

# Africa's Economic Transformation

## A Big Data Economic Perspective

Sebastian Krantz\*

Inaugural Dissertation for a PhD in Quantitative Economics  
by the University of Kiel, in Partnership with the  
Kiel Institute for the World Economy

August 21, 2024

### Abstract

Africa is a continent of great economic potential. With a population of 1.5 billion at a median age of 19 today, projected to reach 2.5 billion by 2050, vast natural and mineral resources, yet a share of only ~3% in global GDP and trade, Africa's economic transformation must materialize to provide opportunities for its youth and to foster a more balanced, equitable, and secure world order in the face of shared global challenges. The African Union's Agenda 2063 sets an ambitious path to achieve this, and with the formal enactment of a continental free trade area, a substantial landmark has been passed. However, trade and regional value chains (RVCs) must pick up significantly to generate the desired economic transformation, supported by industrialization and increases in productivity. For this, macroeconomic and financial stability are needed alongside industrial/RVC policies and large-scale investments in infrastructure and human capital. This doctoral dissertation contributes to our understanding of these critical ingredients. It documents the continent's progress in macroeconomic stability and investigates its drivers. It also dissects Africa's integration into global and regional value chains, eliciting progress, benefits, potentials, and challenges towards/of regional economic integration. Last but not least, it zooms in on the continent's infrastructure, utilizing geospatial big data and modern structural and empirical methods to provide evidence on local and global investment potentials at unprecedented spatial detail and scale. By combining rigorous quantitative economics and (causal) machine learning with the richest data on global production, trade, infrastructure, and economic geography available at the time of writing, it produces very detailed and substantive evidence on critical aspects of Africa's present and future economic transformation. It thus enhances our academic understanding of the continent, its economic potential and challenges, but also informs policies to accelerate economic progress and transformation at scale.

**Keywords:** Africa, economic transformation and development, infrastructure, roads, spatially optimal investments, regional integration, trade, GVCs, RVCs, EAC, macroeconomic stability, growth, volatility, structural change, big data, partial and general equilibrium, causal ML, explainable AI

**JEL Classification:** F14; F15; O11; O18; R42; R10; O10; O11; E30; E60

---

\*Kiel Institute for the World Economy  
Address: Haus Welt-Club, Duesternbrooker Weg 148, D-24105 Kiel  
E-mail: sebastian.krantz@ifw-kiel.de or sebastian.krantz@graduateinstitute.ch  
Website: [sebastiankrantz.com](http://sebastiankrantz.com) and [ifw-kiel.de/experts/sebastian-krantz](http://ifw-kiel.de/experts/sebastian-krantz)

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Macroeconomic Stabilization</b>	<b>6</b>
2.1	Africa's Great Moderation . . . . .	7
<b>3</b>	<b>Regional and Global Economic Integration</b>	<b>66</b>
3.1	Africa's Regional and Global Integration . . . . .	67
3.2	Patterns of Global and Regional Integration in the East African Community . . . . .	86
<b>4</b>	<b>Infrastructure for Trade and Structural Transformation</b>	<b>137</b>
4.1	Mapping Africa's Infrastructure Potential with Geospatial Big Data and Causal ML . . . . .	138
4.2	Optimal Investments in Africa's Road Network . . . . .	198
<b>A</b>	<b>Acknowledgements</b>	<b>284</b>
<b>B</b>	<b>Declarations</b>	<b>286</b>

# **Chapter 1**

## **Introduction**

Africa, a vast continent of 30.4 million km<sup>2</sup>, home to 1.5 billion diverse peoples in 54 recognized states, the median of which is 19 years of age today, has grown economically at an average rate of 4.1% between 2001 and 2023 and is projected by the IMF's World Economic Outlook (WEO) to grow at rates around 4.2% in 2024-2029 ([Krantz \(2023b\)](#), own calculations). This performance must be compared to growth of only 2.1% in 1980-2000. In per-capita terms, the growth rate was ~0% in 1980-2000, rose to ~1.9% in 2001-2023, and is projected to remain at that level through 2029. Thus, Africa is growing, and Africans are becoming wealthier. They have also become healthier and more educated, with an increase in life expectancy from 54.4 years in 2000 to 62 years in 2021 and an increase in expected years of schooling from 7.9 to 10.8 ([Krantz \(2023b\)](#), UNDP data). Despite a large COVID shock reducing human development in 2020 and 2021, these long-term increases in human capital contribute to sustaining future per-capita growth.

Another reason to assume more sustained growth in Africa are the enhanced business conditions in many countries, supported by a long process of macroeconomic stabilization, i.e., a persistent decline in the volatility of real per-capita growth and inflation rates within African economies. The first paper in this dissertation ([Krantz, 2023a](#)) documents this African Great Moderation in key macroeconomic aggregates and investigates its correlates within and across countries. The findings suggest that concurrent global moderation, improved macroeconomic policy frameworks/economic institutions, and domestic financial deepening have contributed towards macroeconomic stabilization in Africa. Classical structural change, which increased the share of the more stable service sector, and changes in agricultural technologies, inducing a decline in agricultural output volatility, also played a role. The more stable macroeconomic environment inspires greater business confidence and more secure private investing in Africa. Its preservation and fortification, despite high public debt levels, is an essential ingredient towards successful economic transformation on the continent.

Yet, with a high annual population growth of currently 2.3% and a total population projection to reach 2.5 billion by 2050, greater economic impetus and transformation are needed to generate significant income increases that can only be supported by complex economic activities.

Such transformation has long been called for. "Africa must unite" is the mantra and title of a 1963 book by Kwame Nkrumah ([Nkrumah et al., 1963](#)), the pan-Africanist first prime minister of Ghana, an African champion in infrastructure and industrialization, and a co-founder of the Organization of African Unity (OAU) established in the same year. Commemorating the OAU, in 2013, its successor organization, the African Union (AU), and African heads of state signed the [50th Anniversary Solemn Declaration](#)<sup>1</sup>, re-dedicated Africa towards the attainment of the Pan African Vision of "An integrated, prosperous and peaceful Africa, driven by its own citizens, representing a dynamic force in the international arena". With [Agenda 2063](#)<sup>2</sup> the AU formulated a concrete manifesto towards the attainment of this vision within 50 years, guided by 10-year implementation plans and [15 flagship projects](#)<sup>3</sup> including a continental free trade area (AfCFTA), African commodities strategy, single air transport market (SAATM), integrated high-speed train network, the Grand Inga Dam, a pan-African E-network and E-university, an African economic forum, African financial institutions, and free movement of people.

As emphasized, key flagship projects revolve around advancing continental economic integration and building the necessary infrastructure. With the entry into force of the African Continental Free Trade Agreement (AfCFTA) in May 2019 and its ratification by 47 African states as of July 2024, a significant flagship project has been instigated. World Bank Estimates suggest that by 2035, AfCFTA will boost African incomes by 7% (450B USD), total African exports by 29%, and inner-African exports by 81% ([World Bank Group, 2020](#)). Yet, trading under the agreement has started sluggishly. Currently, only 8 countries trade a handful of products under the Guided Trade Initiative, and many tariff reductions are outstanding. Countries' hesitation to start trading under the agreement may partly be explained by concerns about national industries and value addition.

Thus, more evidence is needed for the optimal utilization of the agreement and the planned formulation of an African Commodities Strategy. Towards this end, a detailed and forward-looking

---

<sup>1</sup><https://au.int/documents/20130613/50th-anniversary-solemn-declaration-2013>

<sup>2</sup><https://au.int/en/agenda2063/overview>

<sup>3</sup><https://au.int/en/agenda2063/flagship-projects>

analysis of the continent's current integration into global and regional production and trade is highly informative. The second part of this dissertation conducts such analysis utilizing the most detailed data on the global economy available to date. Notably, the EMERGING Multi-Region Input-Output (MRIQ) Tables (Huo et al., 2022) cover 245 economies in 135 sectors, incorporating IO tables for 23 African economies representing 84% of African GDP, and sectoral GDP for 27 African economies. The first paper (Krantz, 2024a), a rather short working paper, analyzes African economic integration through trade, global and regional value chains (GVCs and RVCs) and provides the continental context for the second paper. It finds that inner-African trade has increased steadily and highlights precious stones and metals, mining (petroleum), petrochemicals, and food processing as high-potential sectors driving RVCs, with scope for further RVC expansion and deepening. The second paper (Krantz, 2024d) zooms in on earlier efforts of regional integration and presents a rigorous case study of global and regional integration in the East African Community (EAC). It suggests that regional integration is more beneficial than global integration but has distributional side effects, leading to a loss of competitiveness in smaller manufacturing sectors and favouring the regional manufacturing hegemon (Kenya). Hence, African countries should align domestic industrial strategies with trade liberalization and deeper engagement in trading under AfCFTA and jointly strive for industrial policy coordination.

Continental trade and competitive production and value chains in Africa will also require significant investments in infrastructure, including roads, railways, and ports for transportation, power, communications, and education for production, as well as health and public services for the general well-being of workers. However, many African governments are under financial strain to invest in infrastructure projects, with GDP-weighted average general government gross-debt levels at 64%. The African Development Bank estimates suggest that Africa's infrastructure needs amount to \$130-170 billion (2018 USD) a year, with a financing gap in the range of \$68-108 billion (African Development Bank, 2018). International efforts such as the EU's Global Gateway Initiative have so far also failed to generate significant infrastructure investments, and Chinese investments carry high interest rates. This implies a tight public resource allocation problem both across infrastructure sectors and across space, e.g., where exactly do additional roads, power lines, schools, communications towers, etc., generate the highest economic returns? Very little evidence exists for the spatial dimension of the resource allocation problem, let alone both dimensions combined.

To provide evidence addressing these challenges, the third part of this dissertation collects and analyzes very detailed geospatial data on infrastructure, economic activity, and household welfare. The first paper (Krantz, 2024b) builds a granular 'Africa Infrastructure Database' comprising ~15.1 million individual objects (places of interest) categorized into 47 economic categories (26 simplified ones) and ~4.4 million km of network infrastructure (mainly roads, waterways, power lines, and railways). It jointly analyzes this data with nonparametric and ML methods, focusing on uncovering spatial patterns in infrastructure allocation and complex relationships between infrastructure and household welfare. It then employs novel 'causal ML' methods to estimate the marginal effects of different types of infrastructure on household welfare, both overall and spatially, followed by an examination of the correlates and spatial patterns of these marginal effects. Spatial analysis is done at 9.7km resolution for >100,000 populated locations in Africa. Findings imply that the local benefits of additional infrastructure are highly variable and context-specific. The results broadly suggest that 'hard infrastructure,' such as paved roads, power, transport, and communications, is more beneficial in cities, whereas 'social infrastructure,' such as education, health, public services, and utilities, is more critical in rural areas. Market access and agglomeration effects are important forces governing these returns. Descriptive analysis further reveals that infrastructure in Africa is concentrated in urban areas and often inefficiently allocated. African cities exhibit marked heterogeneity in infrastructure, public services, and economic activities.

Predictive ML models suggest that roads are the overall most significant infrastructure predictor of household welfare in Africa. Many policy reports also highlight the economic significance of roads in Africa, e.g., the African Economic Outlook 2024 (African Development Bank, 2024) estimates that transportation (roads) accounts for 72.9% of the estimated infrastructure financing needs until 2030<sup>4</sup>, and a World Bank report (Foster & Briceño-Garmendia, 2010) notes a low paved road

---

<sup>4</sup>See also blog post at <https://www.afdb.org/en/news-and-events/scaling-financing-key-accelerating-africas-structural-transformation-73244>

density of 31 paved road km per 100km<sup>2</sup> of land in Africa compared to 134km in other low-income countries and urges Sub-Saharan African countries to spend 1% of GDP on roads. Roads are also the most complex infrastructure, connecting people and locations rather than providing a local service. ML approaches only observing local quantities of roads are thus not sufficient to analyze their marginal benefits, let alone the benefits of greater investments across multiple locations.

This motivated an additional paper addressing Africa's road network and spatially optimal investments into it (Krantz, 2024c). Using a routing engine to compute 144 million routes between >12,000 locations ~50km apart enables a precise appraisal of the road network's local and global efficiency. The paper then considers 447 cities with more than 100,000 people and 53 international ports and derives a network graph from fastest car routes between them, comprising 315,000km of transport roads. On this network, it characterizes both market access and welfare-maximizing investments, including cost-benefit analysis for individual links and larger investment packages. Importantly, it also examines the effects of cross-border frictions on market access and optimal road investments and takes into account trade through ports. Findings imply that cross-border frictions and trade elasticities significantly shape optimal road investments. Reducing frictions yields the greatest benefits, followed by road upgrades and new construction. Sequencing matters, as reduced frictions generally increase investment returns. Returns to upgrading key roads/links are large, even under frictions. Due to the combination of multiple kinds of data into a unified spatial framework, including novel strategies to develop accurate network representations, cost-benefit analysis, analysis of border frictions, and the use of quantitative spatial models with endogenous infrastructure, this paper is the largest and most sophisticated work included in this dissertation. Notably, it is able to connect local road investments with continental economic outcomes and computes economically optimal spatial road investment allocations at continental scale.

In summary, this dissertation includes four and a half articles that expound, in great detail, on critical elements of Africa's economic transformation, supported by ambitious data collection and empirical rigour. These articles individually contribute to different economic literatures on macroeconomic moderation, global and regional value chains, economic returns to infrastructure investments, and spatially optimal infrastructure investments, as elucidated further in the respective articles. A common contribution is that all articles are, in several respects, the most detailed and data-intensive examinations of the African context within their respective literatures. They also include several methodological advances; in particular, the papers on infrastructure develop and combine methodologies to jointly examine very rich and heterogeneous geospatial infrastructure data in relation to local and continental economic outcomes. My hope is that these works of economic literature not only contribute to a deeper academic understanding of the continent but also inspire concrete policies and, in the case of the infrastructure work, technological solutions and approaches to utilize granular geospatial data for economic infrastructure policymaking.

## References

- African Development Bank. (2018). Africa's infrastructure: Great potential but little impact on inclusive growth. *African Economic Outlook*. Retrieved from [https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African\\_Economic\\_Outlook\\_2018\\_-\\_EN.pdf](https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African_Economic_Outlook_2018_-_EN.pdf)
- African Development Bank. (2024). Driving africa's transformation: The reform of the global financial architecture. *African Economic Outlook*. Retrieved from <https://www.afdb.org/en/documents/african-economic-outlook-2024>
- Foster, V., & Briceño-Garmendia, C. (2010). *Africa's infrastructure: a time for transformation*. World Bank. Retrieved from <https://documents1.worldbank.org/curated/en/246961468003355256/pdf/521020PUB0EPI1101Official0Use0Only1.pdf>
- Huo, J., Chen, P., Hubacek, K., Zheng, H., Meng, J., & Guan, D. (2022). Full-scale, near real-time multi-regional input–output table for the global emerging economies (EMERGING). *Journal of Industrial Ecology*, 26(4), 1218–1232.
- Krantz, S. (2023a). Africa's great moderation. *Journal of African Economies*, ejad021. Retrieved from <https://doi.org/10.1093/jae/ejad021> doi: 10.1093/jae/ejad021
- Krantz, S. (2023b). *The Kiel Institute Africa Monitor*. Kiel Institute for the World Economy. Retrieved from <https://africamonitor.ifw-kiel.de/>
- Krantz, S. (2024a). Africa's regional and global integration. *SSRN Working Paper*. Retrieved from <https://ssrn.com/abstract=4929189>
- Krantz, S. (2024b). Mapping Africa's infrastructure potential with geospatial big data and causal ML. *SSRN Working Paper*. Retrieved from <https://ssrn.com/abstract=4537867>
- Krantz, S. (2024c). Optimal investments in Africa's road network. *Kiel Working Papers*(2272). Retrieved from <https://www.ifw-kiel.de/publications/optimal-investments-in-africas-road-network-33157/>
- Krantz, S. (2024d). Patterns of global and regional integration in the East African Community. *Review of World Economics, Forthcoming*. Retrieved from <https://www.ifw-kiel.de/de/publikationen/patterns-of-global-and-regional-integration-in-the-east-african-community-25992/>
- Nkrumah, K., Arrigoni, R., & Napolitano, G. (1963). *Africa must unite*. Heinemann London.
- World Bank Group. (2020). *The African continental free trade area: Economic and distributional effects*. Washington, DC: World Bank. Retrieved from <https://www.worldbank.org/en/topic/trade/publication/the-african-continental-free-trade-area>

## **Chapter 2**

# **Macroeconomic Stabilization**

# Africa's Great Moderation

Sebastian Krantz\*

October 23, 2023

## Abstract

Over the past 30 years, African economies have experienced remarkable improvements in macroeconomic conditions, characterized by higher and more stable real per-capita growth rates, and lower and more stable inflation. This paper documents the persistent decline in macroeconomic volatility at the aggregate and sectoral levels and seeks to provide explanations. Sectoral analysis shows a particularly strong reduction of growth volatility in agriculture and, to a lesser extent, in services. Classical structural change only explains a small fraction of the moderation. Analysis of further factors yields that changes in structural characteristics such as institutions, trade intensity and diversification, natural resource dependence, or conflict incidence do not explain the moderation. On the positive side, the paper provides evidence to suggest that changes in the external environment, improved macroeconomic policy frameworks, and 'softer' structural improvements, such as the deepening of the domestic financial sector, were important in reducing macroeconomic volatility on the continent.

**Keywords:** macroeconomic stability and resilience, growth, inflation, volatility, structural change, macroeconomic policy

**JEL Classification:** O11; E30; E60

## 1 Introduction

In both academic literature and policy discourse, Africa has long been conceived as a continent of unstable macroeconomic conditions, where a majority of countries suffer from volatile growth rates, high and volatile inflation, and a multitude of other macroeconomic problems, including fiscal spending, debt levels, and exchange rate management. But, as documented by [Calderon & Boreux \(2016\)](#), [Rodrik \(2018\)](#), and others, starting around 1995, many African economies have experienced real growth rates above 5%, well above the levels of the 1970s and '80s. Whereas real growth has slowed a bit again to around 3-4% from 2012 onwards, the past 30 years from around 1990 show a much more persistent and pronounced process of macroeconomic moderation in Africa, where real growth volatility was cut in half, and inflation volatility is less than a third of its initial level.

This paper investigates macroeconomic volatility in Africa over the last 30 pre-COVID years (1990-2019). It documents the decline of macroeconomic volatility at the aggregate and sector levels and then seeks to draw links to changes in production, external conditions, macroeconomic policy, and structural characteristics of African economies. It draws from a variety of empirical methods spanning time series analysis, sectoral decompositions of volatility, panel regressions to assess policy changes, and machine learning models to assess a wide variety of structural characteristics. The paper contributes to a broad literature on the causes and consequences of macroeconomic volatility in developing countries such as [Ramey & Ramey \(1995\)](#), [Rodrik \(1999\)](#), [Easterly et al. \(2001\)](#), [Acemoglu et al. \(2003\)](#), [Auffret \(2003\)](#), [Koren & Tenreyro \(2007\)](#), [Loayza et al. \(2007\)](#), [Malik & Temple \(2009\)](#), [Papageorgiou & Spatafora \(2012\)](#). It differs from most of this literature by endeavoring a comprehensive examination of developments on the African continent.

Declining volatility in macroeconomic aggregates, popularized as 'The Great Moderation' by [Bernanke \(2004, 2012\)](#), has also been widely studied in economic literature. Most studies focus

---

\*Kiel Institute for the World Economy, Research Center International Development  
Address: Haus Welt-Club, Duesternbrooker Weg 148, D-24105 Kiel  
E-mail: sebastian.krantz@ifw-kiel.de

on the US, such as McConnell & Perez-Quiros (2000), Blanchard & Simon (2001), Ahmed et al. (2004) and Galí & Gambetti (2009), but some authors such as Horan (2006) and Schmidt-Hebbel (2009) also examine the wider global context. Burger (2008) and Du Plessis & Kotzé (2010) discuss the Great Moderation in South Africa, the only country on the continent that has received much attention in this literature. Schmidt-Hebbel (2009) notes that moderation generally occurred with a significant lag in developing countries. The literature has not reached a consensus on the causes of moderation, and the paper provides evidence that the causes of moderation in Africa may be different from the causes studied in advanced and emerging economies.

The paper establishes two main findings. First, only a small fraction of the stark decline in macroeconomic volatility in Africa can be explained by structural change, i.e., the service sector's rise. Other changes in the structure of production and trade, conflict incidence, and political institutions also only assume secondary roles. Secondly, evidence suggests that changes in the external environment, improved internal policy frameworks, and 'softer' structural improvements, such as the deepening of the financial sector, were important in reducing volatility on the continent. Overall, the results are stronger on the negative side, providing compelling evidence that macroeconomic moderation in Africa was largely a within-sector phenomenon, with minor roles for structural change and economic diversification. Structural characteristics such as institutions and resource/commodity dependence remain important for explaining volatility differences in a cross-section of African economies, but they cannot explain the persistent and profound process of macroeconomic stabilization in Africa over the last 30 years.

The paper proceeds as follows: Section 2 characterizes broad trends in the volatility of real per-capita growth rates and CPI inflation in Africa and the world and develops the stylized facts that motivate the analysis. Section 3 goes down to the sector level and analyzes growth volatility in Africa from the production side, seeking to quantify the contribution of different sectors to the moderation and the role of structural change. Section 4 examines changes in the external environment faced by African economies, the domestic financial sector, and macroeconomic policies. Section 5 complements this analysis by assessing a broad range of structural characteristics of African economies. Key contributions from the literature referenced above are reviewed within these sections to contextualize the analysis. Section 6 summarizes the findings and concludes.

## 2 Aggregate Relationships and Trends

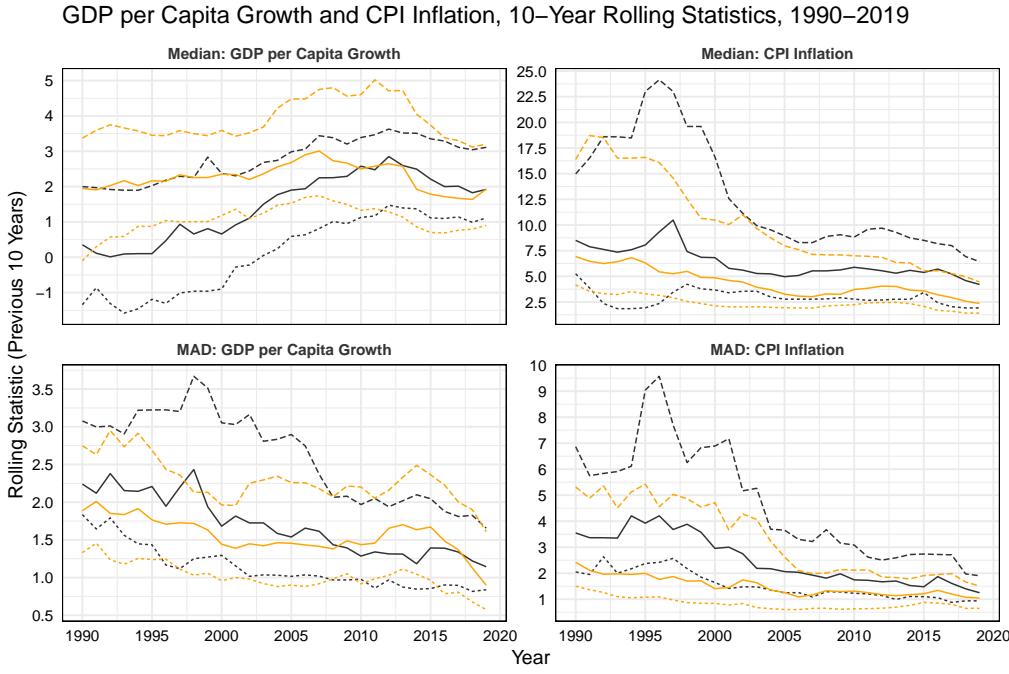
The decline in macroeconomic volatility in the median African country since 1990 has been profound. Figure 1 shows 10-year rolling medians and median absolute deviations (MADs) of real per capita growth and inflation in Africa and the rest of the world (ROW), using data from 1980 where available, calculated at the country-level and aggregated across countries for each year using quartiles. While the whole world has experienced a sizeable macroeconomic moderation in terms of lower inflation and lower volatility of real growth and inflation, this moderation has been particularly strong in Africa, which experienced larger declines in growth and inflation volatility alongside higher growth rates. The bulk of the African transformation occurred between 1995 and 2012, with median per capita growth almost zero in 1986-95, rising to 2.8% in 2003-12. At the same time, the MAD of growth fell from 2.2% to 1.3%, median inflation fell from ~8% to ~5%, and the MAD of inflation fell from 3.8% to 1.8%. After 2003-12, per capita growth in Africa slowed down to 2% in the most recent decade (in line with ROW) alongside further improvements in inflation and volatility, with median inflation coming as low as 4.2% and the MAD dropping to 1.25% in the 2010-2019 period. These trends are robust to weighting countries by GDP or population, as shown in Appendix Figure C3.

Documenting the decline in US output volatility, Blanchard & Simon (2001) run a rolling autoregression of the GDP growth rate over a 20-quarter window to gauge whether the decrease in volatility is due to a decrease in the persistence of shocks, as measured by the AR1 coefficient. They also add a crisis dummy (NBER recessions) to control for large shocks. Blanchard & Simon (2001) find that the US decline in output volatility is due to declines in the magnitude of shocks, reflected in the volatility of the residual, and this holds also when controlling for NBER recessions.

Figure 2 shows the results of a similar analysis conducted for Africa, where I have estimated

autoregressions at the country level using a 15-year rolling window and again aggregated the results quartiles. The crisis definition is adapted from the IMF country risk assessment for LICs (Syed et al., 2017; IMF, 2021), a crisis having occurred if the 2-year average level of real output per capita post-shock ( $t$  and  $t+1$ ) falls below the pre-shock 3-year average level, and output per capita growth is negative in the year of the shock ( $t$ ) (see Figure C6).

Figure 1: Volatility Over Time



*Notes:* 10-year rolling statistics are calculated at the country level and aggregated across countries using quartiles. Countries with less than 15 years of data were removed: in Africa, Liberia, Somalia, South Sudan, and Zimbabwe, leaving 50 African countries in the sample. The ROW category includes 119 countries. Robustness checks see Figures C3–C5.

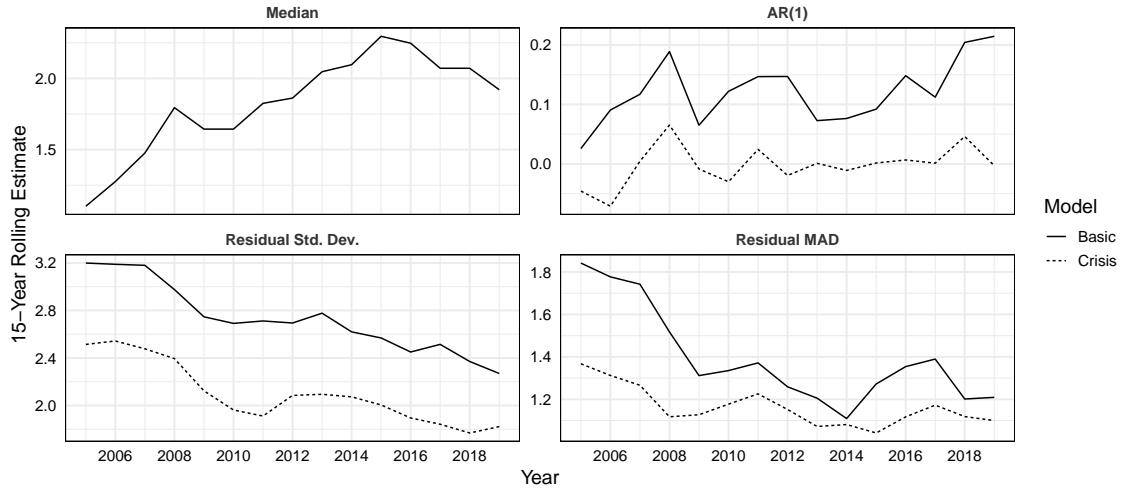
Figure 2 indicates that more or less the same conclusions hold for Africa. The persistence of real output growth even appears to have increased slightly,<sup>1</sup> thus the decline in volatility is primarily associated with factors captured in the residual of the model.

To uncover some heterogeneity, Table 1 compares changes for countries at different income levels, computing statistics over the 1990–2004 and 2005–2019 periods. In Africa, LICs experienced the largest growth acceleration from 1.1% in 1990–04 to 2.25% in 2005–19, alongside a remarkable stabilization of growth from a MAD of 2.43 down to 1.12, and a similar stabilization of inflation from a MAD of 4.34 down to 1.95. LMICs experienced a similar but slightly weaker development in both Africa and ROW. The aggregate statistics for Africa and ROW confirm the results of Figure 1, that in terms of real growth and volatility, Africa performed very similarly to ROW in the 2005–19 period. The only real difference remains the higher median level of inflation at 5.2% in Africa, compared to 2.8% in ROW. This difference is explicable by the Balassa-Samuelson effect.

Appendix Table C4 shows the same table but aggregated across countries using the MAD, indicating that African economies also converged in terms of volatility between these two periods. In ROW, volatility convergence is only evident for inflation. Appendix Figure C2 further shows the extent of moderation in all individual African countries, indicating that 2/3 of countries experienced a moderation in growth volatility, and nearly 90% experienced a moderation in inflation.

It remains to quantify the extent to which improvements in macroeconomic stability are also associated with larger growth and lower inflation within individual countries. Table 2 reports robust regressions of the difference in medians on the difference in the MADs of growth and inflation.

<sup>1</sup>This increase in persistence is even more visible in World Bank data, see Figure C7 in the Appendix.

Figure 2: AR1 Analysis of GDP per Capita Growth in Africa à la [Blanchard & Simon \(2001\)](#)

Data Source: IMF World Economic Outlook, October 2021

*Notes:* 15-year rolling autoregressions of GDP per capita growth on its lag are run for each African country, with or without a crisis dummy (definition following [Syed et al. \(2017\)](#) as described above, see Figure C6), and aggregated across countries using quartiles. The upper panels show the median 15-year growth rate and AR(1) coefficient across countries. The bottom panels show the Std. Dev. and MAD of the residuals. Figure C7 reports the same exercise with GDP per capita estimates from the World Bank.

Table 1: Volatility Over Time and Income Group

Area	Period	N	Per Capita Growth			Inflation		
			Median	MAD	IQR	Median	MAD	IQR
<b>Africa</b>	<b>1990-04</b>	<b>50</b>	<b>1.103</b>	<b>1.950</b>	<b>4.352</b>	<b>6.510</b>	<b>3.110</b>	<b>6.400</b>
	<b>2005-19</b>	<b>50</b>	<b>2.158</b>	<b>1.443</b>	<b>2.884</b>	<b>5.214</b>	<b>1.764</b>	<b>3.406</b>
Low income	1990-04	21	1.096	2.432	4.876	6.787	4.341	8.776
	2005-19	21	2.250	1.121	2.257	5.668	1.950	6.186
Lower middle income	1990-04	20	1.128	1.721	3.952	6.014	2.365	5.087
	2005-19	20	2.212	1.742	3.415	4.812	1.612	3.271
Upper middle income	1990-04	8	1.366	1.947	3.403	6.126	2.217	3.363
	2005-19	8	1.319	1.614	3.765	3.902	1.362	2.574
High income	1990-04	1	0.483	2.973	7.555	2.229	1.723	2.772
	2005-19	1	3.028	2.077	3.891	2.858	2.210	3.815
<b>ROW</b>	<b>1990-04</b>	<b>118</b>	<b>2.278</b>	<b>1.658</b>	<b>3.350</b>	<b>5.086</b>	<b>1.874</b>	<b>4.002</b>
	<b>2005-19</b>	<b>119</b>	<b>2.043</b>	<b>1.446</b>	<b>3.060</b>	<b>2.806</b>	<b>1.201</b>	<b>2.538</b>
Low income	1990-04	4	1.533	1.508	2.894	15.861	6.298	24.794
	2005-19	4	2.391	1.164	2.153	8.456	2.515	5.834
Lower middle income	1990-04	23	1.760	1.217	3.094	8.528	3.973	7.423
	2005-19	23	3.106	0.921	2.103	5.296	1.590	3.253
Upper middle income	1990-04	39	2.504	2.231	4.339	8.410	4.286	14.062
	2005-19	40	2.821	1.898	3.655	4.072	1.492	3.184
High income	1990-04	52	2.476	1.501	3.166	2.395	0.970	1.871
	2005-19	52	1.484	1.323	2.997	1.912	0.864	1.704

Data Source: IMF WEO, October 2021. Real GDP per capita growth is calculated using the constant national currency series (NGDPRPC), and inflation is based on average national consumer price indices (PCPIPCH).

*Notes:* Statistics calculated at country-level and aggregated across countries using the median. Countries with < 9 obs. for growth or inflation in 1990-04 or 2005-19 were excluded, in Africa Liberia, Somalia, South Sudan, and Zimbabwe.

The coefficients imply a negative correlation between volatility and growth and a positive relationship between inflation volatility and median inflation, which is sizeable for LICs and LMICs in Africa and ROW alike. Table 2 thus provides strong evidence that macroeconomic stabilization in these countries is associated with better macroeconomic performance.<sup>2</sup> These relationships

<sup>2</sup>See also Appendix Tables C2, C3, and Figures C3 and C9 for further cross-sectional analysis and correlations.

Table 2: Output and Inflation Volatility: 2005-19 – 1990-04 Difference

Area	N	GDP/Capita			Inflation		
		$\beta$	P( $\beta \neq 0$ )	$R^2$	$\beta$	P( $\beta \neq 0$ )	$R^2$
<b>Africa</b>	<b>50</b>	<b>-0.294</b>	<b>0.004</b>	<b>0.164</b>	<b>0.469</b>	<b>&lt;0.001</b>	<b>0.632</b>
Low income	21	-0.429	0.025	0.244	0.709	<0.001	0.747
Lower middle income	20	-0.111	0.379	0.043	0.322	<0.001	0.715
Upper middle income	8	0.218	<0.001	0.943	-0.108	0.743	0.020
<b>ROW</b>	<b>118</b>	<b>-0.126</b>	<b>0.018</b>	<b>0.046</b>	<b>0.713</b>	<b>&lt;0.001</b>	<b>0.890</b>
Lower middle income	23	-0.432	0.039	0.182	0.920	<0.001	0.761
Upper middle income	39	-0.035	0.756	0.003	0.701	<0.001	0.915
High income	52	-0.087	0.247	0.027	0.250	<0.001	0.364

Data Source: IMF WEO, October 2021. See also note to Table 1.

Notes: Regressions of the difference in medians on the difference in MADs of the country-series between the 1990-2004 and the 2005-2019 periods are run using a robust MM estimator following Koller & Stahel (2011), available in R package *robustbase* (Maechler et al., 2021). Table C3 shows corresponding cross-sectional results.

have been studied empirically, starting with Ramey & Ramey (1995), though mostly in a pure cross-country setting. Among the more detailed analyses, Hnatkovska & Loayza (2005) document that the correlation between volatility and growth is negative for low-income countries, basically zero for middle-income countries, and positive for advanced economies - reproduced for ROW in Table C3. Loayza et al. (2007) and Hallegatte & Przyluski (2011) also show that output volatility in developing countries is strongly related to consumption volatility and incurs a high welfare cost.

A final question regards the international synchronization of business cycles: Has the Great Moderation made African and ROW countries' growth and inflation more correlated? To study this, I compute pairwise Pearson's correlations between countries' per capita growth and inflation series, respectively, and compute an eigendecomposition of this matrix. The first eigenvector spans the first principal component, which I regard as a proxy for the international business cycle. The share of its eigenvalue in the sum of all eigenvalues corresponds to its share in the joint variance of the data. I do this separately with series for Africa, ROW, and all countries (World) for periods 1990-2004 and 2005-2019. In addition, I compute the average absolute correlation between African' and ROW countries as a measure of Africa's alignment with ROW. Table 3 reports the results.

Table 3: International Synchronization of Growth/Inflation Rates

Period	Per Capita Growth				Inflation			
	Africa	ROW	World	Corr	Africa	ROW	World	Corr
1990-2004	0.195	0.195	0.187	0.248	0.347	0.413	0.372	0.337
2005-2019	0.236	0.386	0.312	0.267	0.372	0.495	0.443	0.368
Overall (1990-2019)	0.160	0.221	0.178	0.188	0.331	0.382	0.345	0.289

Data Source: IMF WEO, October 2021. See also note to Table 1. Appendix Table C5 shows equivalent estimates using World Bank Data.

Notes: The numbers under 'Africa', 'ROW', and 'World' are the share of the first eigenvalue in the sum of eigenvalues, computed from a pairwise Pearson's correlation matrix of the country-series. They estimate the share of an international business cycle in the joint variance of the data. The 'Corr' column reports the average absolute correlation between African and ROW countries series and measures alignment between Africa and ROW.

International alignment has increased in Africa and ROW, with a more substantial gain in growth rates than in inflation. The gain in alignment in ROW is, however, much larger than in Africa, and the correlation between African and ROW growth and inflation remains low, indicating that the African Moderation has contributed little to international synchronization.

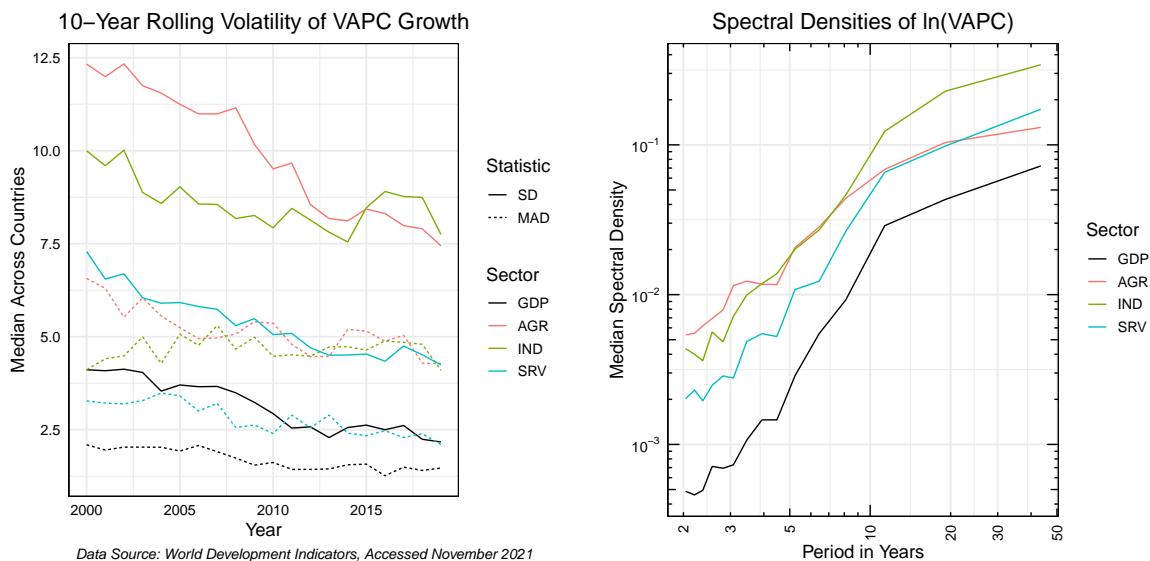
### 3 Decomposing Output Volatility

Structural change in Africa since the 1990s was characterized by an increasing GDP share of the services sector and a declining share of agriculture.<sup>3</sup> This pattern is strongly reflected in the sectors'

<sup>3</sup>See Figure C10.

contribution to GDP per capita growth rates.<sup>4</sup> For volatility analysis, I consider GDP at basic prices. Figure 3 provides a broad view of sectoral volatility, measured by sectoral value-added per capita (VAPC), across time and frequency. The left side of Figure 3 shows that growth volatility in all sectors has decreased over time. In the 1990-2000 decade, the agricultural sector was the most volatile, but volatility decreased rapidly in the 2000s, approaching the volatility of industry. Volatility in services is much lower and has also declined continuously. The right-hand side of Figure 3 shows that agriculture is only more volatile than industry over short-run fluctuations with periods of < 4 years. Industry is the main source of volatility for longer-term variation with > 10-year periods, whereas the volatility of agriculture drops significantly at longer periods. The two sides of Figure 3 can be reconciled by considering, as shown by Shumway & Stoffer (2000) and in Appendix A,<sup>5</sup> that differencing data amounts to a high-pass filter that gives most weight to volatility at periods of 2 years and gradually down-weights higher periods.

Figure 3: Sectoral Volatility Across Time and Frequency



*Notes:* The LHS shows 10-year rolling SDs and MADs of GDP per capita growth at constant 2015 basic prices and its VA components (AGR, IND, SRV). The right side shows the spectral densities of the log levels of VA, computed by removing a linear trend, applying a cosine bell taper to 15% of the data on both sides, computing the periodogram using a Fast Fourier Transform, and smoothing the periodogram using modified Daniell smoothers with widths 3 and 7. The smoothed periodogram ordinates thus obtained at the country level are then aggregated across countries for each period (= 1/frequency) using the median. For more details, see Appendix A.

The relationship between aggregate and sectoral volatility is only linear if sectoral growth rates are independent, which is not true in economic reality. Therefore, I also consider a decomposition of aggregate volatility incorporating sectoral covariances. Formally, let  $Y_t$  be the real GDP per capita at basic prices for a single country in period  $t$ . Aggregate GDP per capita is the sum of sectoral VAPC. Let there be  $K$  sectors indexed by  $k$ , then

$$Y_t = \sum_k y_{kt}. \quad (1)$$

Subtracting Eq. 1 at period  $t - 1$  from Eq. 1, and dividing through by  $Y_{t-1}$ , gives the GDP growth rate in terms of the contribution of sectoral shares, or, after multiplying and dividing by  $y_{k,t-1}$ , as the share-weighted sum of sectoral growth rates

$$\frac{\Delta Y_t}{Y_{t-1}} = \sum_k \frac{\Delta y_{kt}}{Y_{t-1}} = \sum_k \frac{y_{k,t-1}}{Y_{t-1}} \frac{\Delta y_{kt}}{y_{k,t-1}} = \sum_k \theta_{k,t-1} \frac{\Delta y_{kt}}{y_{k,t-1}}, \quad (2)$$

where  $\Delta Y_t = Y_t - Y_{t-1}$ . I now consider the variance of real GDP per capita growth over the entire sample period:  $\text{var}(\% \Delta Y) = E[(\% \Delta Y)^2] - E[\% \Delta Y]^2$ , where  $\% \Delta Y = \Delta Y / Y_{(t-1)}$ . Bienam

<sup>4</sup>See Figure C11, which also shows an increasing role of taxes.

<sup>5</sup>Appendix A contains a frequency domain analysis and discussion concerning volatility harmful to economic activity in Africa and shows that the volatility of growth rates provides an acceptable proxy for such volatility.

Identity gives the variance of a sum of random variables,<sup>6</sup> thus taking the variance of Eq. 2 yields

$$\text{var}(\% \Delta Y) = \sum_{k \in K} \sum_{j \in K} \text{cov} \left( \frac{\Delta y_k}{Y_{(t-1)}}, \frac{\Delta y_j}{Y_{(t-1)}} \right) \approx \sum_{k \in K} \sum_{j \in K} \bar{\theta}_k \bar{\theta}_j \text{cov} (\% \Delta y_k, \% \Delta y_j), \quad (3)$$

where  $\bar{\theta}_k = \frac{1}{T-1} \sum_{t=2}^T \theta_{kt}$  is the average lagged output share of sector  $k$  over the observed period  $T$ . If there is no structural change during the period of observation ( $\theta_{kt} = \bar{\theta}_k \forall t \in T$ ), the right side of Eq. 3 becomes an identity as well.

Computing Eq. 3 over the entire 1990-2019 period (Table C6), confirms that agriculture is the most volatile sector, followed closely by industry. The covariances are negative and significantly smaller, with the largest relationship between agriculture and industry and the smallest between agriculture and services. These patterns are broadly preserved when considering contributions to aggregate volatility. Accounting with robust estimates from Table C6 yields that aggregate volatility in the median African country is composed of 23.6% agriculture, 31.7% industry, and 59% services volatility, and their sum is reduced -14.6% by negative covariances.

I now consider the change in aggregate per-capita growth volatility  $\Delta \text{var}(\% \Delta Y)_\tau$ , computed between two periods  $\tau_1 = 1990 - 2004$  and as  $\tau_2 = 2005 - 2019$ . Following Eq. 3, this equals the sum of changes in the variances and covariances of the sectoral contributions to aggregate growth. To guard against outliers, I also use a comedian-based estimate<sup>7</sup> as a robust alternative to the classical estimator. Another methodological ambiguity regards aggregation. Sectoral shares can either be computed at the country level before aggregation or after aggregating the volatility differences across countries. One would think the former is better, but the data for some countries is of very poor quality. I thus implement both approaches and only report country-level shares aggregated across countries using the median. Together with the choice of covariance estimator, this leads to 6 different estimation strategies, reported in Table 4.

Table 4 shows that the reduction in volatility  $\Delta \text{var}(\% \Delta Y)_\tau$  is due to both reductions in sectoral variances and sectoral covariances.<sup>8</sup> The results differ a bit depending on the methodology: estimates involving Pearson's covariance generally satisfy the equation much more closely but are also most affected by outliers. To generate a representative estimate summarizing the exercise, I compute the median across all 6 strategies and report it in the final row of Table 4.

Table 4: Sectoral Contribution to Moderation in GDP Volatility

CovEst	AggFun	Fit	$\Delta \text{var}(\% \Delta Y)_\tau$	AGR	IND	SRV	$\sum \text{cov}_{jk}$	2AI	2AS	2IS
<i>Sectoral Shares Computed After Aggregation</i>										
Pearson	Mean	100%	-16.18	35%	3%	48%	13%	28%	-6.2%	-8.5%
Comedian	Mean	95%	-6.53	49%	-24%	37%	38%	9.7%	4.9%	24%
Pearson	Median	62%	-5.82	29%	0.84%	40%	31%	2.5%	11%	17%
Comedian	Median	29%	-1.25	45%	0.71%	18%	37%	7.9%	8.7%	20%
<i>Sectoral Shares Computed Before Aggregation</i>										
Pearson	Median	100%	-5.82	23%	14%	46%	17%	3.9%	3.5%	9.6%
Comedian	Median	68%	-1.25	20%	23%	31%	26%	5.9%	14%	6.8%
Median of 6 Estimates:		81%	-5.82	32%	1.9%	38%	28%	6.9%	6.8%	13%

*Notes:* The 'Fit' column signifies how closely Eq. 3 is satisfied. Columns AGR, IND, and SRV give the sectoral contribution to the aggregate volatility reduction in percentage terms, and  $\sum \text{cov}_{jk}$  gives the combined contribution of all covariance terms, which are also individually broken down in columns 2AI, 2AS, and 2IS. Estimates differ depending on the covariance estimator, aggregation function, and whether shares are computed before or after aggregation. The bottom row shows the median of all 6 reported estimates.

<sup>6</sup>

$$\text{var} \left( \sum_k a_k y_k \right) = \sum_{i \in K} \sum_{j \in K} a_i a_j \text{cov}(y_i, y_j) = \sum_k a_k^2 \text{var}(y_k) + 2 \sum_{1 \leq i < j \leq K} a_i a_j \text{cov}(y_i, y_j).$$

<sup>7</sup>The comedian is defined as  $\text{com}(X, Y) = \text{med}((X - \text{med}(X))(Y - \text{med}(Y)))$ . It is not strictly a robust covariance estimator as it does not preserve the relative magnitude of variances and covariances. The implementation available in R package *robustbase* (Maechler et al., 2021) however uses the comedian to estimate the covariance, applying appropriate corrections.

<sup>8</sup>Mostly expressed through already negative covariances becoming more negative, so not a covariance reduction in absolute terms.

The outcome suggests that 32% of the aggregate reduction in per capita growth volatility between  $\tau_1$  and  $\tau_2$  was accounted for by agriculture, 38% by services, and 28% by a reduction in the covariances, of which, abbreviating sectors by their first letter, around 7% are accounted for by AI and AS, and 13-14% by IS. The idiosyncratic reduction in industrial volatility only accounts for 1.9% of the aggregate reduction. The results thus confirm a more than proportional role of agriculture in the African Moderation but also signify a shift towards greater sectoral independence, or rather, sectoral substitutability.<sup>9</sup>

A shortcoming of the results of Table 4, based on the left side of Eq. 3, is that they include the effects of structural change. To examine the contribution of structural change in isolation, I further develop the right side of Eq. 3 and decompose changes in aggregate volatility into changes in sectoral volatilities and changes in sectoral production shares. This is motivated by the consideration that the service sector is substantially less volatile than agriculture and industry, and the share of services in African GDP has been increasing; hence, a quantifiable fraction of the African Moderation must be a direct consequence of structural change. This type of decomposition is well-known in the structural change literature. McMillan et al. (2014) decompose changes in aggregate labor productivity as<sup>10</sup>

$$\Delta LP_t = \Delta \sum_k \theta_{kt} lp_{kt} = \sum_k \theta_{k,t-1} \Delta lp_{kt} + \sum_k \Delta \theta_k lp_{kt}, \quad (4)$$

where  $\theta_{k,t-1} \Delta lp_{kt}$  denotes the sectoral labor productivity changes weighted by sector shares at the beginning of the period, and  $\Delta \theta_k lp_{kt}$  denotes the changes in sectoral shares weighted by final period productivity levels. Applying Eq. 4 to the right side of Eq. 3 yields

$$\Delta var(\% \Delta Y)_\tau \approx \sum_{k \in K} \sum_{j \in K} \bar{\theta}_{kj,\tau-1} \Delta cov(\% \Delta y_k, \% \Delta y_j)_\tau + \sum_{k \in K} \sum_{j \in K} \Delta \bar{\theta}_{kj,\tau} cov(\% \Delta y_k, \% \Delta y_j)_\tau, \quad (5)$$

where  $\tau = (t, \dots, t + N - 1)'$ ,  $N \in 2, \dots, T$  denotes a time-window of size  $N$  over which the covariance is computed, and  $\bar{\theta}_{kj,\tau} = \frac{1}{N-1} \left( \sum_{i=1}^{N-1} \theta_{k,t+i} \times \sum_{i=1}^{N-1} \theta_{j,t+i} \right) \forall k, j$  denotes the product of the average sectoral shares. The first weighted sum of covariances in Eq. 5 thus captures changes in aggregate volatility resulting from changes in volatility within sectors and the second changes due to the shifting of value-added between sectors at different levels of volatility. I estimate Eq. 5 considering again a single difference between periods  $\tau_1$  and  $\tau_2$ , using both classical and comedian estimators, and compute shares before and after aggregating across countries using the median.

Table 5 reports the results. Columns 'Within' and 'Between' give the median value of the respective components in Eq. 5, transformed into shares before aggregation if 'Trans = Share'. 'Fit' indicates how closely Eq. 5 is satisfied, and columns 'Within/Sum' and 'Between/Sum' provide the percentage shares of the two components in their sum i.e. relative to the overall fit, as in Table 4. If 'Trans = Share', these columns are also computed before aggregation. Table 5 shows that in the median country, structural change explains between 3% and 5.6% of the aggregate reduction in per-capita growth volatility in  $\tau_1 \rightarrow \tau_2$ . This result is robust across methodological choices.

A concern may be that the 3-sector setup is too broad to quantify the effects of structural change on aggregate volatility. Thus, I also repeat the exercise with a more detailed dataset used in the structural change literature: the Economic Transformation Database (Kruse et al., 2023) provides a disaggregation into 12 sectors for 21 African countries over the period 1990-2018. It can thus also be split between  $\tau_1$  and  $\tau_2$ . The bottom half of Table 5 reports the results, indicating that even with a finer sectoral disaggregation, the contribution of pure structural change to African moderation is smaller than 5%.

---

<sup>9</sup>Implied by an observed median increase in the negative covariance between AI and AS, and a shift from a small positive IS covariance in  $\tau_1$  to a small negative covariance in  $\tau_2$ .

<sup>10</sup>Equation 4 is derived as:  $\Delta LP_t = \sum_k \Delta (\theta_{kt} lp_{kt}) = \sum_k (\theta_{k,t-1} lp_{kt} - \theta_{k,t-1} lp_{k,t-1}) = \sum_k (\theta_{k,t-1} lp_{kt} - \theta_{k,t-1} lp_{k,t-1} + \theta_{k,t-1} lp_{kt} - \theta_{k,t-1} lp_{kt}) = \sum_k (\theta_{k,t-1} \Delta lp_{kt} + \Delta \theta_k lp_{kt}).$

Table 5: Structural Change and African Moderation: à la [McMillan et al. \(2014\)](#)

CovEst	Trans	$\Delta var(\% \Delta Y)_\tau$	Within	Between	Fit	Within/Sum	Between/Sum
<i>World Bank Data</i> (3-Sectors, 41 Countries)							
Classical	None	-5.823	-2.390	-0.141	43.5%	94.4%	5.58%
Classical	Share	-5.823	0.958	0.019	102%	97.0%	2.98%
Comedian	None	-1.669	-2.506	-0.147	159%	94.5%	5.53%
Comedian	Share	-1.669	0.757	0.045	88.7%	96.9%	3.07%
<i>Economic Transformation Database</i> (12-Sectors, 21 Countries)							
Classical	None	-6.304	-5.002	-0.002	79.4%	100%	0.03%
Classical	Share	-6.304	0.993	0.001	105%	98.1%	1.85%
Comedian	None	-2.775	-3.413	-0.111	127%	96.8%	3.16%
Comedian	Share	-2.775	0.969	0.037	113%	96.9%	3.07%

*Notes:* The decomposition is computed at the country level for 41 African countries according to Eq. 5, comparing 1990-2004 to the 2005-2019 period. The two components are turned into shares if 'Trans = Share', and aggregated across countries using the median. Further descriptions of the columns are provided in the main text above. 13 countries with less than 10 observations for any sectoral growth rate in either period were excluded: Algeria, Angola, the Central African Republic, Djibouti, Equatorial Guinea, Eritrea, Kenya, Liberia, Libya, Madagascar, Somalia, South Sudan, and São Tomé & Príncipe. The ETD of [Kruse et al. \(2023\)](#) records 12 sectors for 21 African countries: BFA, BWA, CMR, EGY, ETH, GHA, KEN, LSO, MAR, MOZ, MUS, MWI, NAM, NGA, RWA, SEN, TUN, TZA, UGA, ZAF, ZMB.

A related exercise, established by [Stock & Watson \(2002\)](#) fixes the sectoral shares but maintains the sectoral growth rates to generate a pseudo-outcome GDP growth series reflecting the absence of structural change. Following their method, I generate two pseudo-outcome series with sectoral shares fixed at their average in periods  $\tau_1$  and  $\tau_2$ , respectively. I then compute various measures of volatility  $\sigma$  (variance, IQR, and MAD) across the two periods and series, yielding two actual (approximately) and two counterfactual volatility estimates. Following [Stock & Watson \(2002\)](#), I compute  $\psi = ([\sigma_{\tau_2}^{\tau_2} - \sigma_{\tau_2}^{\tau_1}] + [\sigma_{\tau_1}^{\tau_2} - \sigma_{\tau_1}^{\tau_1}])/2$ , where  $\sigma_{\tau_2}^{\tau_1}$  is the volatility of GDP in period  $\tau_2$ , computed using the sectoral shares of period  $\tau_1$ . The statistic  $\psi$  thus estimates the change in volatility due to a change in shares by averaging both possible ways of conducting the counterfactual exercise: comparing actual volatility in  $\tau_2$  to volatility in  $\tau_2$  if the sectoral shares are those of  $\tau_1$ , and comparing volatility in  $\tau_1$  with the shares of  $\tau_2$  to actual volatility of  $\tau_1$ . Further, I compute  $\Delta\sigma = \sigma_{\tau_2}^{\tau_2} - \sigma_{\tau_1}^{\tau_1}$  as the actual (approximate) change in GDP volatility. Table 6 reports the results.

