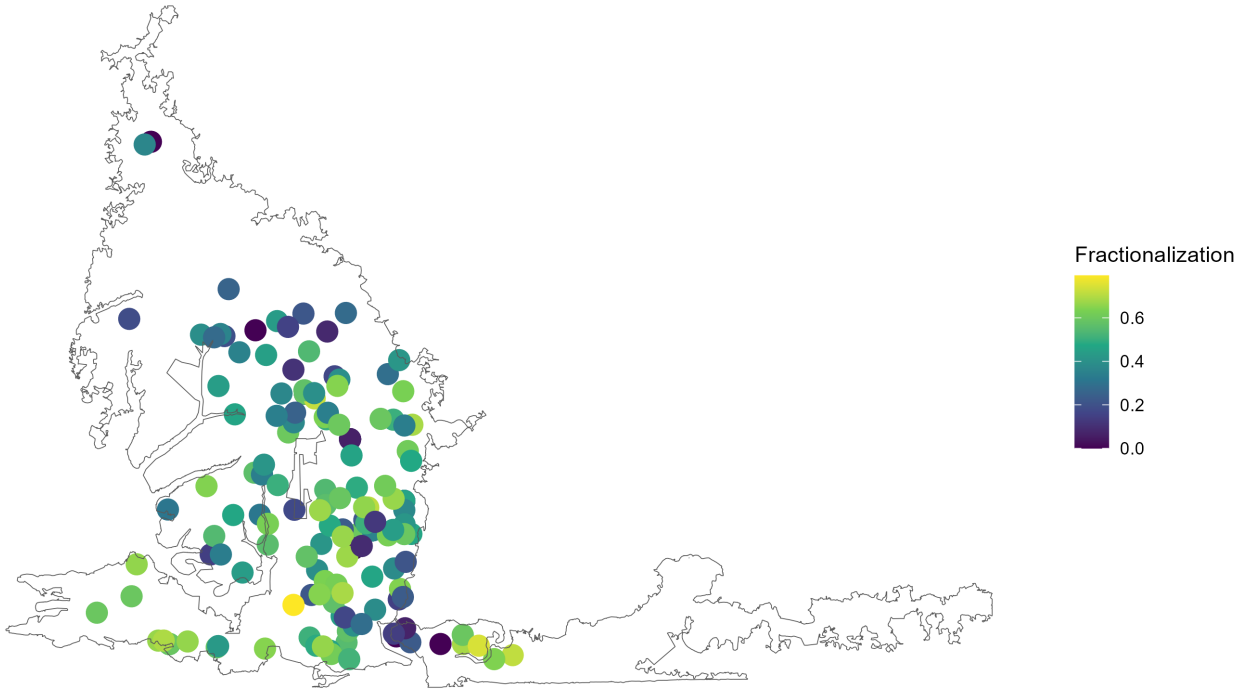


Figure: DHS Clusters within Lagos

Note: Long-Lat points aggregated to Africapolis dataset of cities. DHS clusters cover census enumeration areas (adjacent households)

Limitations of Dataset

For a given region, we know how many people are ethnically related to homeland o.

1. Don't know if born in homeland o, or when migrated
   - Pseudo origin-destination panel: At t, how many people in d are ethnically linked to a homeland o
2. Ethnicities reported in DHS/Census, which are country-specific.
   - South African DHS/Census will code Xhosa, but will code Yoruba as "other" or "Nigeria"
   - Captures internal migrants and ethnic borderlands, but not far away countries

return

Observational Returns in the Quasi-Migration Panel

Does our proxy work?

Sanity check by replicating findings in migration lit with gravity regressions

- Prior work identifies migrants as rural-urban, no details on origins (Henderson et al., 2017; Young, 2013)
- How do wages relate to distance from homeland?
  - Replace wages with housing durables
  - DHS collects consumption of TV, motorcycle, car, etc.
  - Housing quality including floor and wall material

$$Consumption_{odt} = \beta_1 Dist_{od} + \beta_2 CoethShare_{od} + \upsilon_t + \mu_d + \gamma_o \quad (13)$$

return

Consumption and Migration Distance

$$Consumption_{odt} = \beta_1 Dist_{od} + \beta_2 CoethShare_{od} + \upsilon_t + \mu_d + \gamma_o \quad (14)$$

	Durables	House Quality	Durables	House Quality
O-D Distance km	0.096	0.057	0.046	0.014
	[0.008]***	[0.006]***	[0.009]***	[0.008]*
Coethnic Share			-0.232	-0.199
			[0.021]***	[0.019]***
Age	0.004	0.001	0.004	0.001
	[0.000]***	[0.000]***	[0.000]***	[0.000]***
School Years	0.061	0.030	0.061	0.030
	[0.001]***	[0.001]***	[0.001]***	[0.001]***
Mean Dep. Var	0.245	0.263	0.245	0.263
Observations	254,709	273,227	254,709	273,227

Note: Year, origin and destination fixed effects. SE clustered at DHS cluster level.

return

Estimating Equation

Bryan and Morten with origins o and destinations d. Composition in the production equation:

$$Y_{d,t} = \bar{A}_d L_{d,t}^{\alpha} f(\pi_{od})^{\gamma} \tag{15}$$

Where π_{od} is share of d that is from o.