Table 6: Structural Change and African Moderation: à la [Stock & Watson \(2002\)](#)

Statistic	AggFun	$\sigma_{\tau_1}^{\tau_1}$	$\sigma_{\tau_1}^{\tau_2}$	$\sigma_{\tau_2}^{\tau_2}$	$\sigma_{\tau_2}^{\tau_1}$	$\Delta\sigma$	$\psi$	$\frac{100\psi}{\Delta\sigma}$
<i>World Bank Data</i> (3-Sectors, 41 Countries)								
Variance	TrimmedMean (10%)	21.64	26.28	12.00	12.99	-6.67	2.39	-1.76
Variance	Median	17.83	19.40	6.94	7.44	-3.20	0.75	2.20
IQR	TrimmedMean (10%)	4.62	4.73	3.29	3.45	-1.23	-0.04	9.19
IQR	Median	3.90	4.21	3.26	3.21	-0.54	-0.01	5.11
MAD	TrimmedMean (10%)	2.28	2.48	1.70	1.78	-0.54	0.03	-6.64
MAD	Median	2.05	2.24	1.77	1.69	-0.30	-0.02	2.69
<i>Economic Transformation Database</i> (12-Sectors, 21 Countries)								
Variance	TrimmedMean (10%)	15.22	19.29	6.17	4.93	-9.10	1.58	0.97
Variance	Median	8.36	10.52	5.06	4.35	-4.25	0.70	6.49
IQR	TrimmedMean (10%)	4.17	4.40	2.74	2.46	-1.27	0.17	3.98
IQR	Median	3.72	3.51	2.29	2.54	-1.01	0.03	3.51
MAD	TrimmedMean (10%)	2.23	2.21	1.42	1.13	-0.72	0.10	5.32
MAD	Median	2.01	2.04	1.35	1.12	-0.46	0.05	8.76

*Notes:* Decomposition of GDP volatility by fixing sectoral shares and generating counterfactual series as in [Stock & Watson \(2002\)](#).  $\sigma_{\tau_2}^{\tau_1}$  is the volatility of GDP in period  $\tau_2 = 2005-2019$ , computed using the sectoral shares of period  $\tau_1 = 1990-2004$ .  $\psi = ([\sigma_{\tau_2}^{\tau_2} - \sigma_{\tau_2}^{\tau_1}] + [\sigma_{\tau_1}^{\tau_2} - \sigma_{\tau_1}^{\tau_1}])/2$  estimates the average volatility change due to structural change across the two periods.  $\Delta\sigma = \sigma_{\tau_2}^{\tau_2} - \sigma_{\tau_1}^{\tau_1}$  estimates the actual (approximate) change in GDP volatility. Results are aggregated across countries using either the median or a trimmed mean, removing 10% of the observation from both sides.

Overall, the results of Table 6 are very similar to those of Table 5. The counterfactual volatility estimates are very close to the actual ones, confirming that macroeconomic moderation in Africa was largely a within-sector phenomenon. The percentage share of the moderation attributable to structural change by this method, reported in the final column of Table 6, ranges between -6.6 and 9.2 %. The analysis presented in Tables 4 through 6 thus establishes that up to 30% of the aggregate African Moderation of real per-capita growth rates is due to changes in the covariance towards greater sectoral independence/substitution, and ~5% can be attributed to pure structural change, with less volatile sectors like services becoming economically more important. The remaining 65-70% is due to other factors that mostly affected agriculture and services. As noted by Stock & Watson (2002), the increase in services employment could, through more stable incomes and demand, also have stabilizing effects on other sectors, which is not captured by these decompositions. The decompositions might thus underestimate the true general equilibrium effects of structural change on output stability.

In the great moderation literature, a frequently mentioned driver in advanced economies (McConnell & Perez-Quiros, 2000; Blanchard & Simon, 2001; Horan, 2006) is greater production efficiency through inventory management innovations. Given the small role of the industry sector in Africa's moderation, the results suggest that inventory management is unlikely to be an important driver.

Before examining other factors, I validate these results at the country level and uncover some heterogeneity by computing the MAD sector contribution to growth,  $MAD(\Delta y_t/Y_{t-1})$ , and sectoral growth volatility,  $MAD(\%\Delta y_t)$ , for each country. Computing  $MAD(\Delta y_t/Y_{t-1})$  over the whole 1990-2019 period yields a large group of 21 countries has services as the greatest contributor to aggregate volatility, followed by industry (17) and agriculture (13) (see Table C7). This metric thus broadly aligns with the pattern of structural change. Considering the sectoral growth volatility, however, leads to a large reallocation of countries to industry and agriculture, with 22 countries having the largest volatility in agriculture and 28 in industry, and only 1 country (Gabon) in the services category.<sup>11</sup> Table C8 summarizes both metrics for  $\tau_1$  and  $\tau_2$ , and Figure C16 visualizes the movement in  $MAD(\Delta y_t/Y_{t-1})$  for all countries, indicating a nearly ubiquitous and large stabilization of agriculture, as well as a sizeable stabilization of services in most countries. In industry, the developments are very heterogeneous, with some countries like Ghana experiencing greater volatility and others like South Africa experiencing significant stabilization. It is also interesting to compare different regions in Africa. Figure C17 and Table C10 provide a regional summary of sectoral volatility and show that all regions apart from southern Africa experienced a sizeable stabilization in agriculture. Overall, Eastern Africa experienced the largest stabilization in aggregate output, followed by Middle and Western Africa.

Disaggregated analysis hence confirms the results of aggregate analysis, indicating that stabilization of agriculture and services was a shared experience for most African countries since 1990. There is moderate regional heterogeneity, with the more developed regions of Northern and Southern Africa being affected less. Appendix B reports a similar decomposition from the expenditure side of GDP, indicating a decline in the volatility of all expenditure components, especially consumption, investment, and exports.

## 4 External, Financial, and Policy Factors

Since ~70% of the African Moderation cannot be explained by structural change or changes in sectoral covariances, the remainder of the paper examines other contributing factors, including the external economic environment faced by African economies, changes in the financial sector, macroeconomic policies, and other changes in economic or institutional structure (Section 5).

Several papers investigate the great moderation in the US and other advanced economies (AE) along similar lines. Horan (2006) studies the reduction of output volatility in AE, focussing on the competing explanations of better monetary policy, more efficient inventory investment of firms, and lower exposure to global shocks (oil price shocks). He finds that, due to the different onsets

---

<sup>11</sup>See the bottom half of Table C7.

of output moderation in AE, only the former two provide credible explanations for the great moderation. [McConnell & Perez-Quiros \(2000\)](#) and [Blanchard & Simon \(2001\)](#) also find evidence that more countercyclical inventory management has contributed to the stabilization of business cycles in the US. [Ahmed et al. \(2004\)](#) attributes to monetary policy a role in bringing down US inflation volatility. [Schmidt-Hebbel \(2009\)](#) discusses causes of the great moderation in emerging markets and developing economies (EMDE), mentioning stronger policies and better institutions (especially property rights, governance and accountability, and central bank independence) as drivers. He documents that the adoption of inflation-targeting monetary policy (IT) in EME was associated with reduced domestic inflation and exchange rate pass-through. Many developing countries also adopted more sustainable fiscal policies. In South Africa, [Burger \(2008\)](#) provides evidence that better monetary policy and a more efficient financial sector brought down volatility in the 90s, but inventories did not. [Du Plessis & Kotzé \(2010\)](#) also note that a less volatile international environment following South African liberalization in the late 80s and the greater political stability during the post-apartheid late 90s enhanced macroeconomic stability in South Africa.

In the bulk of African economies, little is known about inventory management practices, but I have argued against it based on the small contribution of the industrial sector to African moderation. I will also argue against IT as a driver of moderation in Africa. There have, however, been notable changes in the external environment faced by African economies in this time frame, reflected in better terms of trade (ToT), lower external debt burdens, higher inflows of FDI and remittances, as well as lower volatility of merchandise trade, FDI and remittance inflows. There has also been a gradual process of financial deepening, reflected in broad money, credit to the private sector, national savings, and reserve assets. Finally, there have been changes toward a more stable exchange rate policy and an increased adoption of fiscal rules.

## 4.1 External Environment

Reduced volatility may partly result from a more favorable external environment faced by African economies, permitting both stronger growth and more long-term economic planning and investments. It is notable from Figure 1 that growth rates peaked shortly after 2010, which, thanks to the Heavily Indebted Poor Countries (HIPC) Initiative launched in 1996, is also the period when Africa faced the lowest levels of public and external debt (see Figure C18). African economies also experienced more favorable ToT and higher FDI and remittance inflows after 2010.

Table C11 shows correlations of 10-year rolling averages of these indicators with rolling medians and MADs of per-capita growth and inflation in country-standardized first-differences. ToT and FDI are significantly positively correlated with growth, whereas public and external debt are strongly negatively related to growth. In addition, higher ToT and remittances are associated with lower growth volatility and inflation levels, whereas greater debt stocks correlate with higher volatility and inflation. This indicates that more favorable linkages with the world could have contributed to the increased resilience in African real sectors.

A less volatile external environment may also have directly contributed to less volatile domestic activity. Figure C19 shows that exchange rate, ToT, and merchandise trade volatility have dropped substantially over the sample period, and also FDI and remittance flows became less volatile. Current account volatility also fell after 2010. Table C12 shows corresponding within-country correlations of 10-year rolling volatility measures, indicating that higher exchange rate, ToT, FDI, and remittance volatility are associated with lower growth, higher inflation, and greater macroeconomic volatility. Especially exchange rate volatility is strongly correlated with inflation.

## 4.2 Financial Deepening

Another source of increased resilience in Africa may be domestic financial deepening and increased levels of international reserves to counter external shocks. For example, [Easterly et al. \(2001\)](#) analyze volatility with an emphasis on the financial sector and constate that credit constraints are an important source of volatility in developing countries. Figure C20 indicates that Africa has indeed made some progress in this direction over the past 30 years. Gross National Savings have increased from around 14% of GDP in 1990 to around 18% in 2019, total reserves have risen to

around 100% of external debt or 5.5 months of imports of goods and services, domestic credit to the private sector has risen from 17% to 25% of GDP, broad money from 30% to 40% of GDP, and banks liquid reserves to assets ratio has risen from <20% to >=25%, at least when weighted by GDP or population. Table C13 shows the corresponding correlations in country-standardized first differences, indicating that higher national savings, reserves, domestic credit to the private sector, and broad money correlate positively with economic growth and macroeconomic stability.

### 4.3 Macroeconomic Policies

Improved domestic macroeconomic and financial policies may also have contributed to the African Moderation. I only consider the most important stabilization policies: managing inflation and the exchange rate, macroprudential stringency, and fiscal rules.

#### Inflation Targeting

Africa still has very few inflation targeters. According to the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER) database, only 4 countries currently target inflation: South Africa from 2000, Ghana from 2007, Uganda from 2011, and Seychelles from 2019. Figure C21 shows the inflation rates of these countries, indicating that the IT regimes were adopted when inflation had already stabilized to levels well below 20%. Thus, the adoption of IT did not play a large role in the African Moderation.

#### Exchange Rate Arrangements

To examine the evolution of exchange rate regimes, I take data from Ilzetzki et al. (2019), available for 53 African countries. Figure C22 shows that the share of crawling bands and free-falling/dual markets has declined in Africa since 1992, in favor of crawling peg arrangements. Free floats are also rare, and since 2005, only South Africa has maintained a floating regime.

Table C14 reports 15-year rolling panel fixed-effects regressions of growth and inflation volatility on the exchange regime dummies, using the hard peg as a base category. Results imply that relative to the hard peg, crawling pegs are associated with greater growth stability, whereas more liberal regimes correlate with less stable growth performance. The coefficients on model (3) with country and time fixed effects imply that a crawling peg is associated with a 0.86 percentage point (pp.) decrease in the MAD of real per-capita growth vis-a-vis the hard peg arrangement. For inflation, the hard peg appears to be the most stable regime, but the coefficient on the crawling peg is insignificant, indicating that the inflation cost of switching to a crawling peg is moderate. As elucidated by Bleane et al. (2016), the choice of exchange rate arrangement is endogenous to macroeconomic conditions and policy priorities, and reduced inflation has made pegging more attractive in recent years. Nevertheless, the shift in exchange rate regimes in Africa since 1990 towards crawling pegs and decline in freely-falling and other volatile regimes, evidenced by a strong decline of exchange rate volatility (Figure C19), has likely contributed to Africa's moderation.

#### Macroprudential Regulation

Macroprudential policy received increased attention after the 2008/09 global financial crisis. The IMF adopted a new Institutional View in 2012, recognizing the usefulness of macroprudential measures for macroeconomic stability, particularly in economies with less developed financial markets (Arora et al., 2013; IMF, 2017, 2022). Figure C23 shows indices of total, inflow, and outflow controls across 18 African countries, computed by averaging dummies for restrictions in 10 financial markets, from Fernández et al. (2016) (August 2021 update). Macroprudential measures have eased in the 1995-1997 period,<sup>12</sup> but have remained quite stable around 0.5 for inflow measures and 0.625 for outflow measures afterward. It is, therefore, unlikely that aggregate macroeconomic moderation in Africa is much affected by changes in macroprudential policy.

Some continent-level developments emerge when considering restrictions in the 10 different markets separately. Figure C24 shows 10-year MAs of the overall stringency in 18 African economies,

---

<sup>12</sup>Mainly due to liberalizations in Ethiopia, Nigeria, Uganda, and Ghana.

indicating that bond and guarantee markets, as well as FDI, have become more restricted in recent years, whereas equity, real estate, and commercial credit markets have become less restricted. Table C15 shows 10-year rolling panel-FE regressions of the MADs of GDP per capita growth and inflation on aggregate macroprudential stringency indicators, indicating that macroprudential stringency is negatively associated with both output and inflation volatility. Drawing on the specification with country and time fixed effects, an increase in overall macroprudential stringency by 0.1 is associated with a 0.54 pp. reduction in the MAD of per-capita growth and a 1.22 pp. reduction in the MAD of inflation. Outflow measures have a stronger association with output stability, whereas inflow measures strongly associate with reduced inflation volatility.

### Fiscal Rules

A fourth and important set of stabilization policies are fiscal rules. Global data on fiscal rules adopted since 1985 is available through the IMF Fiscal Rules Dataset (Davoodi et al., 2022b,a). Table 7 compactly summarizes the history of fiscal rules in Africa.<sup>13</sup>

Table 7: A Chronology of Fiscal Rules in Africa

Entity	First Rule	Expenditure (ER)	Revenue (RR)	Budget Balance (BBR)	Debt (DR)
Kenya	1997		1997		1997 (2019)
Cape Verde	1998			1998	1998
WAEMU <sup>a</sup>	2000		2000 (2015)	2000 (2015)	2000 (2015)
Namibia	2001		2010		2001
CEMAC <sup>b</sup>	2002			2002 (2008, 2017)	2002
Botswana	2003	2003 (2006, 2016)			2003
Nigeria	2007				2007
Mauritius	2008				2008 (2010)
Liberia	2009				2009
EAC <sup>c</sup>	2013			2013	2013
Tanzania	2015		2015		2015
Uganda	2016				2016
Rwanda	2019				2019

Data Source: Davoodi et al. (2022b). Rule revisions in parentheses.

<sup>a</sup> Comprising Benin, Burkina Faso, Côte D'Ivoire, Guinea-Bissau, Mali, Niger, Senegal and Togo

<sup>b</sup> Comprising Cameroon, Central African Republic, Chad, Republic of Congo, Equatorial Guinea and Gabon

<sup>c</sup> Comprising Tanzania, Kenya, Rwanda, Uganda, Burundi and South Sudan

Most fiscal rules in Africa can be regarded as weak. Apart from Mauritius and Botswana, no country has instigated a formal enforcement procedure for national rules, and no country has an extra-governmental body to monitor compliance with national rules.

To evaluate the relationship of fiscal rules with macroeconomic stability, Table C16 presents 10-year rolling regressions considering first a dummy indicating the adoption of any fiscal rule, then the total number of rules, and finally a set of dummies for the different types of rules. Adding both country and time-fixed effects lets the within  $R^2$  drop to zero, indicating insufficient time variation in the fiscal rules to control for global events. The models with country-fixed effects, however show a meaningful and significant negative association of fiscal rules with both growth and inflation volatility. When disaggregating the set of rules, only the coefficient on the Budget Balance Rule (BBR) is negative and significant in the growth volatility regression. The coefficient size implies that a BBR is associated with around 1 pp. lower MAD of growth. For inflation, both Revenue Rules and BBRs have large negative coefficients. Debt Rules (DRs) are also negatively related to growth/inflation volatility, with insignificant coefficients of 0.23/0.54.

Table C17 further shows that the existence and number of rules implemented correlate positively with the current account (CAB) and government budget balance (GBB), and negatively with the level of gross government debt (GGD). When disaggregating rules, BBRs are associated with an approx. 4.2 pp. improvement in the CAB (in % of GDP), and a 7 pp. improvement in the GBB. DRs appear to be strongly associated with GGD, at effect sizes up to 70-80 pp. lower debt to GDP.

These coefficients are not to be interpreted as causal since fiscal rules are often a commitment device of governments that already engage in sound macroeconomic practices. They nevertheless

<sup>13</sup>Figure C25 also shows an aggregate timeline of fiscal rules adoption in Africa by type and issuing authority.

suggest that instigating a rule is an effective device for these governments.

Having considered four different types of macroeconomic policies in Africa over the past 30 years, it appears that only the shift towards crawling pegs from crawling bands and freely falling arrangements and the adoption of fiscal rules in an increasing number of countries could have contributed to the large macroeconomic moderation in growth and inflation. Inflation-targeting monetary policy has only been taken up by four countries at a point when their inflation levels were already low and stable, and macroprudential policies, while potentially effective in curbing macroeconomic volatility, show no aggregate trend over most of the period under consideration. This assessment is, of course, incomplete. For example, it is possible that central banks have become more effective in targeting monetary aggregates without shifting to inflation targeting or that many countries have run financial sector, trade, or agricultural policies that contributed to macroeconomic stability. Above all, the issue of policy endogeneity to macroeconomic conditions precludes drawing too wide-ranging conclusions about policy efficacy.

## 5 Structural Factors

An assessment of macroeconomic volatility and moderation would be incomplete without reference to structural characteristics of an economy, such as political and economic institutions, diversification in production and trade, economic openness, the incidence of conflicts and disasters, geography, human capital, etc. Significant literatures in economics have evaluated the effects of these factors in different contexts. For example, [Acemoglu et al. \(2003\)](#) analyze the effects of long-term institutional development on macroeconomic stability and find that countries that inherited more 'extractive' institutions from their colonial past are more likely to experience high volatility and economic crises. They argue that poor institutions cause volatile and distortionary macroeconomic policies, which act as a proximate cause for volatility. [Rodrik \(1999\)](#) relates the lack of persistent growth in developing countries to social conflicts fuelled by inequality, ethnic fractionalization, and weak institutions. [Malik & Temple \(2009\)](#) examine the structural determinants of output volatility in developing countries with Bayesian methods. They find a significant role of market access: remote countries are more likely to have undiversified exports, high levels of export concentration, high ToT volatility, and high output volatility. [Auffret \(2003\)](#) finds positive effects of natural disasters on consumption volatility in the Caribbean region. [Abdullahi & Suardi \(2009\)](#) examine the effects of financial and trade liberalization on output and consumption growth volatility in Africa and show that trade liberalization increases volatility, whereas financial liberalization decreases volatility through greater efficacy of consumption smoothing. They also find that financial depth and institutional quality interact negatively with trade and financial openness.

A significant literature has also evaluated the link between economic and trade diversification and macroeconomic volatility ([Papageorgiou & Spatafora, 2012](#); [Papageorgiou et al., 2015](#); [Moore & Walkes, 2010](#); [Koren & Tenreyro, 2007](#); [Romeu & da Costa Neto, 2011](#); [Farshbaf, 2012](#); [Jansen et al., 2009](#)), reaching a consensus that more diversified economies show lower volatility in variables such as GDP, consumption, investment, and exports, and are more resilient to external shocks. A key channel is that diversification involves LICs shifting resources from sectors where prices are highly volatile and correlated, such as mining and agriculture, to less volatile and correlated sectors, such as manufacturing and services, resulting in greater stability.

The effects of capital flows and transfers have also been heavily studied. [Singh et al. \(2011\)](#) provide a careful macroeconomic study of remittances in Sub-Saharan Africa (SSA), and find that remittances vary counter-cyclically with GDP per capita, consistent with the hypothesis that remittances can help mitigate economic shocks.

### 5.1 Cross-Sectional Analysis

In the following, I present an attempt to rank these factors by relevance for predicting volatility in a cross-section of African economies during the 1990-2019 period. For this, I selected 98 predictors jointly available for 49 African economies (excluding Djibouti, Liberia, Somalia, South Sudan, and Zimbabwe), with a total of 2.5% missing values. These include the vast majority of characteristics studied in the literature referenced above and also the external environment and financial sector

indicators studied in Sections 4.1 and 4.2. I group these 98 indicators into 19 topics, listed in Table 8, and, with statistical details, in Table C18. I then use a Random Forests (RF) machine learning model following Breiman (2001) to predict the volatility of per-capita growth and inflation and determine the importance of different predictors, both individually and at the topic level.<sup>14</sup>

Table 8: Indicator Topics for Cross-Sectional Prediction

#	Topic	Indicators
1	Institutions	9
2	Business Environment	4
3	Production Shares	2
4	Climate & Agriculture	8
5	Trade Intensity and Composition	7
6	Trade Diversification	4
7	Exchange Rate and Terms of Trade	5
8	Financial & Aid Flows	5
9	Financial Sector	6
10	Debt & Reserves	4
11	Population	6
12	Health	5
13	Education	5
14	Natural Disasters & Conflict	6
15	Geography & Accessibility	7
16	Natural Resources	2
17	Poverty & Inequality	3
18	Religion & Ethnicity	4
19	Others	6
SUM		98

*Notes:* 98 indicators, available for a cross-section of 49 African countries (excluding Djibouti, Liberia, Somalia, South Sudan, and Zimbabwe), are classified into 19 topics. See Table C18 for details.

To rank predictors individually, I fit a regression forest of 100,000 highly de-correlated trees, grown to full size, with only 3 out of 98 predictors randomly chosen at each split. I determine each predictor's importance by randomly permuting that predictor's observations and measuring the increase in the Out-of-Bag (OOB) Mean Squared Prediction Error (MSE) caused by the permutation in percentage terms. Figure C28 shows the top 30 predictors of the MAD of GDP per capita growth over the 1990-2019 period. Surprisingly, despite the high-dimensional dataset, the model only explains 28% of the OOB variance in the outcome variable. There are 10 predictors whose permutation increases the MSE by more than 2%; among these, there are 3 institutions, 2 business environment, and 2 remittance variables. The other top 10 variables are natural resource rents as a fraction of GDP, the share of industry in GDP, and natural disaster deaths. Among the variables that decrease predictive accuracy by more than 1% are also oil rents, trade with LMICs as a share of GDP, the MAD of FDI, total reserves, the cereal yield, human rights and level of democracy, the MAD of ToT growth, and the trade share of GDP.

Ranking topics (Table 8) is challenging, as topics are multi-dimensional and correlated. A first approach is to use the model underlying Figure C28, permute all predictors within a topic, and measure the decrease in predictive power. A problem with this method is that it does not attain the predictive performance of a model fit without those predictors. Thus, another approach is fitting different models, excluding topics and comparing their performance to the baseline model. This method can, however, also be criticized if different topics are correlated, as predictors in other topics will capture some variation of predictors in the excluded topic. One possibility to limit this is to project all other predictors on the predictors of the excluded group and use the residuals to fit

<sup>14</sup>Initially, the RF model is used to predict the 2.5% missing values in the predictor dataset by an iterative algorithm called 'MissForest' developed by Stekhoven & Bühlmann (2012). Most predictors have no missing values (see Table C18), and no predictor has more than 8 missing values.

a new model.<sup>15</sup> In the face of ambiguity, I implement all 3 methods and compute the average rank based on the increase in MSE from permutation/exclusion/partialling out the topical predictors. Table C19 reports the results. Overall, institutions emerge as the most important topic, followed by financial flows, trade intensity and composition, the financial sector, business conditions, natural resource intensity, natural disasters, and conflict.

The exercise is repeated, in Figure C29 and Table C20, with the MAD of CPI inflation. Exchange rate pass-through plays a dominant role in many African economies, followed by indicators of fragility and conflict, and institutions. Table C20 shows that excluding exchange rate variables worsens the model fit by 17.6%, whereas excluding most other topics increases the fit by 1-3%. Apart from the exchange rate, conflict/fragility and institutions, business conditions, population dynamics, trade intensity, trade diversification, and the financial sector are important predictors of inflation volatility.

## 5.2 Time-Variation in Structural Factors

The comparison of changes in these factors with the documented changes in volatility over the 1990-2019 period is of great importance in the scope of this paper but challenging as many indicators are either (nearly) time-invariant or lack historical data to trace them back to 1990. Particularly, survey-based variables measuring the quality of the business environment and financial access lack historical coverage. Figure C27 shows some institutional and business variables over the time period, indicating no positive change in the Worldwide Governance Indicators but significant improvements in business conditions and economic institutions in the recent years since measures became available. Restricting the analysis to variables with the necessary history thus provides an incomplete perspective of changes within African economies in the past 30 years.

Of the 98 variables considered in the cross-section, 70 have some time variation to be considered for analysis of changes.<sup>16</sup> Not included are mainly geography, religion, and ethnicity variables, static agricultural characteristics, and some institutions and business indicators with low time coverage. The analysis is then repeated on a cross-section of first-differences for 49 African economies, obtained by subtracting the median of the 70 indicators over the 1990-2004 period from the 2005-2019 median and relating this to the difference in the MADs of PCGDP growth and CPI inflation. Figure C30 and Table C21 show the results for PCGDP. It turns out that predicting changes in macroeconomic volatility over time is very challenging. The RF model in Figure C30 explains 0% of the variance in the change of the MAD of GDP per capita growth between 1990 and 2019 OOB (the in-sample  $R^2$  is 98%, indicating overfitting). With some hyperparameter tuning, the OOB  $R^2$  can be increased to 4%, but this is still poor. It is nevertheless noteworthy that 2 financial sector variables are among the top 5 predictors that increase the MSE by close to 1%. The other 3 variables are GDP per person employed, life expectancy, and population, which proxy for changes in the labor force and in human capital. Table C21 confirms the importance of the financial sector as well as social characteristics such as population dynamics, health, and education, alongside institutions and 'Others' which includes GDP per person employed, gross national savings, and the Human Development Index.

Overall, the result is a negative one. This could be due to relating changes in the medians to changes in the MAD of volatility, which throws away a lot of potentially useful variation, but, as shown in Appendix D, employing less robust measures such as the standard deviation of growth and time-averages of predictors, does not produce models with higher predictive power. Thus, the results strongly suggest that the bulk of the African Moderation in growth volatility is not due to changes in hard structural factors like institutions, trade intensity, and diversification, conflict intensity, poverty, and inequality, or natural resource rents, which make up the bulk of the predictor space, and thus the variables randomly sampled at each split to build a predictive model. These factors continue to be important in explaining different levels of baseline volatility between African countries (as shown above), but they do not explain the African Moderation. This finding is confirmed by the model selecting financial sector and human development variables as the most

---

<sup>15</sup>The projection is done using linear regression, e.g.  $Z(Z'Z)^{-1}Z'X$  where  $Z$  is a set of topical predictors and  $X$  the set of remaining predictors. If the set of topical predictors  $Z$  were large, this projection could also be made using an RF model, but with <10 predictors in  $Z$  the RF is not a sensible modeling choice.

<sup>16</sup>The first column in Table C18 in the Appendix shows which variables are included in the panel.

important time-varying predictors of growth moderation in Africa.

Analogous results for inflation are provided in Figure C31 and Tables C22 and C28. The reduction in exchange rate volatility (Figure C19) is the strongest correlate of inflation moderation in Africa, followed, with some distance, by changes in institutions, natural disasters, trade composition, human capital, and external debt. With an OOB  $R^2$  of 2.8%, the overall result is also negative.

These results are broadly robust against choices of outcome variables. Robustness checks using the standard deviation and IQR of per-capita growth and inflation and alternative per-capita growth and inflation series from the World Bank are provided in Appendix D.

## 6 Summary and Conclusion

Macroeconomic data for the past 30 years (1990-2019) show a large, broad-based, and persistent improvement in macroeconomic conditions in African economies, characterized by less volatile real per capita growth and CPI inflation rates, alongside higher average growth and lower inflation levels. The improvement in macroeconomic conditions is such that, apart from inflation, where the ROW median was at ~2% in 2010-2019 vs. ~4% in Africa, the median African country has caught up with ROW. The bulk of this "Great African Moderation" took place between 1995 and 2012, during which macroeconomic conditions improved ~2 times more rapidly in Africa than in ROW. A particularly large stabilization occurred in African LICs. Disaggregated analysis at the country and sector level reveals that the majority of countries have experienced large declines in the volatility of agricultural VA, and sizeable declines in the volatility of services VA. In parallel, there were more heterogeneous developments in the industrial sector, where several countries like Ghana or Tunisia incurred increases in volatility.

At the regional level, Eastern and Northern Africa experienced the greatest decline in agricultural volatility. This was accompanied, in Eastern Africa, by sizeable declines in services and industrial volatility, while Northern Africa experienced an increase in industrial volatility and a small reduction in services (Table C10). Western and Southern Africa experienced smaller improvements in agriculture and services stability, but both also incurred slight increases in industrial volatility. Overall, Eastern Africa region shows the most remarkable macroeconomic stabilization and is on par with North Africa in terms of macroeconomic stability. The relatively weaker stabilization of Western Africa is an interesting avenue for further research.<sup>17</sup>

Sectoral decompositions of the change in aggregate per-capita growth volatility show that 60-70% is accounted for by changes in the volatility of agriculture and services VA, and around 30% can be explained by changes in the covariance structure of production, with all 3 broad sectors becoming less complementary to each other. Classical structural change accounts for ~5% of the aggregate output moderation and works via VA shifting to the less volatile service sector. This is a small effect, but, as noted by Stock & Watson (2002), a large and stable services sector can have stabilizing effects on other sectors through income and demand channels, so the contribution of more broadly conceived structural change may be larger. This also invites further research.

The second part of this paper investigated changes in the external environment, financial deepening, domestic macroeconomic policies, and structural factors. Results suggest that growth and moderation in Africa were likely benefited by lower levels of domestic and external debt, higher FDI and remittance inflows, and improved terms of trade (ToT). Externally induced volatility, such as volatility of the exchange rate, ToT, the merchandise trade balance, FDI, and remittances, also decreased over the period. A gradual deepening of the financial sector, as evidenced by higher levels of reserves held by the central bank as well as by commercial banks, more domestic credit to the private sector and broad money as a share of GDP, and an increase of gross national savings by 4-5% of GDP, likely also played a role.

In terms of macroeconomic policies, several countries stabilized the exchange rate trough a crawling peg, such that (crawling) pegs constituted more than 80% of African arrangements in

---

<sup>17</sup>This may be associated with specific factors worthwhile investigating, such as larger effects of climate change on agricultural production, greater volatility in crop prices, or greater political instability in the region.

2019, from 25% crawling bands and nearly 20% freely falling (i.e. very volatile) arrangements in 1990. Exchange rate volatility is the most important correlate of inflation volatility in Africa and has more than halved since 1990 (Figure C19). Macroprudential policy, while potentially effective for reducing volatility, shows no aggregate trend since 1998, except within certain financial markets such as bonds and commercial credits. Another potentially important development was the adoption of fiscal rules by a significant number of African countries from 1997 onwards. The adoption of such rules, particularly budget balance and debt rules, has a large and statistically significant negative relationship with macroeconomic volatility. Inflation-targeting monetary policy, on the other hand, has only been adopted by 4 African countries at a time when inflation was already low and is thus an insignificant contributor to the African Moderation.

Examining a broad set of 'structural' characteristics shows that these factors can explain around 30% of the cross-sectional variation of growth volatility between African countries from 1990-2019. The quality of institutions appears to be the most important factor affecting structural per-capita growth volatility, followed by the intensity of financial flows, the characteristics of the financial sector, trade intensity, and composition. The business environment, natural resource extraction, and disaster and conflict incidence were also found to be important. Inflation volatility is heavily influenced by exchange rate volatility, conflict, and the institutional and business environment. This suggests that stabilizing the exchange rate and maintaining a strong institutional environment are important to keeping inflation low. The prediction of differences in volatility over the period with changes in these factors yields that the African Moderation cannot be predicted by them. The results nevertheless suggest a role of financial depth and human capital development for growth stabilization and exchange rate management for inflation stabilization.

Overall, the paper succeeds to a greater extent in showing what did *not* cause the African Moderation; that is, it was not, at large, a byproduct of classical structural change, not caused by reduced volatility in the industrial sector (e.g. via improved inventory management), or by monetary policy shifting to inflation targeting, and also not heavily influenced by other changes in economic structure, diversification, conflict incidence, or institutions. On the positive side, the results provide evidence for a role of changes in the external environment faced by African economies, greater resilience of the financial sector, and macroeconomic policy, particularly exchange rate and fiscal management. The analysis also suggests that improvements in human capital and the business environment played a role, but the evidence presented in this paper is very weak.

These findings provide a basis for further research that investigates in more detail the causes and consequences of macroeconomic moderation in Africa and the role of policy for macroeconomic stabilization. Further significant changes in policies or institutions may have taken place that are not easily measurable. For example, central banks might have become much better over time at targeting macroeconomic aggregates or implementing macroprudential policies. The role of global factors such as US monetary policy, global financial markets, and commodity price volatility for African Moderation could also be investigated in more detail. It is also not clear in which ways improvements in the business environment, as evident in the Doing Business Rankings for Africa and the Logistics Performance Index, interact with broader macroeconomic stabilization and domestic financial deepening.

## References

- Abdullahi, A., & Suardi, S. (2009). Macroeconomic volatility, trade and financial liberalization in africa. *World Development*, 37(10), 1623–1636.
- Acemoglu, D., Johnson, S., Robinson, J., & Thaicharoen, Y. (2003). Institutional causes, macroeconomic symptoms: volatility, crises and growth. *Journal of Monetary Economics*, 50(1), 49–123.
- Ahmed, S., Levin, A., & Wilson, B. A. (2004). Recent us macroeconomic stability: good policies, good practices, or good luck? *Review of Economics and Statistics*, 86(3), 824–832.
- Arora, V., Habermeier, K., Ostry, J. D., & Weeks-Brown, R. (2013). The liberalization and management of capital flows: An institutional view. *Revista de Economia Institucional*, 15(28), 205–255.
- Auffret, P. (2003). High consumption volatility: The impact of natural disasters? *World Bank Policy Research Working Paper*(2962). Retrieved from <https://documents1.worldbank.org/curated/en/880931468769133698/pdf/multi0page.pdf>
- Bernanke, B. (2004). The great moderation. Eastern Economic Association, Washington, D.C.
- Bernanke, B. (2012). The great moderation. In E. F. Koenig, R. Leeson, & G. A. Kahn (Eds.), *The taylor rule and the transformation of monetary policy* (chap. 6). Hoover Institution, Stanford University. Retrieved from <https://EconPapers.repec.org/RePEc:hoo:bookch:4-6>
- Blanchard, O., & Simon, J. (2001). The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity*, 2001(1), 135–174. doi: 10.1353/eca.2001.0013
- Bleaney, M., Tian, M., & Yin, L. (2016). Global trends in the choice of exchange rate regime. *Open Economies Review*, 27, 71–85.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Burger, P. (2008). The changing volatility of the south african economy. *South African Journal of Economics*, 76(3), 335–355.
- Calderon, C., & Boreux, S. (2016, aug). Citius, altius, fortius: Is growth in Sub-Saharan Africa more resilient? *Journal of African Economies*, 25(4), 502–528. doi: 10.1093/jae/ejw006
- Davoodi et al. (2022a). Fiscal rules and fiscal councils: Recent trends and performance during the covid-19 pandemic. *IMF Working Paper*, 22/11. Retrieved from <https://www.imf.org/-/media/Files/Publications/WP/2022/English/wpiea2022011-print-pdf.ashx>
- Davoodi et al. (2022b). *Fiscal rules dataset: 1985–2021*. International Monetary Fund, Washington, D.C. Retrieved from <https://www.imf.org/external/datamapper/fiscalrules/map/map.htm>
- Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. *Unpublished Qualifying Paper, Harvard University, Cambridge, MA*.
- Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4), 549–560.
- Du Plessis, S., & Kotzé, K. (2010). The great moderation of the south african business cycle. *Economic History of Developing Regions*, 25(1), 105–125.
- Easterly, W., Islam, R., & Stiglitz, J. E. (2001). Shaken and stirred: Explaining growth volatility. In B. Pleskovic & N. Stern (Eds.), *Annual world bank conference on development economics 2000* (pp. 191–211). Washington, DC: World Bank. Retrieved from <http://hdl.handle.net/10986/14010> (License: CC BY 3.0 IGO)
- Farshbaf, A. (2012). Does geographical diversification in international trade reduce business cycle volatility. *Job Market Paper, Department of Economics, USC*.
- Fernández, A., Klein, M. W., Rebucci, A., Schindler, M., & Uribe, M. (2016). Capital control measures: A new dataset. *IMF Economic Review*, 64(3), 548–574.

- Galí, J., & Gambetti, L. (2009). On the sources of the great moderation. *American Economic Journal: Macroeconomics*, 1(1), 26–57.
- Hallegatte, S., & Przyluski, V. (2011). Managing volatility: A vulnerability exercise for low-income countries. *IMF Policy Papers*(4540). Retrieved from <https://www.imf.org/en/Publications/Policy-Papers/Issues/2016/12/31/Managing-Volatility-A-Vulnerability-Exercise-for-Low-Income-Countries-PP4540>
- Hnatkovska, V., & Loayza, N. (2005). Volatility and growth. *Managing Economic Volatility and Crises: A Practitioner's Guide*, Cambridge University Press New York, 101–136.
- Horan, S. M. (2006). What Caused the Great Moderation? Some Cross-Country Evidence. *CFA Digest*, 36(2), 102–102. doi: 10.2469/dig.v36.n2.4134
- Ilzetzki, E., Reinhart, C. M., & Rogoff, K. S. (2019). Exchange arrangements entering the twenty-first century: Which anchor will hold? *The Quarterly Journal of Economics*, 134(2), 599–646.
- IMF. (2017). Increasing resilience to large and volatile capital flows: The role of macroprudential policies. *IMF Policy Papers*(060217). Retrieved from <https://www.imf.org/en/Publications/Policy-Papers/Issues/2017/07/05/pp060217-increasing-resilience-to-large-and-volatile-capital-flows>
- IMF. (2021). How to Assess Country Risk: The Vulnerability Exercise Approach Using Machine Learning. *IMF Technical Notes and Manuals*(2021/003). Retrieved from <https://www.imf.org/en/Publications/TNM/Issues/2021/05/07/How-to-Assess-Country-Risk-50276>
- IMF. (2022). Review of the institutional view on the liberalization and management of capital flows. *IMF Policy Papers*, 2022(008).
- Jansen, M., Lennon, C., & Piermartini, R. (2009). *Exposure to external country specific shocks and income volatility* (WTO Staff Working Papers No. ERSD-2009-04). World Trade Organization (WTO), Economic Research and Statistics Division. Retrieved from <https://EconPapers.repec.org/RePEc:zbw:wtowp:ersd200904>
- Koller, M., & Stahel, W. A. (2011). Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis*, 55(8), 2504–2515.
- Koren, M., & Tenreyro, S. (2007). Volatility and development. *The Quarterly Journal of Economics*, 122(1), 243–287.
- Kruse, H., Mensah, E., Sen, K., & de Vries, G. (2023). A Manufacturing (Re)Naissance? Industrialization in the Developing World. *IMF Economic Review*, 71(2), 439–473. Retrieved from <https://doi.org/10.1057/s41308-022-00183-7> doi: 10.1057/s41308-022-00183-7
- Loayza, N. V., Ranciere, R., Servén, L., & Ventura, J. (2007). Macroeconomic volatility and welfare in developing countries: An introduction. *The World Bank Economic Review*, 21(3), 343–357.
- Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., ... Anna di Palma, M. (2021). robustbase: Basic robust statistics [Computer software manual]. Retrieved from <http://robustbase.r-forge.r-project.org/> (R package version 0.93-9)
- Malik, A., & Temple, J. R. (2009, 11). The geography of output volatility. *Journal of Development Economics*, 90, 163–178. doi: 10.1016/j.jdeveco.2008.10.003
- Maronna, R. A., Martin, D. R., Yohai, V. J., & Salibian-Barrera, M. (2018). *Robust statistics: Theory and methods (with r)* (2nd ed.). New York: John Wiley & Sons Ltd. doi: 10.1002/9781119214656
- McConnell, M. M., & Perez-Quiros, G. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review*, 90(5), 1464–1476.
- McMillan, M., Rodrik, D., & Verdúzco-Gallo, I. (2014). Globalization, structural change, and productivity growth, with an update on africa. *World development*, 63, 11–32.

- Moore, W., & Walkes, C. (2010). Does industrial concentration impact on the relationship between policies and volatility? *International Review of Applied Economics*, 24(2), 179–202.
- Papageorgiou, C., & Spatafora, N. (2012). Economic diversification in lics: Stylized facts and macroeconomic implications. *IMF Staff Discussion Note*.
- Papageorgiou, C., Spatafora, N., & Wang, K. (2015). Diversification, growth, and volatility in asia. *World Bank Policy Research Working Paper*(7380).
- Ramey, G., & Ramey, V. A. (1995). Cross-country evidence on the link between volatility and growth. *The American Economic Review*, 85(5), 1138–1151.
- Rodrik, D. (1999). Where did all the growth go? external shocks, social conflict, and growth collapses. *Journal of Economic Growth*, 4(4), 385–412.
- Rodrik, D. (2018). An african growth miracle? *Journal of African Economies*, 27(1), 10–27.
- Romeu, R., & da Costa Neto, M. N. C. (2011). Did Export Diversification Soften the Impact of the Global Financial Crisis? *IMF Working Papers*(2011/099). Retrieved from <https://ideas.repec.org/p/imf/imfwpa/2011-099.html>
- Schmidt-Hebbel, K. (2009). Great moderation and inflation targeting in the world. In *What drives prices in egypt?: An analysis in light of international experience* (pp. 9–36). American University in Cairo Press. Retrieved 2023-10-23, from <http://www.jstor.org/stable/j.ctt15m7jjr.6>
- Shumway, R. H., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). New York: Springer.
- Singh, R. J., Haacker, M., Lee, K. W., & Goff, M. L. (2011, 3). Determinants and macroeconomic impact of remittances in sub-saharan africa. *Journal of African Economies*, 20, 312-340. doi: 10.1093/jae/ejq039
- Stahel, W. A. (1981). Robust estimation: Infinitesimal optimality and covariance matrix estimators. *unpublished doctoral dissertation, ETH, Zurich, Switzerland*.
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest: non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.
- Stock, J. H., & Watson, M. W. (2002). Has the business cycle changed and why? *NBER macroeconomics annual*, 17, 159–218.
- Syed, M. M. H., Ahuja, M. A., & Wiseman, K. (2017). Assessing Country Risk: Selected Approaches. *IMF Technical Notes and Manuals*(2017/008). Retrieved from <https://ideas.repec.org/p/imf/imftnm/2017-008.html>
- Todorov, V., & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3), 1–47. Retrieved from <https://www.jstatsoft.org/article/view/v032i03>
- Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17. doi: 10.18637/jss.v077.i01

## Appendix

The Appendix consists of 4 parts. Part A provides a spectral analysis to examine the qualities of the volatility of the growth rate of a GDP per capita as a proxy for adverse economic volatility. Part B examines GDP composition and volatility from the expenditure side. Part C provides additional tables and figures referred to in the main text of the paper. Part D is provided [separately online](#),<sup>18</sup> and includes a few additional (detailed) tables and figures, and robustness checks for the machine learning analysis of structural factors in Section 5, using alternative outcome measures.

### A. Spectral Analysis

As noted by [Gelb \(1979\)](#), many measures of instability used in economic literature are arbitrary and emphasize volatility at certain frequencies without rigorous justification. [Shumway & Stoffer \(2000\)](#) show that first-differencing amounts to a high-pass filter that gives most weight to volatility at frequencies of 2 years and gradually down weights volatility at lower frequencies. [Gelb \(1979\)](#) suggests considering the full frequency spectrum and devising a weighting scheme emphasizing the relative importance of certain frequencies above others to generate an indicator. This Idea is formalized by [Tsui \(1988\)](#), who also shows that various common trend-cycle estimates (including first-differences) can be regarded as special cases of a weighting function  $f(\omega)$  applied to the spectral density. This analysis follows [Tsui \(1988\)](#) and proposes a weighting function based on the empirical relationship of volatility at different frequencies with average growth rates.

In the first step, the spectral density of fluctuations needs to be computed for all country GDP per Capita series, shown in Figure C1 (the IMF series is used). [Gelb \(1979\)](#) notes that in the presence of strong trends, spectral density estimates on the raw series are often highly misleading because no periodic component fits the trend well, so variance from this very low-frequency phenomenon (the trend) "spills over" onto higher frequency components. Therefore all country series are first detrended using a linear trend on the log-level series, i.e. we consider all variation that lets countries depart from growing at a constant rate in per-capita terms. The spectral density can be approximated by the periodogram given by

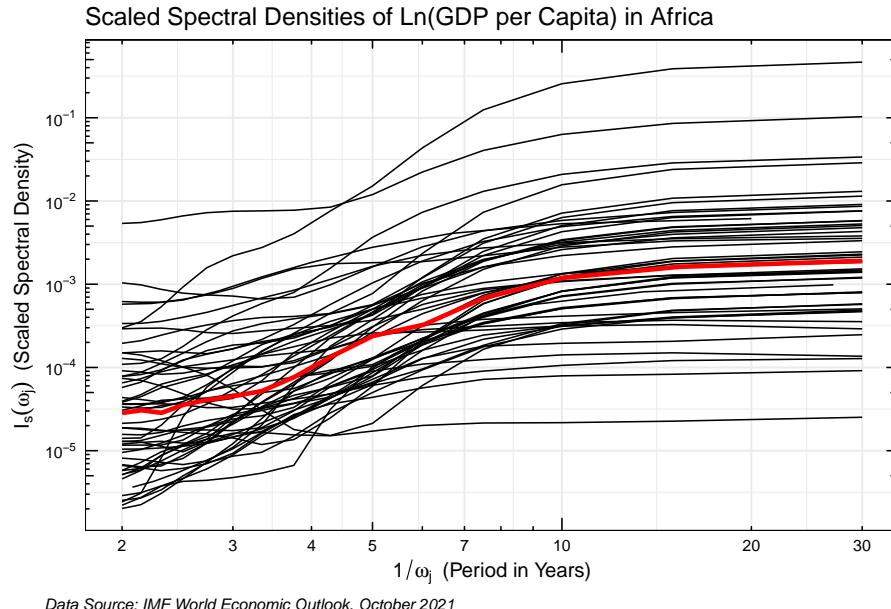
$$I(\omega_j) = |d(\omega_j)|^2 \quad \text{where} \quad d(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-2\pi i \omega_j t} \quad (6)$$

is the complex-valued coefficient of the Discrete Fourier Transform at fundamental frequency  $\omega_j = j/n$  for  $j \in 0, \dots, n - 1$  of the series  $x_t$  observed for  $n$  periods. To aid interpretation we consider the scaled periodogram  $I_s(\omega_j) = \frac{4}{n} I(\omega_j)$ , such that the sum of the periodogram ordinates over all frequencies  $\sum_{\omega_j} I_s(\omega_j)$  equals the squared amplitude of the signal  $x_t$ , and furthermore  $\sum_{\omega_j} I_s(\omega_j) = 2 \text{ var}(x_t)$ , such that the power of the scaled periodogram at each frequency  $\omega$  can be considered as twice the contribution of that frequency to the overall variance of  $x_t$ .<sup>19</sup> A further issue is that the periodogram is not a consistent estimator of the spectral density. A frequently employed solution is smoothing the periodogram with Daniell smoothers to produce more consistent estimates. Another technique to improve the periodogram as a spectral estimator is tapering, which reduces the effect of frequencies outside the estimated interval. To reach consistent spectral estimates at the country level, I apply a cosine bell taper of 15% and smooth the periodogram with two modified Daniell smoothers of widths 3 and 7 (period-years), which are convolved to produce the final spectral estimates. The scaled densities thus estimated for all countries are shown in Figure A1, where the red line denotes the median across all country spectra.

<sup>18</sup><https://www.dropbox.com/s/57fact7ilsq0upc/Appendix%20D.pdf?dl=0>

<sup>19</sup>It is a property of sine and cosine waves that the squared amplitude equals twice the variance. For details see [Shumway & Stoffer \(2000\)](#).

Figure A1: Estimated Country Spectral Densities and Median Spectral Density



It is evident that the spectra of different countries are quite heterogeneous, with about 3 orders of magnitude lying between the least and most-volatile countries at each frequency, but an overall decrease in spectral power with higher frequencies is common to all countries. The median estimate in Figure A1 shows that on average volatility at low frequencies with periods of 20-years+ is around 2 orders of magnitude larger than year-to-year changes in output (2-year period).

To determine whether volatility at certain frequencies is harmful to growth in African economies, I compute the cross-sectional correlation of the spectral density with median GDP per capita growth in the 1990-2019 period, for each fundamental frequency  $\omega_j$ . Figure A2 reports these correlations in the top half, and the bottom half shows corresponding regression coefficients, which also take into account the differing magnitudes of volatility at different frequencies that Figure A1 made evident.

Figure A2: Correlation of Spectral Density and Median Per-Capita Growth, Africa 1990-2019

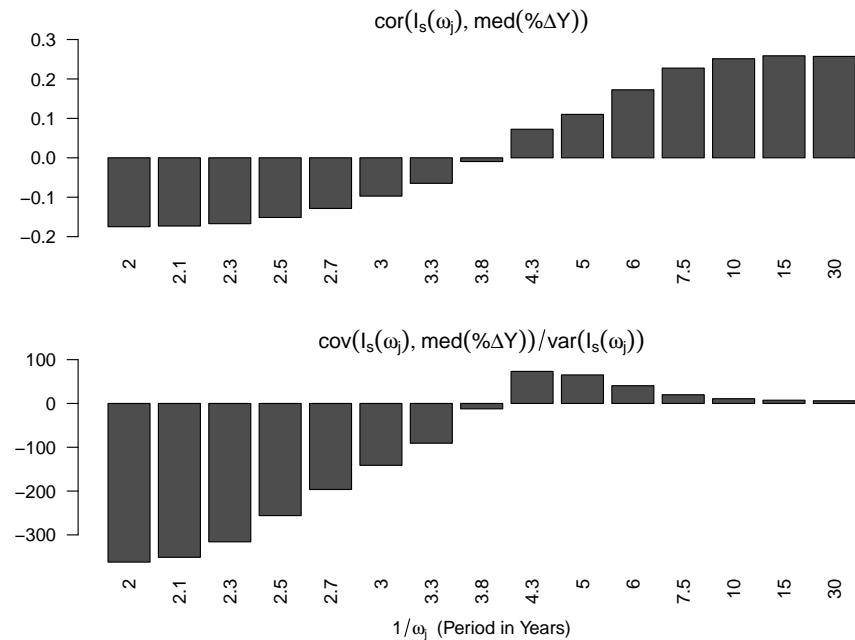


Figure A2 exhibits an astonishingly clear pattern, with a strong negative correlation of about -0.2 between economic growth and volatility at high frequencies of 0.5 (period 2 years), which then gradually tends to zero at frequencies of 0.25 (period 4 years), and turns positive up to about 0.3 for lower frequency variation with 10 to 30-year periods. Thus African data indeed show that short-term volatility with periods of up to 4 years is associated with lower growth, whereas volatility at longer periods is an indication of healthy growth.

This suggests that a high-pass filter like computing the growth rate might do reasonably well to extract fluctuations harmful to growth. I use the regression coefficients in the bottom half of Figure A2 to create an optimal discrete high-pass filter  $f(\omega_j)$  in the spirit of [Tsui \(1988\)](#), that captures volatility harmful for growth. The filter simply consists of the absolute values of all negative regression coefficients on the frequency bands, setting positive coefficients to zero. Multiplying the spectral density estimate for each country with this filter and summing up the weighted spectral ordinates gives the power of the filtered spectrum, which provides a summary statistic of the harmful volatility in each country. Formally, I define a harmful volatility index (HVI) as

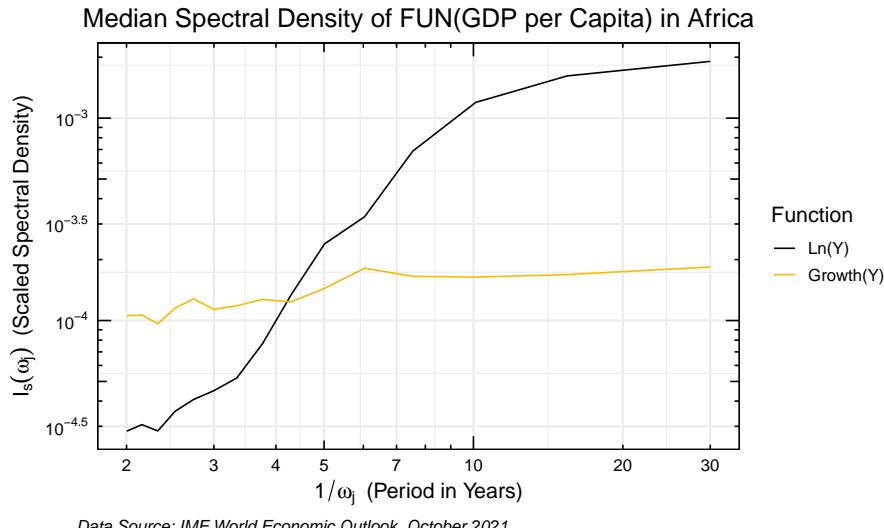
$$\text{HVI} = \sum_j f(\omega_j) \times I_s(\omega_j) \quad \text{where} \quad (7)$$

$$f(\omega_j) = -\beta_{\omega_j} \times 1[\beta_{\omega_j} < 0] \quad \text{and} \quad (8)$$

$$\beta_{\omega_j} = \frac{\text{cov}(I_s(\omega_j), \text{med}(\% \Delta Y))}{\text{var}(I_s(\omega_j))}. \quad (9)$$

Before comparing the HVI to some statistic computed on the growth rate, I wish to determine to what extent computing a growth rate itself resembles the transformation induced by applying  $f(\omega_j)$  to the data. Figure A3 shows that computing the growth rate indeed works like a high-pass filter that, relative to the natural log baseline, accentuates volatility at periods lower than 4.2 years and dampens volatility at higher periods.

Figure A3: Spectral Densities of Growth Rate and Natural Log of GDP per Capita



Dividing the growth spectrum by the log spectrum yields the discrete filter that, if multiplied with the log spectrum, yields the same effect as computing a growth rate (i.e. differencing the log-level series) in the time domain. I call this derived first-difference filter  $f^\Delta(\omega_j)$ . To compare  $f^\Delta(\omega_j)$  to the optimal empirical filter  $f(\omega_j)$  based on regressions against median per capita growth, I scale both filters so that the weights/coefficients on all frequencies  $\omega_j$  sum to 1. Figure A4 shows the outcome.

Figure A4: First-Difference Filter and Regression-Based Filter

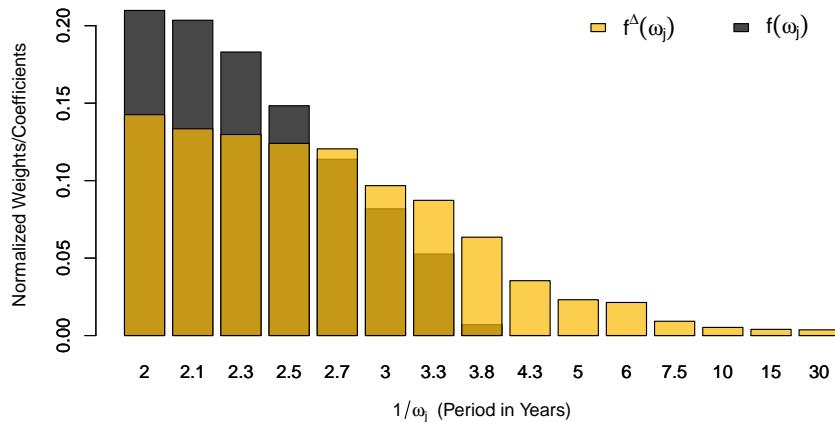
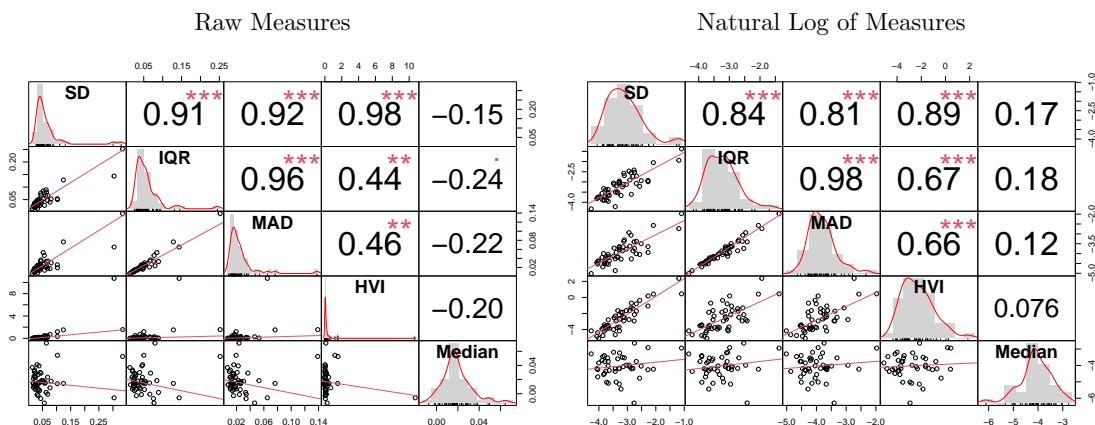


Figure A4 indicates that the first-difference filter  $f^{\Delta}(\omega_j)$  broadly resembles the optimal empirical filter  $f(\omega_j)$  for extracting volatility harmful to growth. Compared to the latter, first-differencing provides a smoother transformation of the data, that puts less weight on high-frequency volatility, but therefore keeps some of the low-frequency volatility as well. Since finding an optimal filter  $f(\omega_j)$  to extract harmful economic volatility is likely always going to be a complex empirical task, and the resulting filter is prone to be highly dependent on the data and methodology used to estimate it, a simpler methodology such as computing first-differences and then applying some statistic to summarise the volatility in the differenced series is preferable to ensure the transparency and reproducibility of research. Below I consider the 3 summary statistics used in this paper: the standard deviation (SD), interquartile range (IQR), and median absolute deviation (MAD) of the growth rate of GDP per capita, and compare them to the HVI index (Eq. 7) and median per capita growth, computed for each African country using data from 1990 through 2019. The data are correlated, and a regression line is fit, using a robust MM estimator following Yohai (1987) and Koller & Stahel (2011), with a high breakdown point of 0.5, ensuring that outliers don't influence the estimates. Figure A5 shows charts including these robust fits, a robust correlation coefficient derived from the fit, and empirical volatility distributions estimated by a histogram and a gaussian kernel density.

Figure A5: Volatility Measures and Median GDP per Capita of 51 African Economies, 1990-2019



The left side of Figure A5 shows that the HVI is positively correlated with all 3 volatility measures derived from the growth rate, particularly with the SD. All volatility measures are also negatively correlated with the median growth rate. Since a few countries such as Libya, Guinea-Bissau, Eritrea, and Rwanda have very high levels of volatility (due to conflicts during this period), the empirical volatility distributions are right-skewed. As indicated on the left side, the negative

correlation of the IQR and MAD of growth with median growth is stronger compared to the HVI and the SD, which may be the effect of outliers having a stronger effect on the SD and HVI.<sup>20</sup> The right side of Figure A5 therefore also shows a version of the chart where the natural log was applied to all measures. This gives nicer scatterplots and density estimates but also lets the relationship between volatility and median growth turn positive (albeit insignificant), for all measures apart from the HVI where the correlation is zero. This change in the sign of correlations is explicable as some of the countries affected by conflict in 1990–2019, such as Rwanda and Guinea-Bissau, also experienced high average growth throughout this period, and may exert a stronger influence on the MM estimates after taking the log.

To conclude, the discussions in this section highlighted that when dealing with a difficult-to-measure phenomenon such as economic volatility, three things are important: precise measurement of (harmful) volatility, robustness against outliers, and a simple, reproducible, and data-independent methodology. This paper endorsed robust statistics such as the IQR and the MAD, computed on the growth rate of the series, to measure economic volatility. The analysis conducted in this section shows that computing the growth rate provides a decent approximation to an optimal empirical filter, applied to the spectral density to extract volatility harmful to economic development in Africa and that computing the IQR or MAD of the growth rate provides an acceptable and robust summary measure of this volatility, comparable to the power of the optimally filtered spectrum (the HVI). The IQR and MAD of the growth rate thus sufficiently meet the joint aims of precision, robustness, and simplicity. At the country level, the MAD is preferred to the IQR as it is more robust.

## References

- Gelb, A. (1979). On the definition and measurement of instability and the costs of buffering export fluctuations. *The Review of Economic Studies*, 46(1), 149–162.
- Koller, M., & Stahel, W. A. (2011). Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis*, 55(8), 2504–2515.
- Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.
- Tsui, K. Y. (1988). The measurement of export instability: a methodological note. *Economics Letters*, 27(1), 61–65.
- Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, 642–656.

---

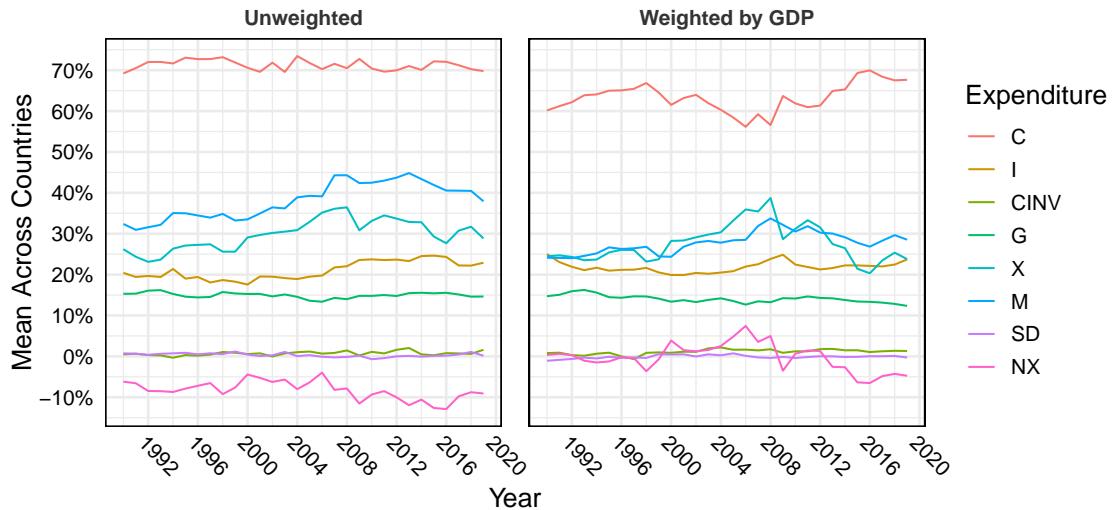
<sup>20</sup>The Fast Fourier Transform underlying the smooth spectral estimates used to produce the HVI is not robust against outliers.

## B. A Brief Look at Expenditure on GDP

Figure B1 provides a detailed breakdown of expenditure shares in GDP, averaged across countries. CINV denotes changes in inventories, SD are statistical deviations, and exports (X) and imports (M) are provided alongside net exports (NX). CINV is very small in the median African country.

Figure B1: GDP Shares: Expenditure Side

### Expenditure Shares in GDP in Africa

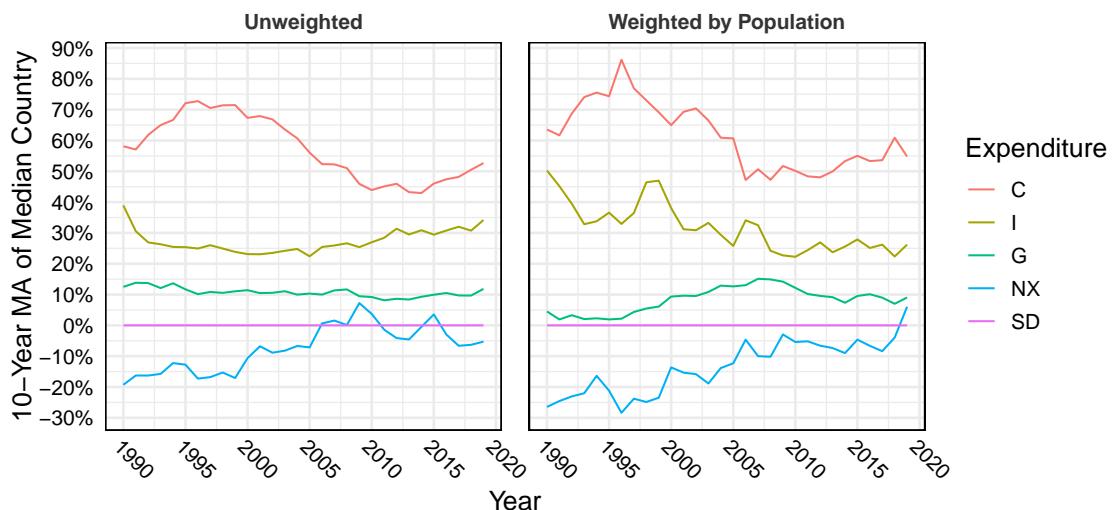


Data Source: World Development Indicators, Accessed November 2021

Figure B2 shows smoothed contributions of major expenditure components to GDP per capita growth, analogous to Figure C11 on the production side. It is evident that consumption growth declined in importance until around 2013 and increased a bit again thereafter.

Figure B2: Contributions to GDP Growth: Expenditure Side

### Expenditure Shares in Average GDP per Capita Growth in Africa

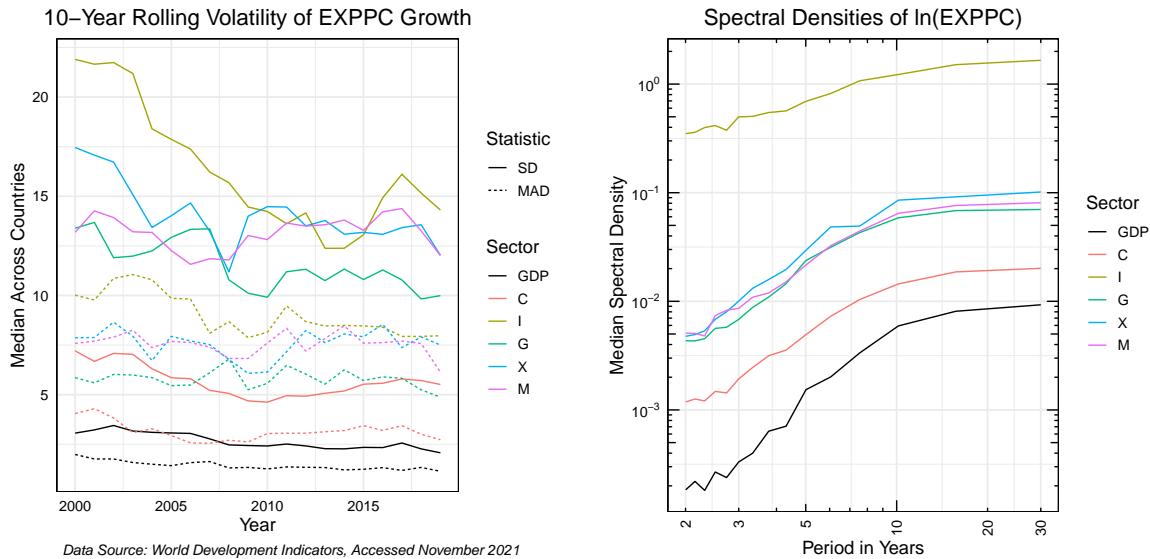


Data Source: World Development Indicators, Accessed November 2021

Overall the shares are relatively stable. Investment has increased slightly, climbing from ~20% in 2005 to ~25% in 2012. Exports and imports also both increased gradually until 2012 and then began to fall, with a greater decline in the export share, yielding a higher aggregate trade deficit.

The left panel of Figure B3 shows the aggregate decline in volatility, which, on the expenditure side, is accounted for by declines in the volatility of all components before 2010, with the volatility of trade and investment remaining high thereafter. Particularly investment volatility (which includes CINV in this disaggregation), declined strongly. The right panel of Figure B3 shows that from a frequency domain perspective, investment is the most volatile component at all frequencies (exempting net exports). Consumption is the least volatile component and approaches the volatility of GDP at lower frequencies.

Figure B3: Expenditure Volatility Across Time and Frequency



*Notes:* The LHS shows 10-year rolling SDs and MADs of the growth rate of GDP per capita at constant 2015 prices and its expenditure components ( $GDP = C + I + G + X - M$ ). For the RHS see the note to Figure 3 and Appendix A.

Table B1 provides a covariance matrix analogous to Table C6. This shows large negative covariances of imports with absorption and exports, indicating the endogeneity of net exports and the difficulties to account aggregate changes in volatility from the expenditure side. Linking production and expenditure side data is also difficult without detailed breakdowns, but the large declines in agriculture and service sector volatility are likely reflected on the expenditure side in the decline in consumption volatility, but also in declining volatility of the merchandise trade balance.

Table B1: Expenditure Volatility and Contribution to Aggregate Volatility, 1990-2019

Data	Sector:	C	I	G	X	M	C	I	G	X	M
GDP Share ( $\bar{\theta}_k$ )		0.701	0.227	0.152	0.336	-0.416	0.698	0.220	0.149	0.338	-0.406
<i>Cov.:</i>											
Expenditure	C	52.22					33.69				
Growth	I	-0.93	528.25				1.87	272.76			
( $\Delta VA/VA_{t-1}$ )	G	0.57	11.42	229.63			3.79	26.07	114.99		
	X	-7.82	14.84	-5.95	269.80		-14.96	5.14	-7.61	212.40	
	M	-26.78	-92.13	-16.58	-93.44	242.37	-20.57	-120.07	-29.83	-63.39	166.73
Expenditure	C	22.95					18.23				
Contribution	I	0.27	14.43				0.47	11.57			
( $\Delta VA/GDP_{t-1}$ )	G	0.12	0.56	3.41			-0.05	0.47	2.15		
	X	-1.09	0.79	-0.18	14.21		-2.02	0.56	-0.08	12.88	
	M	-5.36	-6.73	-0.91	-6.76	22.06	-5.08	-5.54	-1.51	-6.95	16.84

*Notes:* Since sectoral growth rates can be very volatile, I employ both a classical (Pearson) and robust covariance estimator with a high breakdown point (0.5) based on [Stahel \(1981\)](#) and [Donoho \(1982\)](#). The choice of methods was informed by [Maronna et al. \(2018\)](#) and available implementations in various R packages. The Stahel-Donoho robust covariance estimator is implemented by the package *rrcov* ([Todorov & Filzmoser, 2009](#)). Covariance terms are aggregated across countries using the median, whereas sectoral shares are aggregated with the mean. Average shares for each country are computed using all but the first observation following Eq. 3. The shares reported above "Robust" are computed by taking the median share for each country, and aggregating across countries using the mean.

Table B2 shows a decomposition of the reduction in GDP volatility between  $\tau_1 = 1990-2004$  and  $\tau_2 = 2005-2019$ , based on the LHS of Eq. 3, analogous to Table B2 in the paper. Due

to the difficulty with exports accounting, I only report results where shares are computed at the country-level and aggregated across countries using the median. The results imply that the expenditure-side shares in the moderation are roughly consistent with their share in aggregate volatility, reported in Table B1, with consumption, investments and exports having a higher than proportional share, consistent with Figure B3. The results are quite noisy through, even when aggregated across countries using the median; for example the sign of the covariance contribution from the expenditure side is not robust to the choice of covariance estimator.

Table B2: Sectoral Contribution to Moderation in GDP Volatility

CovEst	AggFun	Fit	$\Delta var(\% \Delta Y)_\tau$	C	I	G	X	M	$\sum cov_{jk}$
Pearson	Median	100%	-6.65	48%	16%	3.8%	7.3%	11%	16%
Comedian	Median	55%	-1.14	73%	32%	5.1%	27%	12%	-35%

*Notes:* The 'Fit' column signifies how closely Eq. 3 is satisfied. Columns C-M give the sectoral contribution to the aggregate volatility reduction in percentage terms, and  $\sum cov_{jk}$  gives the combined contribution of all covariance terms. Shares are computed at the country-level, and aggregated using the median.

## C. Additional Tables and Figures

Figure C1: Log10 GDP per Capita for 54 African Economies in Constant USD, 1990-2019

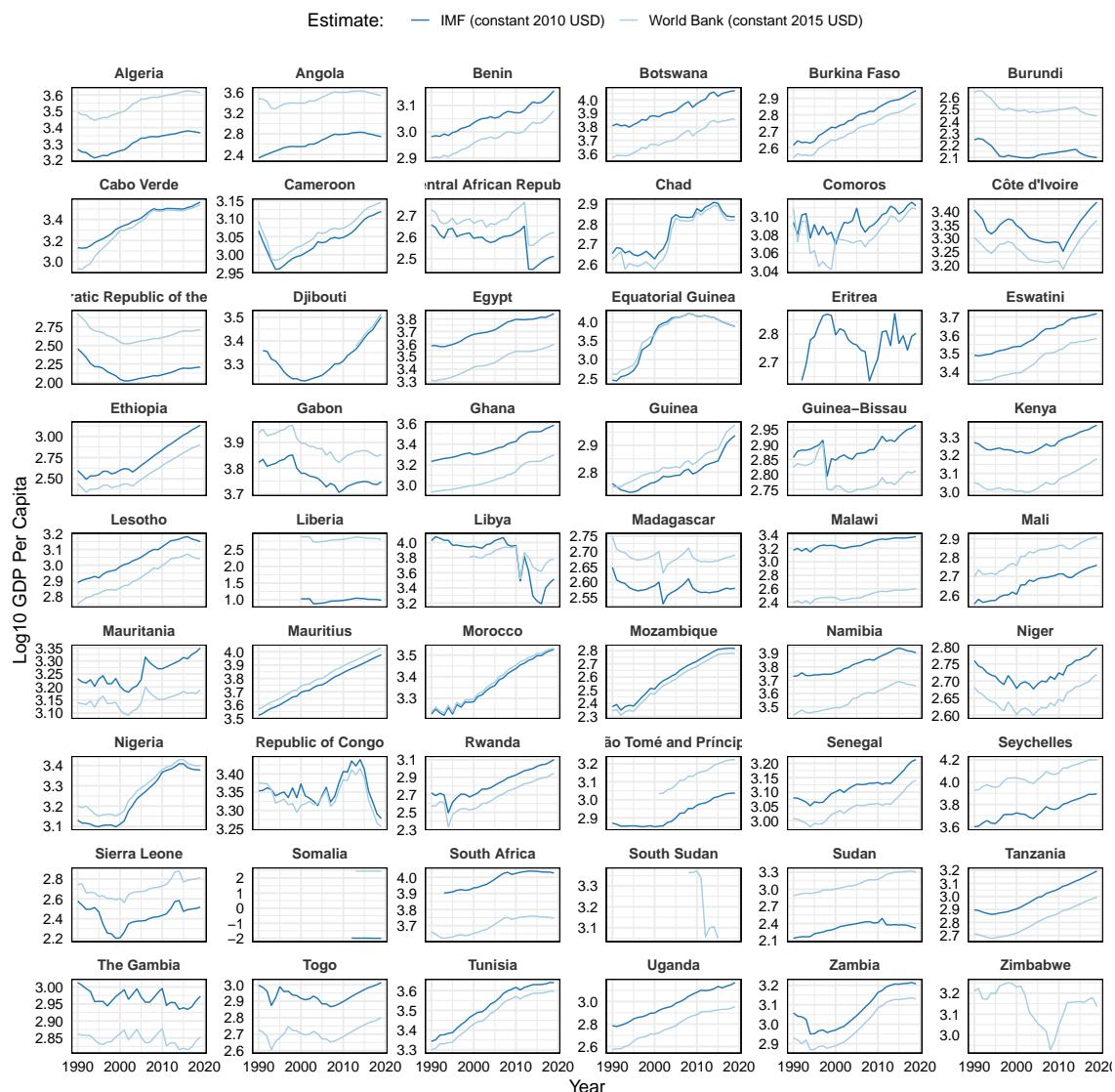
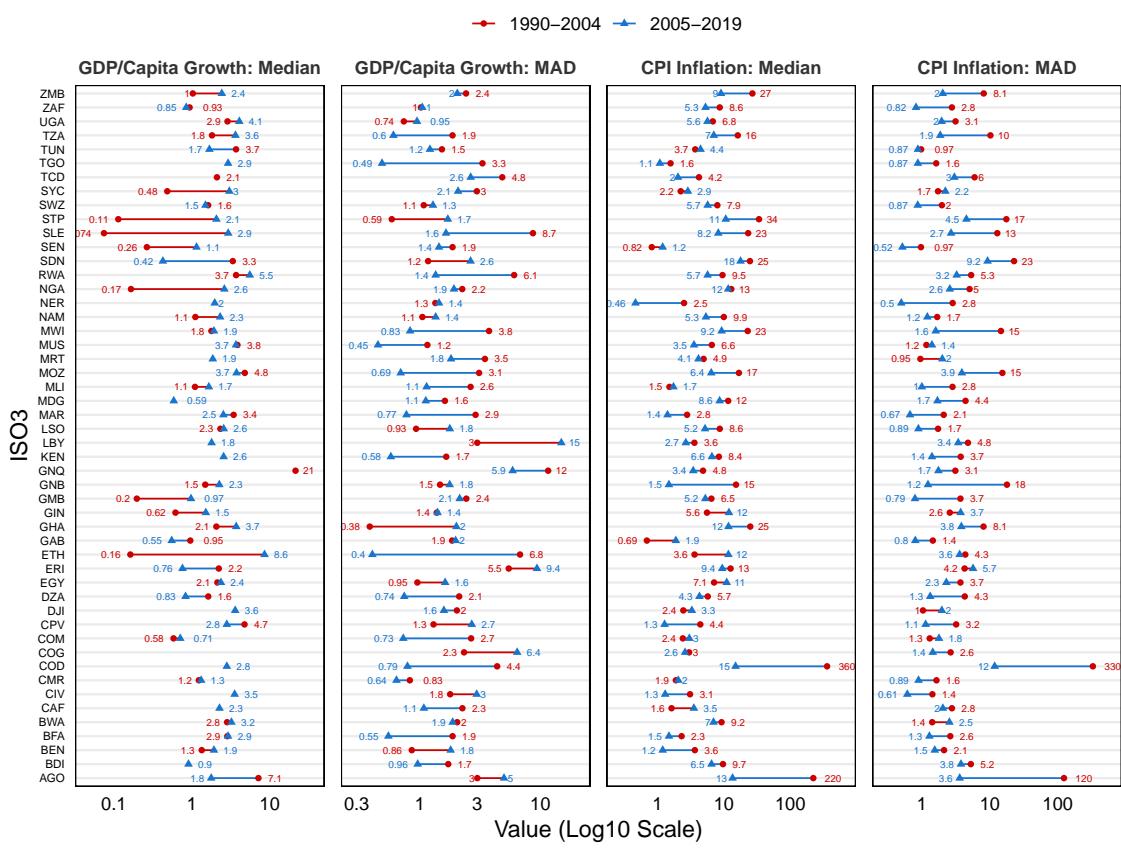


Table C1: Aggregate Volatility of 51 African Countries: 1990-2019

ISO3	Country	Income	GDP Per Capita Growth			CPI Inflation		
			Median	MAD	IQR	Median	MAD	IQR
DZA	Algeria	Upper middle	1.124	1.197	2.359	4.642	1.740	4.925
AGO	Angola	Lower middle	1.851	5.291	8.479	30.269	22.229	190.916
BEN	Benin	Low	1.422	1.246	2.320	2.140	1.766	3.978
BWA	Botswana	Upper middle	3.006	2.100	3.908	8.067	2.109	3.876
BFA	Burkina Faso	Low	2.883	1.476	2.601	1.804	2.025	3.812
BDI	Burundi	Low	-0.274	1.547	3.068	8.294	4.474	8.363
CPV	Cabo Verde	Lower middle	4.092	1.609	4.714	2.944	2.050	4.551
CMR	Cameroon	Lower middle	1.271	0.774	1.481	2.022	0.987	1.988
CAF	Central Afr. Rep.	Low	0.843	1.753	3.991	2.782	1.993	3.787
TCD	Chad	Low	-0.071	3.422	6.621	3.930	4.615	6.987
COM	Comoros	Lower middle	0.618	1.576	2.925	2.704	1.683	3.214
CIV	Côte d'Ivoire	Lower middle	-0.691	3.404	7.570	2.298	1.523	3.067
COD	Dem. Rep. o. Congo	Low	0.185	3.322	8.901	27.230	25.014	326.361
DJI	Djibouti	Lower middle	1.520	2.586	4.819	2.629	1.366	2.887
EGY	Egypt	Lower middle	2.133	1.258	2.389	9.727	3.186	5.367
GNQ	Equatorial Guinea	Upper middle	5.262	13.878	25.209	4.380	2.134	4.118
ERI	Eritrea	Low	1.457	7.647	12.841	10.288	5.501	10.631
SWZ	Eswatini	Lower middle	1.506	1.146	2.197	7.469	1.818	3.302
ETH	Ethiopia	Low	7.154	2.375	7.976	9.024	5.614	11.314
GAB	Gabon	Upper middle	0.664	1.940	4.110	1.448	1.220	2.249
GHA	Ghana	Lower middle	2.333	1.063	1.915	15.291	4.902	13.191
GIN	Guinea	Low	1.345	1.280	2.548	9.592	5.204	11.816
GNB	Guinea-Bissau	Low	1.493	1.402	2.710	3.277	3.886	13.204
KEN	Kenya	Lower middle	1.292	1.464	3.397	7.324	2.230	6.072
LSO	Lesotho	Lower middle	2.326	1.100	2.503	7.043	2.010	3.880
LBY	Libya	Upper middle	-0.577	6.512	14.369	3.122	3.704	7.868
MDG	Madagascar	Low	0.433	1.383	2.551	9.100	2.824	5.332
MWI	Malawi	Low	1.863	2.294	4.394	10.460	2.828	15.386
MLI	Mali	Low	1.493	1.412	3.018	1.563	2.617	5.526
MRT	Mauritania	Lower middle	1.682	2.670	5.467	4.715	1.547	2.772
MUS	Mauritius	Upper middle	3.702	0.666	1.326	5.164	1.936	3.644
MAR	Morocco	Lower middle	2.743	1.542	2.917	1.576	0.939	2.312
MOZ	Mozambique	Low	3.935	2.257	4.526	12.531	8.445	13.257
NAM	Namibia	Upper middle	1.782	1.756	3.442	6.727	2.590	4.673
NER	Niger	Low	-0.120	2.424	4.382	0.952	1.821	2.887
NGA	Nigeria	Lower middle	1.521	2.463	4.909	11.837	3.253	5.849
COG	Republic of Congo	Lower middle	-1.317	4.648	7.723	2.790	1.860	3.586
RWA	Rwanda	Low	5.427	1.942	5.033	6.374	3.907	7.907
SEN	Senegal	Lower middle	1.053	1.842	3.601	1.082	0.894	1.857
SYC	Seychelles	High	2.909	3.665	6.798	2.630	1.846	3.433
SLE	Sierra Leone	Low	1.415	2.604	5.344	13.312	7.259	16.186
ZAF	South Africa	Upper middle	0.911	1.079	2.289	5.980	1.368	3.713
SDN	Sudan	Lower middle	2.325	1.822	4.635	20.161	13.356	38.351
STP	São Tomé & Príncipe	Lower middle	0.658	1.103	2.032	13.830	6.478	23.908
TZA	Tanzania	Low	3.143	0.939	2.023	7.561	3.312	10.897
GMB	The Gambia	Low	0.585	2.546	5.186	5.306	1.743	2.736
TGO	Togo	Low	1.532	2.112	6.524	1.348	1.226	3.446
TUN	Tunisia	Lower middle	2.490	1.505	2.825	4.092	1.104	2.029
UGA	Uganda	Low	3.340	1.094	2.025	5.970	2.446	7.148
ZMB	Zambia	Lower middle	1.696	2.585	4.367	18.147	9.063	17.567
ZWE	Zimbabwe	Lower middle	-0.840	4.289	10.358	0.641	5.470	12.797

Notes: Excluding Liberia, Somalia, and South Sudan. Data Source: IMF WEO October 2021.

Figure C2: The Great Moderation by Country: 50 African Countries



Data Source: IMF World Economic Outlook, October 2021

## Section 2: Aggregate Relationships and Trends

Figure C3: Volatility in Africa Over Time

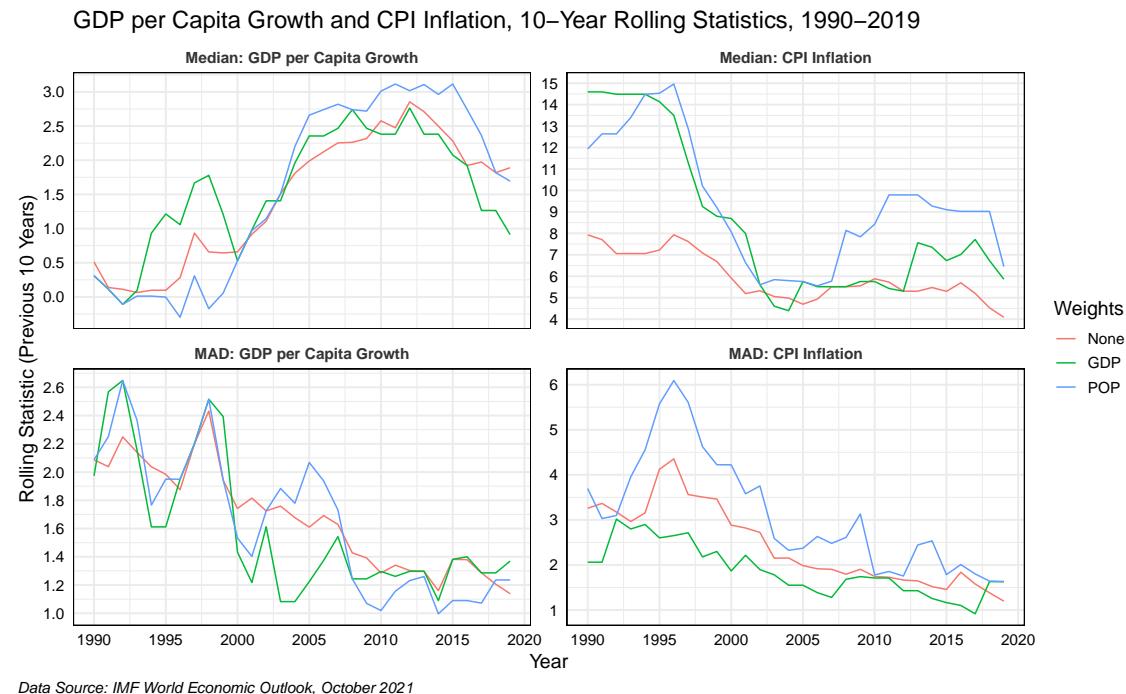


Figure C4: Figure 1 with World Bank WDI Data

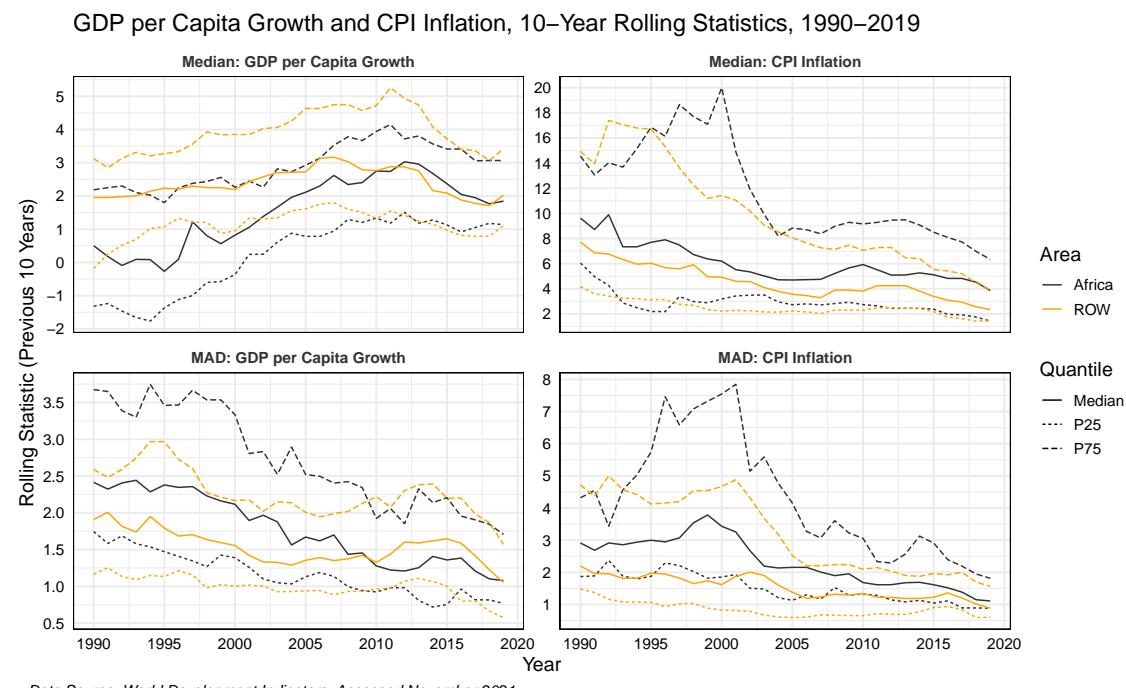
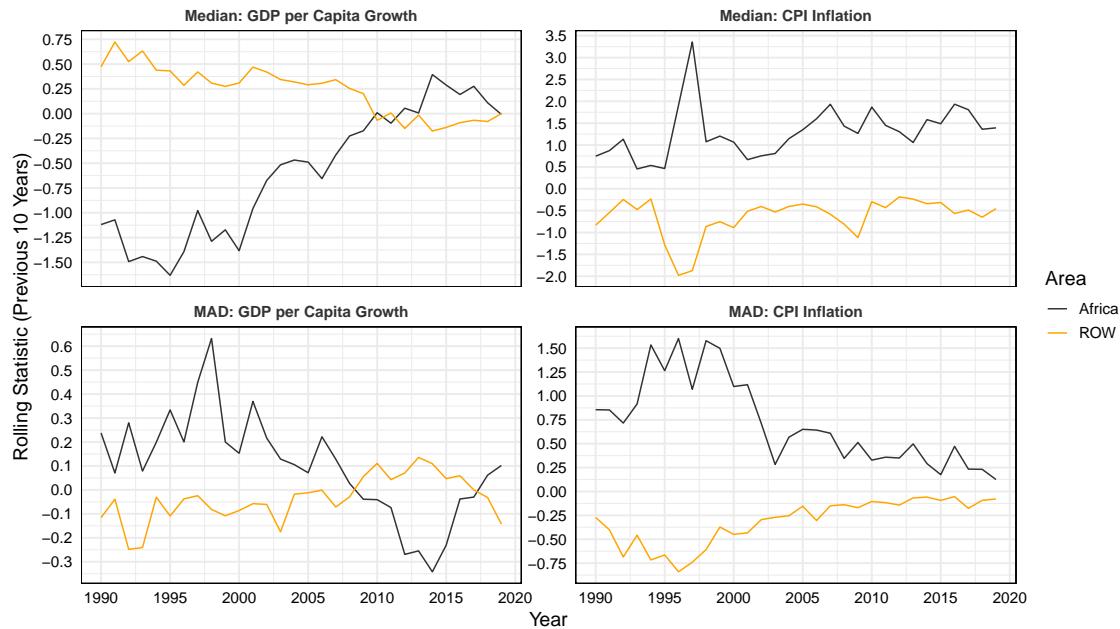


Figure C5: Figure 1 with Time-Medians Subtracted from Rolling Statistics

IMF WEO Data

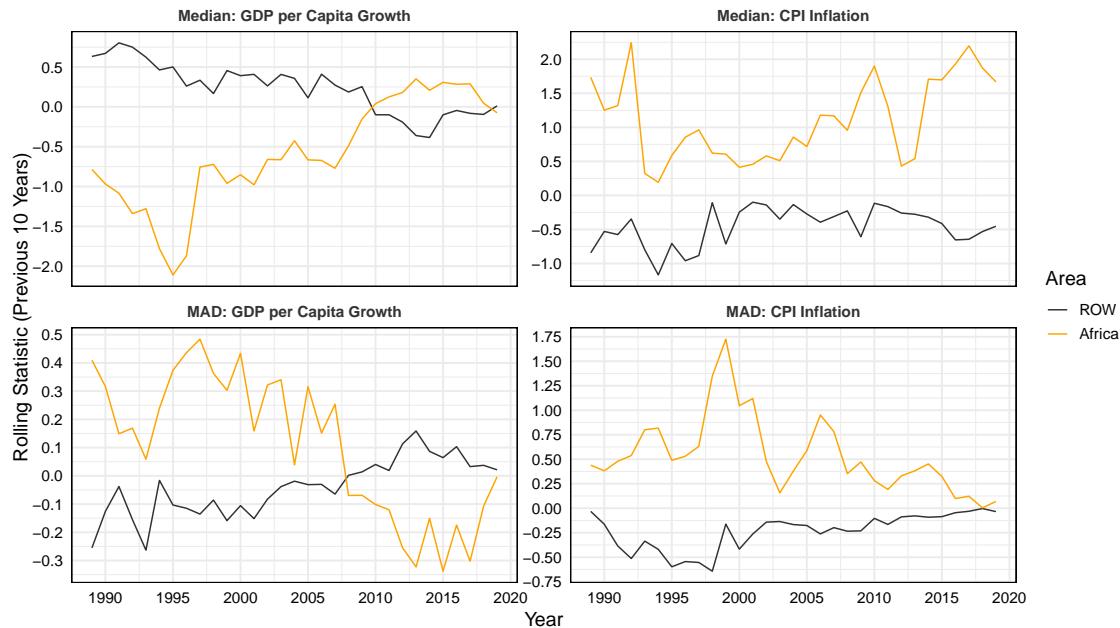
## GDP per Capita Growth and CPI Inflation, 10-Year Rolling Statistics, 1990–2019



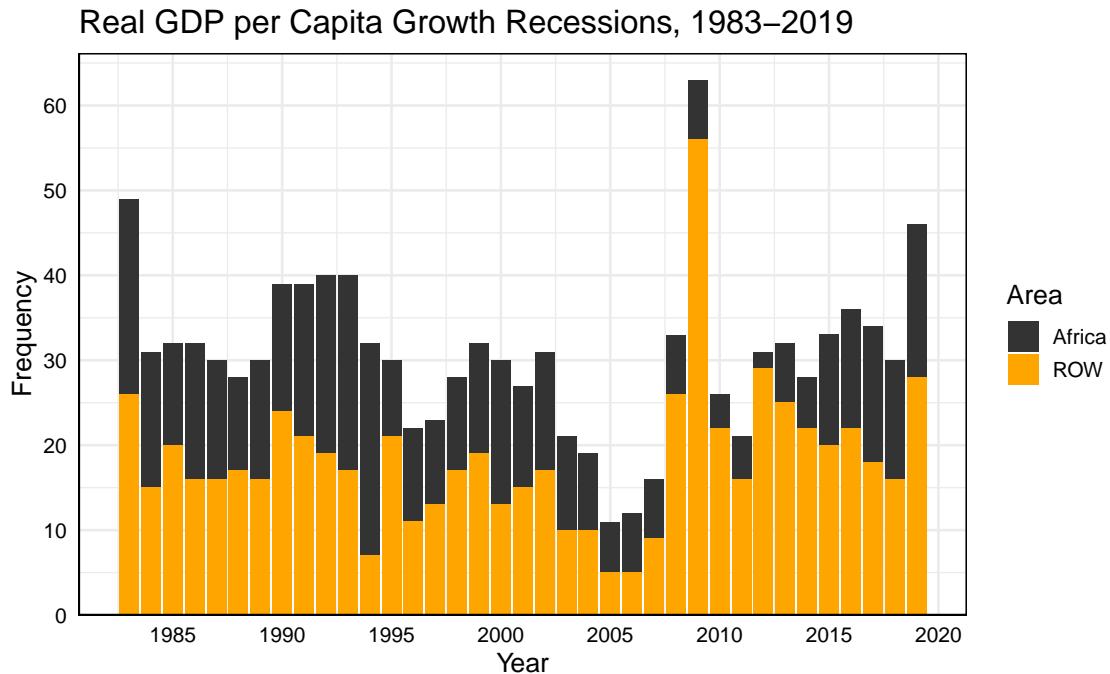
Data Source: IMF World Economic Outlook, October 2021

World Bank WDI Data

## GDP per Capita Growth and CPI Inflation, 10-Year Rolling Statistics, 1990–2019



Source: World Development Indicators, 2021

Figure C6: Growth Recessions Following [Syed et al. \(2017\)](#)

Data Source: IMF World Economic Outlook, October 2021

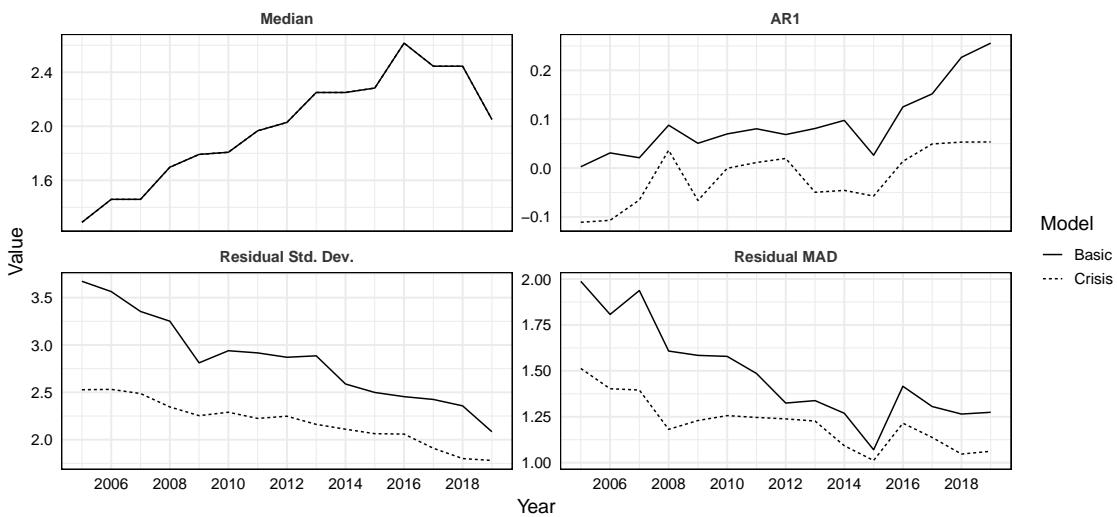
Figure C7: AR1 Analysis à la [Blanchard & Simon \(2001\)](#) with World Bank Data

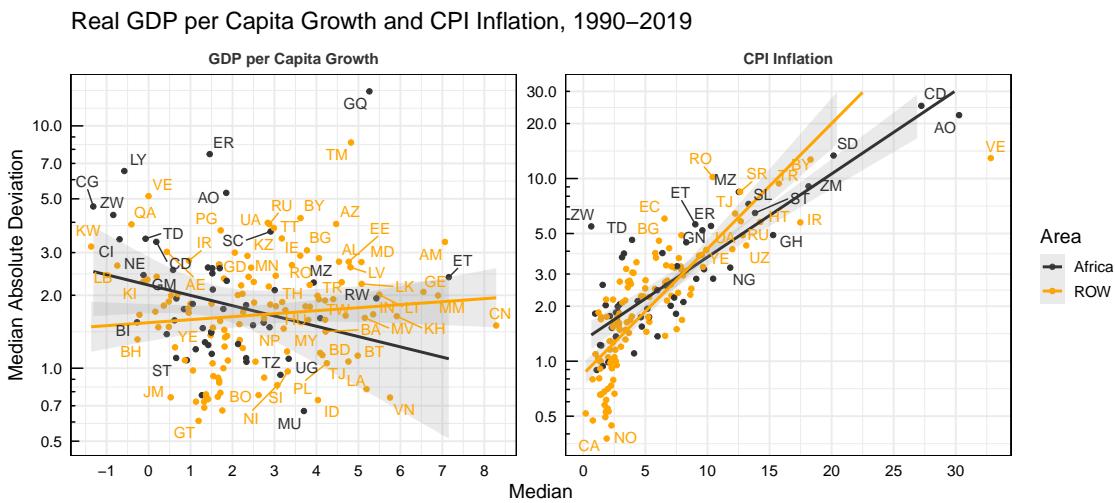
Table C2: Real Per Capita Growth and Inflation Performance in Africa, 1990-2019

Area	N	Per Capita Growth			Inflation		
		Median	MAD	IQR	Median	MAD	IQR
<b>Africa</b>	<b>51</b>	<b>1.506</b>	<b>1.822</b>	<b>3.991</b>	<b>5.306</b>	<b>2.230</b>	<b>4.925</b>
Low income	21	1.457	1.942	4.382	6.374	3.312	7.907
Lower middle income	21	1.521	1.609	3.601	4.715	2.010	3.880
Upper middle income	8	1.453	1.848	3.675	4.903	2.023	3.997
High income (SYC)	1	2.909	3.665	6.798	2.630	1.846	3.433
<b>ROW</b>	<b>124</b>	<b>2.370</b>	<b>1.724</b>	<b>3.532</b>	<b>3.549</b>	<b>1.754</b>	<b>4.318</b>
Low income	5	2.755	1.172	2.642	12.007	4.102	10.610
Lower middle income	24	2.985	1.654	3.085	6.996	2.733	5.856
Upper middle income	43	2.752	2.173	4.207	5.015	2.581	6.057
High income	52	1.758	1.616	3.178	2.285	1.024	2.112

Data Source: IMF WEO, October 2021. Real GDP per capita growth is calculated using the constant national currency series (NGDPRPC), and inflation is based on average national consumer price indices (PCPIPCH).

Notes: Statistics are calculated at the country-level, and aggregated across countries using the median. Countries with < 20 obs. for growth or inflation in 1990-2019 were excluded - in Africa Liberia, Somalia and South Sudan.

Figure C8: Empirical Relationship Between Levels and Volatilities



Data Source: IMF World Economic Outlook, October 2021

Table C3: Output and Inflation Volatility

Area	N	GDP/Capita			Inflation		
		$\beta$	P( $\beta \neq 0$ )	R <sup>2</sup>	$\beta$	P( $\beta \neq 0$ )	R <sup>2</sup>
<b>Africa</b>	<b>51</b>	<b>-0.187</b>	<b>0.035</b>	<b>0.083</b>	<b>0.375</b>	<b>&lt;0.001</b>	<b>0.596</b>
Low income	21	-0.048	0.596	0.016	0.396	<0.001	0.608
Lower middle income	21	-0.569	0.002	0.389	0.252	<0.001	0.810
Upper middle income	8	-0.837	0.130	0.370	0.058	0.689	0.026
<b>ROW</b>	<b>124</b>	<b>0.043</b>	<b>0.322</b>	<b>0.008</b>	<b>0.361</b>	<b>&lt;0.001</b>	<b>0.820</b>
Low income	5	-0.076	0.228	0.444	0.365	0.246	0.427
Lower middle income	24	-0.105	0.341	0.041	0.252	<0.001	0.632
Upper middle income	43	0.068	0.343	0.022	0.383	<0.001	0.800
High income	52	0.125	0.055	0.066	0.427	<0.001	0.478

Data Source: IMF WEO, October 2021. See also footnote to Table C2.

Note: A regression of the medians on the MADs of the country-series is run using a robust MM estimator following Koller & Stahel (2011). Available in R package *robustbase* (Maechler et al., 2021).

Figure C9: Empirical Relationship Between Levels and Volatilities in Africa

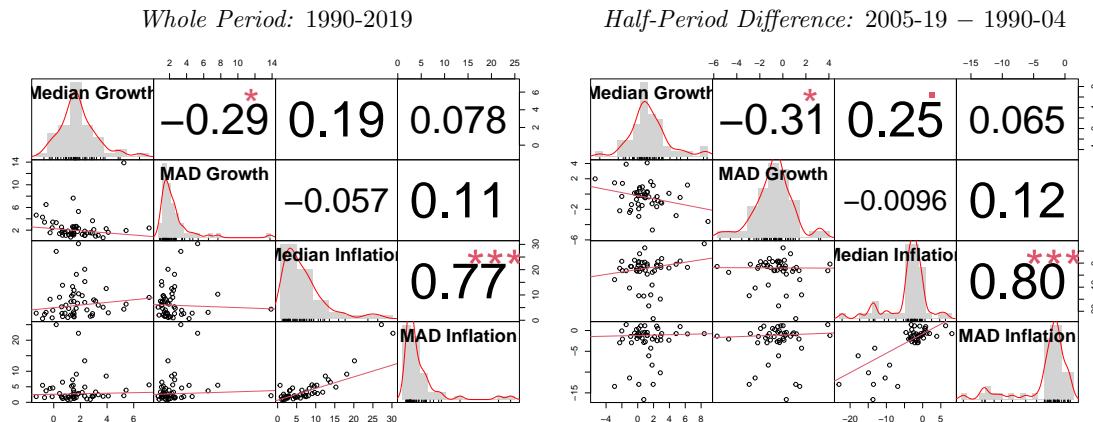


Table C4: Variation (MAD) Between Countries Over Time by Income Group

Area	Period	Per Capita Growth			Inflation		
		N	Median	MAD	IQR	Median	MAD
<b>Africa</b>	<b>1990-04</b>	<b>50</b>	<b>1.166</b>	<b>0.834</b>	<b>2.122</b>	<b>3.868</b>	<b>1.568</b>
	<b>2005-19</b>	<b>50</b>	<b>0.806</b>	<b>0.586</b>	<b>0.986</b>	<b>3.175</b>	<b>0.876</b>
Low income	1990-04	21	1.119	1.049	2.069	4.783	1.612
	2005-19	21	0.742	0.427	0.549	3.639	1.051
Lower middle income	1990-04	20	1.342	0.670	1.605	3.405	1.340
	2005-19	20	0.564	0.492	1.138	3.457	0.733
Upper middle income	1990-04	8	0.948	0.845	1.402	2.522	0.843
	2005-19	8	0.878	0.719	1.997	1.307	0.465
<b>ROW</b>	<b>1990-04</b>	<b>118</b>	<b>1.141</b>	<b>0.585</b>	<b>1.233</b>	<b>3.245</b>	<b>1.245</b>
	<b>2005-19</b>	<b>119</b>	<b>1.144</b>	<b>0.683</b>	<b>1.235</b>	<b>1.375</b>	<b>0.511</b>
Low income	1990-04	4	0.691	0.675	1.137	5.708	2.747
	2005-19	4	1.941	0.551	0.849	1.244	1.212
Lower middle income	1990-04	23	1.247	0.401	1.043	2.105	2.237
	2005-19	23	1.921	0.426	1.203	1.421	0.581
Upper middle income	1990-04	39	1.382	0.783	1.708	5.367	2.796
	2005-19	40	1.535	0.618	0.986	1.642	0.580
High income	1990-04	52	0.866	0.402	0.928	0.869	0.386
	2005-19	52	0.914	0.613	1.344	0.530	0.329

Data Source: IMF WEO, October 2021. Real GDP per capita growth is calculated using the constant national currency series (NGDPRPC), and inflation is based on average national consumer price indices (PCPIPCH).

Notes: Statistics calculated at country-level and aggregated across countries using the MAD. Countries with < 9 obs. for growth or inflation in 1990-04 or 2005-19 were excluded, in Africa Liberia, Somalia, South Sudan, and Zimbabwe.

Table C5: International Synchronization of Growth/Inflation Rates: World Bank Data

Period	Per Capita Growth				Inflation			
	Africa	ROW	World	Corr	Africa	ROW	World	Corr
1990-2004	0.199	0.221	0.202	0.245	0.386	0.404	0.364	0.327
2005-2019	0.252	0.388	0.321	0.275	0.415	0.516	0.476	0.400
Overall (1990-2019)	0.162	0.234	0.191	0.189	0.374	0.395	0.356	0.298

Data Source: IMF WEO, October 2021. See also note to Table 1.

Notes: The numbers under 'Africa', 'ROW' and 'World' are the share of the first eigenvalue in the sum of eigenvalues, computed from a pairwise Pearson's correlation matrix of the country-series. They estimate the share of an international business cycle in the joint variance of the data. The 'Corr' column reports the average absolute correlation between African and ROW countries series and measures alignment between Africa and ROW.

### Section 3: Decomposing Output Volatility

Figure C10: Production side GDP Shares

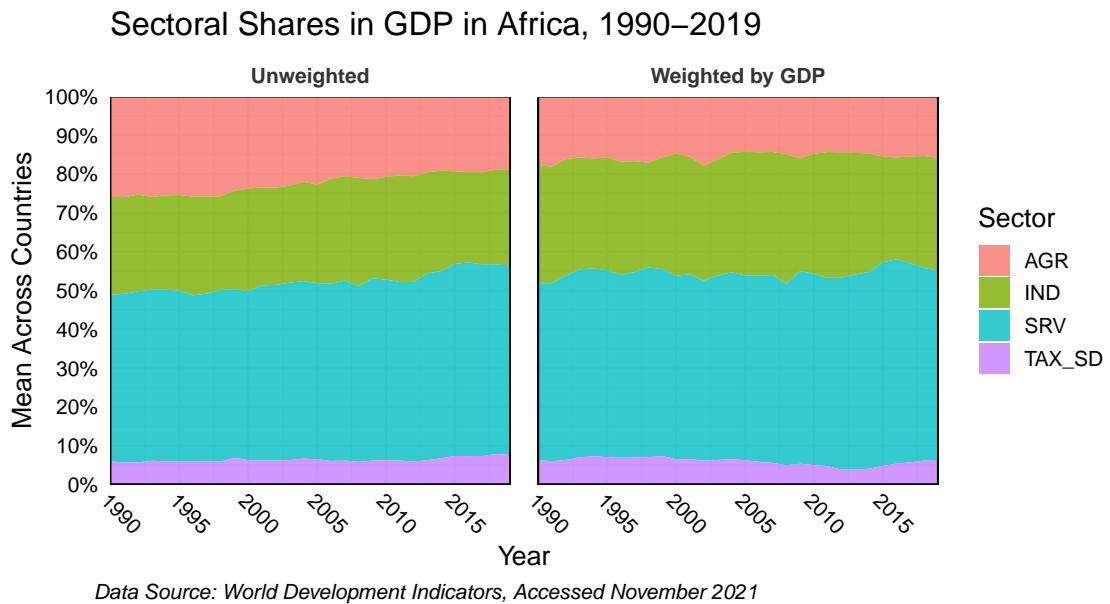


Figure C11: Production side GDP Growth Shares

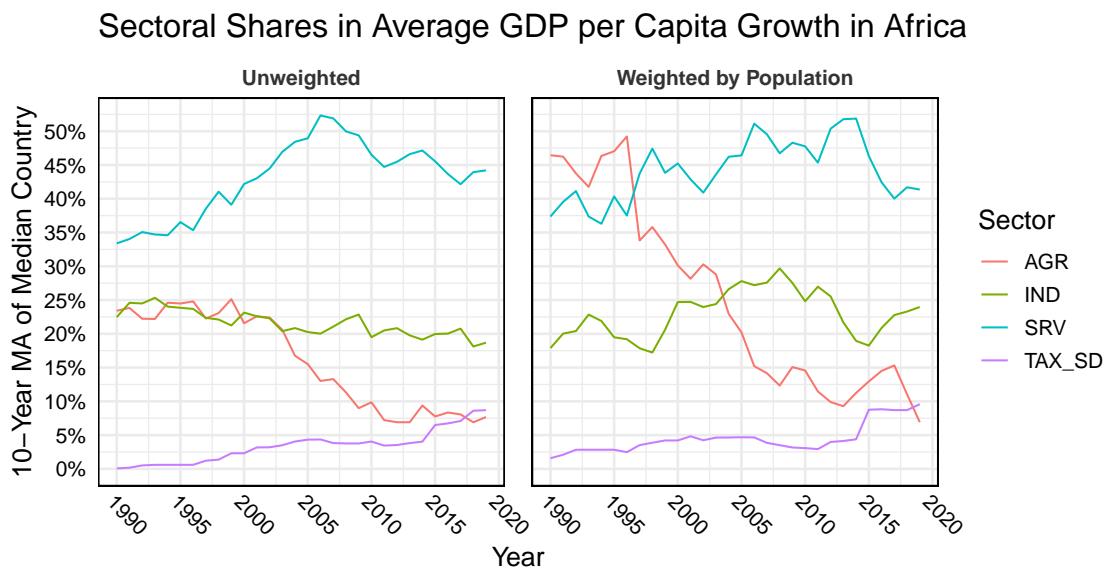


Table C6: Sectoral Volatility and Contribution to Aggregate Volatility, 1990-2019

Data	Sector:	AGR	IND	SRV	AGR	IND	SRV
Sector Share ( $\bar{\theta}_k$ ):		0.232	0.276	0.492	0.226	0.276	0.494
	Covariance:	<b>Classical</b>			<b>Robust (SDE)</b>		
Sector	AGR	126.95			69.01		
Growth	IND	-12.77	124.60		-5.67	64.73	
( $\Delta VA/VA_{t-1}$ )	SRV	-2.64	-9.86	52.93	-1.34	-1.78	25.90
Sector	AGR	6.40			2.14		
Contribution	IND	-0.50	5.79		-0.30	2.85	
( $\Delta VA/GDP_{t-1}$ )	SRV	-0.21	-0.66	8.44	-0.06	-0.38	5.39

*Notes:* Since sectoral growth rates can be very volatile, I employ both a classical (Pearson) and robust covariance estimator with a high breakdown point (0.5) based on Stahel (1981) and Donoho (1982). The choice of methods was informed by Maronna et al. (2018) and available implementations in various R packages. The Stahel-Donoho robust covariance estimator is implemented by the package *rrcov* (Todorov & Filzmoser, 2009). Covariance terms are aggregated across countries using the median, whereas sectoral shares are aggregated with the mean. Average shares for each country are computed using all but the first observation following Eq. 3. The shares reported above "Robust" are computed by taking the median share for each country, and aggregating across countries using the mean.