Analog to Diamond (2016) where productivity function of high and low skill worker mix $f(H, L)$

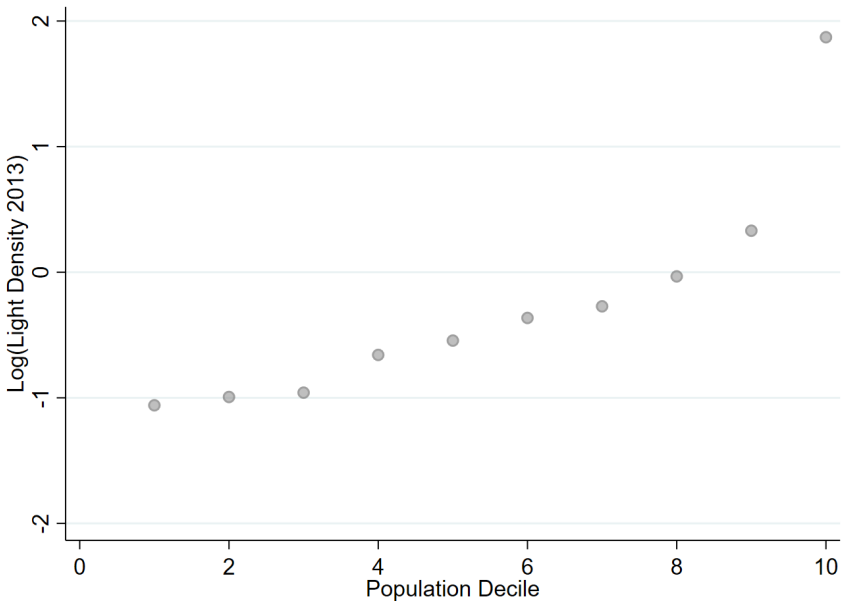
We could put it in the workers' migration decision:

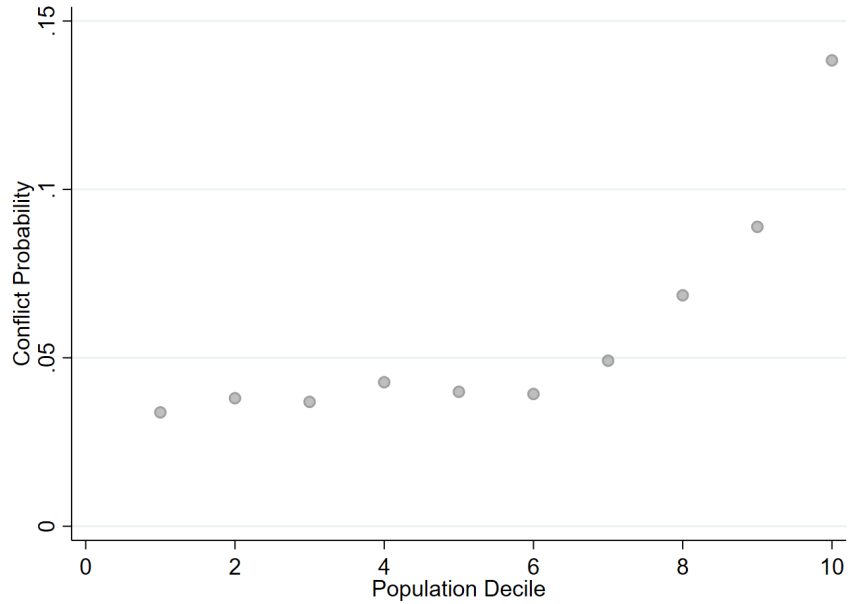
- Linguistic distance as a migration cost (Wang, 2024)
- Composition as an endogenous amenity

return

Agglomeration and Congestion in African Cities

Growth and conflict increasing in city population size.