Figure C12: Rolling Covariances of Sectoral Growth Rates/Contribution

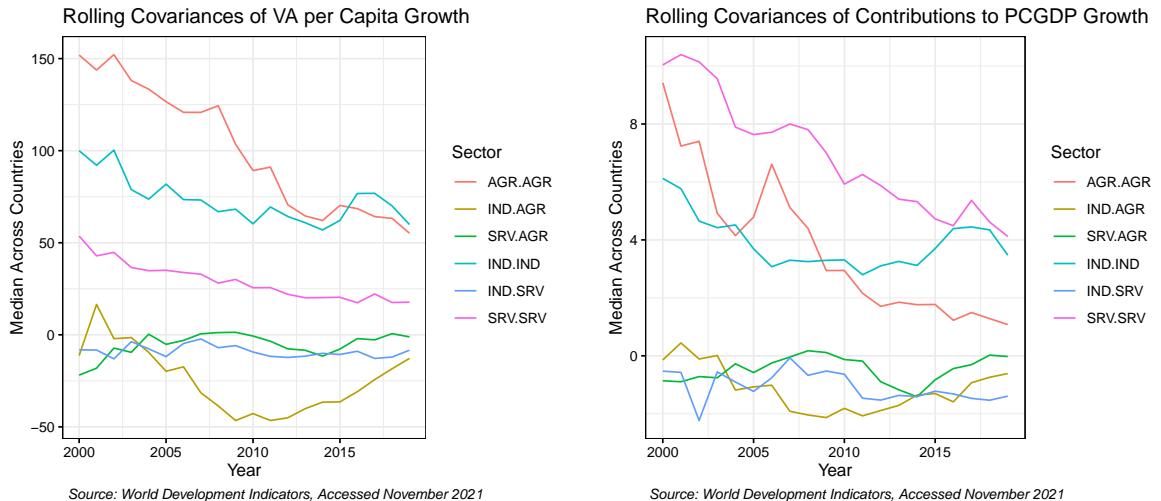


Figure C13: Production side GDP Shares: ETD Data

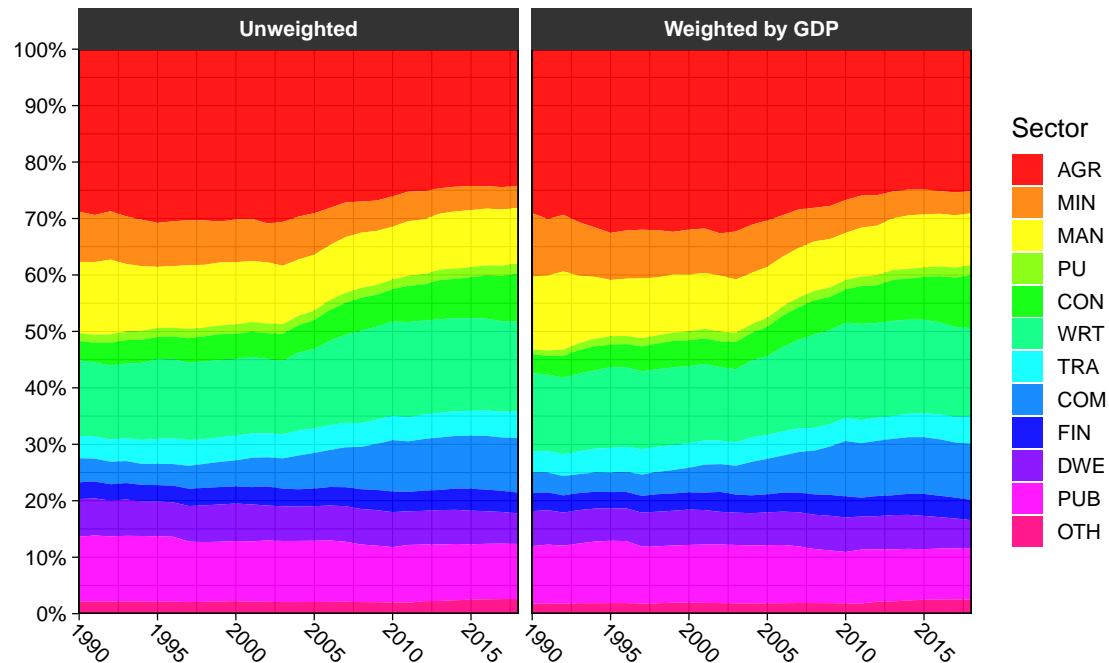


Figure C14: Sector Volatility and Contribution to Aggregate Volatility

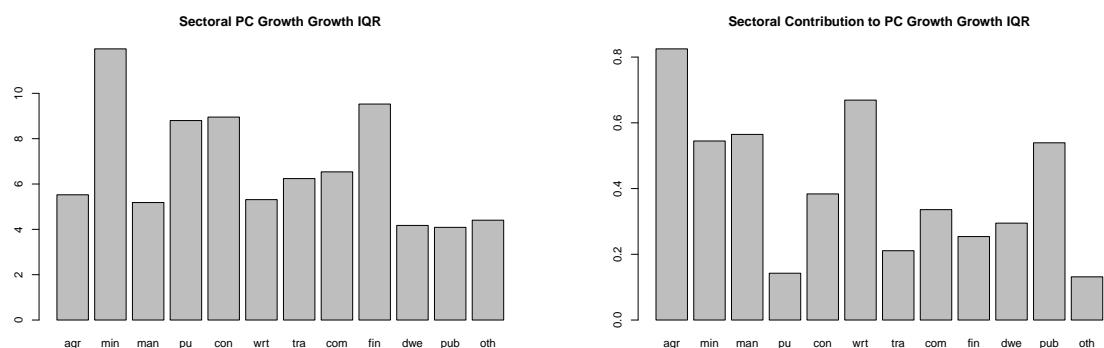


Figure C15: Rolling MADs of Sectoral Growth Rates/Contribution

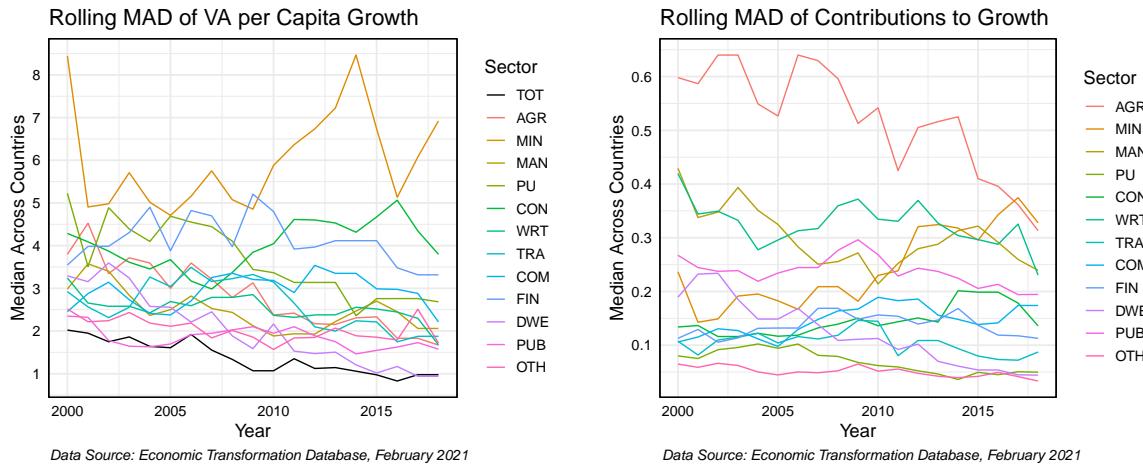


Table C7: Country Classification by Largest Sectoral Volatility

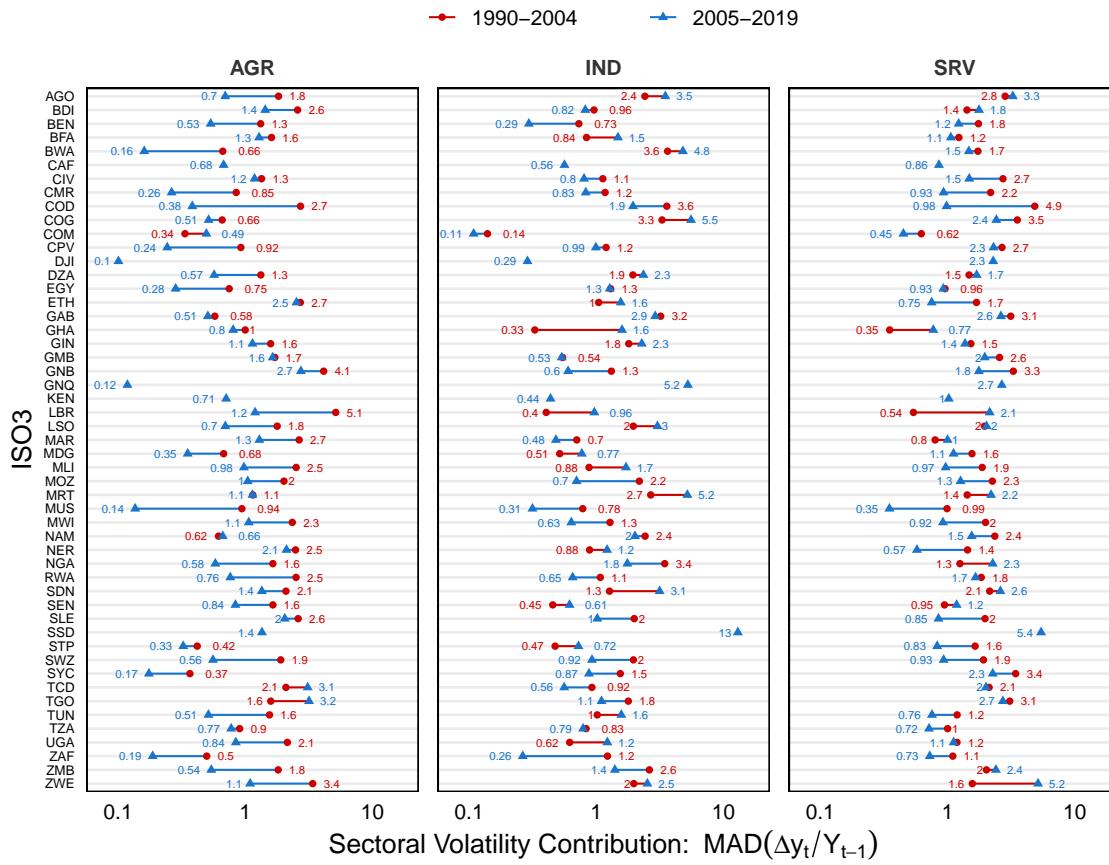
Metric	AGR (13)	IND (17)	SRV (21)
Sector Volatility Contribution: MAD( $\Delta y_t / Y_{t-1}$ )	BDI, BFA, ETH, GNB, LBR, MAR, MLI, NER, SEN, SLE, TCD, UGA, ZWE	AGO, BWA, COD, COG, DZA, EGY, GAB, GIN, GNQ, LSO, MRT, MUS, NGA, SSD, SWZ, TUN, TZA	BEN, CAF, CIV, CMR, COM, CPV, DJI, GHA, GMB, KEN, MDG, MOZ, MWI, NAM, RWA, SDN, STP, SYC, TGO, ZAF, ZMB
Metric	AGR (22)	IND (28)	SRV (1)
Sector Growth Volatility: MAD(% $\Delta y_t$ )	AGO, BFA, CAF, CMR, COM, CPV, DJI, DZA, GHA, GIN, GMB, GNB, KEN, MAR, MUS, SEN, SWZ, SYC, TUN, ZAF, ZMB, ZWE	BDI, BEN, BWA, CIV, COD, COG, EGY, ETH, GNQ, LBR, LSO, MDG, MLI, MOZ, MRT, MWI, NAM, NER, NGA, RWA, SDN, SLE, SSD, STP, TCD, TGO, TZA, UGA	GAB

Table C8: Aggregate Sectoral Growth Stabilization

Period:	1990-2019			1990-2004			2005-2019		
Sector:	AGR	IND	SRV	AGR	IND	SRV	AGR	IND	SRV
<i>Statistic: Median Across Countries (and Periods)</i>									
MAD( $\Delta y_t / Y_{t-1}$ )	1.06	1.17	1.57	1.63	1.21	1.74	0.71	1.02	1.38
MAD(% $\Delta y_t$ )	5.69	5.46	3.14	6.56	5.74	3.95	5.05	4.75	2.65
<i>Share of Countries Above the 1990-2019 Cross-Country-Period Median</i>									
MAD( $\Delta y_t / Y_{t-1}$ )	0.57	0.49	0.51	0.61	0.47	0.51	0.33	0.47	0.43
MAD(% $\Delta y_t$ )	0.47	0.47	0.61	0.49	0.51	0.57	0.45	0.43	0.37

*Note:* The 1990-2019 statistics are medians across country-level MADs for both the 1990-2004 and 2005-2019 periods. This more accurately reflects the median volatility between these two periods, since country-level MADs calculated over the entire 1990-2019 period are much closer to the 2005-2019 MADs.

Figure C16: Sectoral Volatility Contribution by Country

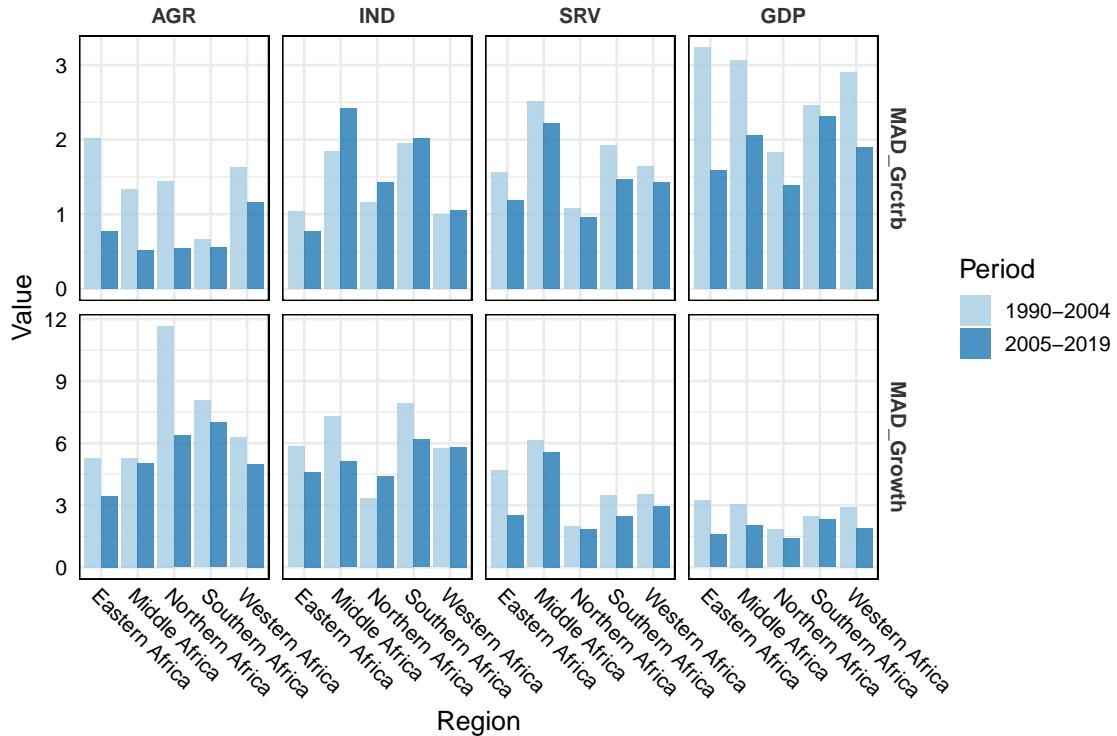


Data Source: World Development Indicators, Accessed November 2021

Table C9: Regions in Africa (51 Countries with Sectoral Data)

Region	Countries ISO3
Eastern Africa	BDI, COM, DJI, ETH, KEN, MDG, MOZ, MUS, MWI, RWA, SSD, SYC, TZA, UGA, ZMB, ZWE
Middle Africa	AGO, CAF, CMR, COD, COG, GAB, GNQ, SDN, STP, TCD
Northern Africa	DZA, EGY, MAR, TUN
Southern Africa	BWA, LSO, NAM, SWZ, ZAF
Western Africa	BEN, BFA, CIV, CPV, GHA, GIN, GMB, GNB, LBR, MLI, MRT, NER, NGA, SEN, SLE, TGO

Figure C17: Sectoral Growth Risk by Region



Data Source: World Development Indicators, Accessed November 2021

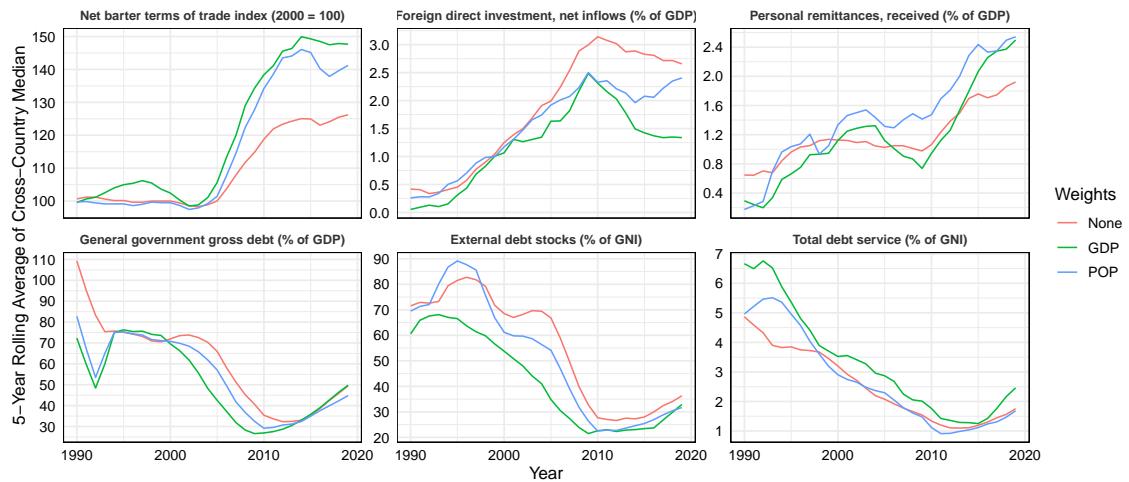
Table C10: Sectoral Growth Stabilization By Region

Region	Period	N	(1) MAD( $\Delta y_t / Y_{t-1}$ )			(2) MAD(% $\Delta y_t$ )		
			AGR	IND	SRV	AGR	IND	SRV
Eastern	1990-2004	13	2.01	1.04	1.57	5.28	5.84	4.69
Eastern	2005-2019	16	0.77	0.78	1.19	3.45	4.62	2.54
Middle	1990-2004	8	1.34	1.84	2.51	5.28	7.30	6.12
Middle	2005-2019	10	0.51	2.42	2.21	5.04	5.11	5.57
Northern	1990-2004	4	1.44	1.15	1.07	11.65	3.36	1.98
Northern	2005-2019	4	0.54	1.42	0.96	6.37	4.40	1.85
Southern	1990-2004	5	0.66	1.95	1.92	8.06	7.94	3.50
Southern	2005-2019	5	0.56	2.01	1.47	7.00	6.18	2.48
Western	1990-2004	16	1.63	1.00	1.64	6.28	5.74	3.55
Western	2005-2019	16	1.16	1.06	1.43	5.01	5.80	2.98

Note: Statistics were aggregated across countries using the median.

## Section 4: External, Financial, and Policy Factors

Figure C18: External Environment: Selected Indicators  
Terms of Trade, FDI, Remittances and Debt in Africa, 1990–2019



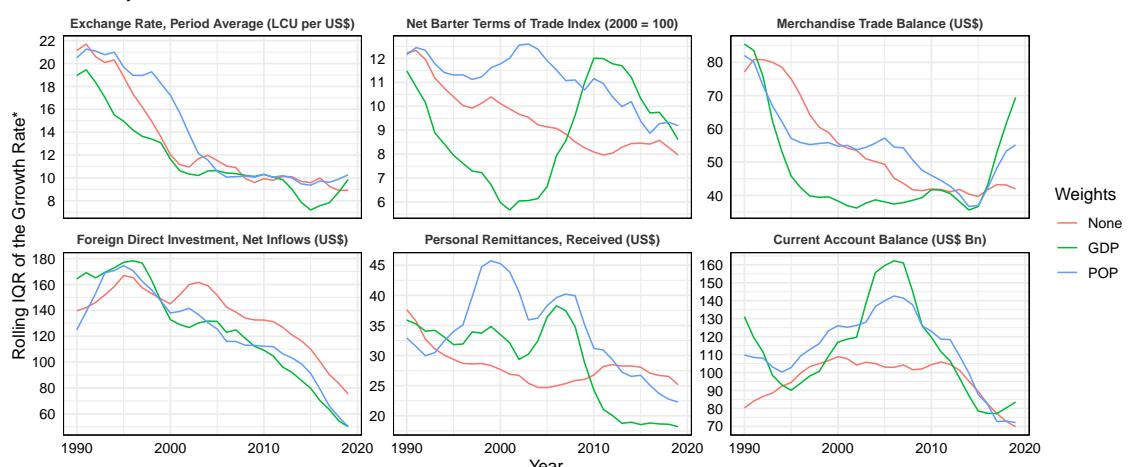
Data Source: IMF and World Bank. Accessed through the africamonitor API.

Table C11: Correlations with External Environment Indicators

Mean:	ToT	FDI	REM	GGDT	EDT	EDS
Median PC Growth	.065*	.177*	.053	-.271*	-.208*	-.045
MAD PC Growth	-.116*	-.052	-.074*	.091*	.104*	.039
Median Inflation	-.057*	-.039	-.083*	.117*	.187*	.084*
MAD Inflation	-.077*	-.031	-.017	.098*	.067*	.035

Notes: A 10-year MA with data from 1981 is used to smooth the variables shown in Figure C18 (in % of GDP/GNI terms), and 10-year rolling medians and MADs for per-capita growth and inflation. These rolling series are then standardized within each country, and first-differenced. Pairwise Pearson's correlations are computed on these first differences across all countries. A star denotes significance at the 5% level.

Figure C19: External Environment Volatility: Selected Indicators  
Volatility of the Growth Rate of Selected External Variables in Africa, 1990–2019



Data Source: IMF and World Bank. Accessed through the africamonitor API.

\*Note: Plots show a 5-year MA of the cross-country (weighted) median of a 10-year rolling IQR of the growth rate of the series.

Table C12: Correlations with External Environment Volatility Indicators

MAD:	E_PA	ToT	TB	FDI	REM	CAB
Median PC Growth	-.173*	-.202*	-.116*	-.034	-.115*	.000
MAD PC Growth	.043	.263*	.093*	.206*	.265*	.153*
Median Inflation	.911*	.299*	.058	.134*	.243*	.048
MAD Inflation	.915*	.295*	.032	.132*	.321*	.060*

Notes: 10-year rolling medians and MADs of the growth rates of the data from 1981 are computed for each country and related through pairwise Pearson's correlations across all countries. A star denotes significance at the 5% level.

Figure C20: Reserves and Financial Depth: Selected Indicators

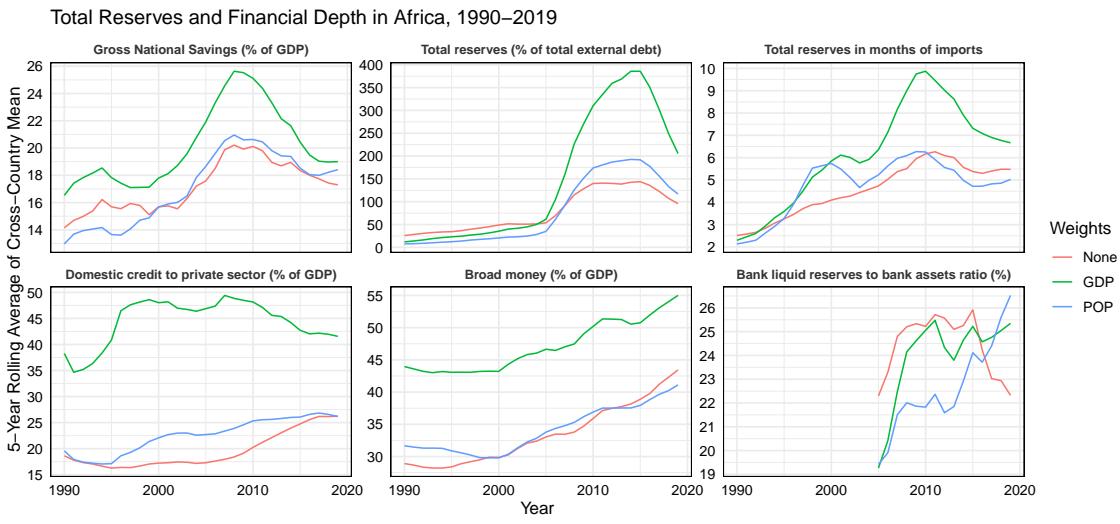


Table C13: Correlations with Financial Indicators

Mean:	GNS	TR_EDT	TR_MIM	PSC	BM	BLR_A
Median PC Growth	.073*	.249*	.093*	.086*	.072*	-.017
MAD PC Growth	-.024	-.077*	.060	-.106*	-.078*	.035
Median Inflation	-.056*	-.160*	-.050	-.098*	-.090*	-.011
MAD Inflation	-.004	-.035	.006	-.074*	-.078*	-.126*

Notes: A 10-year MA with data from 1981 is used to smooth the variables shown in Figure C20, and 10-year rolling medians and MADs for per-capita growth and inflation. These rolling series are then standardized within each country, and first-differenced. Pairwise Pearson's correlations are computed on these first differences across all countries. A star denotes significance at the 5% level.

Figure C21: Inflation Targeting in Africa

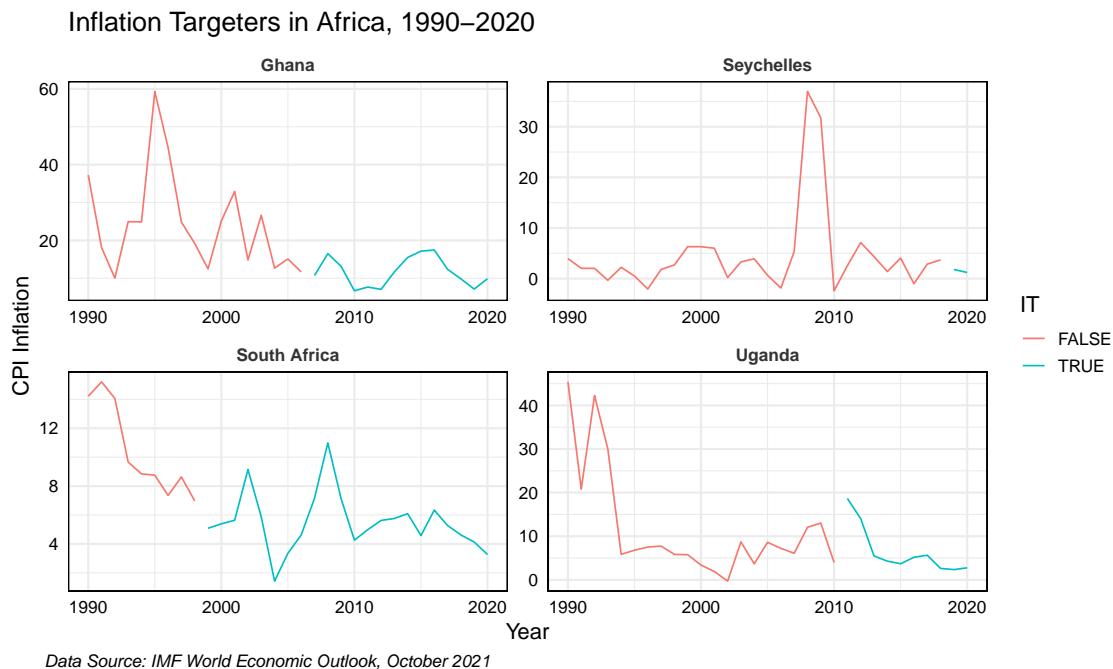
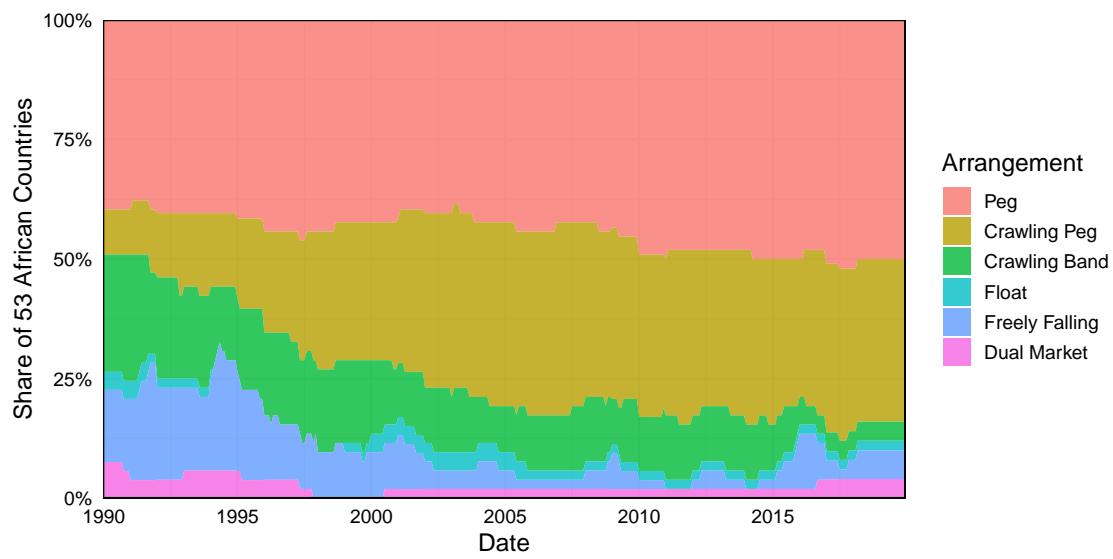


Figure C22: Exchange Rate Regimes in Africa, 1990-2019



*Notes:* The figure shows the 'coarse' exchange rate regime classification from Ilzetzki et al. (2019) with 6 categories. The share of 53 African economies (excl. South Sudan) with different regimes is computed for each year from 1990–2019.

Table C14: Exchange Rate 15-Year Rolling Panel-Dummy-Regressions, 1990-2019

Dependent Variable:	MAD Real GDP/Capita Growth (%)			MAD Inflation (%)		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variables</i>						
Crawling Peg	-0.5017*** (0.1213)	-1.030** (0.4212)	-0.8594* (0.4743)	-0.2853 (0.4360)	5.676 (10.57)	6.576 (10.68)
Crawling Band	-0.7126** (0.2486)	0.2733 (0.6678)	-0.0654 (0.5721)	-0.5408 (2.155)	19.72** (8.871)	18.62* (9.523)
Float	-1.227*** (0.1489)	1.325 (1.067)	0.7436 (0.7344)	-1.756 (1.161)	46.70* (22.06)	46.38* (21.93)
FF + DM	0.3907*** (0.1044)	1.288*** (0.3280)	0.4496 (0.2781)	46.42** (17.21)	117.5*** (26.13)	114.7*** (27.07)
<i>Fixed-effects</i>						
Country	—	52	52	—	52	52
Year	—	—	15	—	—	15
<i>Fit statistics</i>						
Observations	751	751	751	759	759	759
R <sup>2</sup>	0.026	0.733	0.743	0.198	0.474	0.477
Within R <sup>2</sup>	0.030	0.010	—	0.254	0.220	—

Driscoll & Kraay (1998) (L=1) standard-errors in parentheses      Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1  
 Avg. Country Group Sizes: Peg: 22.8, Crawling Peg: 17, Crawling Band: 6.4, Float: 1, FF: 2.9, DM: 0.9

Notes: 15-year MAs of the exchange regime dummies on data from 1990-2019 (retaining 15 observations per country)  
 are regressed onto 15-year rolling MAs of GDP per capita growth and CPI inflation. Data from WEO, Oct. 21.

Figure C23: Macroprudential Measures in Africa

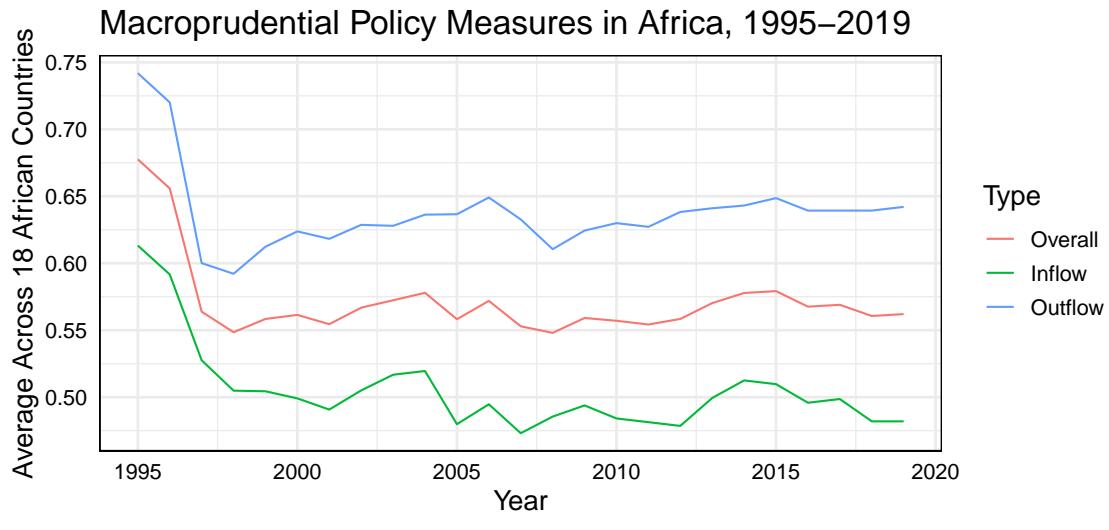


Figure C24: Disaggregated Macroprudential Measures in Africa

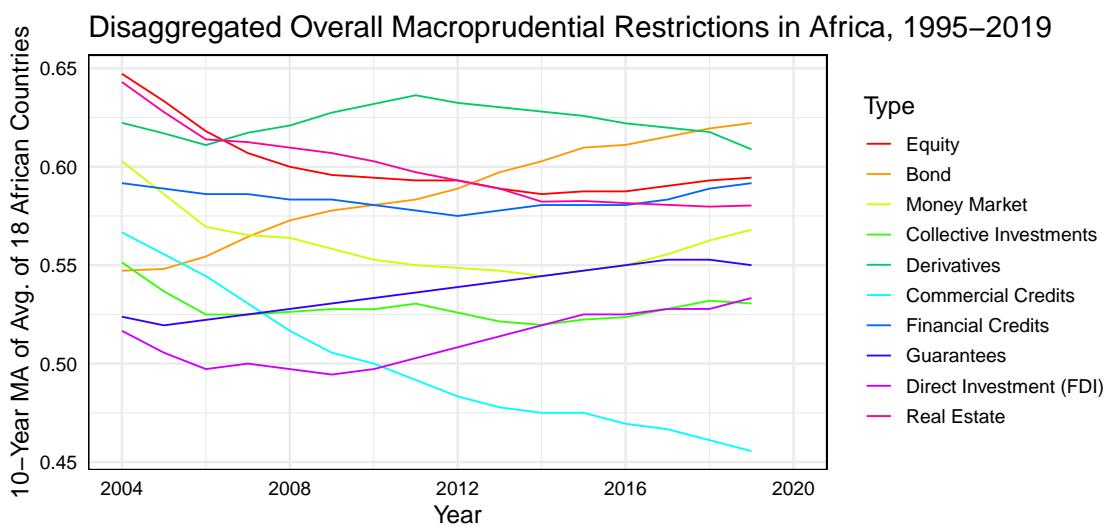


Table C15: Macroprudential Policy: 10-Year Rolling Panel-Regressions, 1995-2019

Dependent Variables:	MAD Real GDP/Capita Growth (%)		MAD Inflation (%)			
Model:	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variables</i>						
Overall Measures	0.1724 (0.1264)	-5.095*** (0.7261)	-5.368*** (0.7969)	2.497 (1.966)	-10.23* (4.871)	-12.17*** (3.644)
R <sup>2</sup>	0.004	0.512	0.602	0.007	0.344	0.388
Within R <sup>2</sup>		0.153	0.196		0.004	0.007
Inflow Measures	0.3058*** (0.0654)	-0.5994 (0.7604)	-2.215*** (0.5578)	9.942* (5.142)	-7.188*** (2.314)	-21.25*** (2.018)
Outflow Measures	-0.0717 (0.1187)	-4.096*** (1.034)	-3.050*** (0.9308)	-5.008** (1.978)	-3.473 (3.277)	5.768 (4.398)
R <sup>2</sup>	0.006	0.520	0.603	0.035	0.344	0.392
Within R <sup>2</sup>		0.167	0.197		0.004	0.013
<i>Fixed-effects</i>						
Country	—	18	18	—	18	18
Year	—	—	16	—	—	16
Observations	288	288	288	287	287	287

Driscoll &amp; Kraay (1998) (L=2) standard-errors in parentheses

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Notes: 10-year rolling MADs of GDP per capita growth and CPI Inflation from the WEO Oct. 21 are regressed onto 10-year MADs of overall, inflow and outflow measures taken from the macroprudential database of Fernández et al. (2016) (August 2021 update) and available for 18 African economies: Algeria, Angola, Burkina Faso, Côte d'Ivoire, Egypt, Ethiopia, Ghana, Kenya, Kingdom of Eswatini, Mauritius, Morocco, Nigeria, South Africa, Tanzania, Togo, Tunisia, Uganda, and Zambia.

Figure C25: The Adoption of Fiscal Rules in Africa

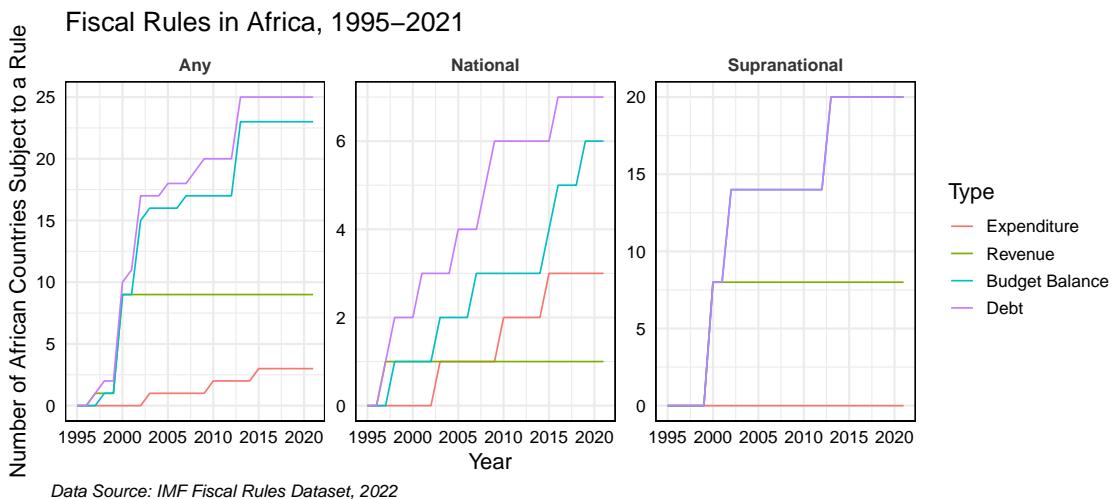


Figure C26: Important Macroeconomic and Fiscal Aggregates, 1990-2019

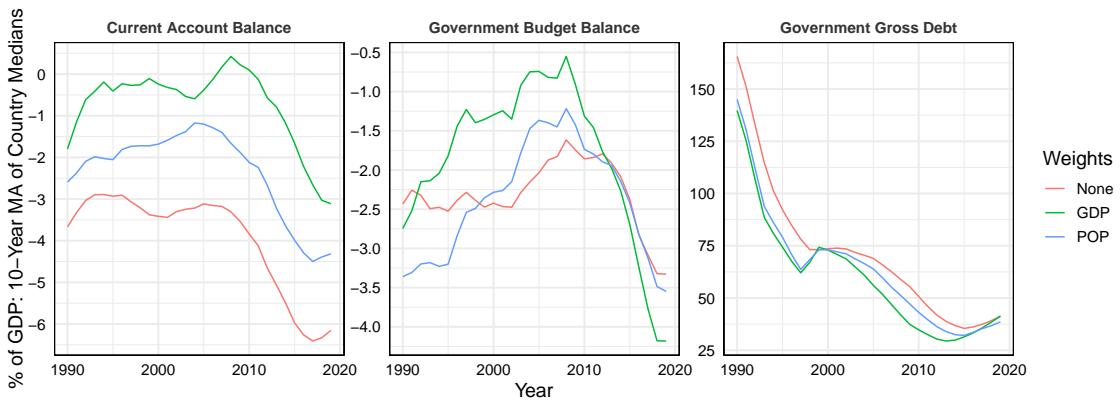


Table C16: Fiscal Rules: 10-Year Rolling Panel-Regressions with Data from 1990-2019

Dependent Variables:	MAD Real GDP/Capita Growth (%)			MAD Inflation (%)		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Any Rule	-0.4450 (0.3045)	-0.8644*** (0.1213)	0.7599* (0.3743)	-1.832*** (0.2925)	-1.886*** (0.2921)	0.5127 (0.4207)
R <sup>2</sup>	0.008	0.710	0.732	0.148	0.448	0.504
Within R <sup>2</sup>		0.070	0.011		0.163	0.003
N. Rules	-0.2174** (0.0998)	-0.3574*** (0.0463)	0.2266** (0.0831)	-0.6190*** (0.1196)	-0.8204*** (0.1126)	-0.0546 (0.1158)
R <sup>2</sup>	0.015	0.710	0.731	0.123	0.461	0.503
Within R <sup>2</sup>		0.070	0.007		0.184	< 0.001
<i>Rule Dummies</i>						
ER	-1.012*** (0.0984)	0.3297* (0.1843)	0.4739** (0.2173)	-0.8629*** (0.2966)	0.2318 (0.1978)	0.4147 (0.2470)
RR	-1.115*** (0.0951)	0.4285** (0.1581)	0.5313*** (0.1674)	0.0267 (0.1617)	-1.890*** (0.1814)	-1.559*** (0.1607)
BBR	0.6600*** (0.1413)	-0.9849*** (0.1603)	-0.4113** (0.1843)	-0.2630 (0.1701)	-0.7664** (0.3150)	-0.1291 (0.3817)
DR	-0.2568 (0.2932)	-0.2262 (0.2273)	0.6215* (0.3534)	-1.497*** (0.2762)	-0.5366 (0.4100)	1.158* (0.5931)
R <sup>2</sup>	0.050	0.716	0.733	0.145	0.474	0.524
Within R <sup>2</sup>		0.088	0.016		0.203	0.042
<i>Fixed-effects</i>						
Country	—	25	25	—	25	25
Year	—	—	21	—	—	21
Observations	512	512	512	509	509	509

Driscoll & Kraay (1998) (L=2) standard-errors in parentheses *Signif. Codes:* \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

*Notes:* 10-year rolling MADS of GDP per capita growth and CPI Inflation from the WEO Oct. 21, are regressed onto 10-year MAs of fiscal rule dummies from Davoodi et al. (2022b). 26 African countries are recorded to have introduced fiscal rules since 1990 (see Table 7). 3 sets of regressions are run with an 'Any Rule' dummy indicating the presence of a fiscal rule, an ordinal 'N. Rules' variable, obtained as the sum of dummies for 4 different types of rules, and with the 4 dummies: expenditure (ER), revenue (RR), budget balance (BBR), and debt (DR).

Table C17: Fiscal Rules: Panel-Regression in 30-Year Panel with Data from 1990-2015

Dependent Variables:	Current Account Balance			Government Budget Balance			Government Gross Debt		
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Any Rule	0.3748 (1.011)	0.1935 (1.078)	1.155 (1.558)	7.645* (4.415)	8.740* (5.080)	3.671 (3.112)	-32.86*** (9.081)	-37.18*** (7.913)	-38.99*** (9.958)
R <sup>2</sup>	0.0003	0.260	0.306	0.013	0.143	0.201	0.057	0.434	0.506
Within R <sup>2</sup>	< 0.001	< 0.001	0.001		0.016	0.001		0.089	0.046
N. Rules	0.0977 (0.2358)	0.0135 (0.3457)	0.1347 (0.3715)	2.478 (1.529)	2.800 (1.749)	-0.5224 (0.9959)	-10.02*** (2.554)	-9.074*** (2.256)	-2.196 (1.959)
R <sup>2</sup>	< 0.001	0.260	0.306	0.010	0.138	0.200	0.039	0.396	0.483
Within R <sup>2</sup>	< 0.001	< 0.001	< 0.001		0.010	< 0.001		0.029	< 0.001
ER	4.603*** (1.078)	-6.162*** (2.061)	-5.470** (2.118)	-0.5070 (1.492)	-13.63** (6.335)	-11.48* (6.287)	-26.55*** (5.545)	43.71*** (7.603)	51.35*** (6.510)
RR	-0.2078 (1.662)	-3.274 (2.140)	-4.546* (2.283)	-1.307 (0.9910)	-17.77* (9.362)	-25.20** (11.35)	-1.093 (3.214)	28.14*** (9.968)	8.131 (4.862)
BBR	3.246** (1.577)	3.576*** (1.142)	4.238*** (1.270)	2.788** (1.291)	7.615 (5.273)	7.374 (4.772)	-7.840 (4.624)	45.23*** (9.988)	51.17*** (9.414)
DR	-3.138 (2.140)	-0.9916 (1.437)	0.1215 (1.592)	5.647 (3.375)	10.22* (5.305)	7.303* (3.979)	-18.55* (10.28)	-83.44*** (12.17)	-71.83*** (11.66)
R <sup>2</sup>	0.014	0.274	0.322	0.013	0.160	0.225	0.049	0.469	0.542
Within R <sup>2</sup>	0.018	0.024			0.035	0.032		0.145	0.115
<i>Fixed-effects</i>									
Country	—	26	30	—	26	30	—	26	30
Year	—	—	26	—	—	26	—	—	26
Observations	749	749	749	673	673	673	586	586	586

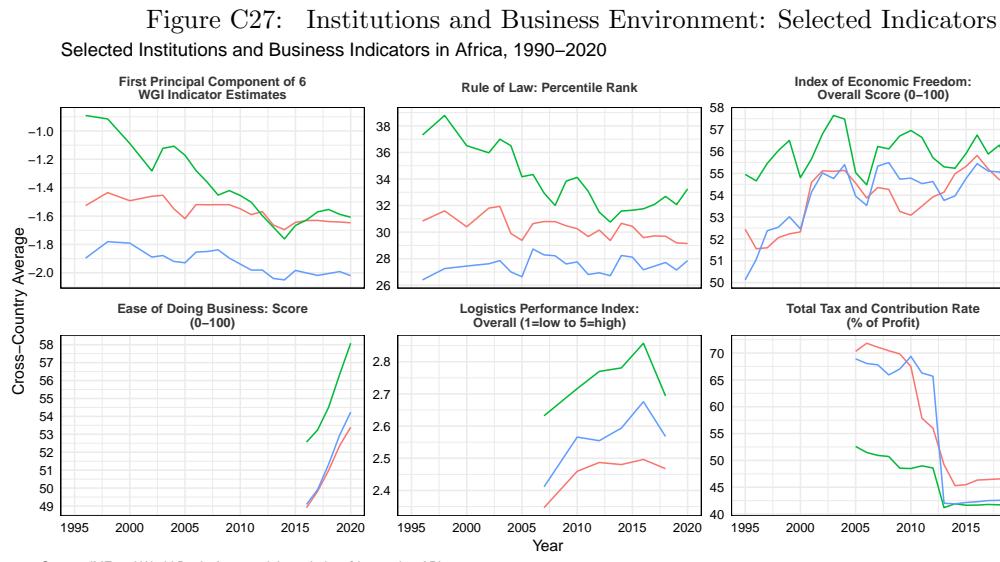
Driscoll & Kraay (1998) (L=2) standard-errors in parentheses, dependent variables in % of GDP. *Signif. Codes:* \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

## Section 5: Structural Factors

Table C18: Summary Statistics of Predictors in Cross-Sectional and Panel Analysis

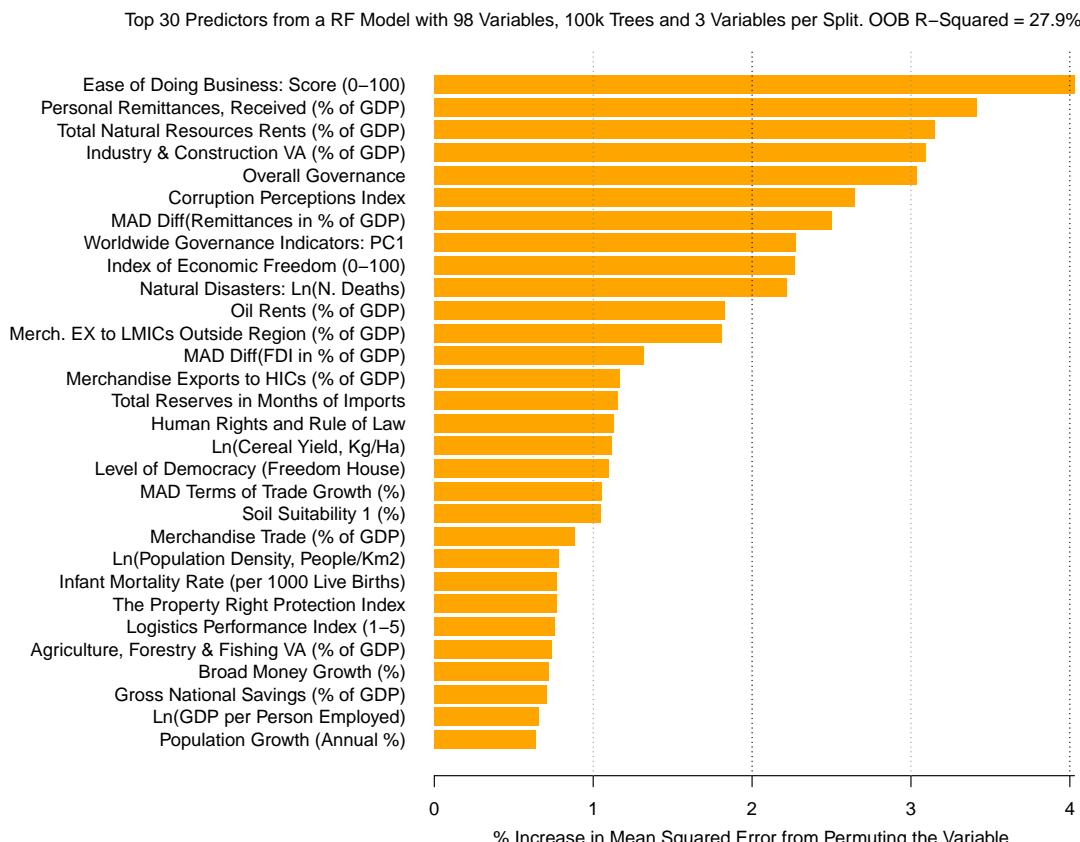
Panel	Topic / Variables	N	Ndist	Mean	Median	SD	Min	Max
<i>Institutions</i>								
	Overall Governance	49	49	49.67	48.57	12	26.35	78.06
X	Worldwide Governance Indicators: PC1	49	49	-1.47	-1.49	1.39	-3.97	1.88
X	Human Rights and Rule of Law	49	49	7	7.17	1.53	3.76	9.79
X	Level of Democracy (Freedom House)	49	37	5.27	5	2.64	0.75	10
X	50-Year Average Freedom House Ratings	48	29	2.44	2.5	0.5	1	3
X	Regime Durability	47	45	12.51	10.28	9.77	2	39
	Colonial Origin: British	49	2	0.39	0	0.49	0	1
	Colonial Origin: French	49	2	0.39	0	0.49	0	1
	Corruption Perceptions Index	48	44	33.77	32.69	11.35	17	62.25
<i>Business Environment</i>								
	Ease of Doing Business Score (0-100)	49	47	50.71	50.25	10.99	21.15	77.70
	Logistics Performance Index (1-5)	46	40	2.48	2.47	0.24	2.03	3.50
X	Index of Economic Freedom (0-100)	49	46	53.76	54.80	7.30	34.60	73
X	The Property Right Protection Index	49	49	49.79	50.00	2.12	46.49	54.38
<i>Production Shares</i>								
X	Agriculture, Forestry & Fishing VA (% of GDP)	49	49	20.89	22.11	13.18	1.33	51.19
X	Industry & Construction VA (% of GDP)	49	49	26.72	22.84	14.73	11.52	77.54
<i>Climate &amp; Agriculture</i>								
X	Permanent Cropland (% of Land Area)	49	49	4.01	0.74	7.66	0.002	40.62
X	Ln(Cereal Yield, Kg/Ha)	47	47	7.03	7.08	0.61	5.45	8.88
X	Annual Average Rainfall	49	49	82.27	82.98	52.86	2.78	206.82
X	Annual Average Temperature	49	49	24.31	24.48	3.34	12.67	28.87
	% 1995 Pop. in Tropics (Af+Am+Aw)	44	28	46.39	43.04	42.42	0	100
X	% of Cropland Equipped for Irrigation	47	46	7.75	2.11	15.79	0.05	99.81
	Irrigation Suitability 1 (%)	44	44	4.12	3.64	2.81	0.16	13.41
	Soil Suitability 1 (%)	44	44	9.37	7.43	7.79	0.15	32.01
<i>Trade Intensity and Composition</i>								
X	Merchandise Trade (% of GDP)	49	49	52.39	46.59	25.36	19.22	128.68
X	Agricultural Raw Materials Exports (% of GDP)	47	47	0.96	0.34	1.43	0	7.01
X	Manufactures Exports (% of GDP)	47	47	5.78	1.37	9.08	0.001	35.54
X	Ores and Metals Exports (% of GDP)	47	47	2.49	0.21	4.92	0.0003	22.52
X	Merchandise Exports to HICs (% of GDP)	49	49	13.59	9.90	12.18	0.61	53.02
X	Merch. EX to LMICs Outside Region (% of GDP)	49	49	3.02	1.94	3.87	0.05	17.34
X	Merchandise Imports from HICs (% of GDP)	49	49	15.07	12.76	9.23	3.69	53.41
<i>Trade Diversification</i>								
X	Herfindahl Index of Bilateral Trade (X+M)	48	48	0.13	0.09	0.10	0.04	0.47
X	Theil Index of Bilateral Trade (X+M)	48	48	2.03	1.87	0.45	1.39	3.29
X	Herfindahl Index of Exports by Product	47	47	0.30	0.27	0.24	0.03	0.92
X	Theil Index of Exports by Product	47	47	3.41	3.32	0.94	1.43	5.46
<i>Exchange Rate and Terms of Trade</i>								
X	Exchange Rate Growth (%)	49	35	5.32	2.90	7.37	-0.01	44.28
X	MAD Nominal Exchange Rate Depreciation (%)	49	37	10.36	7.94	9.17	0.91	63.23
X	Net Barter Terms of Trade Index (2000 = 100)	49	49	113.10	110.16	20.36	65.92	162.62
X	Terms of Trade Growth (%)	49	46	0.44	0	2.12	-4.37	9.00
X	MAD Terms of Trade Growth (%)	49	49	9.14	7.45	5.70	0.84	24.36
<i>Financial &amp; Aid Flows</i>								
X	Net FDI Inflows (% of GDP)	49	49	2.52	2.05	2.15	0.04	9.69
X	MAD Diff(FDI in % of GDP)	49	49	1.98	1.29	2.26	0.07	12.88
X	Personal Remittances, Received (% of GDP)	49	45	2.98	1.14	6.04	0	39.72
X	MAD Diff(Remittances in % of GDP)	49	45	0.46	0.21	0.75	0	4.77
X	Net ODA Received (% of GNI)	49	49	7.79	6.73	6.47	0.20	23.07
<i>Financial Sector</i>								
X	Broad Money (% of GDP)	49	49	33.79	23.17	25.87	10.11	132.60
X	Broad Money Growth (%)	49	49	8.05	8.05	3.46	-0.03	17.16
X	MAD Broad Money Growth (%)	49	49	14.36	13.26	5.47	2.90	33.26
X	Domestic Credit to Private Sector (% of GDP)	49	49	20.87	13.13	22.95	2.37	118.17
X	Bank Liquid Reserves to Bank Assets Ratio (%)	46	46	21.70	18.83	14.46	3.50	59.29
	Bank/MM Account (% of Population Ages 15+)	42	42	30.47	28.44	19.75	6.71	82.21

Panel	Topic / Variables	N	Ndist	Mean	Median	SD	Min	Max
<i>Debt &amp; Reserves</i>								
X	General Government Gross Debt (% of GDP)	48	48	56.24	52.95	31.71	12.92	198.71
X	External Debt Stocks (% of GNI)	45	45	62.83	55.79	35.70	12.71	196.62
X	Total Debt Service (% of GNI)	45	45	2.85	2.08	2.34	0.62	12.54
X	Total Reserves in Months of Imports	41	41	4.62	2.99	6.03	0.07	28.19
<i>Population</i>								
X	Ln(Population)	49	49	15.76	16.11	1.62	11.33	18.74
X	Population Growth (Annual %)	49	49	2.39	2.57	0.78	0.61	4.10
X	Urban Population (% of Total Population)	49	49	39.38	38.02	16.88	9.26	82.12
X	Ln(Population Density, People/Km <sup>2</sup> )	49	49	3.68	3.88	1.33	0.85	6.40
X	Age Dependency Ratio (% of Work. Age Pop.)	49	49	83.26	87.86	15.71	45.86	106.47
X	International Migrant Stock (% of Population)	49	49	3.13	2.26	3.46	0.15	15.55
<i>Health</i>								
X	Life Expectancy at Birth, Total (Years)	49	49	57.22	55.66	7.81	43.98	74.13
X	Infant Mortality Rate (per 1000 Live Births)	49	49	63.75	67.15	26.01	12.20	125.65
X	% of People using Basic Sanitation Services	49	49	38.27	31.45	26.67	5.81	98.17
	% Pop. at Risk of Malaria, 2005	44	14	75.21	100	40.01	0	100
	Malaria Ecology (Sachs, 2003)	47	47	10.16	7.51	8.53	0	31.55
<i>Education</i>								
X	Human Capital Index	49	49	0.53	0.55	0.17	0.16	0.82
X	Mean Years of Schooling	49	36	4.37	4.30	1.93	1.30	8.90
X	Expected Years of Schooling	49	39	9.12	9	2.60	3.70	15.40
X	Adult Literacy (% of People Ages 15+)	49	49	62.83	67.09	19.91	23.00	93.00
	% of Pop. Speaking Major European Language	48	11	3.37	0	12.88	0	70.00
<i>Natural Disasters &amp; Conflict</i>								
X	Natural Disasters: Ln(N. Homeless)	49	45	9.39	10.91	3.81	0	13.83
X	Natural Disasters: Ln(N. Deaths)	49	49	6.84	7.27	2.12	1.39	10.13
X	Natural Disasters: Ln(Damage in USD)	49	39	8.69	10.34	5.27	0	15.69
X	Ln(ACLED Fatalities, 1997-2019)	44	44	7.49	7.60	2.50	1.39	11.88
	Societal Violence Scale Index (1-5)	48	13	3.44	3.58	0.92	1.50	5
X	State Fragility Index	47	43	14.36	15.04	4.88	1.33	23.33
<i>Geography &amp; Accessibility</i>								
	Geogr. Predicted Trade (FR 1999)	48	48	-3.07	-3.08	0.50	-3.98	-2.16
	% Area 100km from Coast/Sea-Nav. River	44	32	20.61	12.09	26.19	0	100
	Sub-Saharan Africa Dummy	49	2	0.90	1	0.31	0	1
	Landlocked Dummy	48	2	0.27	0	0.45	0	1
	Internal Distance Based on Area	48	48	232.83	205.27	159.59	8.02	595.40
	Latitude in Degrees	48	47	2.52	4.85	17.41	-33.93	36.83
	Longitude in Degrees	48	47	15.08	14.12	20.75	-23.50	57.50
<i>Natural Resources</i>								
X	Total Natural Resources Rents (% of GDP)	49	49	10.88	7.37	10.58	0.01	42.04
X	Oil Rents (% of GDP)	49	20	4.48	0	10.16	0	40.52
<i>Poverty &amp; Inequality</i>								
X	% Poor at \$1.90 a Day (2011 PPP)	46	44	39.11	41	23.95	0.40	85.75
X	Poverty Gap at \$1.90 a Day (2011 PPP) (%)	46	45	15.96	15.03	12.46	0.10	51.70
X	Gini Index	46	44	43.37	41.68	7.87	31.45	63
<i>Religion &amp; Ethnicity</i>								
	Religion: Muslim, 1980	45	38	32.82	16.20	36.80	0	99.40
	Religion: Protestant, 1980	45	35	11.52	4.90	13.25	0	50
	Religion Fractionalization, 2000	49	49	0.47	0.58	0.27	0.003	0.86
	Ethnic Fractionalization, 2000	48	47	0.62	0.71	0.25	0	0.93
<i>Others</i>								
X	Index of Globalization	49	49	44.31	43.29	8.33	28.37	62.70
X	Human Development Index	49	49	0.49	0.48	0.12	0.30	0.76
	Ln(GDP per Capita 1960)	42	41	6.57	6.51	0.54	5.52	7.94
X	Ln(GDP per Person Employed)	47	47	9.16	8.96	1.04	7.51	11.16
X	Access to Electricity (% of Population)	49	49	41.00	34.75	29.66	5.05	99.40
X	Gross National Savings (% of GDP)	48	48	17.38	15.81	9.14	2.31	37.98



## Cross-Sectional Results

Figure C28: RF Predicting the MAD of PCGDP Growth of 49 African Economies in 1990-2019



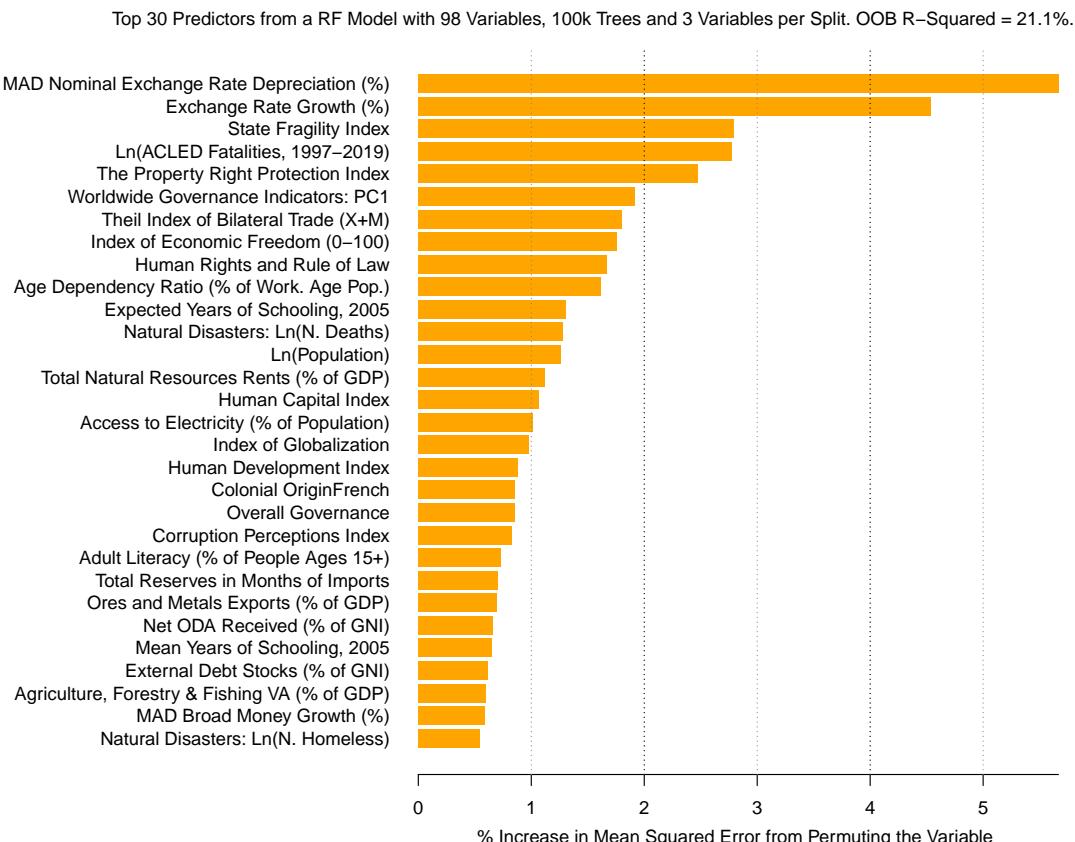
*Notes:* The Random Forest model is fit following Breiman (2001), using the *ranger* R package (Wright & Ziegler, 2017). Each tree is fit on a bootstrap sample of the data, randomly choosing 3 variables at each split. The Out-of-Bag (OOB)  $R^2$  is computed using each tree to predict only the data excluded from the bootstrap sample used to grow the tree, and averaging the predictions from all trees, yielding an OOB fit used to compute the  $R^2$ . Table C18 summarises the variables. 2.5% missing values were imputed beforehand following Stekhoven & Bühlmann (2012) with R package *missRanger*.

Table C19: RF Ranking of Indicator Topics: Predicting MAD PCGDP Growth, 1990-2019

Method: Topic	Permutation %ΔMSE	Exclusion %ΔMSE	Residual Fit %ΔMSE	Combined Avg. Rank
	Rank	Rank	Rank	
Institutions	88.03	1	6.38	4
Financial & Aid Flows	73.47	2	4.29	3
Trade Intensity and Composition	57.63	4	1.90	1
Financial Sector	65.98	3	0.93	2
Business Environment	50.67	5	2.83	10
Natural Resources	28.81	12	3.20	5
Natural Disasters & Conflict	29.17	11	1.91	9
Production Shares	31.11	9	2.22	14
Population	31.17	8	0.75	8
Exchange Rate and ToT	29.18	10	0.39	7
Climate & Agriculture	45.07	6	-0.02	15
Health	16.62	16	1.47	12
Others	28.06	13	-0.21	6
Geography & Accessibility	20.78	14	-0.06	11
Trade Diversification	20.78	15	-0.12	13
Debt & Reserves	32.33	7	-0.57	19
Education	13.00	17	1.38	18
Poverty & Inequality	7.27	19	1.41	17
Religion & Ethnicity	10.90	18	0.55	16

*Notes:* This table reports 3 methods to rank topics of predictors using Random Forests. *Permutation* permutes all predictors in a topic and calculates the percent increase in the Out-Of-Bag (OOB) MSE. *Exclusion* excludes all predictors from the topic, fits a new model, and obtains the percent increase in MSE to the full model. *Residual Fit* is like *Exclusion*, but additionally partials out the predictors from a topic using multivariate linear regression, and fits a new model using the residual predictors from all other topics.

Figure C29: RF Predicting the MAD of CPI Inflation of 49 African Economies in 1990-2019



*Notes:* See notes to Figure C28. The variables are summarised in Table C18.

Table C20: RF Ranking of Indicator Topics: Predicting MAD CPI Inflation, 1990-2019

Method: Topic	Permutation %ΔMSE	Exclusion Rank	Residual Fit %ΔMSE	Combined Avg. Rank
Exchange Rate and ToT	86.54	1	17.56	1
Natural Disasters & Conflict	72.25	2	5.16	2
Institutions	63.26	3	0.72	5
Business Environment	58.39	4	1.04	4
Population	44.53	5	-1.23	8
Trade Intensity and Composition	43.54	6	-2.22	13
Trade Diversification	38.31	9	1.20	3
Financial Sector	43.36	7	-4.69	19
Education	12.74	17	-0.91	6
Geography & Accessibility	33.51	10	-1.43	10
Others	38.37	8	-2.75	18
Poverty & Inequality	23.94	13	-2.31	15
Health	20.53	15	-2.31	14
Natural Resources	18.40	16	-1.57	11
Debt & Reserves	26.00	12	-1.98	12
Religion & Ethnicity	12.42	18	-1.39	9
Production Shares	9.84	19	-0.97	7
Financial & Aid Flows	22.82	14	-2.72	17
Climate & Agriculture	26.20	11	-2.51	16

Notes: See notes for Table C19 and explanations provided in the text.

## Results on Cross-Section of Differences

Figure C30: RF Predicting the MAD-Difference of PCGDP Growth of 49 African Economies

Top 30 Predictors from a RF Model with 70 Variables, 100k Trees and 3 Variables per Split. OOB R-Squared = -2.1%.

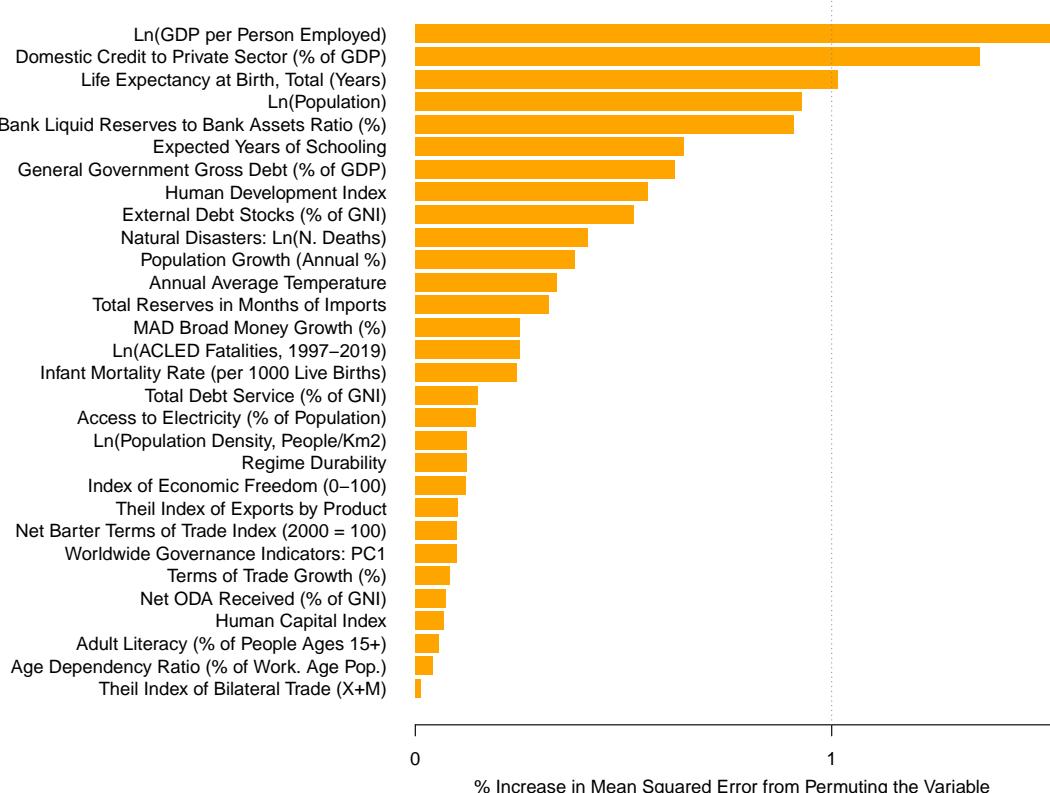


Table C21: RF Ranking of Indicator Topics: Predicting the MAD-Difference of PCGDP Growth

Method: Topic	Permutation %ΔMSE	Exclusion %ΔMSE	Residual Fit %ΔMSE	Combined Avg. Rank
Rank	Rank	Rank	Rank	
Others	63.47	3	2.29	1
Financial Sector	68.48	2	7.05	2.00
Population	44.62	4	0.37	3.33
Health	34.72	6	0.37	4.67
Institutions	23.60	14	0.47	7.33
Education	24.29	13	0.30	7.67
Trade Intensity and Composition	74.42	1	-1.25	8.00
Climate & Agriculture	30.06	8	-0.28	8.33
Debt & Reserves	28.69	9	-1.38	9.00
Trade Diversification	33.82	7	-0.89	10.00
Natural Disasters & Conflict	39.17	5	-0.09	10.67
Exchange Rate and ToT	27.29	10	-0.05	11.33
Financial & Aid Flows	24.47	12	-0.15	11.67
Poverty & Inequality	16.19	16	0.42	12.00
Natural Resources	25.97	11	-1.52	12.33
Business Environment	11.88	17	0.26	12.67
Production Shares	16.27	15	-0.92	14.33

Figure C31: RF Predicting the MAD-Difference of CPI Inflation of 49 African Economies

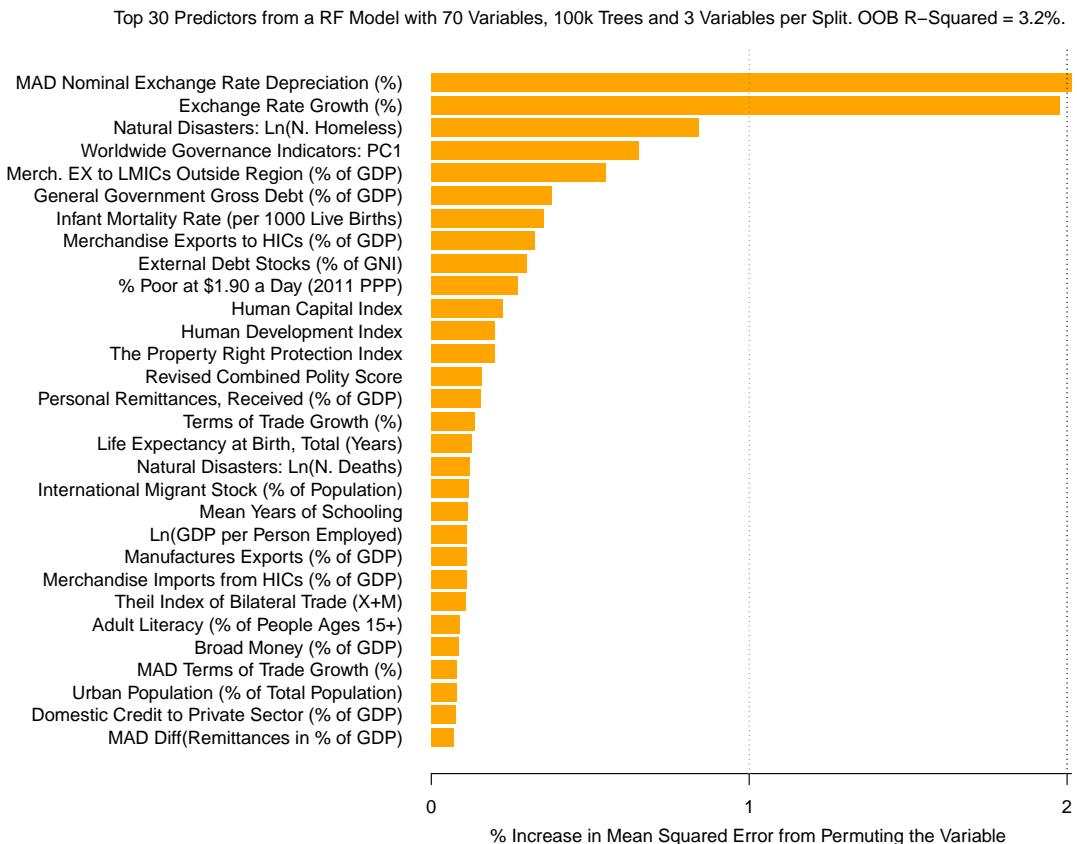


Table C22: RF Ranking of Indicator Topics: Predicting the MAD-Difference of CPI Inflation

<i>Method:</i> Topic	Permutation		Exclusion		Residual Fit		Combined
	%ΔMSE	Rank	%ΔMSE	Rank	%ΔMSE	Rank	Avg. Rank
Exchange Rate and ToT	62.34	2	7.87	1	36.45	1	1.33
Population	68.14	1	-0.01	7	2.10	3	3.67
Debt & Reserves	61.76	3	-1.12	15	2.89	2	6.67
Institutions	22.22	13	0.55	2	1.35	6	7.00
Natural Disasters & Conflict	37.33	9	0.26	4	0.79	9	7.33
Trade Intensity and Composition	39.68	7	0.29	3	-1.69	13	7.67
Production Shares	38.81	8	0.05	6	0.41	10	8.00
Education	45.36	4	-0.03	8	-1.80	15	9.00
Natural Resources	4.70	17	0.18	5	1.43	5	9.00
Climate & Agriculture	32.01	10	-0.84	14	1.25	7	10.33
Financial Sector	42.28	5	-0.05	9	-5.75	17	10.33
Health	25.41	11	-0.37	10	-0.32	11	10.67
Poverty & Inequality	19.66	14	-1.28	16	1.61	4	11.33
Trade Diversification	23.90	12	-0.56	12	-1.20	12	12.00
Business Environment	5.95	16	-0.67	13	1.23	8	12.33
Financial & Aid Flows	42.13	6	-1.29	17	-2.60	16	13.00
Others	13.40	15	-0.42	11	-1.72	14	13.33

**Cross-Sectional Prediction: With First 2 Principal Components for Each Topic**

Table C23: Percent Variance Explained by First 2 Principal Components

Topic	% Variance Explained			
	N	PC1	PC2	Total
Institutions (excl. Colonial Origin)	7	66.86	16.29	83.16
Business Environment	4	76.24	13.38	89.62
Production Shares	2	82.30	17.70	100.00
Climate & Agriculture	8	34.76	21.51	56.27
Trade Intensity and Composition	7	35.04	20.03	55.07
Trade Diversification	4	51.14	27.62	78.75
Exchange Rate and ToT	5	42.81	34.21	77.02
Financial & Aid Flows	5	40.12	34.34	74.46
Financial Sector	6	40.30	24.56	64.86
Debt & Reserves	4	38.84	27.70	66.54
Population	6	39.68	24.03	63.71
Health	5	73.36	12.26	85.63
Education	5	73.77	18.01	91.79
Natural Disasters & Conflict	6	52.65	18.79	71.44
Geography & Accessibility	7	37.78	28.31	66.09
Natural Resources	2	93.18	6.82	100.00
Poverty & Inequality	3	68.30	30.29	98.59
Religion & Ethnicity	4	60.32	24.71	85.04
Others	6	69.91	11.03	80.94
Average	5.05	56.70	21.66	78.37

Table C24: RF Ranking of Indicator Topics: PC12 Predicting MAD PCGDP Growth, 1990-2019

<i>Method:</i> Topic	Permutation %ΔMSE	Exclusion %ΔMSE	Residual Fit %ΔMSE	Combined Avg. Rank
	Rank	Rank	Rank	
Financial Sector	114.50	1	2.48	3
Production Shares	61.84	4	4.47	1
Institutions	35.02	6	2.84	2
Financial & Aid Flows	102.31	2	1.65	4
Natural Resources	30.86	7	0.88	6
Trade Intensity and Composition	68.14	3	-1.36	14
Population	19.52	10	0.88	5
Trade Diversification	13.92	12	-0.42	8
Natural Disasters & Conflict	35.68	5	-1.10	13
Business Environment	23.65	9	-2.32	17
Exchange Rate and ToT	12.02	13	-1.87	16
Debt & Reserves	17.17	11	-2.60	18
Climate & Agriculture	11.98	14	-1.37	15
Geography & Accessibility	9.80	15	-0.38	7
Poverty & Inequality	3.57	19	-0.72	10
Education	8.60	16	-0.45	9
Religion & Ethnicity	5.53	18	-0.91	11
Health	6.07	17	-1.02	12
Others	25.18	8	-3.33	19
			-3.09	18
				15.00

Table C25: RF Ranking of Indicator Topics: PC12 Predicting MAD CPI Inflation, 1990-2019

<i>Method:</i> Topic	Permutation %ΔMSE	Exclusion %ΔMSE	Residual Fit %ΔMSE	Combined Avg. Rank
	Rank	Rank	Rank	
Exchange Rate and ToT	225.75	1	19.91	1
Institutions	58.74	4	1.30	2
Natural Disasters & Conflict	61.13	2	0.61	3
Business Environment	43.39	6	-0.01	5
Population	15.61	12	-0.01	4
Geography & Accessibility	19.32	9	-0.79	9
Natural Resources	59.37	3	-4.29	18
Others	20.09	8	-1.99	14
Financial Sector	17.26	11	-0.73	8
Trade Diversification	10.41	14	-0.22	6
Poverty & Inequality	44.91	5	-4.60	19
Health	11.58	13	-2.06	15
Climate & Agriculture	5.27	19	-0.30	7
Education	5.58	18	-0.86	10
Financial & Aid Flows	17.57	10	-1.54	12
Religion & Ethnicity	9.38	15	-3.13	17
Debt & Reserves	38.12	7	-2.63	16
Trade Intensity and Composition	9.08	17	-1.44	11
Production Shares	9.08	16	-1.97	13
			-1.67	17
				15.33

**Panel Prediction: With First 2 Principal Components for Each Topic**

Table C26: Percent Variance Explained by First 2 Principal Components

Topic	N	% Variance Explained		
		PC1	PC2	Total
Institutions	3	47.87	40.35	88.22
Business Environment	2	61.01	38.99	100.00
Production Shares	2	73.69	26.31	100.00
Climate & Agriculture	5	35.72	21.57	57.29
Trade Intensity and Composition	7	38.79	23.46	62.26
Trade Diversification	4	43.20	22.48	65.69
Exchange Rate and ToT	5	40.17	28.43	68.61
Financial & Aid Flows	5	39.57	33.53	73.10
Financial Sector	5	36.74	27.05	63.79
Debt & Reserves	4	46.07	28.98	75.05
Population	6	44.68	22.30	66.98
Health	3	58.92	32.95	91.87
Education	4	40.36	28.70	69.06
Natural Disasters & Conflict	5	31.21	20.61	51.82
Natural Resources	2	80.75	19.25	100.00
Poverty & Inequality	3	68.51	27.15	95.66
Others	5	29.37	23.56	52.93
Average		4.12	48.04	27.39
				75.43

Table C27: RF Ranking of Indicator Topics: PC12 Predicting MAD-Difference of PCGDP Growth

Method: Topic	Permutation %ΔMSE	Exclusion Rank	Residual Fit %ΔMSE	Combined Avg. Rank
Others	83.95	1	1.98	1
Exchange Rate and ToT	45.31	4	4.34	6
Financial Sector	35.92	7	2.24	3
Natural Resources	63.17	3	0.55	2
Institutions	44.60	5	1.21	4
Natural Disasters & Conflict	76.57	2	1.47	17
Financial & Aid Flows	12.61	15	0.72	7
Health	25.13	9	-1.76	5
Debt & Reserves	26.06	8	-0.91	8
Trade Intensity and Composition	36.89	6	-0.87	12
Population	19.11	10	-0.24	11
Business Environment	13.99	13	-0.07	13
Climate & Agriculture	7.50	17	-0.13	9
Production Shares	17.46	12	-0.70	14
Trade Diversification	17.49	11	-0.67	16
Poverty & Inequality	11.56	16	0.09	15
Education	13.49	14	-1.91	10

Table C28: RF Ranking of Indicator Topics: PC12 Predicting MAD-Difference of CPI Inflation

<i>Method:</i> Topic	Permutation		Exclusion		Residual Fit		Combined
	%ΔMSE	Rank	%ΔMSE	Rank	%ΔMSE	Rank	Avg. Rank
Exchange Rate and ToT	106.41	1	8.15	1	25.13	1	1.00
Debt & Reserves	43.52	5	0.15	6	5.97	2	4.33
Production Shares	96.40	2	1.52	2	0.86	12	5.33
Poverty & Inequality	6.23	13	0.25	4	2.58	6	7.67
Climate & Agriculture	38.44	6	-0.15	8	0.97	11	8.33
Natural Resources	8.82	12	-0.20	10	3.66	3	8.33
Health	72.33	3	-0.68	15	1.28	8	8.67
Education	46.45	4	-1.51	17	2.98	5	8.67
Population	36.27	7	1.11	3	-2.30	17	9.00
Natural Disasters & Conflict	11.64	11	0.02	7	1.22	9	9.00
Institutions	13.15	10	-0.64	14	2.33	7	10.33
Financial Sector	24.62	8	-0.48	12	0.42	13	11.00
Business Environment	5.09	15	-1.01	16	3.55	4	11.67
Trade Diversification	13.46	9	-0.64	13	0.13	14	12.00
Financial & Aid Flows	4.99	16	-0.40	11	0.98	10	12.33
Trade Intensity and Composition	5.93	14	-0.19	9	-0.73	15	12.67
Others	2.60	17	0.25	5	-0.91	16	12.67

## **Chapter 3**

# **Regional and Global Economic Integration**

# Africa's Regional and Global Integration

Sebastian Krantz

August 16, 2024

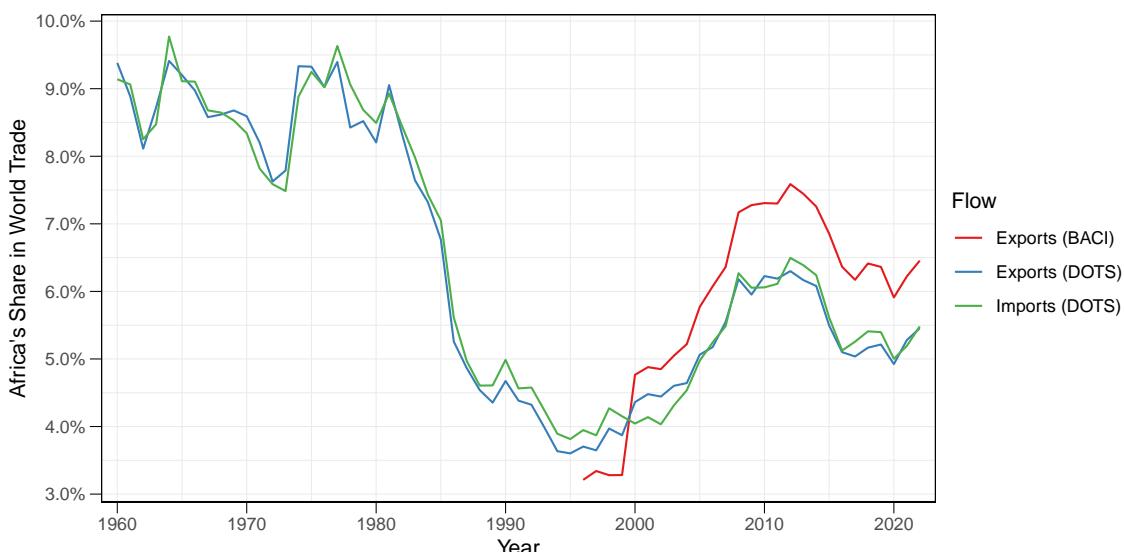
## Abstract

This short paper examines Africa's regional and global integration through trade, global and regional value chains (GVCs and RVCs) at the aggregate and sector levels using detailed trade data and the EMERGING MRIO tables. It finds that the share of Africa's trade with itself is increasing and that precious stones and metals, petrochemicals, mining, and processed foods are driving RVCs, with high potential for further integration. Most RVC trade is within regional economic communities (RECs), particularly inside SADC, implying opportunities to expand RVCs in other RECs and establish cross-REC RVCs. The continent's upstreamness in GVCs has decreased in many sectors, suggesting a trend towards greater local value-addition.

## 1 Gross Trade Flows

Africa's share of world trade is widely acknowledged to be low. According to a recent report by the UN Economic Commission for Africa (UNECA),<sup>1</sup> it is less than 3% of global trade, and mainly driven by merchandise trade. A careful inspection of two widely used databases on merchandise trade, CEPII's BACI ([Gaulier & Zignago, 2010](#)) (HS 1996 version) and the IMF's Direction of Trade Statistics (DOTS) ([IMF General Statistics Division, 1993](#)) database, visualized in Figure 1, suggests a higher African share of 5.5-6.5% in global merchandise trade. Interestingly, the share was high at 9% until 1980, then saw a rapid decline to less than 4% in 1995 and a subsequent rise to above 6% in 2012. The trade slowdown in the 80s and 90s is congruent to the extended period of high inflation, low commodity prices, debt distress, structural adjustment, and political instability commonly referred to as the "lost decade(s)" for Africa. The trade spurt in the 2000s, on the other hand, is aligned with a growth spurt supported by higher commodity prices and economic reforms, commonly referred to as "Africa Rising" and analyzed in [Calderón & Boreux \(2016\)](#), [Rodrik \(2018\)](#), and [Krantz \(2023\)](#) among others. Thus, Africa's share of world trade co-varies to a large extent with its aggregate macroeconomic performance.

Figure 1: Africa's Share of Global Merchandise Trade

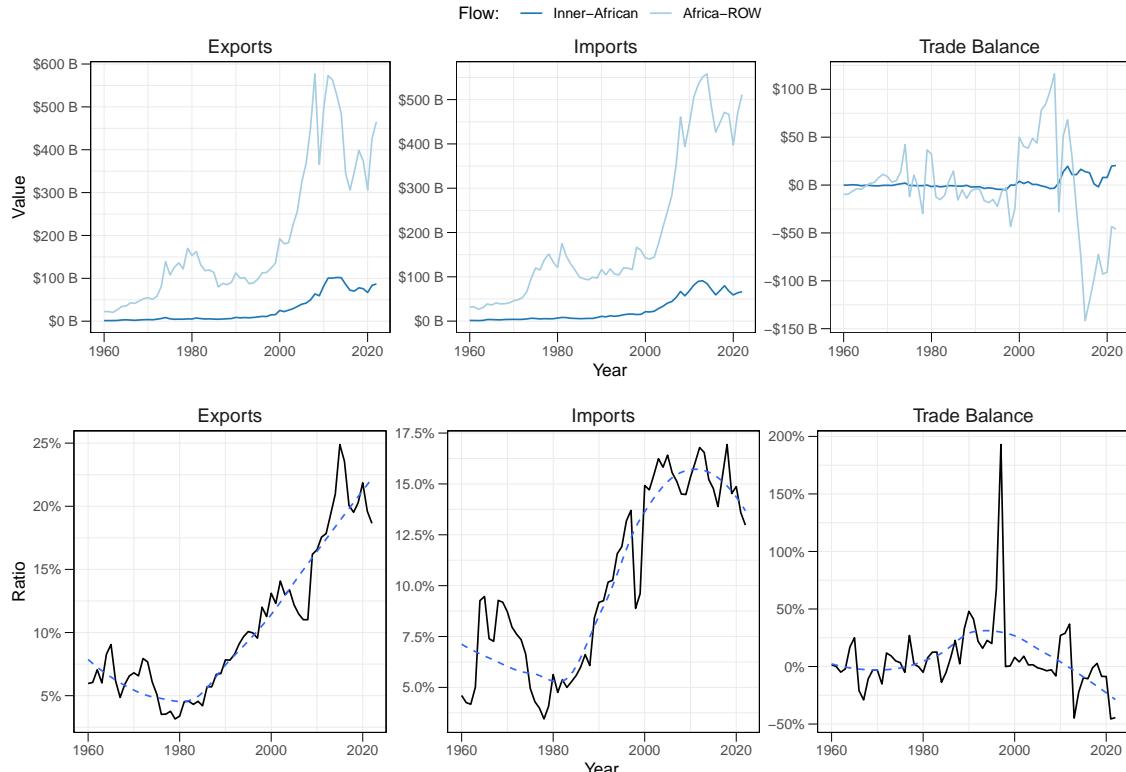


*Notes:* Figure shows Africa's share in global merchandise trade from different databases. The DOTS is an aggregate database derived from official sources, whereas BACI is a product-level database derived from COMTRADE.

<sup>1</sup><https://www.uneca.org/stories/african-countries-trading-more-outside-the-continent-than-amongst-themselves%2C-eca-report>

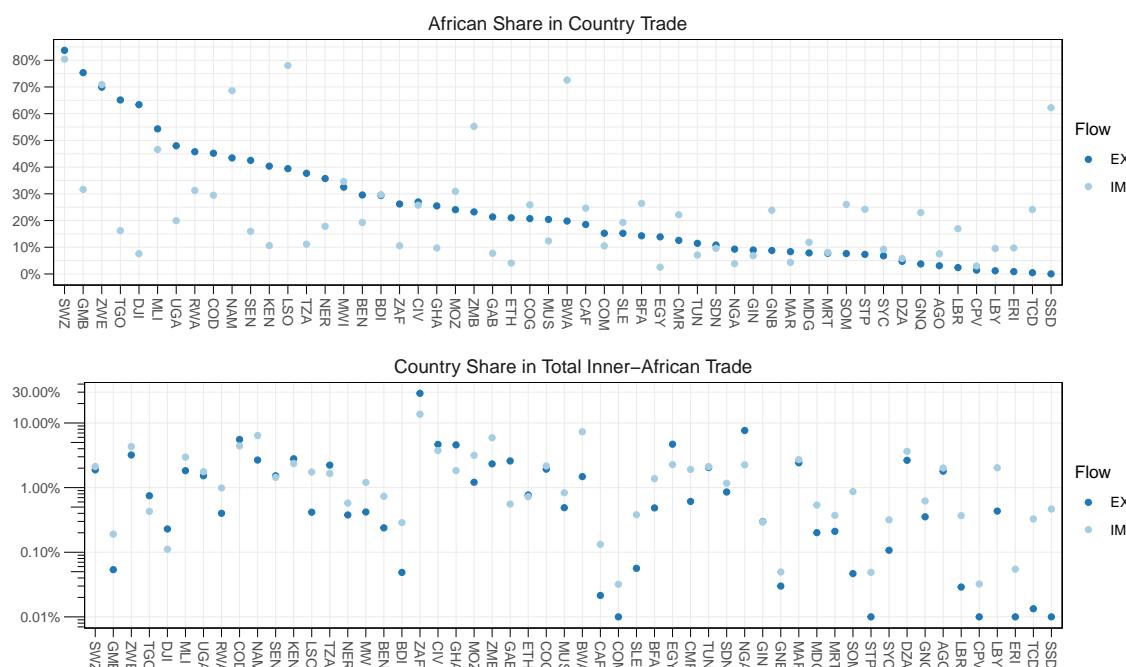
There is also notable variation in the extent of African trade with itself versus with the rest of the World (ROW). Figure 2 shows the inner-African and Africa-ROW trade levels in constant 2015 USD according to DOTS, and below it the ratio of these flows together with a lowess trend. Evidently, the ratio (in percentage terms) of inner-African trade to Africa-ROW has risen considerably since 1980, from 5% in 1980 to 20% for Exports and 15% for imports in 2020. The trade balance of Africa with itself has been positive since 2010 and negative with ROW.

Figure 2: African Trade with Itself and ROW



*Notes:* Figure shows inner-African and Africa-ROW trade flows according to DOTS data (top panel), and the ratio of inner-African to Africa-ROW trade including a smooth lowess trend (bottom panel).

Figure 3: Inner-African Trade Share in Total African Trade by Country: 2010-2022 Averages



*Notes:* Using DOTS data, Figure shows the African share in countries' trade, and countries' share in inner-African trade.

Figure 3 further disaggregates inner-African trade shares by country, expressed as a share of total country exports/imports rather than a ratio to the Africa-ROW trade as in Figure 2. Figure 3 uncovers considerable heterogeneity in countries' share of trade with African partners (top panel) and countries' total share of inner-African trade (bottom panel). With shares of 30% on both metrics, South Africa is the most significant regional trader, followed by Nigeria and Congo.

A further level of heterogeneity in African trading is the share of inner-African trade within or between regional economic communities (REC), which have played a significant role in the continent's economic development and regional integration agenda up to this point. To facilitate regional trade analysis, Table 1 provides a mutually exclusive classification of countries into either their most important economic union or the union of closest geographic proximity.

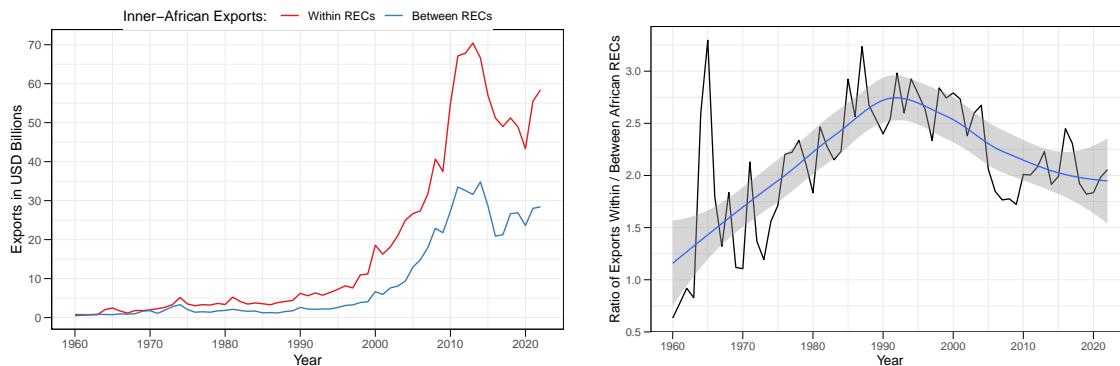
Table 1: Classification of African Countries into RECs

REC	Description	Countries
AMU+EGY	Arab Maghreb Union + Egypt	DZA, EGY, LBY, MRT, MAR, TUN, ESH
CEMAC+STP	Economic Community of Central African States + São Tomé and Príncipe	CMR, CAF, TCD, COG, GNQ, GAB, STP
EAC	East African Community	BDI, COD, KEN, RWA, SSD, TZA, UGA
ECOWAS	Economic Community of West African States	BEN, BFA, CPV, CIV, GMB, GHA, GIN, GNB, LBR, MLI, NER, NGA, SEN, SLE, TGO
IGAD-EAC	Intergovernmental Authority on Development, excluding EAC Members	DJI, ERI, ETH, SOM, SDN
SADC-COD	Southern African Development Community, excluding the DRC (now EAC)	AGO, BWA, COM, SWZ, LSO, MDG, MWI, MUS, MOZ, NAM, SYC, ZAF, ZMB, ZWE

*Notes:* Table provides mutually exclusive classification of countries into RECs. The REC name reflects deviations from official membership.

Figure 4 decomposes inner-African exports between and within RECs. The ratio of within- to between-REC exports shown on the RHS indicates that the period from 1960-1990 was characterized by increasing trade within regional blocks - from 1.2 times greater in 1960 to 2.75 times greater than between-REC trade in 1990. After 1990, trade between RECs picked up again and appeared to stabilize at a ratio of 2 in 2020. Thus, trade in RECs in Africa is currently 2 times greater than trade between RECs.

Figure 4: Inner-African Trade Between and Within RECs

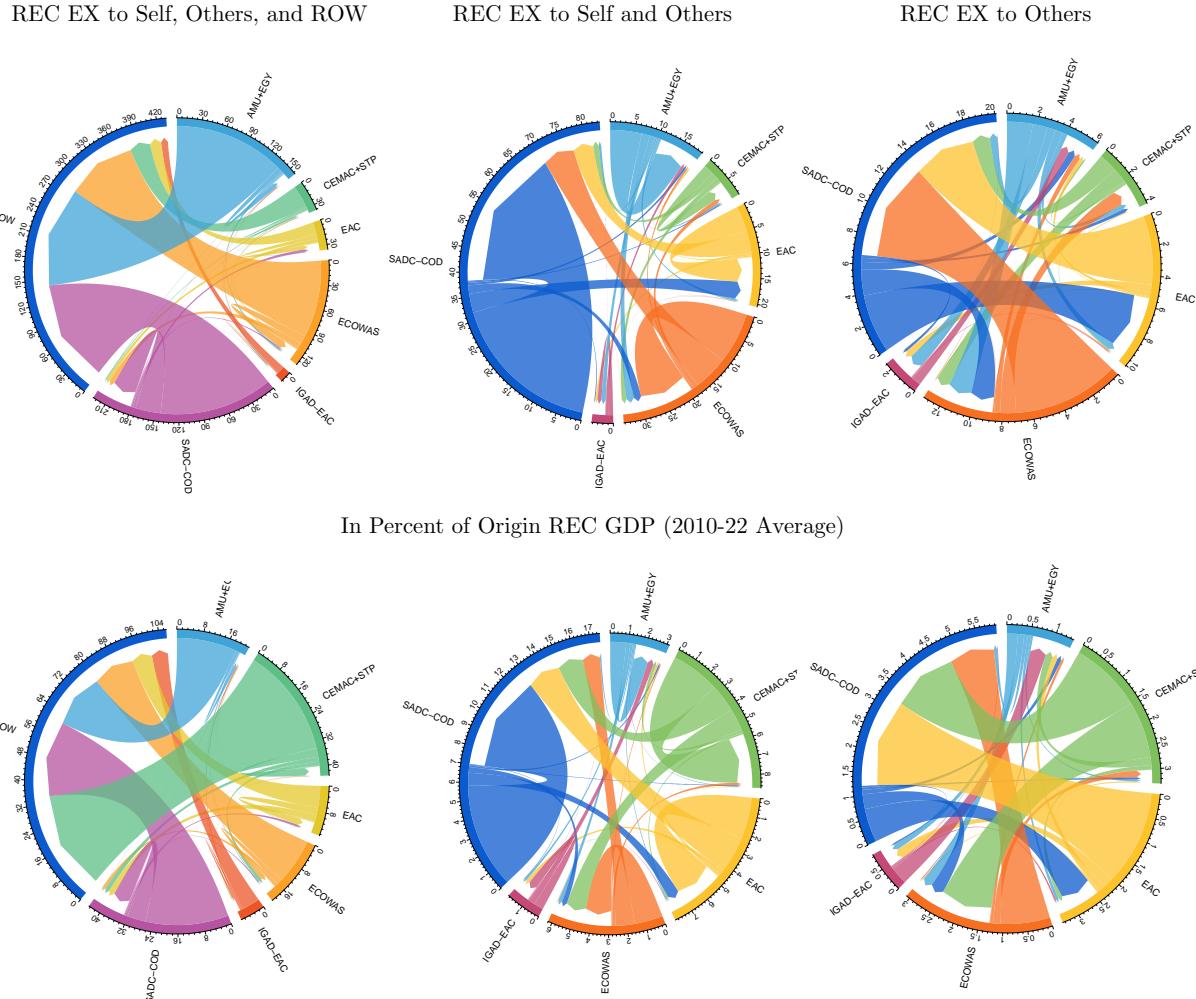


*Notes:* Figure plots inner-African trade within and between RECs according to DOTS data (LHS) and their ratio (RHS).

At the REC level, it is feasible to visualize the trade flows. Figure 5 does this in 3 ways, depending on whether exports to ROW and within the REC are included. The LHS indicates that the AMU+EGY, SADC-COD, and ECOWAS have the largest exports to ROW. SADC-COD has the largest inner-REC exports, at a volume of 32 billion USD'15, followed, with some distance, by ECOWAS with inner-REC exports above 10.6 billion. The AMU+EGY and EAC have inner REC exports of 6.5 and 5.2 billion, respectively. However, these RECs have different economic sizes, so their total trade volumes may not be indicative of their extent of regional integration. Averaging the GDP in constant 2015 USD of the RECs over the same period (2010-22) yields a GDP of 736

billion for AMU+EGY, 673 billion for ECOWAS, 545 billion for SADC-COD, 225 billion for the EAC, 168 billion for IGAD-EAC, and 82 billion for CEMAC+STP. The bottom panel of Figure 5 shows trade flows in percent of the origin REC GDP.

Figure 5: (Inner-)African Trade (Exports) by RECs: 2010-2022 Averages in 2015 USD Billions



*Notes:* Figure shows migration flow diagrams visualizing REC-level exports. The top panel provides three different diagrams depending on whether inner-REC exports and exports to ROW are included. The bottom panel provides the same diagrams but with flows in percent of the exporting REC's GDP. Exports data is taken from DOTS, and GDP is from the World Development Indicators. Both are averaged between 2010 and 2022 to smooth temporal variation.

Surprisingly, CEMAC+STP has the greatest export penetration at 41% of GDP, of which 34% is destined for ROW, 2.7% is internal trade, 1.5% for SADC-COD, and 1.2% for ECOWAS. It is followed by SADC-COD, with a total export penetration of 33.2%, of which 26.2% for ROW. SADC-COD maintains the largest inner-REC trade share at 5.8% of GDP, and also exports 0.7% and 0.3% of GDP to the EAC and ECOWAS, respectively. The EAC remains the greatest inter-REC trader, at 6.9% of its GDP exported to ROW, 2.3% regional trade, and a large share of 2% of GDP is exported to SADC-COD, followed, with some distance, by 0.23% of GDP exports to IGAD-EAC. ECOWAS has an export penetration of 17.2%, of which 14.0% is for ROW, 1.6% regional, 1.1% for SADC-COD and 0.16% for CEMAC+STP. The AMU+EGY trades 20.3% of GDP, but 18.9% with ROW at only 0.9% regional trade, 0.23% for ECOWAS, and 0.13% for IGAD-EAC. IGAD-EAC has the lowest export penetration at 6.6% of GDP, of which 5.7% is for ROW, 0.5% regional, and 0.4% for the AMU+EGY. Overall, the absence of direct east-west trade is striking in this picture. The EAC does not trade meaningful quantities with ECOWAS or CEMAC+STP, indicating the existence of large trade barriers in central Africa. North-South trade is also scarce, likely reflecting physical and cultural barriers and the proximity of the AMU+EGY to Europe and the Middle East. The exports between AMU+EGY and SADC+COD, valued at 500-600 million and below 0.1% of GDP from both sides, are also insignificant in relative magnitude.

## 1.1 Trade Flows by Sector

To analyze trade flows by sector, I employ a 18-sector classification adapted from [Huo et al. \(2022\)](#), comprising 11 goods producing and 7 services sectors. The classification, summarized in Table 2, maps to the 2-digit HS2002 codes (see Appendix Table 6). Whereas [Huo et al. \(2022\)](#) distinguish 10 goods producing sectors, I add an 11th sector called PSM - "Precious stones & base metals incl. compounds" to distinguish minimally processed mining outputs that play an important role in inner-African trade from raw mining, petrochemical, and metal products. In the original classification of [Huo et al. \(2022\)](#), most of these products are subsumed into petrochemicals (PCM). Appendix Table 6 provides a mapping. I loosely refer to 2-digit HS codes as industries.

Table 2: Sector Classification

Code	Sector Definition of <a href="#">Huo et al. (2022)</a> + Created PSM Sector
<i>Goods Producing Sectors (EMERGING and BACI)</i>	
AFF	Agriculture, hunting, forestry & fishing
FBE	Food production, beverages & tobacco
PCM	Petroleum, chemicals & non-metallic mineral products
PSM	Precious stones & base metals incl. compounds
MIN	Mining & quarrying
TEX	Textiles, leather & wearing apparel
WAP	Wood, paper & publishing
MPR	Metal & metal products
ELM	Electrical & machinery
TEQ	Transport equipment
MAN	Manufacturing & recycling
<i>Service Sectors (Only EMERGING)</i>	
EGW	Electricity, gas & water
SMH	Sale, maintenance & repair of vehicles; fuel; trade; hotels & restaurants
TRA	Transport
PTE	Post & telecommunications
CON	Construction
FIB	Financial intermediation & business activity
PAO	Public administration; education; health; recreation; other services

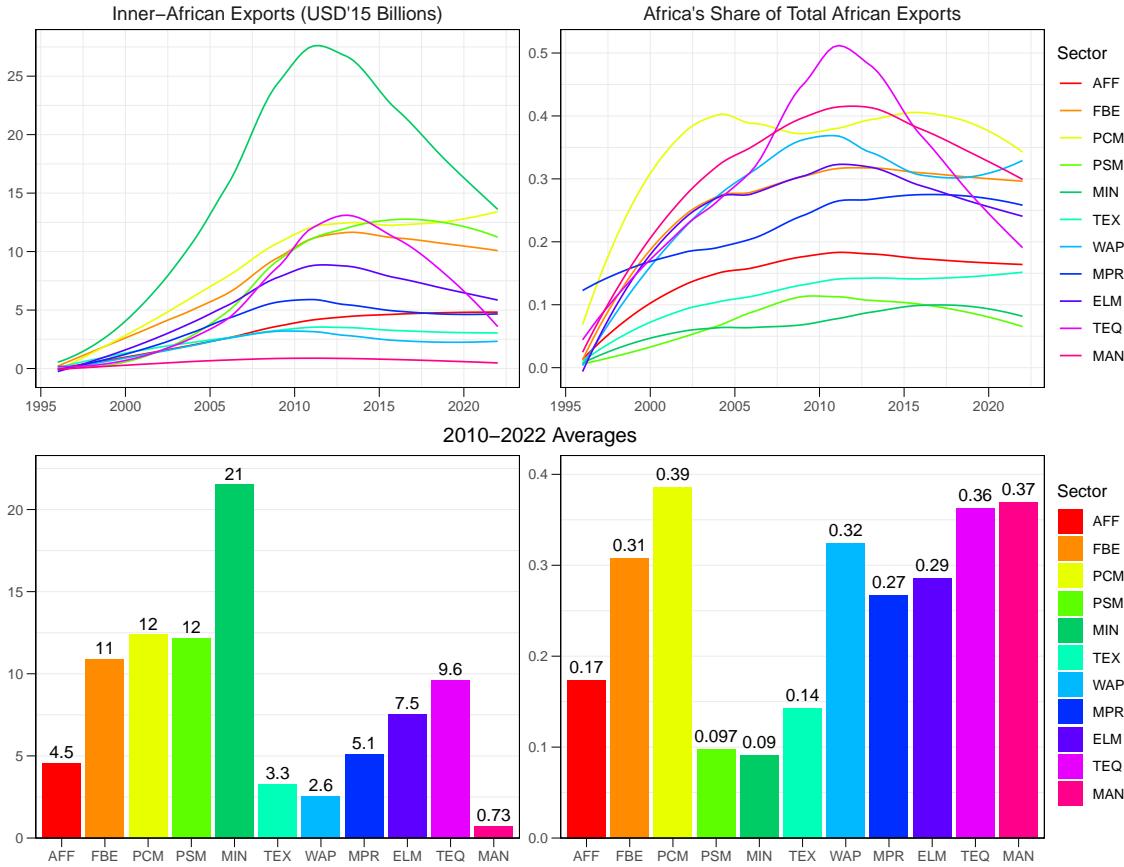
*Notes:* Table shows 18 sector classification applied to EMERGING and BACI. See Appendix Table 6 for details.