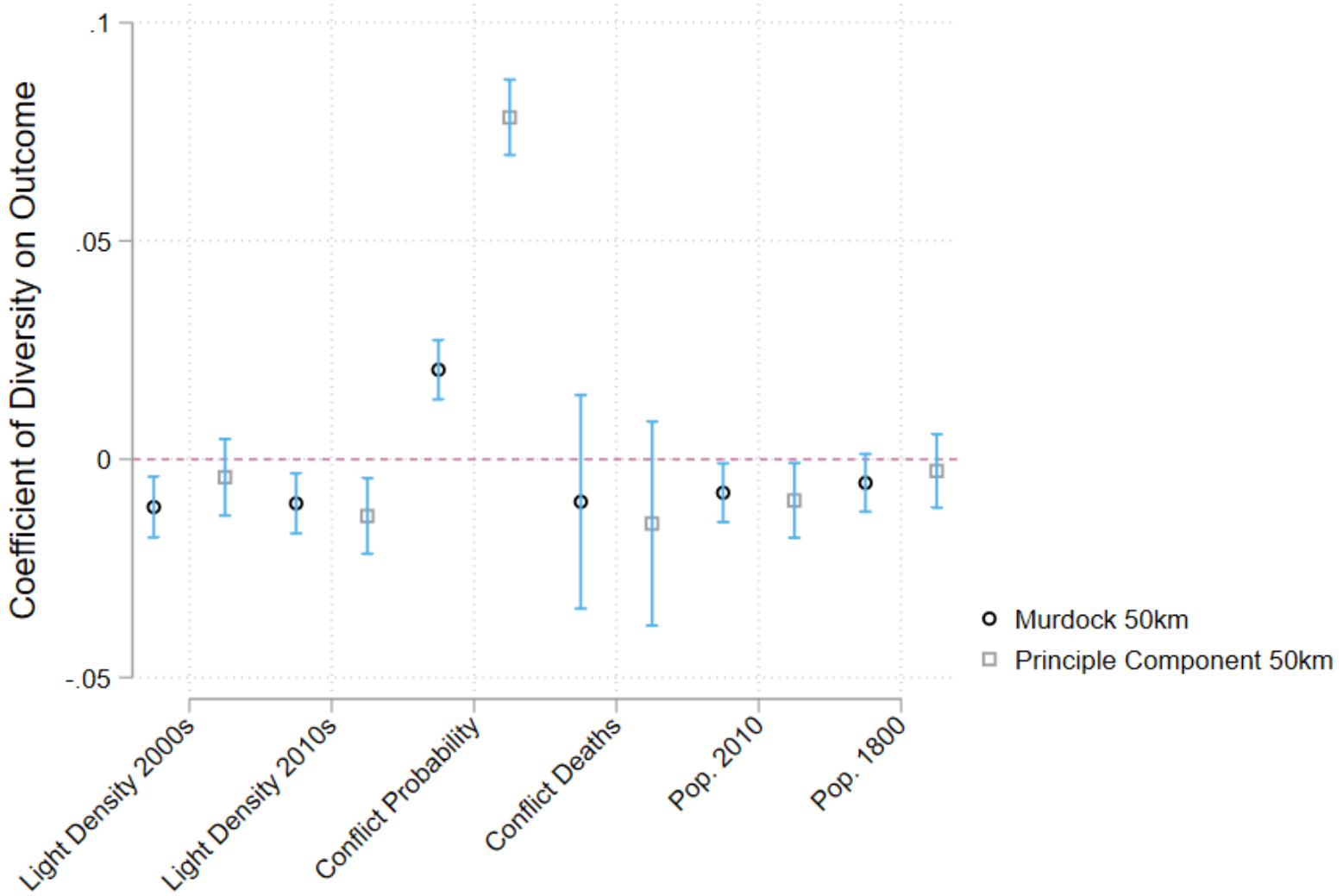




Development and Diversity

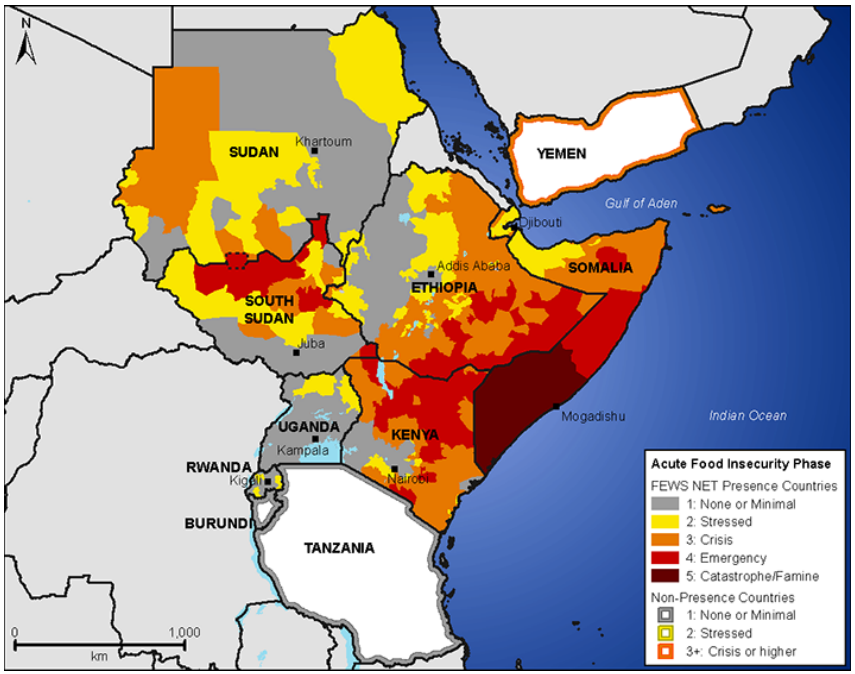
OLS regression of diversity indices on development outcomes.

$y = \alpha + \beta D + X + \upsilon_s$



Rural-Urban Push Shocks

Claim: Migrants move to cities as refuge from disasters (unsubstantiated)



(a) Drought Conditions, 2011

(b) Migrant Flows, 2011

Figure: Ex. East Africa Drought, 2011

return