Employing this classification, Figure 6 plots inner-African goods trade (BACI) by sector, using a lowess-smoother to curb some high-frequency volatility in trade flows. The bottom half shows 2010-2022 averages, indicating that petrochemicals (PCM), precious stones and base metals (PSM), and mining (MIN) together make up almost 50% of inner-African trade - amounting to 90.2 billion USD'15. Mining exports, driven by petroleum, peaked in 2011 and declined to less than 15 billion in recent years, whereas PCM exports increased slightly to almost 14 billion. The RHS shows that only 9%-10% of African mining and PSM exports were inner-African, compared to 39% of PCM exports. This suggests that PCM has a leading role in inner-African trade expansion.

Following these three, processed foods and beverages (FBE) accounts for 12% of inner-African trade, amounting to 11 billion USD'15. Also here, inner-African exports peaked in 2013 but remain above 10 billion. The African share in exports of 31% for FBE is also very high. Thus, FBE has a similarly important role in inner-African trading to PCM. Of the other sectors, transport equipment (TEQ), metal products (MPR), electrical machinery (ELM), and agriculture (AFF) have inner-African exports between 4.5 and 9.6 billion, at African shares between 17 and 36%. TEQ exports also peaked in 2013 and have declined below 5 billion since then.

To validate the importance of different sectors for inner-African trade, I also consider the extensive margin of regional trading in these sectors. Figure 7 shows all African countries that exported more than 100 million USD'15 per year to Africa during the 2010-22 period by sector. The counts indicate the number of countries. Notably, in the FBE sector, 22 countries had inner-African exports above 100M, comparable to mining (23), followed by PCM (17) and PSM (14).

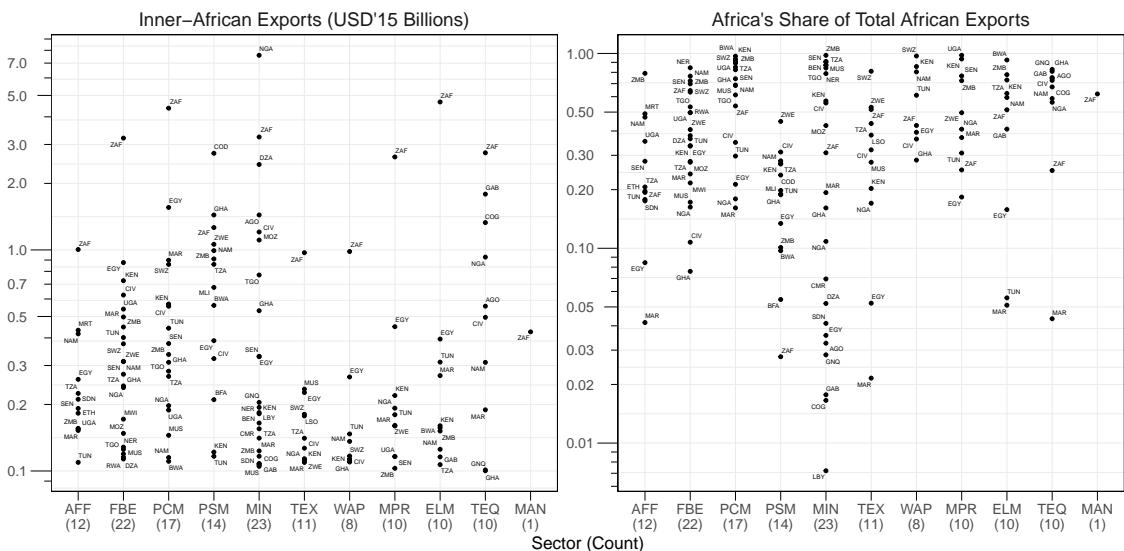
Figure 6: Inner-African Exports by Sector: Billions of Constant 2015 USD



Notes: Figure shows inner-African exports by sector (LHS) and as a share of total African exports (RHS) using BACI.

In all sectors apart from mining (mainly petroleum) and PSM, South Africa (ZAF) is the largest inner-African exporter. In FBE and PCM, >50% of ZAF exports are to African countries, versus <3% in PSM. Nigeria has the largest overall sectoral export flow at above 7 billion USD'15 annual mining (petroleum) exports to Africa. This only comprises 11% of total Nigerian mining exports. Exempting ZAF and Nigerian mining, other flows above 1 billion USD'15 are PCM exports by Egypt, PSM exports by the DRC, Ghana, Zimbabwe, and Namibia, mining exports of Algeria, Angola, Cote d'Ivoire, and Mozambique, and TEQ exports by Gabon, the Republic of Congo, and Nigeria. In the FBE sector, Egypt, Kenya, Cote d'Ivoire, Uganda, and Morocco have exports between 500M and 1B USD'15. In PCM, this category includes Morocco, Swaziland, Kenya, and Cote d'Ivoire; in PSM, Zambia, Tanzania, Mali, and Botswana; and in mining, Togo and Ghana.

Figure 7: Inner-African Exports &gt; 100M USD'15 by Sector and Country: 2010-22 Averages



Notes: Figure shows sectoral inner-African exports > 100M USD'15 (LHS) and their share in country exports (RHS).

It remains to analyze sectoral exports by RECs. Table 3 does so in a very compact way by reporting the average 2010-22 African exports of each REC in billions of USD'15, and, in parentheses, the share of Africa in overall sectoral exports by the REC, followed by the own-REC share in the REC's African exports. The final row additionally reports these metrics for total exports across sectors. Drawing from this last row, according to the BACI database SADC-COD exports 40.9 billion USD'15 to Africa, which comprises 20% of the total exports of this REC, and 79% of these African exports are within the REC. The EAC, in contrast, only exports 9.15 billion to Africa, constituting 30% of total EAC exports, and 51% of these exports are regional. Of the sectors with sizeable regional trade volumes, AFF, FBE, PCM, MPR, ELM, and TEQ generally have African export shares above 10%. In FBE and PCM, all RECs except for CEMAC+STP and IGAD-EAC have African export volumes above 1 billion USD'15, indicating a high potential for driving regional trade integration. In most RECs, 50% or more of these exports are regional.

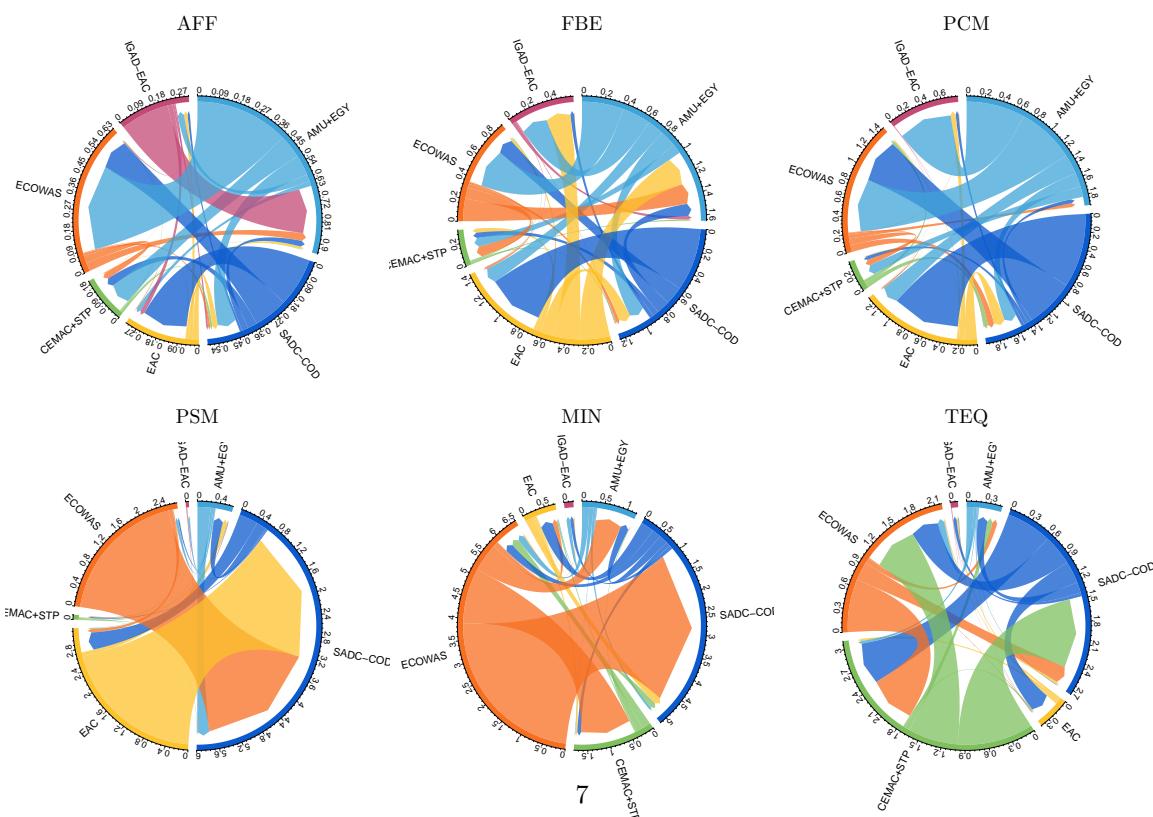
Table 3: Inner-African Exports by Sector and REC: 2010-22 Averages

Sector	AMU+EGY	CEMAC+STP	EAC	ECOWAS	IGAD-EAC	SADC-COD
AFF	0.99 (12% 34%)	0.03 (12% 47%)	0.5 (18% 88%)	0.65 (14% 87%)	0.41 (16% 40%)	1.96 (26% 78%)
FBE	1.9 (30% 47%)	0.12 (13% 70%)	1.67 (37% 58%)	1.82 (16% 80%)	0.13 (10% 63%)	5.26 (50% 82%)
PCM	3.03 (20% 40%)	0.18 (28% 45%)	1.07 (88% 79%)	1.88 (42% 88%)	0.02 (35% 51%)	6.19 (59% 75%)
PSM	0.61 (6.1% 46%)	0.05 (1.6% 10%)	3.74 (22% 27%)	2.83 (11% 8.1%)	0.02 (0.62% 1.7%)	4.89 (6.9% 88%)
MIN	3.13 (3.7% 89%)	0.58 (2.4% 32%)	0.5 (19% 46%)	10.89 (14% 46%)	0.11 (4.2% 99%)	6.29 (11% 86%)
TEX	0.4 (2.9% 58%)	0.02 (10% 29%)	0.31 (30% 60%)	0.54 (20% 67%)	0.05 (10% 16%)	1.95 (42% 94%)
WAP	0.52 (45% 49%)	0.1 (5.2% 20%)	0.24 (59% 79%)	0.27 (22% 86%)	0.01 (18% 29%)	1.41 (46% 86%)
MPR	0.88 (22% 62%)	0.1 (31% 51%)	0.44 (87% 84%)	0.55 (48% 87%)	0 (20% 42%)	3.09 (23% 69%)
ELM	1.02 (7.6% 54%)	0.25 (45% 39%)	0.36 (64% 61%)	0.4 (44% 74%)	0.05 (21% 24%)	5.46 (52% 74%)
TEQ	0.38 (7% 57%)	3.3 (72% 47%)	0.24 (77% 61%)	1.77 (55% 42%)	0.02 (30% 51%)	3.91 (30% 62%)
MAN	0.1 (12% 44%)	0.01 (29% 50%)	0.08 (63% 62%)	0.05 (46% 69%)	0 (33% 28%)	0.49 (61% 85%)
SUM	12.97 (8% 57%)	4.75 (13% 44%)	9.15 (30% 51%)	21.65 (16% 51%)	0.82 (7.8% 48%)	40.9 (20% 79%)

Notes: Table shows the REC exports to Africa, and, in parentheses, Africa's share in total REC exports and the own-REC share in exports to Africa.

To examine the between-REC trade, Figure 8 visualizes the between-REC exports in AFF and the 5 largest sectors. The largest between-REC flows are in mining and PSM, valued at 7.8 and 8.3 billion USD'15, respectively. These are dominated by Nigeria and the DRC. Next up, TEQ, valued at 4.5 billion, shows meaningful trade between CEMAC+STP, SADC-COD, and ECOWAS. This is followed by PCM at 3.9 billion, FBE at 3.1 billion, and AFF at 1.5 billion. In these sectors, SADC-COD and the AMU+EGY are the largest inter-regional exporters, but in FBE, the EAC and ECOWAS are also important, and in AFF, IGAD-EAC exports much to AMU+EGY. The ELM and MPR sectors, omitted here, are dominated by SADC-COD (South African) exports.

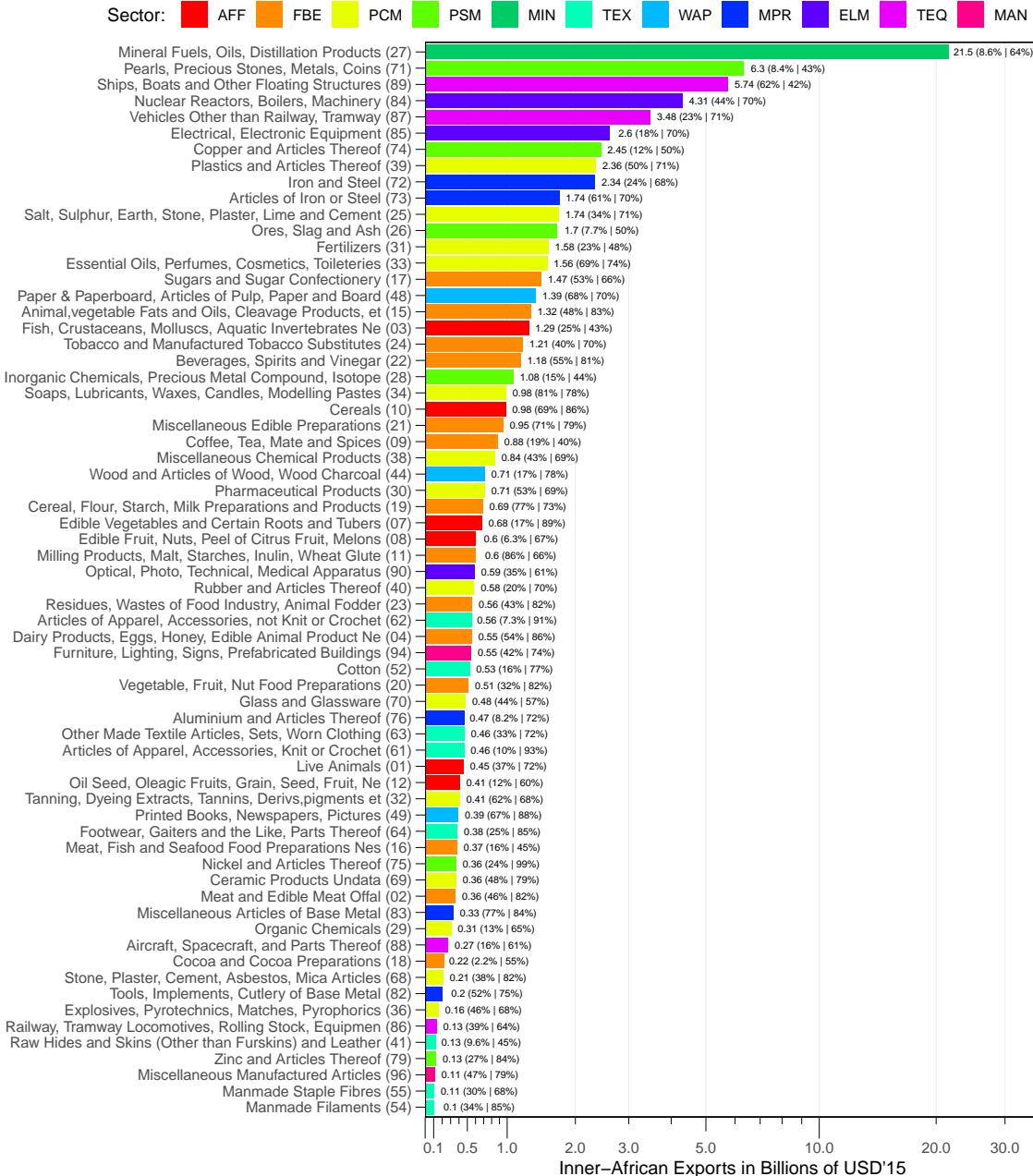
Figure 8: Between-REC Trade (Exports): 2010-2022 Averages in 2015 USD Billions



## 1.2 Trade Flows by Industry

Concluding gross trade analysis, I also examine inner-African trade at the industry level to better understand what is being traded in different sectors. Figure 9 provides an overview at the 2-digit HS02 level, showing inner-African exports for all significant industries, including, in parentheses, the percentage in total African exports and the percentage traded inside RECs, as in Table 3.

Figure 9: Inner-African Exports > 100M USD'15 by Industry (HS02 2-Digit): 2010-22 Averages



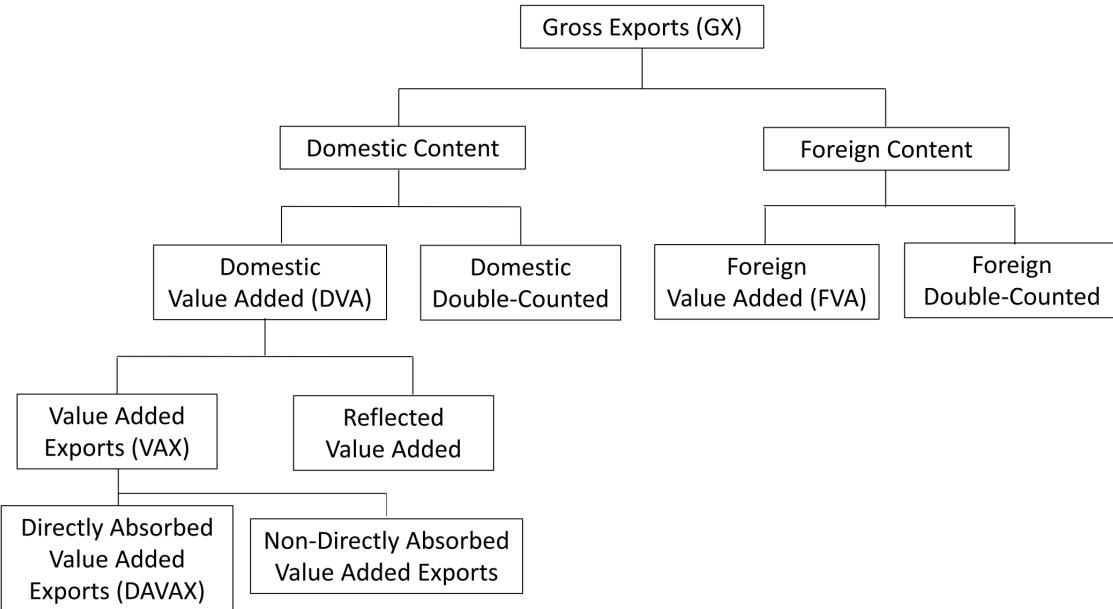
Notes: Figure shows total inner-African exports by 2-digit HS2002 industries valued above > 100M USD'15 (BACI 2010-2022 Average), and, in parentheses, its percentage in total African exports and the percentage of it traded inside RECs.

Figure 9 indicates that the mining sector only includes petroleum. With 21.5 billion USD'15 annually over the last decade, the petrol industry is by far the largest in inner-African trade but only accounts for 8.6% of total African petroleum exports. It is followed by precious stones and metals at 6.3 billion. Other large industries in inner-African trade are boats, machinery, vehicles, electrical equipment, copper, plastics, and iron/steel at annual volumes above 2 billion USD'15. Following these, many industries with volumes between 0.5 and 2 million are in the PCM, FBE, and AFF sectors. Exempting AFF, these industries all involve low to moderate levels of processing and have African shares in total exports above 20%. Typically, >60% of this trade is within RECs.

## 2 Value Chains

While gross trade flows provide useful information about direct economic relationships, their value-added (VA) content partly stems from previous production stages performed in other countries. Koopman et al. (2014) first proposed a complete decomposition of a country's gross exports into different VA components. Borin & Mancini (2019) showed that their decomposition is inexact for certain components and proposed a refined version, visualized schematically in Figure 10.

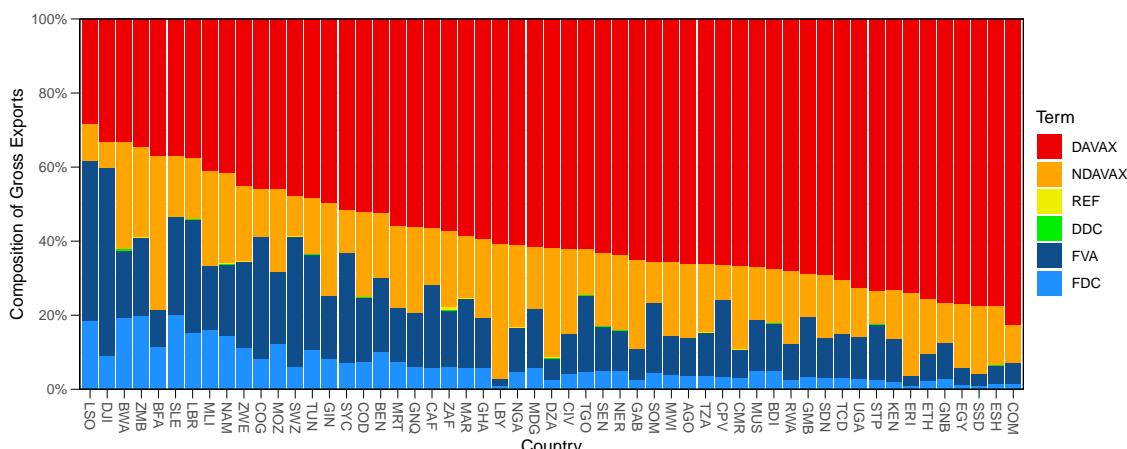
Figure 10: REFINED KOOPMAN WANG WEI DECOMPOSITION OF GROSS EXPORTS



*Notes:* Graphic of Borin & Mancini (2019)'s exports decomposition into VA components. *Source:* Antràs & Chor (2022).

Borin & Mancini (2019) propose  $GX - DAVAX$  as a measure of total GVC-related trade. This can be further broken down in to *backward* GVC participation ( $GX - DVA$ ) comprising non-domestic content in exports, and *forward* GVC participation ( $DVA - DAVAX$ ) capturing re-exported domestic content in exports. Figure 11 shows the decomposition for all African countries based on EMERGING 2015-2019 tables aggregated to the sector level. Since the Domestic Double Counted (DDC) and Reflected Value Added (REF) terms are very small, backward GVC participation is approximately equal to  $FVA + FDC$ , and forward participation is approximately  $NDAVAX$ . For the lion's share of countries, DAVAX still makes up more than 50% of exports, but many countries such as Lesotho, Djibouti, Botswana, Zambia, and Burkina Faso are highly engaged in GVCs. The mode of engagement also differs, with some countries such as Libya, Eritrea, and Burkina Faso mostly participating in forward GVCs, whereas Others like Djibouti, Lesotho, or Swaziland mostly engaged in backward GVCs, partly also due to their small economic size.

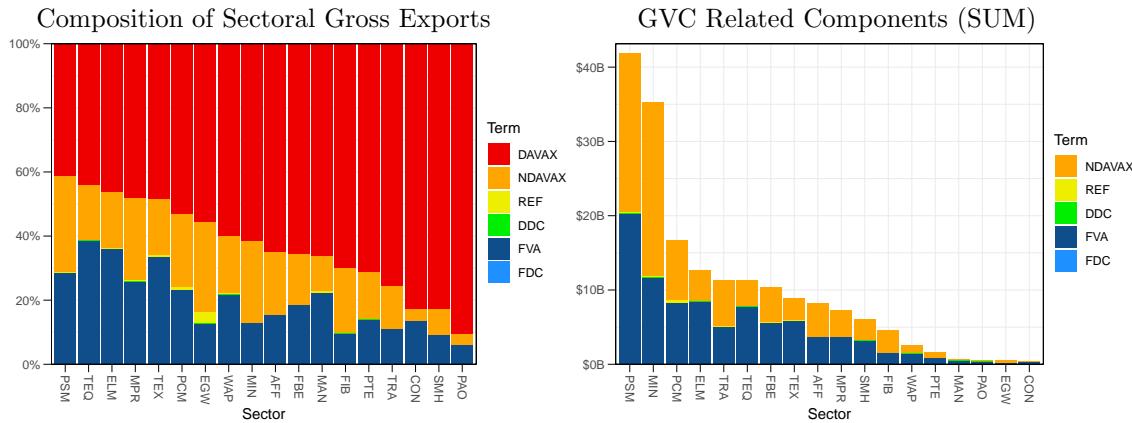
Figure 11: Decomposition of Total Gross Exports by Country: 2015-19 Average



*Notes:* Figure plots decomposition of countries' exports into VA components following Borin & Mancini (2019) (Fig. 10).

The decomposition can also be computed at the sector level, using a source-based exporter perspective following [Borin & Mancini \(2019\)](#), which allows summing items across countries. Figure 12 provides both the decomposition of sector-level gross exports and the total size of GVC-related components. The PSM sector has the greatest GVC export share at nearly 60%, followed closely by core manufacturing sectors (TEQ, ELM, MPG, TEX). The latter are, however, less important in overall size. The RHS of Figure 12 shows that PSM and mining at >\$30B USD of GVC-related exports each, together make up 42.7% of total African GVC-related exports. They are followed, with some distance, by PCM at \$16.7B USD or 9.2% of total African GVC-related exports.

Figure 12: Decomposition of African Countries' Gross Exports by Sector: 2015-19 Average



Notes: Figure shows sectoral exports decomposition into VA terms (Fig. 10) using a source-based exporter perspective.

## 2.1 Regional Value Chains

To study which sectors are important for African RVCs, it is necessary to extract the parts of GVC components accounted for by African partners. Towards this end, the [Leontief \(1936\)](#) inverse is useful. Let  $\mathbf{A}$  be a normalized ICIO table where each element  $a_{oi,uj}$  gives the units of origin country  $o$  and sector  $i$ 's (row) output required for the production of one unit of using country  $u$  and sector  $j$ 's (column) output,  $\mathbf{x}$  the vector of outputs of each country-sector, and  $\mathbf{d}$  a vector of final demand (FD) such that the following productive relationship holds

$$\mathbf{x} = \mathbf{Ax} + \mathbf{d}. \quad (1)$$

Then, following [Leontief \(1936\)](#)

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{d} = \mathbf{Bd}, \quad (2)$$

where the Leontief Inverse is denoted  $\mathbf{B} = (\mathbf{I} - \mathbf{A})^{-1}$ . This matrix is also often called the total requirement matrix since it gives the total productive input requirement from each sector to produce one unit of final output<sup>2</sup>. The direct VA share of each country-sector is given by

$$\mathbf{v} = \mathbf{1} - \mathbf{A}'\mathbf{1}, \quad (3)$$

where  $\mathbf{1} = (1, 1, 1, \dots, 1)'$  is a column-vector of 1's<sup>3</sup>. Let  $\mathbf{V}$  be the matrix with  $\mathbf{v}$  along the diagonal and 0's in the off-diagonal elements. Multiplying Eq. 2 with  $\mathbf{V}$  then gives VA in each country-sector

$$\mathbf{Vx} = \mathbf{V}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{d} = \mathbf{VBd}. \quad (4)$$

The matrix  $\mathbf{VB} = \mathbf{V}(\mathbf{I} - \mathbf{A})^{-1}$  is known as the matrix of VA multipliers or VA shares, which can be used to obtain the amount of VA generated in each country-sector ( $\mathbf{Vx}$ ) when producing to satisfy FD ( $\mathbf{d}$ ). More specifically, it contains the amount of VA by each country-sector (row) to the production of one unit of each country-sector's (column's) output.

<sup>2</sup>Specifically each element in  $b_{oi,uj}$  in  $\mathbf{B}$  gives the output required from country-sector  $oi$  for the production of one unit of the final good in  $uj$ . Thus, the first column of  $\mathbf{B}$  gives all the productive input required from all sectors for the production of one unit of the final good in sector 1, and the first row of  $\mathbf{B}$  gives all the input required from sector 1 to produce one unit of the final good in each sector.

<sup>3</sup>Thus the expression amounts to summing up the entries in each column of  $\mathbf{A}$  (representing the intermediate input shares for 1 unit of output) and subtracting them from 1.

Let  $\mathbf{E}$  be a vector of sector-level gross exports, then  $\mathbf{VBE}$  is the matrix of sector-level exports (columns) decomposed by origin of VA country-sector (rows) with elements  $vbe_{oi,uj}$ . From it, following Hummels et al. (2001) and Baldwin & Lopez-Gonzalez (2015), we can derive simple indicators for the imported exports (I2E) and re-exported exports (E2R) shares.

$$I2E_{uj} = \frac{1}{E_{uj}} \sum_{oi, o \neq u} vbe_{oi,uj} \quad \forall uj \quad (5)$$

$$E2R_{oi} = \frac{1}{E_{oi}} \sum_{uj, u \neq o} vbe_{oi,uj} \quad \forall oi, \quad (6)$$

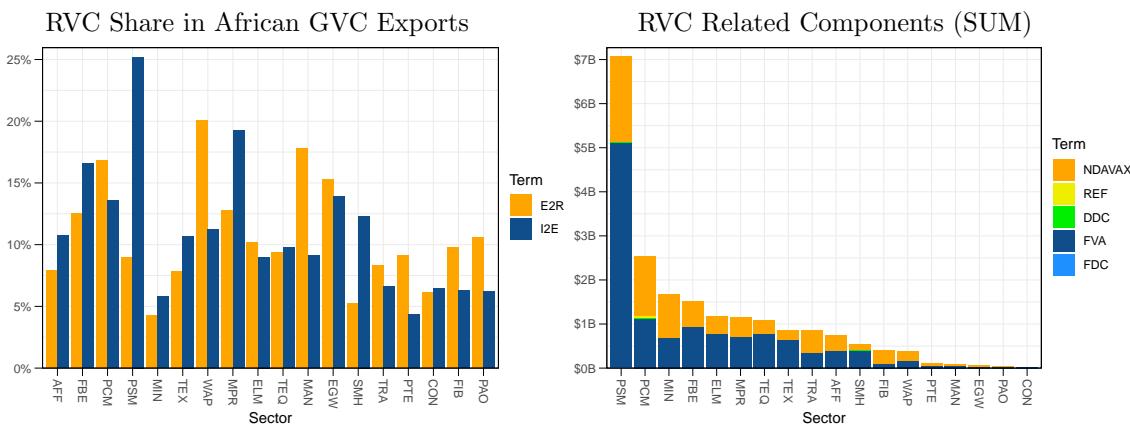
The import measure is equal to the share of backward GVC participation [ $I2E = (GX - DVA)/GC$ ]. The export measure is approximately equal to the forward GVC participation share but imprecise because it includes double-counted components. These are, however, relatively small in Africa. Using these measures, I obtain the shares of African partners in countries' backward and forward GVC participation. Let  $AFR$  denote the set of all African countries, then

$$I2E_{uj}^{AFR} = \sum_{oi \in AFR, o \neq u} vbe_{oi,uj} / \sum_{oi, o \neq u} vbe_{oi,uj} \quad \forall uj \in AFR, \quad (7)$$

$$E2R_{oi}^{AFR} = \sum_{uj \in AFR, u \neq o} vbe_{oi,uj} / \sum_{uj, u \neq o} vbe_{oi,uj} \quad \forall oi \in AFR. \quad (8)$$

are relative shares tracking the African share in countries' forward and backward GVC participation. The LHS of Figure 13 shows  $I2E^{AFR}$  and  $E2R^{AFR}$  computed at the sector level, i.e., the African share in total African GVC-exports within each sector. The RHS applies these shares to the exact measures following Borin & Mancini (2019), multiplying FVA, FDC, and DDC with  $I2E^{AFR}$  and NDAVAX and REF with  $E2R^{AFR}$ . In most sectors, the RVC share is between 5% and 20%. Sectors WAP, MAN, PCM, and EGW have high forward participation shares (E2R), whereas PSM, MPR, FBE, and PCM have high backward participation shares (I2E). In absolute values, PSM alone accounts for 35% of African RVCs, around \$7 billion, followed by PCM (12.4%), MIN (8.3%), and FBE (7.5%). It is unlikely that PSM and MIN RVCs are sophisticated, but they highlight great potential to process raw materials such as gold, diamonds, and other ores and precious stones in African countries with more advanced industrial facilities rather than exporting them directly or via African neighbours. PCM, comprising mainly plastics, salt, cement, fertilizers, oils, and cosmetics, soaps, and lubricants (see Figure 9), and FBE also show high potential for RVC deepening.

Figure 13: African RVCs by Sector: 2015-19 Average

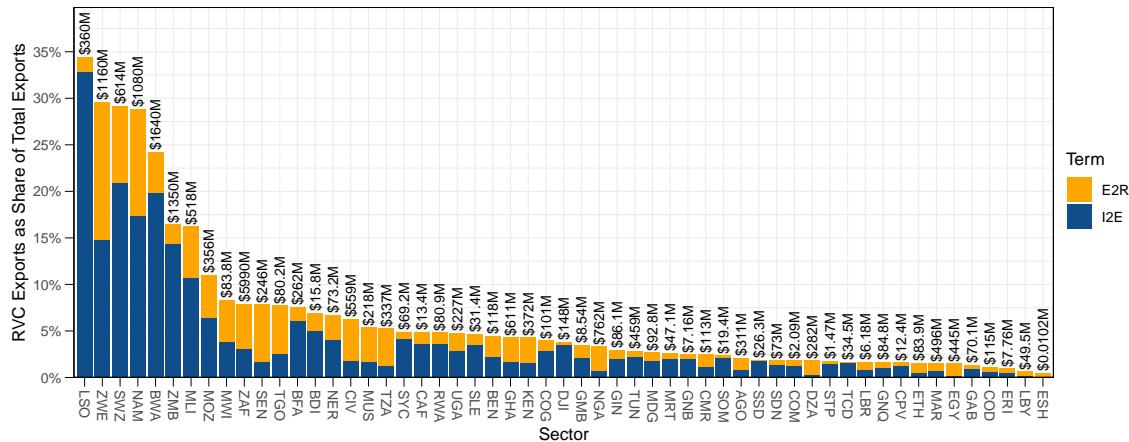


Notes: Figure shows African RVCs by sector, computed via regional backward (I2E) and forward (E2R) GVC shares.

To investigate how countries engage differently in RVCs, I plot their overall and sector-level engagement below. Figure 14 provides the share of RVC exports (both forward/E2R and backward/I2E RVC integration) in country total gross exports. Ostensibly, for most countries, the RVC share is below 5%, but a handful of countries, particularly in SADC, are highly engaged at export shares close to 30%. The overall volume of RVC-related flows is also concentrated in SADC, with South

Africa leading at \$6 billion USD RVC-related exports, followed by Botswana, Zambia, Zimbabwe, and Namibia at between \$1 and \$2 billion. Nigeria follows at \$762 million.

Figure 14: African RVCs by Country: 2015-19 Average



Notes: Figure shows total RVC share (backward (I2E) and forward (E2R)) in country gross exports, including its value.

Table 4 provides a summary of bilateral RVCs by RECs. Evidently, the largest share of RVC-related exports is by SADC-COD countries, followed by ECOWAS, AMU+EGY, and the EAC. In total, 72% of RVC exports are within RECs. Of the remaining 28% between-REC RVC exports, the largest flows are between ECOWAS and SADC-COD, valued at around \$1-1.4 billion, and between the EAC and SADC-COD, valued around \$450-520 million.

Table 4: African RVCs by REC: 2015-19 Average (USD millions)

Exporter	Total	AMU+EGY	CEMAC+STP	EAC	ECOWAS	IGAD-EAC	SADC-COD	
AMU+EGY	<b>1777.9</b>	<b>1038.4</b>		64.7	84.7	238.2	103.8	248.1
CEMAC+STP	<b>418.3</b>		<b>99.1</b>	17.8	123.7	6.6	91.9	
EAC	<b>1173.7</b>		95.9	16.0	<b>527.4</b>	49.1	31.6	453.5
ECOWAS	<b>3380.3</b>	278.2		101.1	51.4	<b>1880.9</b>	22.8	1045.9
IGAD-EAC	<b>331.8</b>	137.5		5.7	38.6	20.6	<b>67.7</b>	61.7
SADC-COD	<b>13321.5</b>	289.5		90.9	516.0	1376.0	57.1	<b>10992.0</b>

I also examine RVCs at the country-sector and REC-sector levels. Figure 15 shows sectoral shares in country-level RVC exports, with the largest RVC exporter to the left of the chart. There is significant heterogeneity in countries' RVC export contents. In many countries, PSM and MIN (green bars) dominate RVC engagement. In some smaller countries like Rwanda, Cape Verde, and Sao Tome, services (SRV) exports dominate. Apart from these, many countries export AFF, FBE, and PCM as part of RVCs. South Africa has the largest and most diversified RVC exports.

Figure 15: African RVCs by Country: 2015-19 Average Sectoral Shares

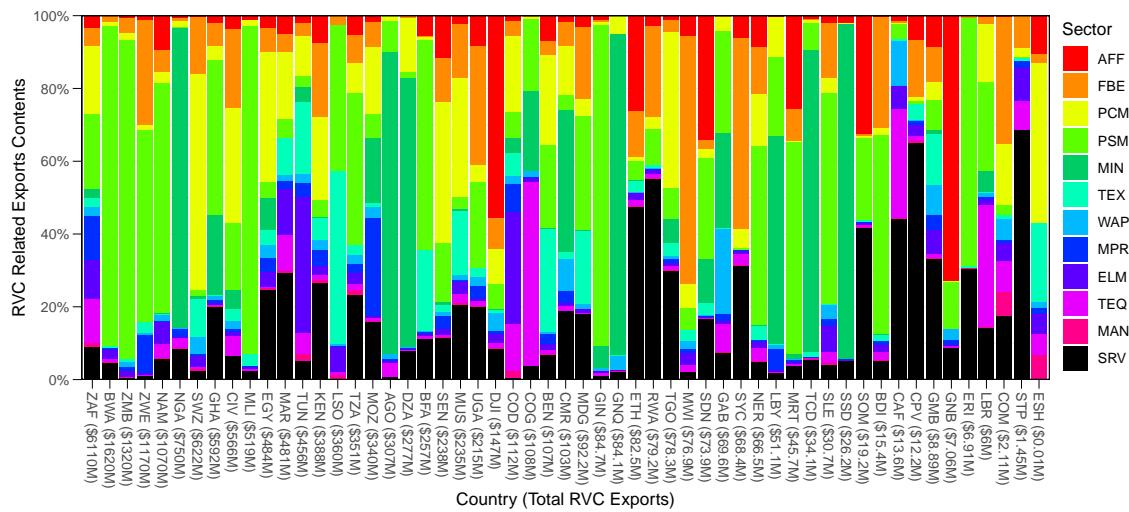


Table 5 gives a breakdown of RVC exports by REC and sector. SADC-COD has the largest RVC exports in all sectors, and in all RECs apart from AMU+EGY and IGAD-EAC, either PSM or MIN is the largest traded sector. However, there are significant differences between RECs. FBE RVCs play a significant role in the EAC, where they make up 17% of all RVC exports. They play a lesser role in ECOWAS and SADC-COD at shares around 7.7%. PCM, on the other hand, is very important in the AMU+EGY (20%) and also in the EAC, SADC, and ECOWAS (12-13%).

Table 5: African RVCs by REC and Sector: 2015-19 Average (USD millions)

Sector	Total	AMU+EGY	CEMAC+STP	EAC	ECOWAS	IGAD-EAC	SADC-COD
AFF	<b>749.8</b> (3.7%)	62.35 (3.5%)	2.65 (0.64%)	70.94 (6%)	108.6 (3.3%)	134.9 (41%)	370.4 (2.8%)
FBE	<b>1596</b> (7.8%)	79.51 (4.4%)	7.5 (1.8%)	205.9 (17%)	221.7 (6.7%)	25.05 (7.6%)	1056 (7.9%)
PCM	<b>2700</b> (13%)	357.7 (20%)	21.6 (5.2%)	153.8 (13%)	386.8 (12%)	17.27 (5.2%)	1763 (13%)
PSM	<b>6986</b> (34%)	101.8 (5.7%)	48.81 (12%)	239 (20%)	1177 (35%)	43.96 (13%)	5375 (40%)
MIN	<b>1774</b> (8.7%)	296 (16%)	184.6 (45%)	29.98 (2.5%)	787.9 (24%)	9 (2.7%)	466.4 (3.5%)
TEX	<b>884.2</b> (4.3%)	160.6 (8.9%)	2.85 (0.69%)	47.19 (4%)	138.4 (4.2%)	6.42 (1.9%)	528.8 (3.9%)
WAP	<b>364.2</b> (1.8%)	36.58 (2%)	32.11 (7.8%)	26.26 (2.2%)	28.5 (0.86%)	7.89 (2.4%)	232.9 (1.7%)
MPR	<b>1133</b> (5.5%)	51.53 (2.9%)	5.39 (1.3%)	42.66 (3.6%)	41.96 (1.3%)	1.57 (0.48%)	989.7 (7.4%)
ELM	<b>1202</b> (5.9%)	251.3 (14%)	3.78 (0.91%)	57.93 (4.9%)	25.44 (0.77%)	5.23 (1.6%)	858.5 (6.4%)
TEQ	<b>1071</b> (5.2%)	77.07 (4.3%)	65.4 (16%)	29.23 (2.5%)	70.99 (2.1%)	4.01 (1.2%)	824.1 (6.2%)
MAN	<b>97.22</b> (0.48%)	14.51 (0.81%)	0.25 (0.06%)	10.13 (0.85%)	1.62 (0.049%)	0.24 (0.074%)	70.47 (0.53%)
SRV	<b>1889</b> (9.2%)	306 (17%)	39.12 (9.4%)	273.9 (23%)	333.9 (10%)	73.9 (22%)	862.2 (6.4%)
SUM	<b>20446</b> (100%)	<b>1795</b> (100%)	<b>414.1</b> (100%)	<b>1186.9</b> (100%)	<b>3322.4</b> (100%)	<b>329.5</b> (100%)	<b>13398</b> (100%)

Notes: Table shows RVC content by REC and sector, including sectoral shares in total REC RVC exports. Computed using VBE shares.

Another critical consideration is geography, as deepening RVCs is facilitated by countries being geographically close. Towards this end, Figure 16 summarizes countries' aggregate engagement. Panel (A) clearly shows that most RVC engagement is in SADC, followed by ECOWAS and the EAC. North Africa is not very engaged in African RVCs. Panel (B) shows the largest RVC sector, which is MIN or PSM for most countries. If these two are excluded in Panel (C), North African countries mostly engage in PCM RVCs, and many countries in eastern, western and central Africa mainly engage in services RVCs. If services are also excluded in Panel (D), IGAD-EAC focuses on agricultural RVCs, the EAC is split between FBE and PCM, and SADC and ECOWAs are more diverse, exempting two geographic clusters: in SADC, FBE is the leading sector in Zimbabwe, Zambia, and Malawi, and in ECOWAS textiles is leading in Benin, Burkina Faso, and Mali.

To provide a detailed spatial analysis of important RVC sectors, Figure 17 shows the total values (in million USD) of RVCs in 6 important sectors. Agricultural RVCs in Panel (A) are quite dispersed, with South Africa and Namibia the most significant participants, but also sizeable participation in the EAC, Ethiopia, and Sudan, as well as Ghana, Morocco, and Senegal. In FBE, shown in Panel (B), Zimbabwe is the leading RVC exporter, with sizeable contributions in all of SADC, but also in the EAC, especially between Uganda and Kenya, as documented in greater detail in Krantz (2024). In West Africa, Ghana contributes significantly to the FBE RVC. The PCM RVC is geographically dispersed, just like the agricultural one. PSM RVCs are heavily concentrated in southern Africa, where Botswana takes lead, but also exist in western Africa, particularly Mali. Mining (petroleum) RVCs in Panel (E) are dominated by Nigeria, followed by Angola and Algeria. Services RVCs, shown in Panel (F), are dispersed, with key contributors Egypt, Morocco, and South Africa, Ghana, and Kenya. I end by examining Africa's (changing) position in GVCs.

## 2.2 GVC Positioning

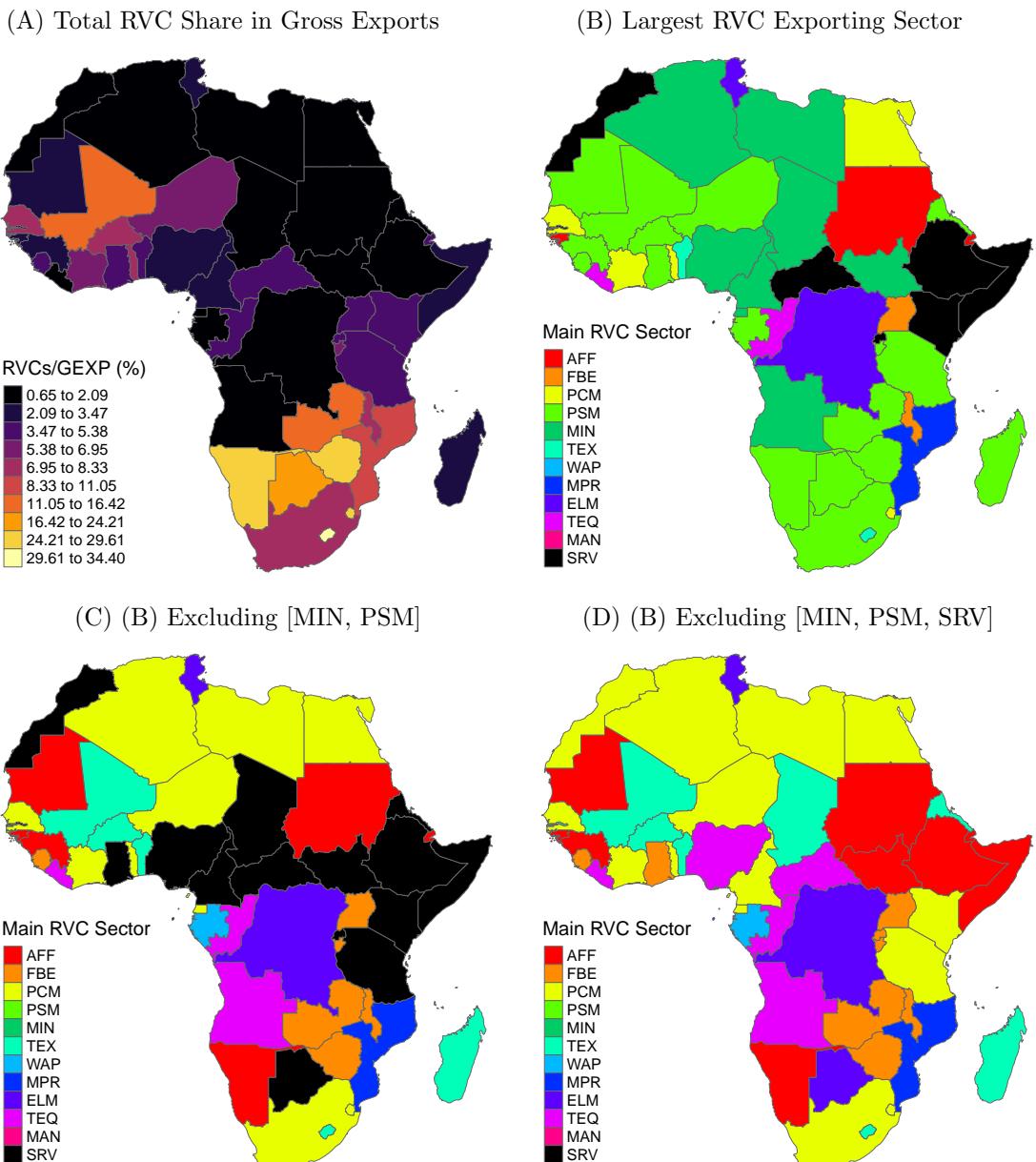
Following Antràs et al. (2012); Antràs & Chor (2022), a common measure of upstreamness  $U_{oi} \in \mathbf{u}$  is obtained by iterating forward the IO model in Eq. 1, multiplying terms by the number of production stages needed to obtain them, and normalizing by gross output. In matrix notation:

$$\mathbf{u}\mathbf{x} = \mathbf{d} + 2\mathbf{Ad} + 3\mathbf{AAD} + 4\mathbf{AAAd} + \dots = (\mathbf{I} - \mathbf{A})^{-2}\mathbf{d}. \quad (9)$$

The index is, by definition, greater than 1, and Antràs et al. (2012) state that it can be interpreted as the dollar amount by which the output of all country-sectors combined increases following a one-dollar increase in the VA of sector  $i$  in country  $o$ . Intuitively, it measures the distance of the production stage performed by sector  $i$  in country  $o$  to the finally demanded product ( $\mathbf{d}$ ).<sup>4</sup>

<sup>4</sup> An equivalent measure of downstreamness ( $\mathbf{d}$ ) can be computed measuring the distance to VA instead of FD (Antràs & Chor, 2022; Miller & Temurshoev, 2017; Mancini et al., 2024), but, for the sake of brevity, this is omitted.

Figure 16: African RVCs Total Engagement and Main Sector: 2015-19 Average



*Notes:* Figure visualizes countries' RVC share in country gross exports and the sector with the largest RVC exports.

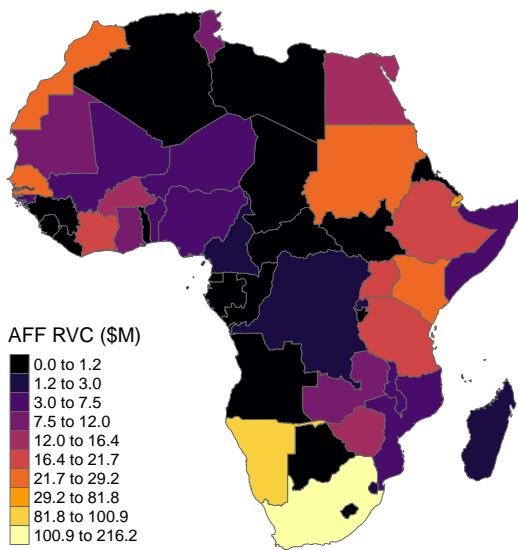
Antràs et al. (2012) further find that  $U$  is positively correlated with physical capital intensity and negatively correlated with skill intensity across US industries, and negatively correlated with rule of law, private credit to GDP, and education across a sample of OECD countries.

Figure 18 shows aggregate upstreamness by country, where sector-level  $U_{oi}$  estimates were averaged using gross export weights. Countries with a highly concentrated export mix in sectors such as mining or PSM that receive a lot of downstream processing, such as Equatorial Guinea, Botswana, South Sudan, and Gabon, are relatively upstream. At the lower end of the spectrum, Kenya, Egypt, and Sao Tome mainly export travel services and FBE, which are close to final demand. GVC positioning does not necessarily imply anything about the state of development or economic diversification, but typically, countries with service-led economies are more downstream. China, for example, is relatively upstream because it heavily exports electrical machinery, which is a GVC-intensive sector with long GVCs. The top US exports, on the other hand, are financial and business services, which place it much more downstream.

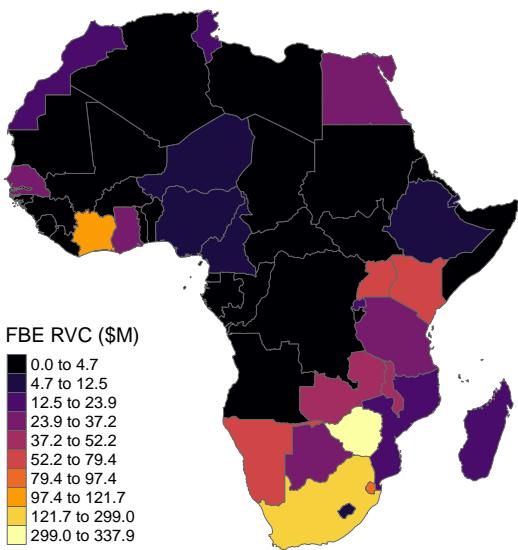
The simplest way of computing this index is as  $\mathbf{d} = \mathbf{1}'\mathbf{B}$ , i.e., it is the column-sum of the Leontief inverse matrix (Miller & Temurshoev, 2017; Antràs & Chor, 2022). It can be interpreted as the total increase in gross output in the world economy that a unit increase in FD in the respective country-sector would generate. At the world level,  $\mathbf{u}$  and  $\mathbf{d}$  are identical and measure the length of GVCs (Mancini et al., 2024).

Figure 17: African RVCs in Key Sectors: 2015-19 Average

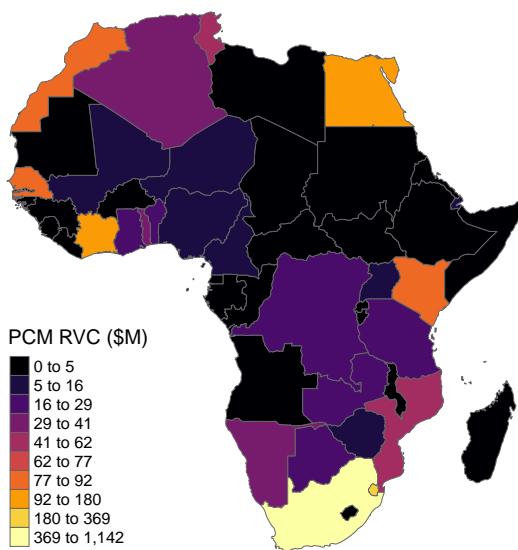
(A) Agriculture, Forestry &amp; Fishing



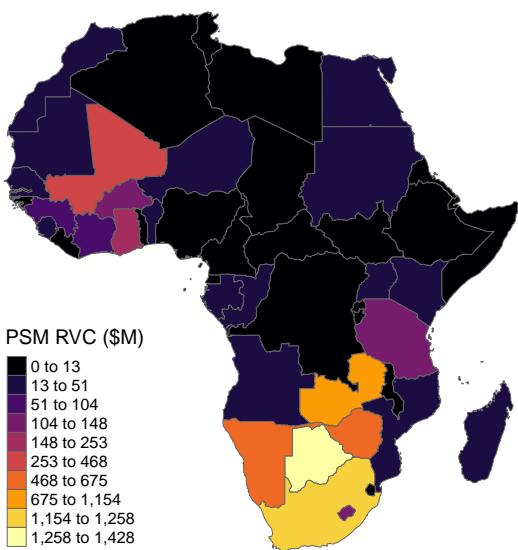
(B) Processed Foods &amp; Beverages



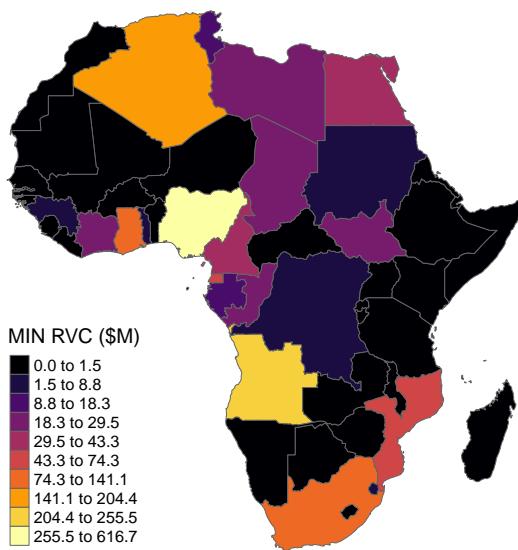
(C) Petrochemicals



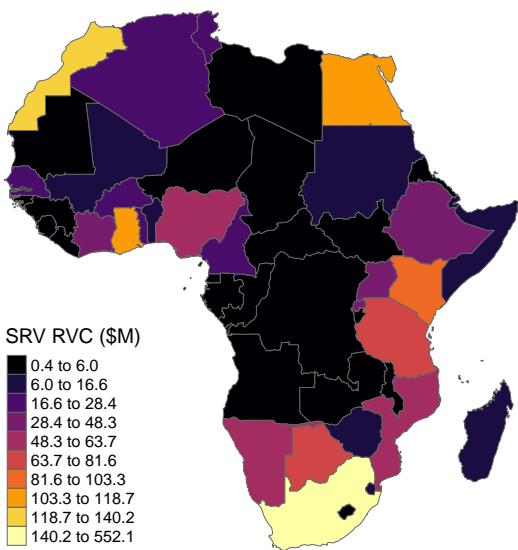
(D) Precious Stones and Metals



(E) Mining (Petroleum)

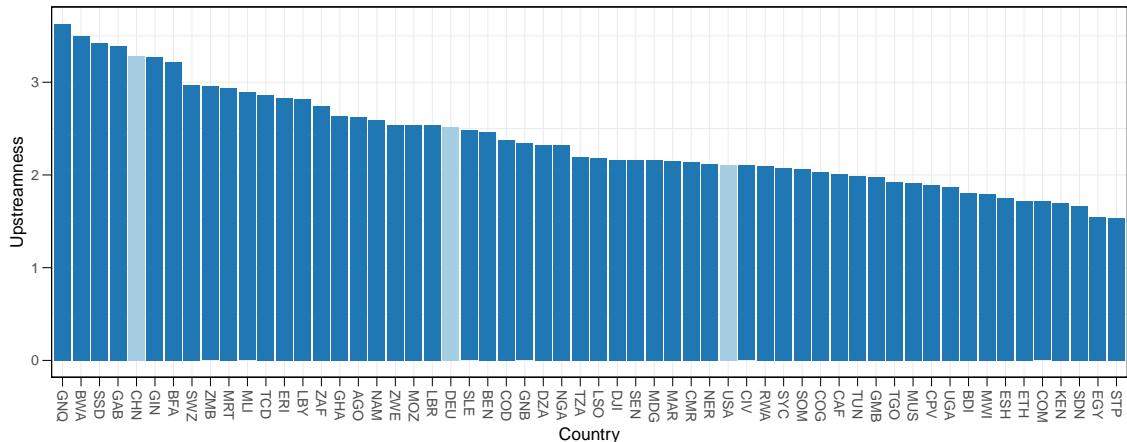


(F) Services (incl. Utilities)



Notes: Figure visualizes countries' RVC exports within key RVC sectors, based on EMERGING 2015-19 MRIO tables.

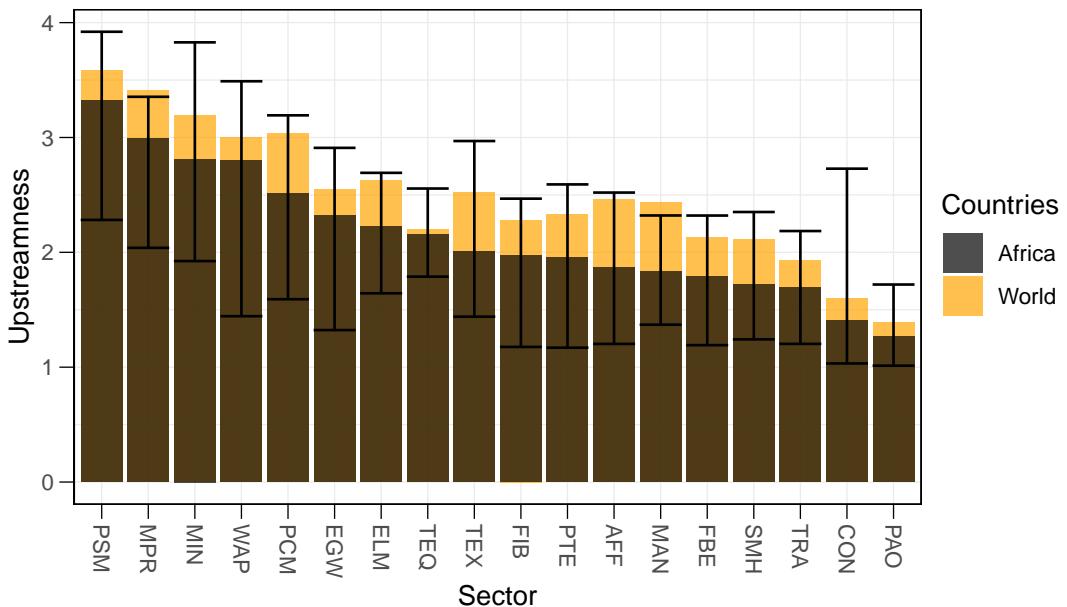
Figure 18: UPSTREAMNESS INDEX FOR AFRICAN COUNTRIES



Notes: Figure shows upstreamness index following Antràs et al. (2012), computed at the sector level and averaged across sectors using sectoral gross exports as weights. Based on EMERGING MRIO tables averaged 2015-2019.

Figure 19 provides an export-weighted average (across countries) of sectoral upstreamness, including weighted 5th and 95th percentile bounds. In all sectors, average World upstreamness is greater, indicating that African products are part of shorter GVCs. The PSM sector is most upstream, together with MPR, MIN, WAP, and PCM. At nearly 3.5x the VA generated in downstream production stages, African countries have significant opportunities to increase local value addition by further processing precious stones and other mining products (e.g., petroleum). At the other end of the spectrum, FBE is the most downstream manufacturing sector at an upstreamness of around 1.8. Thus, for more local VA, African countries should produce and export more foods and beverages and expand local PSM, MIN, and, to a lesser extent, PCM processing.

Figure 19: UPSTREAMNESS INDEX BY SECTOR

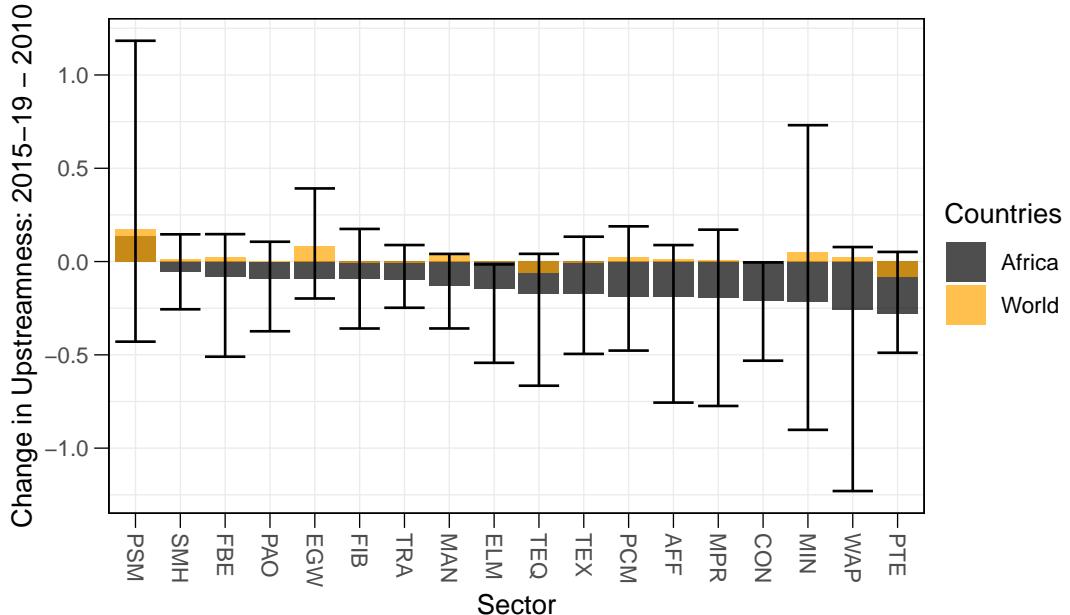


Notes: Figure shows upstreamness index following Antràs et al. (2012), averaged across countries using sectoral gross exports as weights, including weighted 5th and 95th percentiles. Based on EMERGING MRIO tables averaged 2015-2019.

Has there already been any progress in this direction? Figure 20 shows the difference between the 2010 EMERGING estimate and the 2015-19 median estimate, computed at the country level and then averaged across countries using 2015-19 median exports as weights. It is fascinating to see from Figure 20 that all sectors in Africa apart from PSM have moved downstream on average. This stands in contrast to a global trend towards longer manufacturing GVCs. In the world average, all manufacturing sectors apart from TEQ have moved upstream, reflecting this increase in the

length of chains. For Africa, the trend thus suggests greater local value addition in many sectors but may also imply a shift towards processing trade in some sectors.

Figure 20: UPSTREAMNESS INDEX BY SECTOR: DIFFERENCE



*Notes:* Figure shows an export-weighted average of the country-level differences between the 2015-2019 median upstreamness and 2010 upstreamness, computed using EMERGING. Bars give weighted 5th and 95th percentiles.

### 3 Conclusion

This short paper examines Africa's global and regional integration through trade, GVCs, and RVCs, supported by the EMERGING MRIO tables and the DOTS and BACI trade databases. It shows that Africa's share in global merchandise trade is at 5.5-6.5% and fluctuates with Africa's macroeconomic performance. However, the share of the region's trade with itself has increased steadily since 1980. Today, African inner-African trade is ~5 times smaller than Africa-ROW trade, up from ~20 times smaller in 1980. Regional trade intensity differs significantly by country. Particularly countries in SADC are heavily engaged, with South Africa alone accounting for 30% of inner-African trade. Trade within African RECs is 2x larger than trade between RECs.

The largest sectors for inner-African trade are mining (petroleum), petrochemicals (PCM), precious stones and metals (PSM), and processed foods and beverages (FBE). These sectors also drive RVCs, with significant potential for deeper engagement and more local value addition. In particular, countries with advanced processing facilities could process more PSM, mining, and agricultural outputs from countries lacking these facilities. PCM RVCs could also become longer, with deeper processing and more complex products produced across borders. In FBE, an increase in the share of agricultural output being locally processed and efforts to market these products regionally could expand the size and reach of RVCs, which are very localized at this point.

All RECs could engage much more deeply in regional trading and value addition. Establishing competitive cross-REC RVCs will likely require significant transport infrastructure investments. Countries should thus focus on exploiting comparative advantages within their RECs. The sectoral and graphical analysis in this paper provides some suggestions towards this end, such as enhanced agricultural processing in IGAD and the EAC (alongside PCM), more PSM processing, electrical machinery and transport equipment in SADC and central Africa, deeper textile RVCs in western Africa (alongside mining, PCM, and PSM), longer PCM RVCs in the AMU and Egypt, etc.

Africa's overall position in GVCs has shifted slightly downstream, countering a global trend towards longer GVCs. In most sectors, this implies that more local value is added, a trend that could be accelerated by harnessing processing capabilities in other countries through RVCs.

## References

- Antràs, P., & Chor, D. (2022). Global value chains. *Handbook of international economics*, 5, 297–376.
- Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3), 412–416.
- Baldwin, R., & Lopez-Gonzalez, J. (2015). Supply-chain trade: A portrait of global patterns and several testable hypotheses. *The World Economy*, 38(11), 1682–1721.
- Borin, A., & Mancini, M. (2019). *Measuring what matters in global value chains and value-added trade* (Tech. Rep. No. 8804). World Bank Policy Research Working Paper.
- Calderón, C., & Boreux, S. (2016). Citius, altius, fortius: Is growth in sub-saharan africa more resilient? *Journal of African Economies*, 25(4), 502–528.
- Gaulier, G., & Zignago, S. (2010). *Baci: International trade database at the product-level. the 1994-2007 version* (Working Papers No. 2010-23). CEPII. Retrieved from <http://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=2726>
- Hummels, D., Ishii, J., & Yi, K.-M. (2001). The nature and growth of vertical specialization in world trade. *Journal of international Economics*, 54(1), 75–96.
- Huo, J., Chen, P., Hubacek, K., Zheng, H., Meng, J., & Guan, D. (2022). Full-scale, near real-time multi-regional input–output table for the global emerging economies (emerging). *Journal of Industrial Ecology*, 26(4), 1218–1232.
- IMF General Statistics Division. (1993). *Direction of trade statistics*. International Monetary Fund. Retrieved from <https://data.imf.org/?sk=9d6028d4f14a464ca2f259b2cd424b85>
- Koopman, R., Wang, Z., & Wei, S.-J. (2014). Tracing value-added and double counting in gross exports. *American Economic Review*, 104(2), 459–94.
- Krantz, S. (2023). Africa's great moderation. *Journal of African Economies*, ejad021.
- Krantz, S. (2024). *Patterns of global and regional integration in the east african community* (Tech. Rep.). Kiel Working Paper.
- Leontief, W. W. (1936). Quantitative input and output relations in the economic systems of the united states. *The review of economic statistics*, 105–125.
- Mancini, M., Montalbano, P., Nenci, S., & Vurchio, D. (2024). Positioning in global value chains: World map and indicators, a new dataset available for GVC analyses. *The World Bank Economic Review*, lhae005.
- Miller, R. E., & Temurshoev, U. (2017). Output upstreamness and input downstreamness of industries/countries in world production. *International Regional Science Review*, 40(5), 443–475.
- Rodrik, D. (2018). An african growth miracle? *Journal of African Economies*, 27(1), 10–27.

Table 6: EMERGING Sectors (2-Digit HS2002) Mapping to Broad Sectors

HS02	EMERGING Sector Definition	BSC	Broad Sector Definition of Huo et al. (2022)
1	Live Animals	AFF	Agriculture, Hunting, Forestry & Fishing
2	Meat and Edible Meat Offal	FBE	Food Production, Beverages & Tobacco
3	Fish, Crustaceans, Molluscs, Aquatic Invertebrates Ne	AFF	Agriculture, Hunting, Forestry & Fishing
4	Dairy Products, Eggs, Honey, Edible Animal Product Ne	FBE	Food Production, Beverages & Tobacco
5	Products of Animal Origin, Nes	AFF	Agriculture, Hunting, Forestry & Fishing
6	Live Trees, Plants, Bulbs, Roots, Cut Flowers Etc	AFF	Agriculture, Hunting, Forestry & Fishing
7	Edible Vegetables and Certain Roots and Tubers	AFF	Agriculture, Hunting, Forestry & Fishing
8	Edible Fruit, Nuts, Peel of Citrus Fruit, Melons	AFF	Agriculture, Hunting, Forestry & Fishing
9	Coffee, Tea, Mats and Spices	FBE	Food Production, Beverages & Tobacco
10	Cereals	FBE	Food Production, Beverages & Tobacco
11	Milling Products, Malt, Starches, Inulin, Wheat Glute	FBE	Food Production, Beverages & Tobacco
12	Oil Seeds, Oligenic Fruits, Grains, Seed, Fruit, Etc, Ne	AFF	Agriculture, Hunting, Forestry & Fishing
13	Lac, Gums, Resins, Vegetable Saps and Extracts Nes	FBE	Food Production, Beverages & Tobacco
14	Vegetable Plant Materials, Vegetable Products Nes	FBE	Food Production, Beverages & Tobacco
15	Mineral Oil, Vegetable Fats and Oils, Cleaning Products, et	FBE	Food Production, Beverages & Tobacco
16	Meat, Fish and Seafood Food Preparations Nes	FBE	Food Production, Beverages & Tobacco
17	Sugars and Sugar Confectionery	FBE	Food Production, Beverages & Tobacco
18	Cocoa and Cocoa Preparations	FBE	Food Production, Beverages & Tobacco
19	Cereal, Flour, Starch, Milk Preparations and Products	FBE	Food Production, Beverages & Tobacco
20	Vegetable, Fruit, Nut, Etc Food Preparations	FBE	Food Production, Beverages & Tobacco
21	Miscellaneous Edible Preparations	FBE	Food Production, Beverages & Tobacco
22	Beverages, Spirits and Vinegar	FBE	Food Production, Beverages & Tobacco
23	Residues, Wastes of Food Industry, Animal Fodder	FBE	Food Production, Beverages & Tobacco
24	Tobacco and Manufactured Tobacco Substitutes	FBE	Food Production, Beverages & Tobacco
25	Salt, Sulphur, Earth, Stone, Plaster, Lime and Cement	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
26	Ores, Slag and Ash	PSM	Precious Stones & Base Metals Incl. Compounds
27	Mineral Fuels, Oils, Distillation Products, Etc	MIN	Mining & Quarrying
28	Inorganic Chemicals, Precious Metal Compound, Isotope	PSM	Precious Stones & Base Metals Incl. Compounds
29	Organic Chemicals	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
30	Pharmaceutical Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
31	Fertilizers	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
32	Tanning, Dyeing Extracts, Tannins, Derivs.pigments et	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
33	Essential Oils, Perfumes, Cosmetics, Toiletries	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
34	Soaps, Lubricants, Waxes, Candles, Modelling Pastes	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
35	Albuminoids, Modified Starches, Glues, Enzymes	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
36	Explosives, Pyrotechnics, Matches, Pyrophorics, Etc	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
37	Photographic or Cinematographic Goods	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
38	Miscellaneous Chemical Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
39	Plastics and Articles Thereof	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
40	Rubber and Articles Thereof	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
41	Raw Hides and Skins (Other than Furskins) and Leather	TEX	Textiles, Leather & Wearing Apparel
42	Articles of Leather, Animal Gut, Harness, Travel Good	TEX	Textiles, Leather & Wearing Apparel
43	Furskins and Artificial Fur, Manufactures Thereof	TEX	Textiles, Leather & Wearing Apparel
44	Wood and Articles of Wood, Wood Charcoal	WAP	Wood, Paper & Publishing
45	Cork and Articles of Cork	WAP	Wood, Paper & Publishing
46	Manufactures of Plaiting Material, Basketwork, Etc.	WAP	Wood, Paper & Publishing
47	Pulp of Wood, Fibrous Cellulosic Material, Waste Etc	WAP	Wood, Paper & Publishing
48	Paper & Paperboard, Articles of Pulp, Paper and Board	WAP	Wood, Paper & Publishing
49	Printed Books, Newspapers, Pictures Etc	WAP	Wood, Paper & Publishing
50	Silk	TEX	Textiles, Leather & Wearing Apparel
51	Wool, Animal Hair, Horsehair Yarn and Fabric Thereof	TEX	Textiles, Leather & Wearing Apparel
52	Cotton	TEX	Textiles, Leather & Wearing Apparel
53	Vegetable Textile Fibres Nes, Paper Yarn, Woven Fabri	TEX	Textiles, Leather & Wearing Apparel
54	Mannmade Filaments	TEX	Textiles, Leather & Wearing Apparel
55	Mannmade Staple Fibres	TEX	Textiles, Leather & Wearing Apparel
56	Wadding, Felt, Nonwovens, Yarns, Twine, Cordage, Etc	TEX	Textiles, Leather & Wearing Apparel
57	Carpets and Other Textile Floor Coverings	TEX	Textiles, Leather & Wearing Apparel
58	Special Woven or Tufted Fabric, Lace, Tapestry Etc	TEX	Textiles, Leather & Wearing Apparel
59	Integrated, Coated or Laminated Textile Fabric	TEX	Textiles, Leather & Wearing Apparel
60	Knitted or Crocheted Fabrics	TEX	Textiles, Leather & Wearing Apparel
61	Articles of Apparel, Accessories, Knit or Crochet	TEX	Textiles, Leather & Wearing Apparel
62	Articles of Thread, Yarns, not Knit or Crochet	TEX	Textiles, Leather & Wearing Apparel
63	Other Made Textile Articles, Sets, Worn Clothing Etc	TEX	Textiles, Leather & Wearing Apparel
64	Footwear, Gauntlets and the Like, Parts Thereof	TEX	Textiles, Leather & Wearing Apparel
65	Headgear and Parts Thereof	TEX	Textiles, Leather & Wearing Apparel
66	Umbrellas, Walking-Sticks, Seat-Sticks, Whips, Etc	TEX	Textiles, Leather & Wearing Apparel
67	Bird Skin, Feathers, Artificial Flowers, Human Hair	TEX	Textiles, Leather & Wearing Apparel
68	Stone, Plaster, Cement, Asbestos, Mica, Etc Articles	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
69	Ceramic Products Undate	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
70	Glass and Glassware	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
71	Pearls, Precious Stones, Metals, Coins, Etc	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
72	Iron and Steel	PSM	Precious Stones & Base Metals Incl. Compounds
73	Articles of Iron or Steel	MPR	Metal & Metal Products
74	Copper and Articles Thereof	PSM	Precious Stones & Base Metals Incl. Compounds
75	Nickel and Articles Thereof	PSM	Precious Stones & Base Metals Incl. Compounds
76	Aluminium and Articles Thereof	MPR	Metal & Metal Products
77	Lead and Articles Thereof	PSM	Precious Stones & Base Metals Incl. Compounds
78	Zinc and Articles Thereof	PSM	Precious Stones & Base Metals Incl. Compounds
79	Tin and Articles Thereof	MPR	Metal & Metal Products
80	Other Base Metals, Cermets, Articles Thereof	PSM	Precious Stones & Base Metals Incl. Compounds
81	Tools, Implements, Cutlery, Etc of Base Metal	MPR	Metal & Metal Products
82	Miscellaneous Articles of Base Metal	MPR	Metal & Metal Products
83	Nuclear Reactors, Boilers, Machinery, Etc	ELM	Electrical & Machinery
84	Electrical, Electronic Equipment	ELM	Electrical & Machinery
85	Railway, Tramway Locomotives, Rolling Stock, Equipment	TEQ	Transport Equipment
86	Vehicles Other than Railways, Tramway	TEQ	Transport Equipment
87	Aircraft, Spacecraft, and Parts Thereof	TEQ	Transport Equipment
88	Ships, Boats and Other Floating Structures	TEQ	Transport Equipment
89	Optical, Photo, Technical, Medical, Etc Apparatus	ELM	Electrical & Machinery
90	Clocks and Watches and Parts Thereof	ELM	Electrical & Machinery
91	Musical Instruments, Parts and Accessories	ELM	Electrical & Machinery
92	Arms and Ammunition, Parts and Accessories Thereof	ELM	Electrical & Machinery
93	Furniture, Lighting, Signs, Prefabricated Buildings	MAN	Manufacturing & Recycling
94	Toys, Games, Sports Supplies	MAN	Manufacturing & Recycling
95	Miscellaneous Manufactured Articles	MAN	Manufacturing & Recycling
96	Works of Art, Collectors Pieces and Antiques	MAN	Manufacturing & Recycling
97	Commodities not Specified According to Kind	EGW	Electricity, Gas & Water
98	Electricity	EGW	Electricity, Gas & Water
99	Gas Manufacture, Distribution	EGW	Electricity, Gas & Water
100	Water Collection, Purification, and Distribution	MIN	Mining & Quarrying
101	Coal	MIN	Mining & Quarrying
102	Oil	MIN	Mining & Quarrying
103	Gas	MIN	Mining & Quarrying
104	Petroleum, Coal Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
105	Manufacturing Services on Physical Inputs Owned by Others	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
106	Maintenance and Repair Services N.i.e.	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
107	Sea Transport	TRA	Transport
108	Air Transport	TRA	Transport
109	Other Modes of Transport	TRA	Transport
110	Postal and Courier Services	PTE	Post & Telecommunications
111	Courts (Travel)	TRA	Transport
112	Local Transport Services	TRA	Transport
113	Accommodation Services	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
114	Food-Serving Services	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
115	Construction	CON	Construction
116	Direct Insurance	FIB	Financial Intermediation & Business Activity
117	Pension and Standardized Guaranteed Services	FIB	Financial Intermediation & Business Activity
118	Financial Services	FIB	Financial Intermediation & Business Activity
119	Real Estate	FIB	Financial Intermediation & Business Activity
120	Charges for the Use of Intellectual Property N.i.e.	FIB	Financial Intermediation & Business Activity
121	Telecommunications Services	PTE	Post & Telecommunications
122	Computer Services	PTE	Post & Telecommunications
123	Information Services	FIB	Financial Intermediation & Business Activity
124	Research and Development Services	FIB	Financial Intermediation & Business Activity
125	Professional and Management Consulting Services	FIB	Financial Intermediation & Business Activity
126	Engineering	FIB	Financial Intermediation & Business Activity
127	Waste Treatment and De-Pollution Agricultural and Mining Services	PAO	Public Administration; Education; Health; Recreation; Other Services
128	Operational Leasing Services	FIB	Financial Intermediation & Business Activity
129	Other Business Services N.i.e.	PAO	Public Administration; Education; Health; Recreation; Other Services
130	Audiosvisual and Related Services	PAO	Public Administration; Education; Health; Recreation; Other Services
131	Health Services	PAO	Public Administration; Education; Health; Recreation; Other Services
132	Education Services	PAO	Public Administration; Education; Health; Recreation; Other Services
133	Recreation & Other Services	PAO	Public Administration; Education; Health; Recreation; Other Services
134	Government Goods and Services N.i.e.	PAO	Public Administration; Education; Health; Recreation; Other Services

# Patterns of Global and Regional Integration in the East African Community

Sebastian Krantz\*

July 20, 2024

## Abstract

Using detailed global trade and novel Multi-Region Input-Output (MRIO) data, this paper examines the East African Community's (EAC) global and regional integration through trade, global, and regional value chains (GVCs and RVCs). With surgical attention to detail, the first part of the paper dissects key patterns and trends of EAC members' participation in global and regional trade and production networks at the aggregate, bilateral, sectoral, and bilateral-sectoral levels. The second part then provides causal reduced-form evidence for the economic benefits of EAC integration through trade, GVCs, and RVCs at the sector level. Findings imply that the region is moderately integrated into GVCs and RVCs but shows no overall trend towards greater integration. Regional integration is advancing in agriculture and food processing, and Kenya is becoming a more dominant regional supplier of manufactures. Integration through trade and GVCs positively affects economic development in the region, particularly deeper forward GVC linkages in manufacturing. Deepening regional trade and forward linkages yields additional economic benefits vis-a-vis global linkages.

**Keywords:** GVCs, RVCs, EAC, trade, regional integration, economic development

**JEL Classification:** F14; F15; O11

## 1 Introduction

Global Value Chains (GVCs), referring to the internationalization of production networks, have become a central topic in trade and development policy. With the entry into force of the African Continental Free Trade Area (AfCFTA) in May 2019 and some progress towards its full enactment, the potential of a large common market in Africa for increased GVC-related trade, both within Africa and between Africa and the world, is of great interest to economic researchers and policymakers. To gauge the potential implications and distributional side-effects of AfCFTA for trade and GVCs, it is instructive to study smaller efforts of regional integration and creation of common markets in Africa, as has been the case in East Africa with the East African Community (EAC).

(Re-)founded in 2000 by Uganda, Kenya, and Tanzania as a body to facilitate regional cooperation, the EAC quickly became a vehicle for economic integration. A customs union became operational in January 2005, with Kenya, the region's largest exporter, continuing to pay duties on some goods entering other countries on a declining scale until 2010 (EAC Customs Union Protocol, Article 11 and [Aloo \(2017\)](#)). Rwanda and Burundi acceded in 2007, joining the customs union in 2009. The customs union expanded to a common market for goods, labor, and capital effective in 2010. In 2013, the Protocol for the Establishment of the EAC Monetary Union was signed, aiming for a monetary union within 10 years, subject to macro-fiscal convergence criteria. In 2016, the newly founded Republic of South Sudan joined the EAC, and the Democratic Republic of Congo joined in July 2022. Thus, the EAC, particularly the years following the customs union in 2005 and the common market in 2010, provides a small case study in light of AfCFTA's broader aims.

There is, by now, extensive academic and policy literature on the state, determinants, and consequences of integration into GVCs, including for countries at different income levels. As one of the first, [Kummritz & Quast \(2016\)](#) examine patterns of GVC integration in low- and middle-income countries (LMICs) using the OECD TiVA database. They find that LMICs have become an integral part of GVCs and are driving their expansion, with a rising share in both the foreign

---

\*Kiel Institute for the World Economy  
Address: Haus Welt-Club, Duesternbrooker Weg 148, D-24105 Kiel  
E-mail: sebastian.krantz@ifw-kiel.de

content of global value added (VA) exports (9% in 1995 to 24% in 2011) and re-exported exports (9% to 23%). High-income economies use GVCs to outsource low-VA downstream production stages. However, over time, many developing economies move up the value chain.

The 2020 World Development Report (WDR), focusing on GVCs, classifies Africa as primarily a supplier of raw materials, with only a handful of countries (Morocco, Tunisia, Namibia, South Africa, Ethiopia, Kenya, and Tanzania) engaging in limited manufacturing ([World Bank, 2020](#)). At the same time, GDP per capita grows most rapidly when countries enter limited manufacturing GVCs. The report estimates the average benefits from a 1 percent increase in GVC participation to boost per capita income by more than 1 percent, much more than the 0.2 percent income gain from standard trade. To enter GVCs, the report stipulates attracting FDI, improving access to finance, keeping labor costs low, trade liberalization, investments in ICT and transport infrastructure, and political stability. African economies score low in all of these dimensions. In particular, overvalued exchange rates and restrictive labor regulations raise the cost of labor: "Manufacturing labor costs in Bangladesh are in line with its per capita income, but in many African countries, labor costs are more than twice as high." ([World Bank, 2020](#)).

These policy conclusions are broadly echoed in much early and recent academic work. E.g., [Fernandes et al. \(2022\)](#), using a panel with more than 100 countries and a novel identification strategy, show that factor endowments, geography, political stability, liberal trade policies, FDI inflows, and domestic industrial capacity are key determinants of GVC participation, whereas traditional exports are less important. The findings are commensurate with [Antràs & De Gortari \(2020\)](#), which develop a general-equilibrium framework where trade costs imply a concentration of downstream production stages in central locations/countries (close to final demand). [Kowalski et al. \(2015\)](#) also find that proximity to manufacturing hubs in Europe, North America, and East Asia, domestic market size, and the level of development, are key determinants of GVC participation.

[Foster-McGregor et al. \(2015\)](#) provide one of the first comprehensive analyses of GVCs in Africa, using the EORA 25 sector database over 2000-2011. They find that Africa is more involved in GVCs than many other developing regions but mainly supplies primary goods. Downstream involvement is relatively small and shows little improvement in 1995-2011. GVC involvement is also very heterogeneous across African countries, with some relatively successful countries (Tunisia, South Africa) heavily involved in (downstream) GVCs. Inner-African GVCs are also small in most African countries, with several exceptions in southern Africa. The EU is Africa's biggest GVC partner, with increasing shares of (South-)East Asia and other transition countries.

[Kowalski et al. \(2015\)](#) study GVC participation in Africa, the Middle East, and Asia, showing that developing countries reap important benefits from GVC participation through both forward and backward linkages, including enhanced productivity, export diversification, and sophistication. Analyzing export competitiveness, they find that Asia dominates more advanced products such as electronic equipment or motor vehicles. In contrast, African and Middle Eastern regions are competitive in agriculture, food processing, and less advanced manufacturing. While all regions have become more competitive, they find no trend towards GVC-led industrialization in Africa.

[Balié et al. \(2019\)](#) present a careful analysis of bilateral-sectoral GVC linkages in SSA with an emphasis on food processing GVCs and show that SSA's participation in these chains is substantial. This is driven by a handful of countries, including Kenya and Uganda, where the share of agriculture in total GVC participation is 30%, and in Kenya, the food processing sector is at 15%. They further show that bilateral trade policy is a key determinant in shaping SSA's GVC integration in the food sector, with high tariffs detrimental to GVC participation. They also echo [Foster-McGregor et al. \(2015\)](#) that SSA GVC participation is high - at 40%, comparable to China and India. Africa is also the continent with the highest forward integration - around 25% of domestic VA (DVA) produced in SSA are inputs for other countries' exports and over 35% in North Africa.

There has also been some GVC and RVC-related work on African regional economic communities (RECs). Notably, [Obasaju et al. \(2021\)](#) examine the impact of regional integration on upgrading through GVCs (proxied by DVA in exports per capita) in the EAC, Southern African Customs Union (SACU) and Economic Community of West African States (ECOWAS) in 2000-2015. They show that regional integration and FDI are not significant drivers of upgrading but lagged backward

GVC participation is. They also find weak positive effects of regional integration on labor productivity in the EAC and SACU (the communities with stronger trade integration). Regional hegemons (Kenya, South Africa, and Nigeria) have weak backward linkages with other members.

[Tinta \(2017\)](#) studies determinants of GVC participation in ECOWAS and finds that intra-regional trade is not a significant predictor of trade openness, but backward GVC participation is. Further, trade diversification is a key predictor of backward GVC participation. [Engel et al. \(2016\)](#) provide a detailed analysis of GVC integration, position, and performance of SACU members (Botswana, Lesotho, Namibia, South Africa and Eswatini). They show that the SACU region is moderately integrated into GVCs in relatively upstream tasks, but the scale and nature of integration vary by country, with South Africa and Namibia being the most integrated. South Africa remains a moderately important player in global trade networks and an important regional hub. Lesotho shows a rapid increase in GVC integration, Namibia a moderate increase, whereas Botswana and Eswatini appear stagnant or in decline. Overall growth in GVC participation in services is stronger than in manufacturing. South Africa is the only country with strong forward GVC integration and a major source of foreign content for the other members, which are more integrated into RVCs than GVCs. China has grown significantly as a source of foreign content, but the EU remains the predominant partner for forward GVC participation.

[Lwesya \(2022\)](#) studies GVC integration in the EAC with respect to economic upgrading using UNCTAD-Eora data from 2005 to 2018. This analysis is largely complementary to the one carried out in this paper. He computes measures of backward and forward integration and finds that Kenya, Tanzania, and Uganda are relatively better integrated into GVCs, with Kenya having the deepest level of integration, especially in terms of indirect VA and forward integration. Overall, the EAC's participation in GVCs is in upstream low- and middle-VA production activities. Using a cross-country panel regression framework predicting DVA in exports, which includes GVC indicators and other macroeconomic indicators, he finds that domestic credit, foreign direct investment, the quality of institutions, and foreign VA (FVA) have significant positive effects, but observes no such effects for measures of human capital, infrastructure quality, and GDP per capita. The analysis is focused on economic upgrading and does not provide a detailed bilateral and sector-level exposition of the region's integration into GVCs and RVCs. He also does not provide a detailed examination of how different forms of trade and GVC participation affect economic growth, and does not establish economic causality between any of the studied factors and DVA in exports.

This paper adds to our understanding of GVCs and regional integration in the EAC in the following significant respects: (1) it uses better data, including gross trade flows data and the EMERGING MRIO tables, which include IO/SUT/SAM tables for 4 EAC countries; (2) it conducts a detailed examination of EAC members global and regional integration using both gross trade flows and VA content shares, paying close attention to specific bilateral linkages and sector-level patterns; (3) it constructs metrics to track regional integration in VA terms and uses them to measure progress in recent years; (4) It examines the positioning of EAC members and sectors in GVCs and (5) revealed comparative advantage in gross and VA terms; (6) It analyzes the effect of conventional trade, GVC, and RCV integration on GDP using a bilateral-sector-level regression framework with triple fixed effects and instrumental variables for GVC participation following [Kummritz \(2016\)](#). Thus, it presents a rigorous and detailed study of the region's global and regional integration through trade and value chains using the best currently available data and attempts to establish economic causality between different forms of trade and economic growth.

## 2 Data

Most GVC analysis uses Inter-Country Input-Output tables (ICIOs), such as those published by the OECD (TiVA) or the World Input-Output Database (WIOD) ([Timmer et al., 2012](#)). These, however, focus on OECD countries, with very limited coverage of SSA. This paper, therefore, uses two Multi-Region Input-Output (MRIO) databases that are global in scope.

The first is the EORA 26 Global MRIO ([Lenzen et al., 2012, 2013](#)), which has extensive coverage of 189 countries and 26 sectors from 1990-2015 and uses 74 country IOT/SUTs and detailed international macroeconomic and trade data as input. EORA relies on sophisticated methods to impute, harmonize, and interpolate data across countries and time and is thus less

accurate than the OECD or WIOD tables. Particularly for small countries like EAC members, data can be highly distorted. The Kenya 2010 IOT is the only source of EAC national data used in EORA ([Lenzen et al., 2013](#)). A 2021 EORA update added administrative data through 2018 and WEO-based forecasts through 2021. It introduced a large structural break in the time series in 2016, with different macroeconomic totals and GVC indicators for EAC members. The analysis thus emphasizes the initial release through 2015. Since the EAC customs union only became operational in 2005, I consider EORA 26 tables from 2000 onwards. Data from 1990 shows no interesting trends in GVC engagement. EORA is denominated in thousands of current USD at basic prices<sup>1</sup>. Appendix Figure [A1](#) shows the official EORA data quality reports for 6 EAC countries<sup>2</sup>. Despite its shortcomings, EORA has enabled significant research on GVCs in Africa.

Due to the shortcomings of EORA in terms of accuracy and usage of national data for developing countries, I also employ the more recently introduced EMERGING (EM) MRIO tables ([Huo et al., 2022](#)). This impressive effort has created a global MRIO database covering 245 countries and territories in 135 sectors for the years 2015-2019. A recent update (v2) also provides a table for 2010. EM uses 111 national IO/SUT/SAM tables alongside detailed trade and macroeconomic data. In particular, the UN Comtrade database is utilized to the fullest extent to provide greater sectoral detail than EORA. Macroeconomic data from national statistical offices is used where available and reconciled (scaled) using World Bank data. The purpose of the MRIO is to provide greater detail and accuracy for emerging economies than EORA. From EAC countries, EM uses a SAM and sectoral GDP from Uganda up to 2016 and the same information up to 2019 for Rwanda and Kenya. For Tanzania, EM uses a SAM and an IO table up to 2017. For Burundi, Congo (DR), and South Sudan, only international data is available. Thus, EM incorporates, to the greatest extent possible, national data from these EAC countries in a harmonized global MRIO framework. EM is denominated in millions of current USD at basic prices.

The WDR also provides GVC indicators using EORA 2015, and [Mancini et al. \(2024\)](#) provide corresponding GVC positioning indicators following [Fally \(2012\)](#), [Antràs et al. \(2012\)](#) and [Antràs & Chor \(2013, 2018\)](#). These pre-computed indicators are used to verify manually computed indicators. This is important because I aggregate the non-EAC World and/or sectoral resolution for different indicators to lift computational constraints<sup>3</sup> and enable comparisons across databases. In particular, for backward GVC indicators the non-EAC World is aggregated into 11 geographic and trade regions summarised in Table 1, and, in more detail, in Appendix Table [A1](#).

Table 1: REGIONAL AGGREGATION

Region	Description	Countries & Territories	
		EORA	EMERGING
EAC	East African Community	7	7
SSA	Sub-Saharan Africa (Excluding EAC)	38	41
EUU	European Union + UK	28	29
ECA	Europe and Central Asia (Non-EU)	26	29
MEA	Middle East and North Africa	20	21
NAC	North America and Canada	3	13
LAC	Latin America and Caribbean	32	44
ASE	ASEAN	10	10
SAS	South Asia	8	9
CHN	China	3	3
ROA	Rest of Asia	7	14
OCE	Oceania	6	22
SUM:	7 EAC Members + 11 World Regions	188	245

<sup>1</sup>The basic price is the amount receivable by the producer from the purchaser for a unit of a good or service produced, as output minus any tax payable, and plus any subsidy receivable. It excludes any transport charges invoiced separately by the producer.

<sup>2</sup>While global GDP is broadly consistent with representative estimates, the GDP of EAC countries is highly distorted. Most notably, Tanzania's GDP is decreasing in the data. The situation is better for exports, whose level and sectoral composition are roughly consistent with estimates from other sources. Thus, detailed analysis and results from EORA should be treated with great caution, particularly for Tanzania.

<sup>3</sup>EMERGING has 245 countries/territories and 134 sectors, implying 32,830 rows and columns or 1 billion records in the transaction matrix. It is computationally infeasible for me to compute GVC indicators directly on these tables, and also the full EORA database (186 x 26 = 4836 rows and columns observed over 21 years) strains my computing resources for non-trivial GVC indicators.

To verify EM, (Huo et al., 2022) develop a broad sector classification of 17 sectors and mappings it to major global ICIOs (EXIOBASE3rx, OECD-TiVA, EORA, GTAP, and EM). I use these mappings to report results at the sector level. Most GVC indicators are computed at the full sector resolution using STATA's ICIO package (Belotti et al., 2020) and the default source-based exporter perspective (Borin & Mancini, 2019) also used in the WDR, which permits aggregation of GVC indicators across sectors. Table 2 shows the 26 EORA sectors<sup>4</sup> and their mapping to broad sectors. Appendix Table A2 shows the mapping for EM.

Table 2: EORA 26 SECTORS AND MAPPING TO BROAD SECTORS

Code	EORA 26 Sector Definition	Code	Broad Sector Definition of Huo et al. (2022)
AGR	Agriculture	AFF	Agriculture, Hunting, Forestry & Fishing
FIS	Fishing	AFF	Agriculture, Hunting, Forestry & Fishing
MIN	Mining and Quarrying	MIN	Mining & Quarrying
FBE	Foods & Beverages	FBE	Food Production, Beverages & Tobacco
TEX	Textiles and Wearing Apparel	TEX	Textiles, Leather & Wearing Apparel
WAP	Wood and Paper	WAP	Wood, Paper & Publishing
PCM	Petroleum, Chemical and Non-Metallic Mineral Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
MPR	Metal Products	MPR	Metal & Metal Products
ELM	Electrical and Machinery	ELM	Electrical & Machinery
TEQ	Transport Equipment	TEQ	Transport Equipment
MAN	Other Manufacturing	MAN	Manufacturing & Recycling
REC	Recycling	MAN	Manufacturing & Recycling
EGW	Electricity, Gas and Water	EGW	Electricity, Gas & Water
CON	Construction	CON	Construction
MRE	Maintenance and Repair	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
WTR	Wholesale Trade	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
RTR	Retail Trade	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
AFS	Hotels and Restaurants	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
TRA	Transport	TRA	Transport
PTE	Post and Telecommunications	PTE	Post & Telecommunications
FIB	Financial Intermediation and Business Activities	FIB	Financial Intermediation & Business Activity
PAD	Public Administration	PAO	Public Administration; Education; Health; Recreation; Other Services
EHO	Education, Health and Other Services	PAO	Public Administration; Education; Health; Recreation; Other Services
PHH	Private Households	PAO	Public Administration; Education; Health; Recreation; Other Services
OTH	Others	PAO	Public Administration; Education; Health; Recreation; Other Services
REI	Re-Export & Re-Import	PAO	Public Administration; Education; Health; Recreation; Other Services

To complement and verify MRIO table results, I also use gross trade flow data from CEPII's BACI (Gaulier & Zignago, 2010) (HS 1996 version) and the IMF's Direction of Trade Statistics (DOTS) (IMF General Statistics Division, 1993). EM's goods-producing sectors are identical to the 2-digit HS codes so that BACI can be aggregated to match the MRIO databases using the mapping in Table A2. The DOTS database only records aggregate bilateral trade, with imports denominated in Cost Insurance Freight (CIF) terms (including transport and insurance costs).

### 3 Trade

In light of the known macroeconomic inconsistencies in EORA for EAC countries and that VA flows are estimated from gross flows, I begin by examining EAC integration through trade using BACI, DOTS, and gross total and intermediate flows from the EORA and EM MRIO tables.

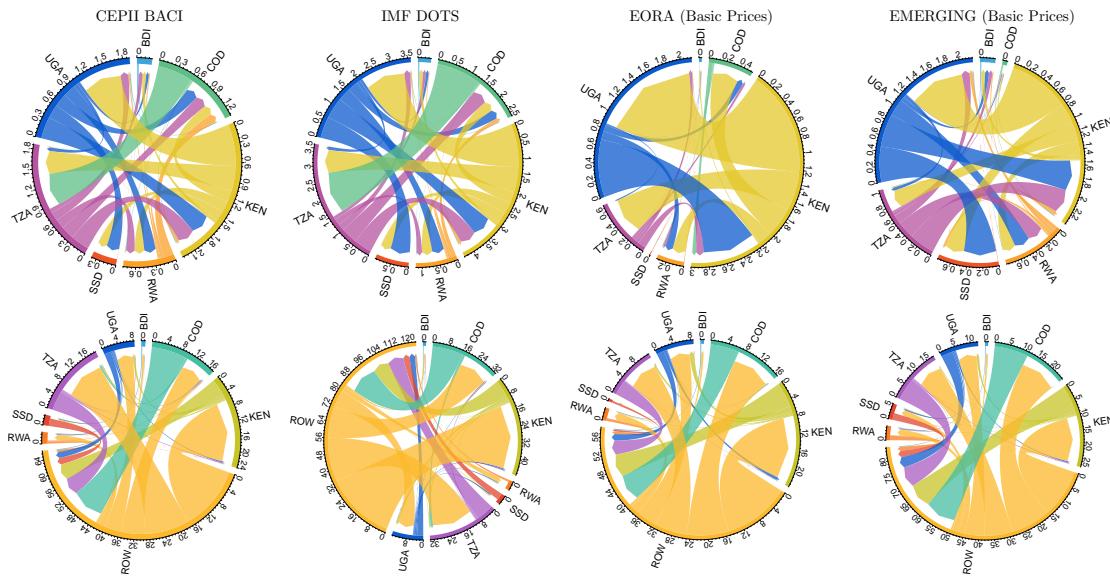
#### 3.1 Gross Trade Flows

Figure 1 shows diagrams of EAC trade flows averaged over 2010-2019. All databases emphasize Kenya as regional trading hegemon, followed by Uganda and Tanzania, but EORA gives disproportional weight to Kenya and shrinks the other countries, whereas EM overemphasizes Rwanda and Uganda a bit and also shrinks Congo. Notably, the large exports from Congo to Tanzania are not reflected in either EORA or EM. An examination of BACI reveals that 88% of this 550 million USD flow is copper and 7.4% precious metals. Congo does not trade with other EAC members to a similar extent.

The bottom half of Figure 1 includes the rest of the world (ROW) as a trading partner. In all databases, trade with ROW dwarfs inner-EAC trade. According to BACI, inner EAC trade is 15 times smaller than EAC trade with ROW, in DOTS 14.9 times, in EORA 17.5 times, and in EM it is 22.9 times smaller. When excluding South Sudan and Congo, the ratios increase to 18.6 (BACI), 19.5 (DOTS), 14.9 (EORA (decrease)), and 19.7 (EM).

<sup>4</sup>3-character sector codes are assigned and used throughout the paper based on the authors discretion, but not provided in the raw data. These codes are purely descriptive and do not correspond to any formal classification.

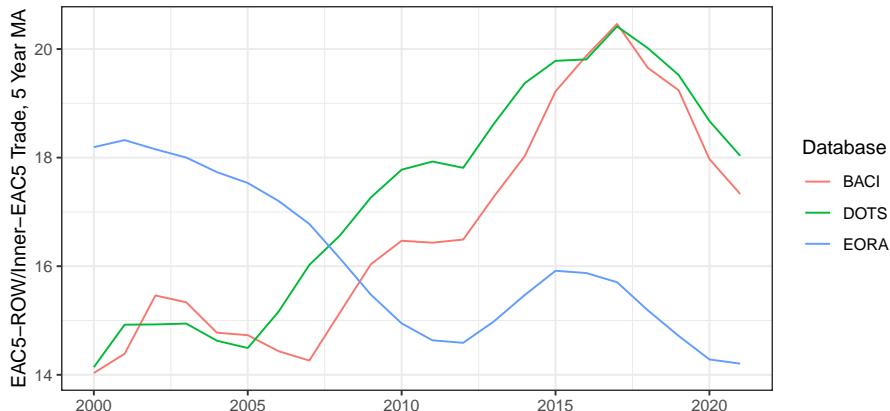
Figure 1: AVERAGE GROSS TRADE FLOWS FROM 2010-2019: 4 DATABASES: USD BILLIONS



*Notes:* Figure shows the mean of bilateral gross trade flows over years 2010-19 recorded in billions of current USD. The top panel shows inner-EAC flows; the bottom panel includes ROW as a trading partner. The circular axis records the total flows (exports + imports) for each partner. Produced using the *migest* R package (Abel, 2023).

Figure 2 shows the evolution of this ratio for the five early EAC members (EAC5), smoothed using a backward-looking 5-year moving average (MA). Up to 2017, trade with ROW has grown faster than inner-EAC5 trade. However, in 2018 and 2019, trade with ROW slowed a bit, and in 2020, the COVID-19 shock strengthened regional trading again. This can be disaggregated further by exports and imports, also considering individual members' EAC5 trade shares.

Figure 2: GROSS ROW-EAC5 TRADE TO INNER EAC5 TRADE RATIO

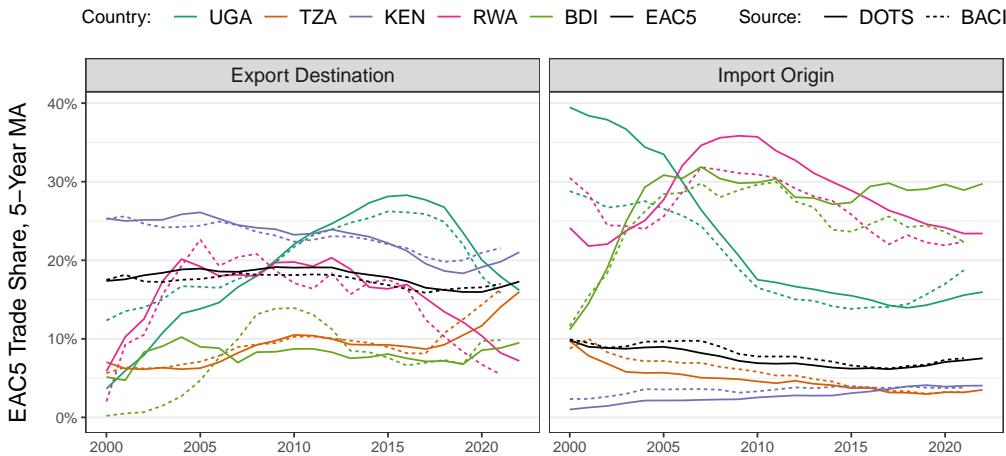


*Notes:* Figure shows the ratio of EAC5  $\leftrightarrow$  ROW to inner-EAC5 trade (exports + imports), smoothed using a backward-looking 5-year MA. The EAC5 includes Tanzania, Kenya, Uganda, Rwanda and Burundi.

Figure 3 shows the EAC5 share in members and total EAC5 exports and imports. Uganda substantially increased the share of its exports destined to EAC5 partners, reaching 28% in 2016 and falling again in recent years. Tanzania also increased its EAC5 export share from about 7% in 2000 to 15% in 2021. Kenya, on the other hand, decreased its EAC export share from 25% in 2000 to 20% in 2021. Rwanda shows a declining trend in both export and import share since 2012. In Burundi, the EAC5 share is constant since 2005. The total EAC5 shows a slight decline in regional export share from 2010 (18%) to 2020 (16%), and a clear decline in the import share from 10% in 2000 to 7.5% in 2020. Both shares slightly increased thereafter, consistent with Figure 2. The imports decline is driven by Uganda and Tanzania, while Kenya increased its EAC5 import share from 1% in 2000 to 4% in 2020. The aggregate pattern echoes Obasaju et al. (2021)'s observation that the regional hegemon (Kenya) has weak backward linkages (imports) with other

REC members. Also, as documented by [Engel et al. \(2016\)](#), the hegemon has strong regional forward linkages (exports), whereas smaller economies (Rwanda, Burundi) are more regionally focused, particularly through high import shares. Over time, this pattern has weakened slightly, but trade integration has not improved in overall terms.

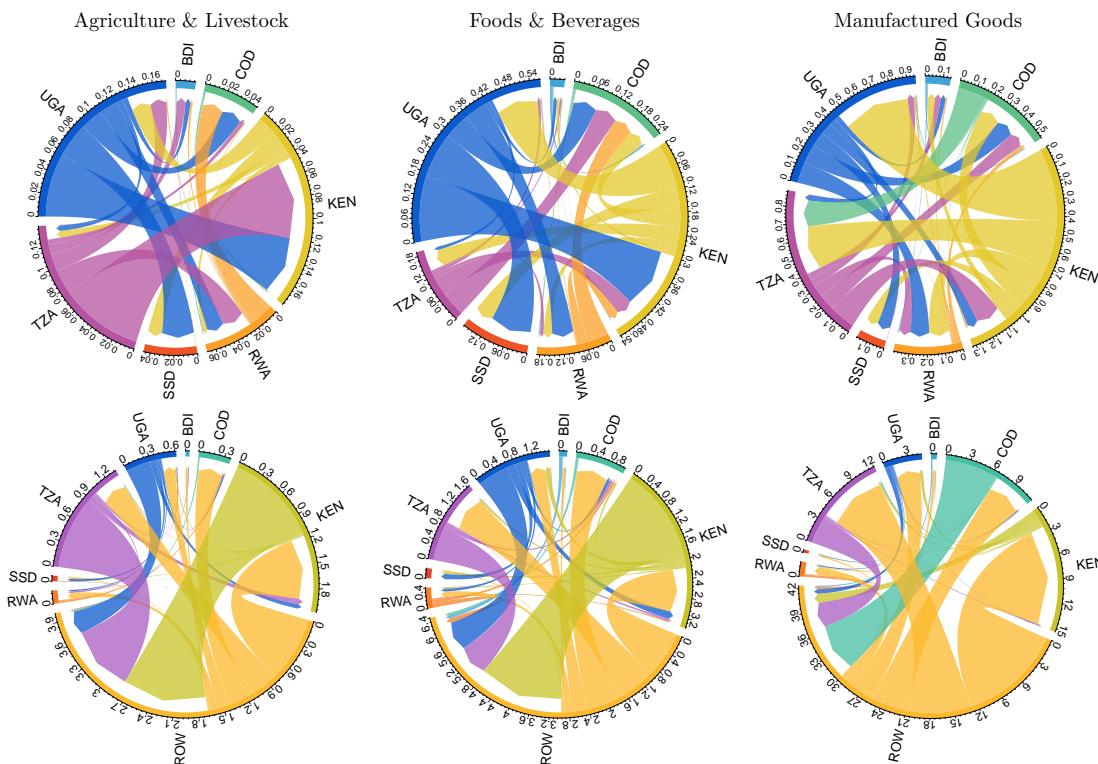
Figure 3: EAC5 SHARE IN MEMBERS GROSS TRADE FLOWS



*Notes:* Figure shows the EAC5 share in members total exports and imports, smoothed using a backward-looking 5-year MA. The black line also shows the EAC5's total share of exports and imports with itself.

When dividing trade flows broadly into agricultural products (AFF), processed foods and beverages (FBE), and manufactured goods, some further heterogeneity emerges. Figure 4 shows these flows using BACI, Appendix Figures B1 and B2 using EORA and EM, respectively.

Figure 4: AVERAGE 2010-2019 BACI TRADE FLOWS BY BROAD SECTOR: USD BILLIONS



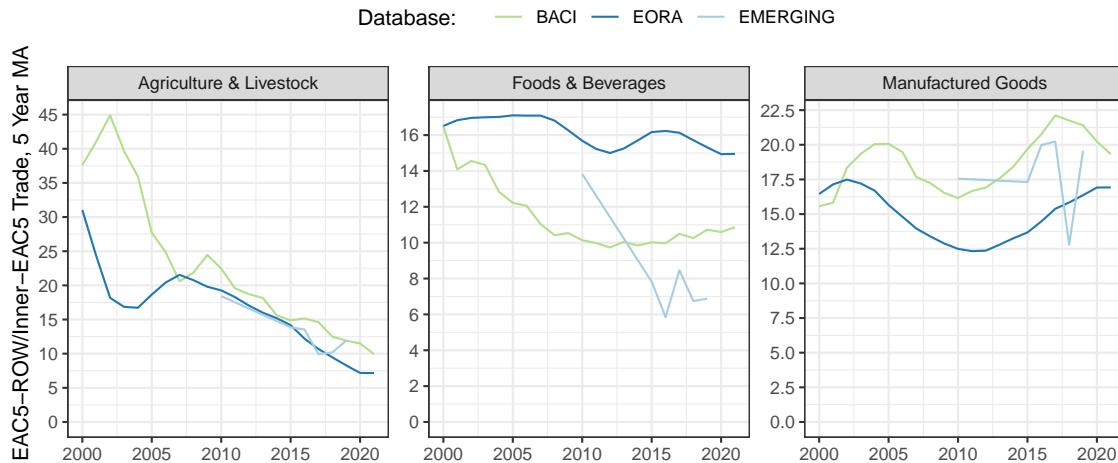
*Notes:* Figure shows the mean of bilateral gross trade flows over the years 2010-19 recorded in billions of current USD according to CEPII BACI. The top panel shows inner-EAC flows; the bottom panel includes ROW as a trading partner. The circular axis records the total flows (exports + imports) for each partner. The broad sectors shown are AFF (left), FBE (middle), and TEX-MAN (right) in Table 2. Produced using the *migest* R package ([Abel, 2023](#)).

According to all databases, Uganda and Tanzania are large regional suppliers of agricultural

produce. All countries have some stakes in FBE, with Uganda supplying the most, followed by Kenya. In manufacturing, Kenya has a distinct lead, followed by Tanzania and Uganda. With ROW, all EAC countries are large agricultural exporters and importers of manufactured products. Kenya is the largest EAC supplier of both agriculture and processed foods to ROW, whereas it only plays a minor supplier role in the EAC. Tanzania supplies large amounts of gold, and Congo large amounts of minerals to ROW, which are subsumed under MPR and PCM in Table 2, making Kenya also the largest EAC exporter of manufactures. The data thus expound differences in the nature of trade both within the EAC and with ROW. Shared capacities exist for FBE, which has also been the focus of policymakers and regional studies. For example [Daly et al. \(2017\)](#) show that Uganda exports diary and maize produce to Kenya for processing, but has also received FDI and begun to upgrade its own food processing sector. The Ugandan Ministry of Finance and Planning ([MoFPED, 2021](#)), IGC Uganda ([Fowler & Rauschendorfer, 2019](#)) and IFPRI ([Van Campenhout et al., 2020](#)) have identified agro-industrialization as an important pillar of growth for the country.

Figure 5 shows corresponding ratios of EAC5-ROW to inner-EAC5 trade, indicating that regional trade in agriculture and, to a lesser extent, FBE, assumes increasing shares of overall EAC trade in these sectors. According to BACI, in 2020, the inner-EAC5 trade in agricultural products was 10 times smaller than EAC5-ROW trade, down from almost 40 times smaller in 2000. Similarly, FBE inner-EAC5 trade was 11 times smaller in 2020, compared to 16 times smaller in 2000. In contrast, the ratio in manufacturing shows an oscillating increase from 15 in 2000 to 20 in 2020. These developments are also reflected in the MRIO databases.

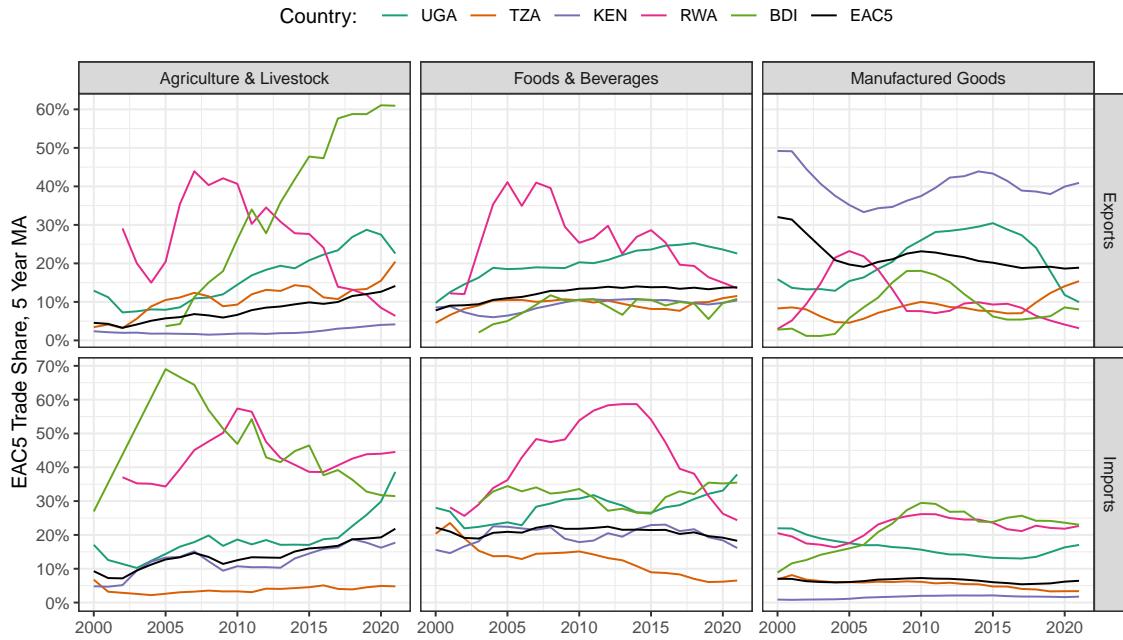
Figure 5: Gross ROW-EAC5 Trade to Inner EAC5 Trade Ratio by Broad Sector



*Notes:* Figure shows the ratio of EAC5  $\leftrightarrow$  ROW to inner-EAC5 trade (exports + imports), smoothed using a backward-looking 5-year MA. The EAC5 includes Tanzania, Kenya, Uganda, Rwanda and Burundi. The broad sectors shown are AFF (left), FBE (middle), and TEX-MAN (right) in Table 2.

Figure 6 again provides a detailed breakdown by exports/imports and individual members. In agriculture, export and import shares both increased: in 2015-20, the EAC5 exported 12.6% of agricultural exports to itself, up from 4.6% in 1995-2000, and imported 19.3%, up from 9.3% in 1995-2000. The FBE export shares also rose from 7.8% to 13.7%, whereas the import share remained constant around 20%. In manufacturing, the opposite is the case, with the EAC5 exports share declining from 32% to 18.6% and the import share remaining roughly constant at around 7%. At the country level, Uganda significantly increased its EAC5 share as an exporter and importer of both agricultural produce and FBE. This development is mirrored, to a lesser extent, by Kenya, which additionally maintains a very high EAC5 share in manufactured exports of around 40%, down from nearly 50% in 2000. This stands in stark contrast to a very small EAC5 import share of less than 1%. Tanzania increased its export share to the EAC5 in all 3 broad sectors while further decreasing its already low import shares in foods and manufactures to around 5%. Rwanda and Burundi have high export and import EAC5 shares in all sectors apart from manufacturing exports. Rwanda strongly decreased its EAC5 agriculture and foods export shares since 2007, approaching the levels of Kenya in 2020, whereas Burundi strongly increased its agricultural export share from almost 0% in 2005 to 60% in 2020, while decreasing its import share from 70% to 30%.

Figure 6: EAC5 SHARE IN MEMBERS GROSS TRADE BY SECTOR USING BACI DATA



*Notes:* Figure shows the EAC5 share in members' total exports and imports, smoothed using a backward-looking 5-year MA. The broad sectors shown are AFF (left), FBE (middle), and TEX-MAN (right) in Table 2.

Considering their different levels of development, this suggests that countries first become regional agricultural exporters and later suppliers of manufactured goods. However, it seems like these manufactures do not cater very well to other members' demands, as evidenced by the declining EAC5 shares in both exports and imports, and thus fail to become a driver of regional integration. The hegemonic position of Kenya as a supplier of manufactures may also crowd out other countries' attempts to increase their regional supply. Thus, gross trade data suggests that EAC regional integration through trade is asymmetric, has progressed mainly via agriculture and FBE, and is stronger in exports. Particularly, the larger economies of Tanzania and Kenya import much more from ROW. Among the EAC5, Tanzania is overall least integrated into regional trading.

### 3.2 Intermediate Flows

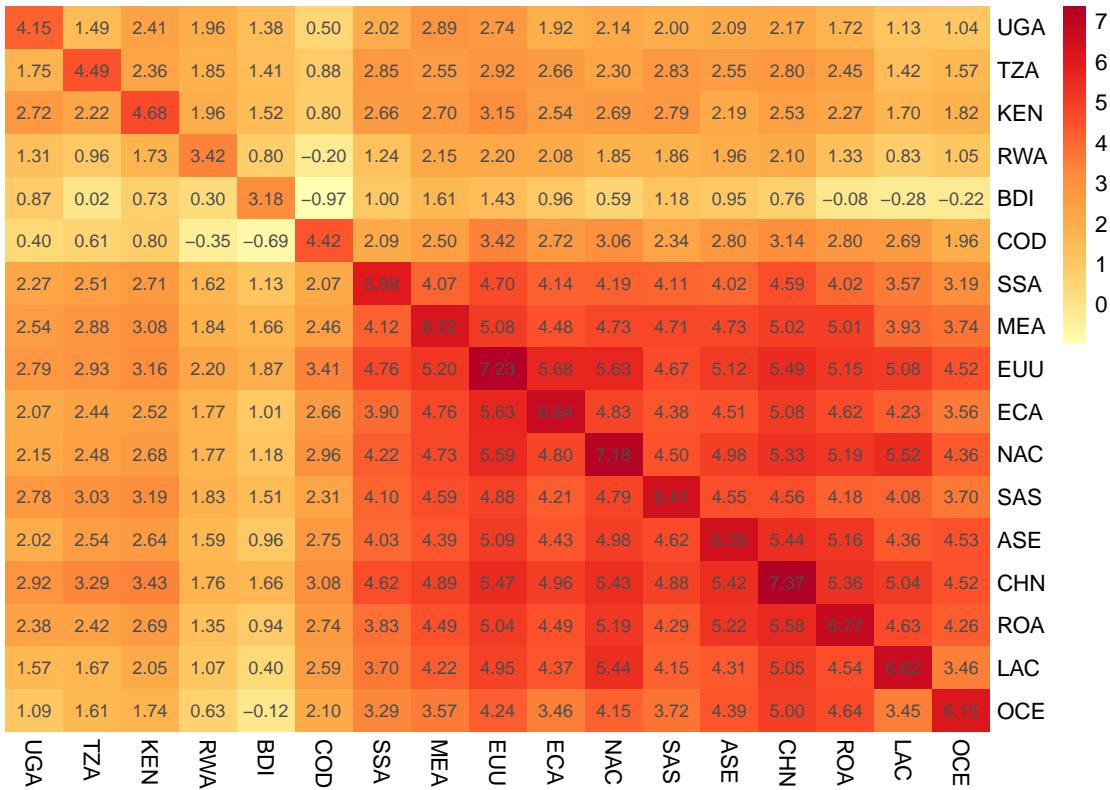
An advantage of MRIO databases is that they record gross trade in both intermediates and final goods. Due to its greater accuracy, I only examine such flows using the EM database, averaged across 2015-2019 to smooth temporal variation. Figure 7 provides an aggregate intermediate flows table. The columns indicate intermediate inputs required by each country or region from each row country or region. Conversely, the rows indicate intermediate quantities supplied.

Among the EAC countries, the table shows a significant supplier role of Kenya, supplying  $10^{2.72} = 524$  million USD to Uganda,  $10^{2.22} = 168$  million USD to Tanzania and  $10^{1.96} = 91$  million USD to Rwanda. Uganda/Tanzania also supplies 258/228 million to Kenya and 90/70 million to Rwanda. Tanzania supplies 56 million to Uganda, Rwanda 54 million to Kenya, and all other inner-EAC intermediate trade is below 35 million.<sup>5</sup> EM estimates intermediate trade with ROW to be 27.3 times greater than inner-EAC trade, composed of intermediate inputs from ROW summing to 15.6 times EAC intermediates trade and EAC inputs to ROW summing to 11.8 times EAC intermediates trade. The largest supplier of intermediates is China, supplying 831m to Uganda, 1937m to Tanzania, and 2704m to Kenya, followed by South Asia supplying 601/1066/1562, respectively, and the EU supplying 620/844/1462. Compared with these, the rest of SSA is relatively insignificant at 184/325/514. In terms of demand for EAC intermediates, the EU is the largest importer, importing 552/828/1402, followed by the Middle

<sup>5</sup>Exempting South Sudan, subsumed in SSA because of data quality concerns, which receives 385 million in intermediates from Uganda and 223 million from Kenya.

East and North Africa (769/352/503), South Asia (101/683/618), the rest of SSA (105/716/462) and China (148/631/340). China, notably, supplies 4.9 times more intermediates than it demands from these three economies. The supply and demand of intermediates with the EU, NAC, and SSA are quite balanced. Overall, Uganda, Tanzania, and Kenya combined demand 1.7 times more inputs from ROW than they supply. It should be noted that Congo, while not really integrated with other EAC members in terms of intermediates, has large and surprisingly balanced intermediate flows with ROW, demanding/supplying 2598/2636 with the EU and 1199/1375 with China.

Figure 7: AGGREGATED EMERGING MRIO TABLE: 2015-2019 AVERAGE  
*Log10 Millions of Current USD at Basic Prices*



*Notes:* Figure shows gross intermediate input flows in log10 USD millions. Rows indicate the sources, and columns the destinations of intermediates. The diagonal sums the domestic IO/regional ICIO table.

Despite its high use of foreign inputs, domestic intermediate inputs corresponding to the diagonal entries are, on average, 4.1 times greater than foreign inputs in EAC countries and 6.3 times greater than EAC inputs to other countries. For the region as a whole, these figures are 4.6 and 6.13, respectively. This is low compared to other major regions, which produce and trade a lot more within themselves. For example, in the EU and North America, own inputs are around 10 times greater than foreign inputs. For China, it is 14 times.

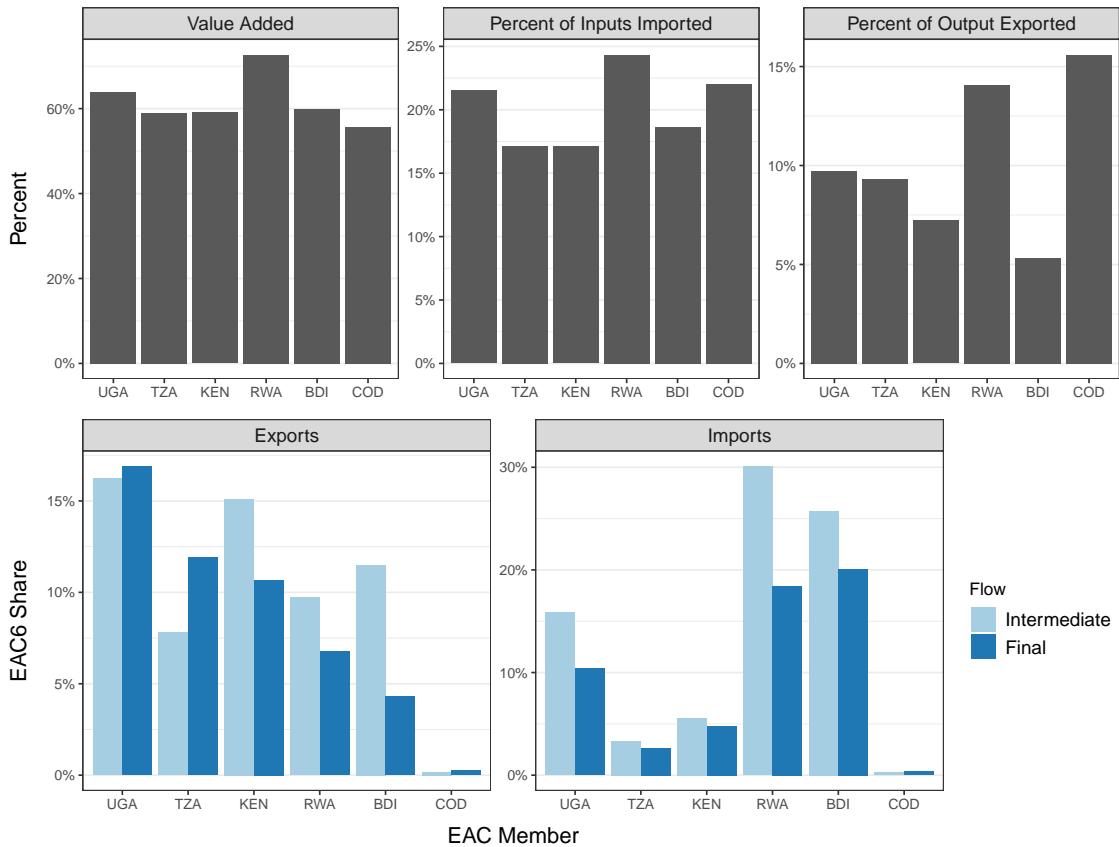
To provide some sector-level detail, Appendix Table B1 records the 50 largest sector-level intermediate flows (excl. Congo). Both with ROW and inside the EAC, the largest intermediate flows are in manufacturing and, in particular, in petrochemicals (PCM), FBE, and, to a lesser extent, textiles (TEX). Kenya is a significant EAC supplier of manufacturing inputs, particularly for PCM, FBE, and metal product (MPR) industries. Kenya also supplies large transport (TRA) (including travel and tourism) intermediates to EU TRA services and agricultural inputs to EU FBE industries. It also supplies large inputs for FBE industries in South Asia. These flows are, on average, 3-4 times larger than its regional intermediate supplies. Uganda supplies PCM to ROW, and FBE and agriculture to Kenyan FBE and TRA industries.

### 3.3 Aggregate Structure of Production and Trade

Figure 8 compactly summarizes the structure of production and trade in the EAC. VA is around 60% of output in all EAC members, apart from Rwanda, where it is 73%. The other components

of output are domestic and imported intermediates, of which, as the second plot shows, between 17 and 24% are imported by different EAC members. Gross output is then either consumed or exported for either intermediate or final use. The RHS of Figure 8 shows that between 5 and 16% of gross output is exported by EAC members.

Figure 8: GROSS DECOMPOSITION OF EAC PRODUCTION AND TRADE



*Notes:* Based on EMERGING and computed using an 2015-2019 average MRIO table.

The bottom panel of Figure 8 decomposes exports and imports by type of flow. It shows that Uganda, Rwanda, and Burundi have significant export and import shares with the EAC for both intermediate and final products. Kenya and Tanzania, on the other hand, export significant amounts to the EAC but only import small shares. With the exception of Tanzanian and Ugandan exports, inner-EAC trade in intermediates is slightly larger than trade in final goods.

## 4 Value Chains

While gross intermediate flows provide useful information about direct productive relationships, they do not reveal how much of the value was added in the supplying country-industry and previous production stages performed by other country-industries. The Leontief decomposition solves this problem by reallocating the value of intermediate inputs to the original producers (Quast & Kummritz, 2015). To guide the further discussion of VA trade flows, I begin with some formal derivations and introduce a consistent notation used throughout this paper.

Let  $\mathbf{A}$  be a normalized ICIO table where each element  $a_{oi,uj}$  gives the units of origin country  $o$  and sector  $i$ 's (row) output required for the production of one unit of using country  $u$  and sector  $j$ 's (column) output,  $\mathbf{x}$  the vector of outputs of each country-sector, and  $\mathbf{d}$  a vector of final demand (FD) such that the following productive relationship holds

$$\mathbf{x} = \mathbf{Ax} + \mathbf{d}. \quad (1)$$

Leontief (1936)'s insight was that one could solve this equation for  $\mathbf{x}$  to get the amount of output

each country-sector should produce given a certain amount of FD

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{d} = \mathbf{B} \mathbf{d}, \quad (2)$$

where the Leontief Inverse is denoted  $\mathbf{B} = (\mathbf{I} - \mathbf{A})^{-1}$ . This matrix is also often called the total requirement matrix since it gives the total productive input requirement from each sector to produce one unit of final output<sup>6</sup>. The direct VA share of each country-sector is given by

$$\mathbf{v} = \mathbf{1} - \mathbf{A}' \mathbf{1}, \quad (3)$$

where  $\mathbf{1} = (1, 1, 1, \dots, 1)'$  is a column-vector of 1's. Let  $\mathbf{V}$  be the matrix with  $\mathbf{v}$  along the diagonal and 0's in the off-diagonal elements. Multiplying Eq. 2 with  $\mathbf{V}$  then gives VA in each country-sector

$$\mathbf{Vx} = \mathbf{V}(\mathbf{I} - \mathbf{A})^{-1} \mathbf{d} = \mathbf{VBd}. \quad (4)$$

The term  $\mathbf{VB} = \mathbf{V}(\mathbf{I} - \mathbf{A})^{-1}$  is known as the matrix of VA multipliers or VA shares, which can be used to obtain the amount of VA generated in each sector ( $\mathbf{Vx}$ ) when producing to satisfy FD ( $\mathbf{d}$ ). More specifically, the matrix  $\mathbf{VB}$  contains the amount of VA by each country-sector (row) to the production of one unit of each country-sector's (column's) output.

#### 4.1 Backward GVC Participation

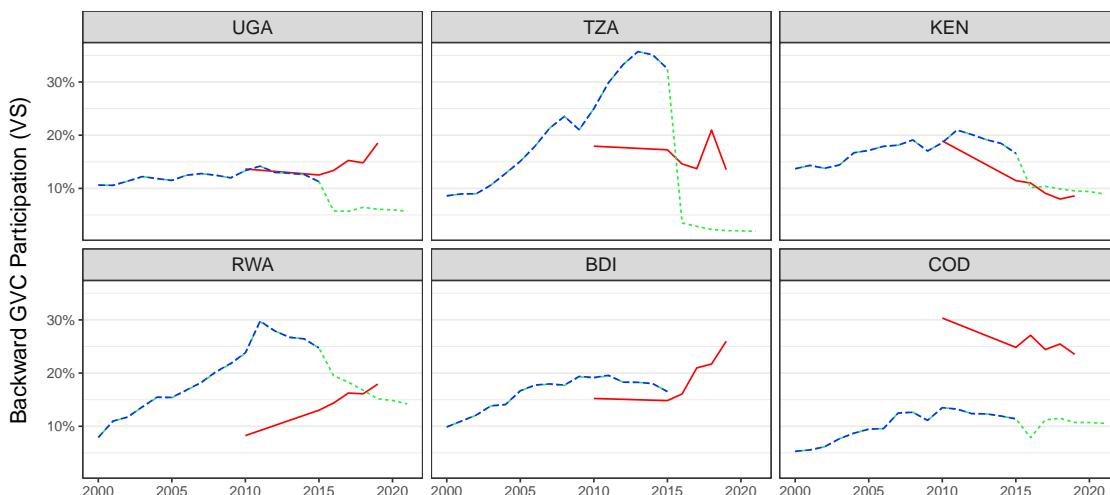
The FVA share in domestic production and exports, termed 'Vertical Specialization' (VS) by [Hummels et al. \(2001\)](#), is the most widely used measure of backward GVC integration. Consider  $\mathbf{VB}$  with elements  $vb_{oi,uj}$ , then VS for a particular country-sector may be expressed as

$$VS_{uj} = \sum_{oi, o \neq u} vb_{oi,uj} \quad \forall uj. \quad (5)$$

Figure 9 shows a time series of VS according to different data sources. The calculated VS measure using EORA21 is identical to the WDR one. The extension of EORA through 2021, as mentioned, introduces a large structural break in 2016, which, in some cases such as Tanzania where VS drops to zero or Burundi where VS rises to above 50% (truncated in Figure 9) is highly unrealistic. EM is the more reliable database for these countries and indicates that for all EAC members, between 8% and 30% of production/exports is foreign content. Furthermore, EM suggests that the smaller economies Burundi, Rwanda, and Uganda have increased their VS, especially in 2015-2019, whereas Kenya and Congo have seen a decline in VS. In Tanzania, VS appears stagnant at ~16%.

Figure 9: EAC BACKWARD GVC PARTICIPATION

Source: — EMERGING — EORA — WDR\_EORA



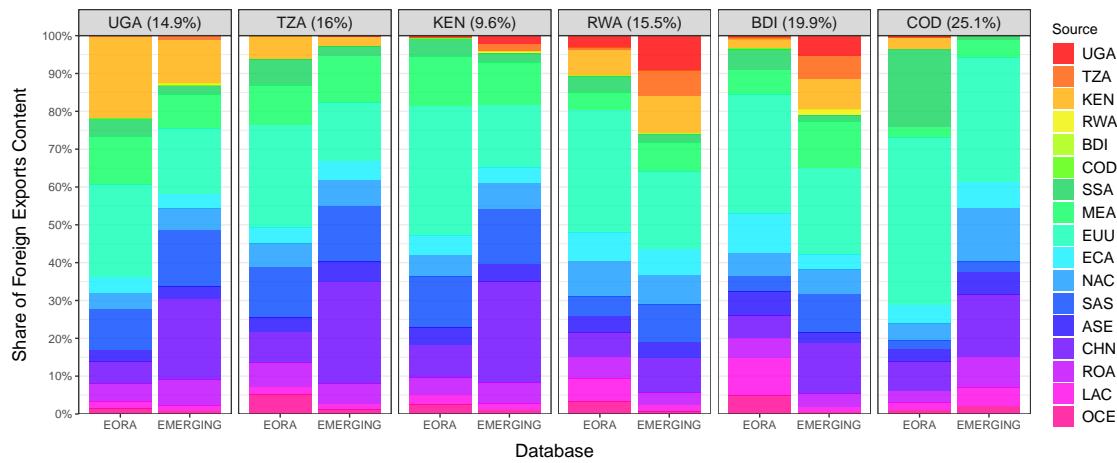
Notes: [Hummels et al. \(2001\)](#)'s index of Vertical Specialization (VS) is the FVA share in gross output and exports.

<sup>6</sup>Specifically each element in  $b_{oi,uj}$  in  $\mathbf{B}$  gives the output required from country-sector  $oi$  for the production of one unit of the final good in  $uj$ . Thus, the first column of  $\mathbf{B}$  gives all the productive input required from all sectors for the production of one unit of the final good in sector 1, and the first row of  $\mathbf{B}$  gives all the input required from sector 1 to produce one unit of the final good in each sector.

Apart from its overall size, the composition of VS is of interest. Figure 10 shows a breakdown of VS by source country/region, averaged, for EORA between 2010 and 2015 and for EM between 2015 and 2019. Congruent to the EAC import share shown in the bottom right panel of Figure 8, only Rwanda, Burundi, and Uganda source a significant fraction of foreign inputs from EAC partners. According to EM, Kenya supplies 11.5% of the foreign content in Ugandan exports, 9.8% in Rwanda, and 8% in Burundi. Uganda also supplies 9.3% of the foreign content in Rwandan exports and 5.5% in Burundi. In absolute values, Uganda supplies slightly more to Kenyan export production (around 23 million USD according to EM, vs. 20.7 million to Rwanda). This is dwarfed by the 115 million that Kenya adds to Ugandan exports.

In total, the EU and China have the greatest shares in EAC VS. The EU supplies 33% of the foreign content of Congolese exports, 23% in Burundi, 21% in Rwanda, 17%, 16%, 15% in Uganda, Kenya, and Tanzania, respectively. China supplies 27% of the foreign content of Kenyan and Tanzanian exports (approx. 300 million USD in both cases), 21% in Uganda, and 16% in Congo. Thus, overall, EAC exports have modest amounts of foreign content, and most of this VS, particularly for major exporters Congo, Kenya, and Tanzania, originates in the EU or China.

Figure 10: EAC BACKWARD GVC PARTICIPATION: SOURCES OF FOREIGN CONTENT  
Average EMERGING 2015-2019 Foreign Content Share in Parentheses



*Notes:* Figure shows a breakdown of VS by source country according to EORA (2010-2015) and EM (2015-2019) averages.

Sectors exhibit great heterogeneity, both in terms of overall foreign content and its composition. Table 3 shows overall VS content shares according to EM. In general, manufacturing sectors have higher foreign content, a pattern emphasized in the WDR, which also notes that a handful of sectors, including electrical machinery (ELM) and transport equipment (TEQ), have driven GVC expansion since 1995. In the average EAC country, these manufacturing sectors have more than 20% foreign content, but there is marked heterogeneity across countries. Notably, in Rwanda, manufacturing sectors have less than 15% foreign content. The highest foreign content sectors by country are ELM in Tanzania (42%), wood and paper (WAP) in Kenya (40%), mining (MIN) and textiles (TEX) in Uganda (29%), sales and repairs (SMH) in Rwanda (25%), petrochemicals (PCM) in Burundi (47%) and TEQ in Congo (36%). Since Burundi and Congo have no IO table, these figures need to be taken with caution. The final columns of Table 3 give FVA in overall sectoral exports by EAC members, including value addition by other members, with and without Congo. These resemble a classical VS distribution centering around ELM and TEQ at ~ 35%.

Figure 11 breaks down the origin of total EAC5 VS and thus provides a sector-level perspective of EAC regional integration. The sectors with the highest EAC5 share are SMH at 14.5% and FBE at 14%. Other sectors with sizeable regional shares are PCM at 8.6%, TEX at 7.4%, AFF at 7.2%, electricity (EGW) at 6.3% and TEQ at 6.2%. This quantitatively highlights the potential of the FBE sector for regional integration but also indicates a failure of regional integration in many core manufacturing sectors. For example, ELM, which has a VS of around 35% according to Table 3, only has a 2.8% regional share. Multiplying these percentages yields that only 1% of the gross exports (and output) in EAC ELM is regional FVA, compared to 1.5% for FBE.<sup>7</sup> Figure 11 thus indicates great potential and challenges in developing regional manufacturing value chains.

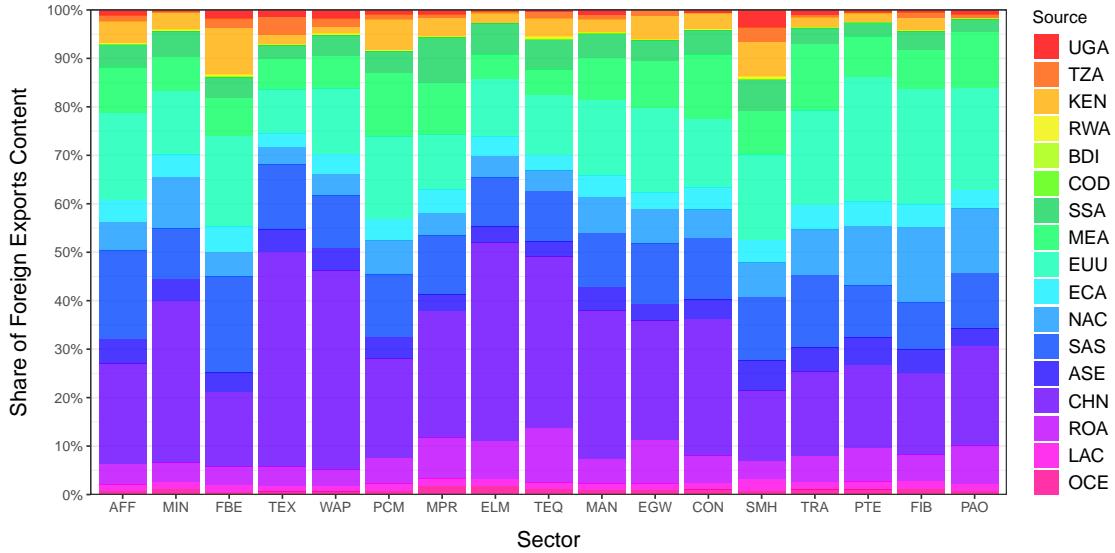
<sup>7</sup>Due to the lower FVA share of 11% in the FBE sector.

Table 3: EAC BACKWARD GVC PARTICIPATION: SECTORAL HETEROGENEITY  
*Average EMERGING 2015-2019 Foreign Content Shares (%)*

sector	UGA	TZA	KEN	RWA	BDI	COD	Mean	Median	EAC6	EAC5
AFF	5.9	4.1	3.7	5.8	15.2	3.3	6.3	5.0	4.2	4.4
MIN	29.2	5.5	0.0	2.0	17.0	4.7	9.7	5.1	4.6	6.8
FBE	22.7	7.6	3.1	20.6	19.3	13.6	14.5	16.4	11.1	10.7
TEX	29.6	17.1	25.2	8.3	8.5	28.6	19.6	21.2	26.1	24.1
WAP	13.7	22.7	39.5	1.4	5.9	20.1	17.2	16.9	24.8	29.2
PCM	23.9	19.8	19.5	10.2	47.1	20.6	23.5	20.2	20.0	19.7
MPR	27.0	26.9	16.2	10.2	39.7	25.0	24.2	26.0	24.3	23.6
ELM	18.7	41.9	30.7	5.2	27.4	34.9	26.5	29.1	34.9	35.2
TEQ	22.7	17.1	19.7	0.0	32.8	36.4	21.4	21.2	34.6	23.9
MAN	23.3	21.8	27.3	0.8	0.4	24.5	16.3	22.6	25.3	25.6
EGW	28.2	3.1	26.2	0.0		2.1	11.9	3.1	16.9	27.1
CON	15.6	13.1	16.2	7.4	5.3	30.6	14.7	14.3	12.9	12.9
SMH	5.6	11.7	9.4	25.2	14.1	15.3	13.5	12.9	10.9	10.9
TRA	7.6	18.5	6.6	4.7	0.1	4.3	7.0	5.7	11.0	11.0
PTE	11.0	21.7	5.4	0.0	0.0	2.5	6.8	4.0	11.9	12.0
FIB	0.4	8.2	0.3	1.9	0.0	4.5	2.5	1.1	1.1	0.9
PAO	5.0	2.2	8.0	0.0	0.0	4.5	3.3	3.4	6.8	7.0

*Notes:* Table reports total foreign content shares (VS) according to the EM 2015-2019 average in percentage terms. These shares are reported for each EAC6 country and for the EAC6 and EAC5 as a whole, which also counts VA by members among each other as FVA, i.e., these are export-weighted averages of individual members VS. The 'Mean' and 'Median' give unweighted EAC6 averages.

Figure 11: EAC5 BACKWARD GVC PARTICIPATION: SOURCES OF VS BY SECTOR  
*Based on Average EMERGING 2015-2019 EAC Exports (Excl. Congo)*



*Notes:* Figure shows a sector-level breakdown of total EAC5 VS by source country, according to EM (2015-2019) averages.

## 4.2 Forward GVC Participation

Apart from VS, which measures backward GVC integration, [Hummels et al. \(2001\)](#), and more formally [Daudin et al. \(2011\)](#), introduced the share of domestic exports that enter foreign countries' exports, termed VS1, as a measure of forward GVC Integration. It is defined as<sup>8</sup>

$$VS1_{oi} = \frac{1}{E_{oi}} \sum_{uj, u \neq o} vbe_{oi,uj} \quad \forall oi, \quad (6)$$

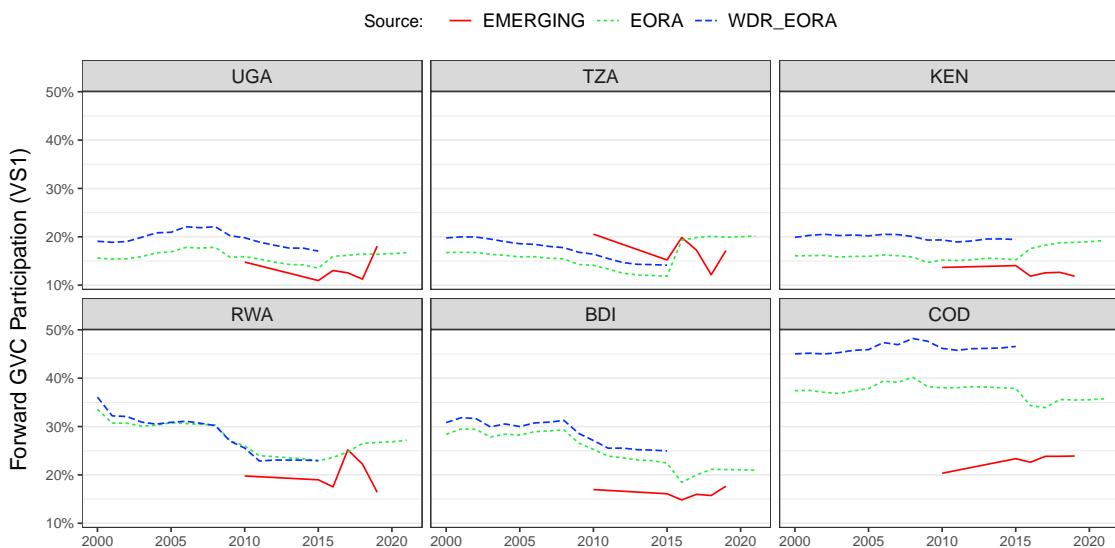
where  $E_{oi}$  are the gross exports of country-sector  $oi$  used to normalize the sum along the rows of **VBE** (excluding domestic sectors, **E** is a diagonal gross exports matrix) which capture the use of VA from a domestic sector  $oi$  in the exports of all foreign sectors  $uj$ . [Borin & Mancini \(2019\)](#) show

<sup>8</sup>For completeness I note that VS can be defined in an analogous way as  $VS_{uj} = \frac{1}{E_{uj}} \sum_{oi, o \neq u} vbe_{oi,uj} \quad \forall uj$ , however, since  $\sum_{oi} vbe_{oi,uj} = 1 \quad \forall uj$ , the exports cancel out and the equation reduces to Eq. 5.

that this measure is biased because it contains double-counted components. They propose (DVA - DAVAX)/E, which is the ratio of DVA (excl. double-counted items) minus directly absorbed DVA in exports (DAVAX) to gross exports as a refined measure of forward GVC participation.

Accurate computation of forward GVC participation requires a full country-level ICIO database. Due to computational constraints, I reduce the number of sectors to 5: AFF, FIB, MIN, MAN (combining 7 manufacturing sectors), and SRV (all other sectors) while preserving the full number of countries and territories (187 for EORA and 245 for EM). Figure 12 shows the corrected measure of forward GVC participation following [Borin & Mancini \(2019\)](#). Evidently, a reduction of the sectoral dimension attenuates aggregate VS1 indicators a bit, but the trends are broadly preserved. All indicators show that commodity exporters such as Congo and Burundi have greater forward GVC integration. EM measures suggest that VS1 has increased slightly in Congo and decreased slightly in Kenya, Rwanda, and Tanzania since 2010, suggesting a slight shift away from commodities in the latter three economies.

Figure 12: EAC FORWARD GVC PARTICIPATION



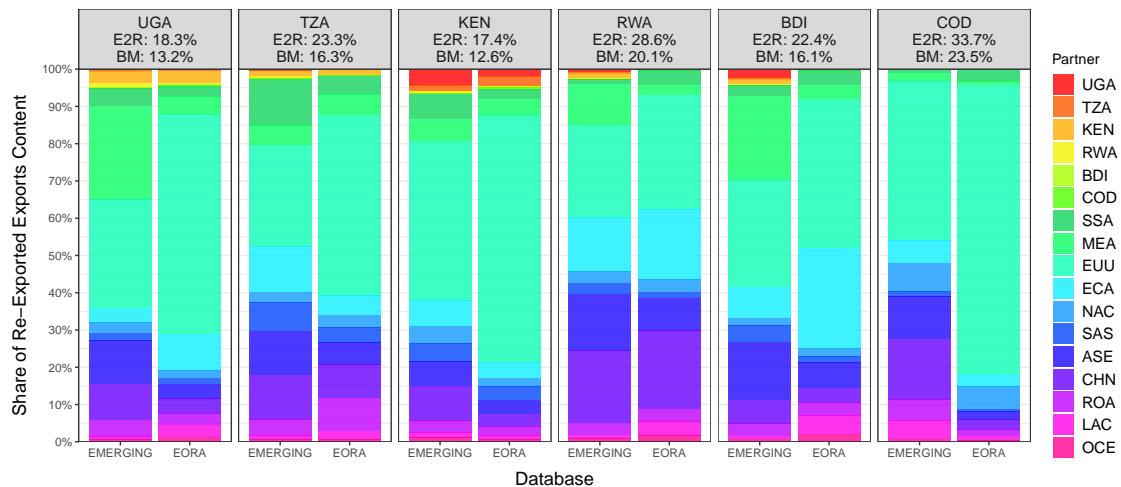
*Notes:* [Borin & Mancini \(2019\)](#)'s index of forward GVC integration (VS1) is the (non-double counted) DVA in exports that is not directly absorbed by the direct importer, divided by gross exports: (DVA - DAVAX)/E.

Since even with 5-sector ICIO tables, bilateral GVC indicators using [Belotti et al. \(2020\)](#)'s ICIO STATA package are extremely time-consuming, I compute the simple VS1 measure following Eq. 6 (also called exports to re-exports (E2R) by [Baldwin & Lopez-Gonzalez \(2015\)](#)) to examine bilateral relationships. Figure 13 offers a breakdown of VS1 by GVC partner. The headers indicate that E2R (Eq. 6) is indeed upward biased vis-a-vis the corrected measure of [Borin & Mancini \(2019\)](#) (BM), but this does not necessitate bias in the GVC partner shares. According to EM, 4.4% of Kenya's VS1 was re-exported by Uganda, and 3.2% of Ugandan VS1 is re-exported by Kenya. Other EAC countries also re-export a small share of their VS1 through Kenya: Burundi (1.25%), Rwanda (1.5%), and Tanzania (1.5%). Burundi and Rwanda export 2.4% and 0.8% of their VS1 through Uganda, respectively. Forward GVC linkages in the EAC are almost an order of magnitude smaller than backward linkages. The major GVC partner for EAC countries is the EU, accounting for 43% of Kenyan and Congolese VS1 and close to 30% of VS1 in the other EAC members. The early literature (e.g., [Foster-McGregor et al. \(2015\)](#), [Kummritz \(2016\)](#)) associates increased VS1 with productive upgrading, which, according to Figure 13, is still in its infancy in EAC RVCs.

Table 4 shows total forward GVC participation by sector, similar to Table 3 for backward GVC participation, and highlights considerable heterogeneity across EAC countries and sectors. In the EAC5 (excl. Congo), around 21% of gross exports in agriculture and manufactured products are re-exported as part of GVCs.

Since EAC forward GVC integration focuses on Uganda and Kenya, I also examine this link at the sector level. Based on EM 2015-19 averages, Kenya exports 81 million USD through Uganda, which amounted to 4.4% of Kenya's VS1 and 0.76% of its gross exports. 54% of these 81 million

Figure 13: EAC FORWARD GVC PARTICIPATION: RE-EXPORTING GVC PARTNERS  
Average EMERGING 2015-2019 Re-Exported Content Shares (VS1)



*Notes:* Figure shows a breakdown of forward GVC integration by GVC partner according to EM 2015-2019 averages. The classical VS1 measure of [Daudin et al. \(2011\)](#) (Eq. 6, also termed E2R) is used to determine each partner's share in total VS1. The headers provide overall VS1 using both E2R and the corrected measure by [Borin & Mancini \(2019\)](#).

Table 4: EAC FORWARD GVC PARTICIPATION: SECTORAL HETEROGENEITY  
Average EMERGING 2015-2019 Re-Exported Content Share (%) (VS1 following BM)

sector	UGA	TZA	KEN	RWA	BDI	COD	Mean	Median	EAC6	EAC5
AFF	16.1	17.0	27.8	13.5	21.0	21.8	19.5	19.0	21.3	21.2
MIN	28.1	14.3	15.7	3.4	6.8	31.1	16.6	15.0	30.7	14.7
FBE	11.3	19.3	11.4	13.2	19.2	15.6	15.0	14.4	13.3	12.9
MAN	22.0	23.4	11.9	42.8	21.6	23.7	24.2	22.7	22.5	21.0
SRV	7.8	10.2	10.0	7.6	8.3	12.1	9.3	9.2	9.6	9.5

*Notes:* Table reports total forward GVC participation (VS1) following [Borin & Mancini \(2019\)](#) using the EM 2015-2019 average in percentage terms. These shares are reported for each EAC6 country and for the EAC6 and EAC5 as a whole, which includes re-exported VA by EAC members among each other. They are thus export-weighted averages. The 'Mean' and 'Median' columns give unweighted EAC6 averages.

are manufactured goods, 20% are services, and 17% are agricultural products. Uganda, on the other hand, exports 30 million USD through Kenya, which amounts to 3.2% of Ugandan VS1 and 0.58% of Ugandan gross exports. Of these 30 million, 45% are agricultural products, 22% services, 16% FBE, and 18% other manufacturing. The links between these two countries account for the bulk of EAC forward GVC integration, summarized compactly by Table 5. Of particular interest in this table is the EAC share in sectoral VS1, which is high at 20.7% for Kenyan manufactures, indicating that about 1/5th of re-exported VA in Kenyan manufacturing is exported by its EAC partners. Other notable figures are the 41%/29% EAC shares in re-exported Rwandan/Kenyan mining exports, which are, however, very small in value.

Table 5: EAC FORWARD GVC INTEGRATION AT THE SECTOR LEVEL  
Average EMERGING 2015-2019 Traditional VS1 Estimates ([Daudin et al., 2011](#))

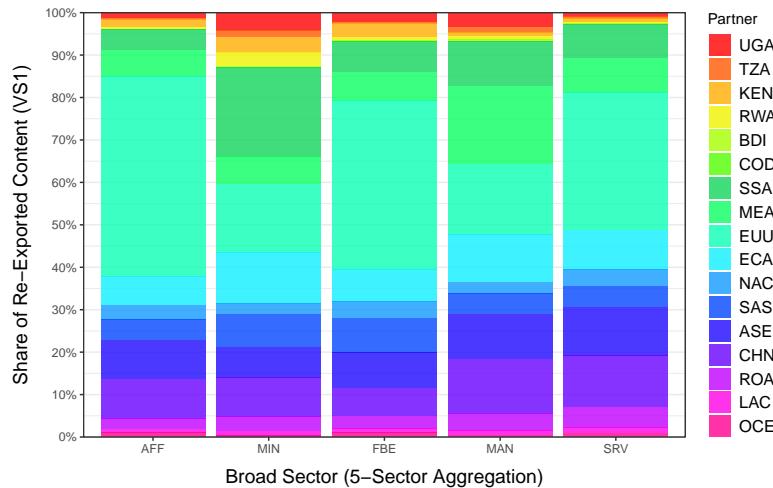
Country	VS1 (Re-Exported By EAC Partners)					Total + EAC Shares			EAC Share in Sectoral VS1				
	AFF	FBE	MAN	MIN	SRV	SUM	VS1	EXP	AFF	FBE	MAN	MIN	SRV
UGA	18.23	6.05	13.52	0.00	11.98	49.78	5.60	0.95	6.36	9.05	6.37	6.79	4.12
TZA	11.06	9.20	12.40	1.15	17.64	51.45	2.80	0.61	3.44	5.89	3.03	8.91	1.91
KEN	19.58	9.24	65.88	0.98	28.33	124.01	6.83	1.17	3.22	6.50	20.74	28.73	3.86
RWA	3.63	3.08	3.39	0.00	3.59	13.69	2.86	0.80	13.98	12.99	1.21	41.04	2.27
BDI	0.37	1.27	0.38	0.01	0.33	2.37	4.37	0.97	5.49	7.49	2.98	3.01	2.24
COD	0.22	0.10	1.33	0.46	0.34	2.45	0.06	0.02	0.07	0.08	0.06	0.05	0.07

*Notes:* VS1 is recorded in million USD, shares in percentage terms. Column 'SUM' gives total country VS1 through EAC partners, and columns 'VS1' and 'EXP' give the share of this in the country's total VS1 and gross exports, respectively.

To complete the picture, Figure 14 shows the sector-level shares in forward GVC partners for the EAC5 (Excl. Congo). The EAC share is highest in mining at 13%, but, Congo being excluded,

mining VS1 comprises only 13 million USD, compared to 2.8/3 billion in AFF/FBE, 7.8 billion in manufacturing, and 5.6 billion in services re-exports. Among these, the EAC has a share of 4% in AFF and 6.7% in both FBE and MAN, indicating that manufacturing accounts for the bulk of GVC forward regional integration. The biggest forward GVC partner in all sectors remains the EU, at shares between 47% for AFF and 16% for MIN and MAN.

Figure 14: EAC FORWARD GVC PARTICIPATION: GVC PARTNERS BY SECTOR  
Based on Average EMERGING 2015-2019 EAC Exports (Excl. Congo)



Notes: Figure shows a sector-level breakdown of total EAC5 VS1 (E2R) by GVC partner using EM (2015-2019) averages.

### 4.3 Trends in EAC Regional Integration in Value Added Terms

While overall EAC GVC integration appears relatively stable, exempting an increase in VS in the smaller economies and a gradual decline in VS1, there may be stronger trends in regional integration relative to overall trade and GVC integration - as evident in gross trade flows. In this section, I thus introduce four metrics to track EAC regional integration through VA in supply chains relative to the members' overall GVC participation. The first metric is the share of FVA in a member's production/exports accounted for by its EAC neighbours. It is defined as

$$VS_{uj}^{EAC} = \frac{1}{VS_{uj}} \sum_{oi \in EAC, o \neq u} v_{b_{oi,uj}} \quad \forall uj \in EAC, \quad (7)$$

where  $VS_{uj}$  is defined as in Eq. 5.  $VS^{EAC}$  is thus a relative measure tracking the EAC share in VS, as shown also in Figure 10, such that the overall EAC VA share in domestic production/exports can be computed as  $VS_{uj}^{EAC} \times VS_{uj} \quad \forall uj$ . I define an analogous measure for VS1 as the proportion of DVA in re-exported exports exported by EAC partner states, also visible in Figure 13

$$VS1_{oi}^{EAC} = \sum_{uj \in EAC, u \neq o} v_{be_{oi,uj}} / \sum_{uj, u \neq o} v_{be_{oi,uj}} \quad \forall oi \in EAC. \quad (8)$$

These two metrics effectively track the role of the EAC in members' GVC participation. They, however, do not account for the import side, i.e., the EAC's role in providing goods and services to members' relative to ROW. I thus compute two additional metrics to capture this aspect of regional integration. The first is the share of EAC VA in members' imports, which I denote by  $VAI_u^{EAC}$ . Consider  $\mathbf{e}_u$  the vector of gross exports to EAC using country  $u \in EAC$  from each country-sector. I then compute the VA origins of these exports to country  $u$  as

$$\mathbf{e}_u^{VA} = \mathbf{VBe}_u, \quad (9)$$

where  $\mathbf{e}_u^{VA}$  denotes the vector, with elements  $e_{oi,u}^{VA}$ , of VA supplied by each country-sector ( $oi$ ) in these imports of country  $u$ . From  $\mathbf{e}_u^{VA}$ , the share of EAC VA is easily computed as

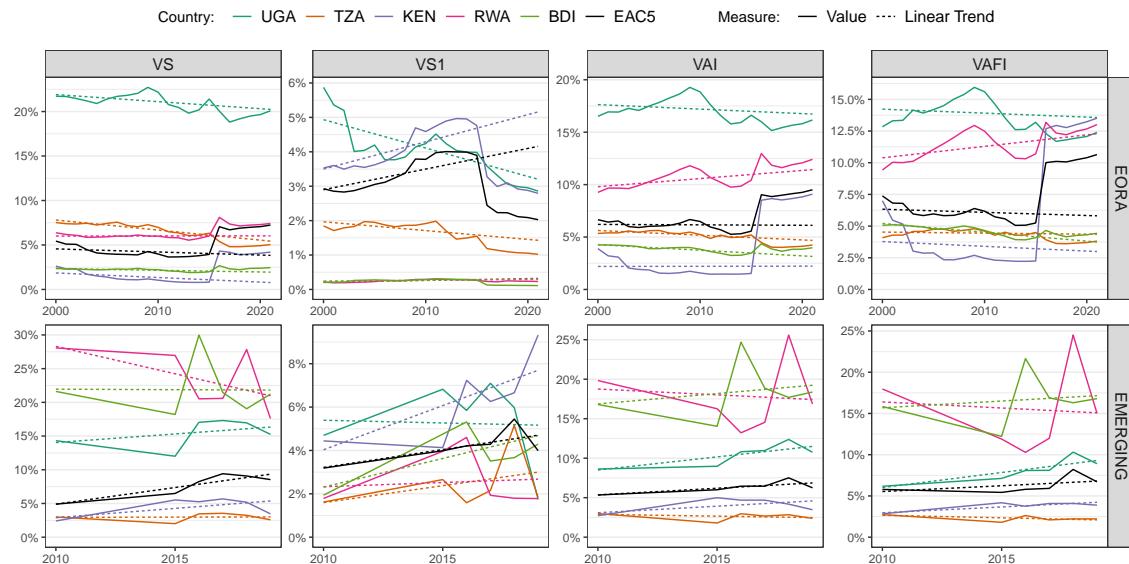
$$VAI_u^{EAC} = \sum_{oi \in EAC, o \neq u} e_{oi,u}^{VA} / \sum_{oi, o \neq u} e_{oi,u}^{VA}. \quad (10)$$

$VAI_u^{EAC}$  is thus a country-level measure of EAC VA in the import mix. It may include intermediates of goods being exported. To single out the EAC share in imported consumption goods, I also consider only exports for final consumption. Let  $\mathbf{fe}_u$  be the final exports to country  $u$  from each country-sector. Then  $\mathbf{fe}_u^{VA} = \mathbf{VB}\mathbf{fe}_u$  denotes these exports in VA terms, and I define

$$VAFI_u^{EAC} = \sum_{oi \in EAC, o \neq u} \mathbf{fe}_{oi,u}^{VA} / \sum_{oi, o \neq u} \mathbf{fe}_{oi,u}^{VA} \quad (11)$$

as the EAC VA share in final goods exported to a particular member  $u$ . I compute these metrics using the MRIO tables with reduced country dimension, except for  $VS1_{oi}^{EAC}$ , where I use the tables with reduced sectoral dimension. Figure 15 plots all metrics, including a weighted linear trend.<sup>9</sup>

Figure 15: EAC5 VA SHARES IN MEMBERS VS, VS1, IMPORTS AND FINAL IMPORTS

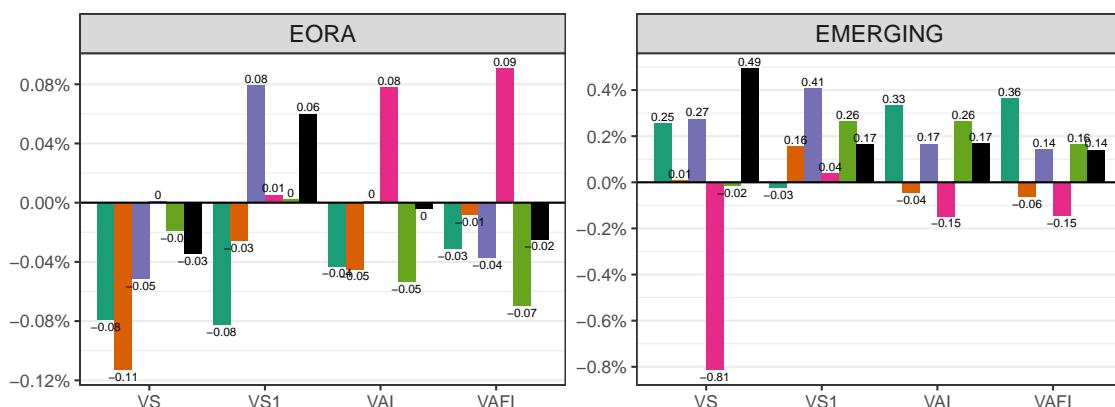


Notes: Figure shows regional EAC regional integration metrics following Eqns. 7-11, including a (weighted) linear trend.

Both databases agree that EAC shares in members' VS1 are substantially lower than in VS, VAI, and VAFI but increased over the period, mainly driven by Kenya. They also agree that the larger economies drive EAC forward linkages, and smaller economies (Rwanda and Burundi) are more important in backward linkages (VS) and as importers of final goods (VAFI). Otherwise, there is not much agreement regarding the direction of the trend. Figure 16 plots the slope coefficients.

Figure 16: WEIGHTED SLOPE ESTIMATES MEASURING THE SPEED OF REGIONAL INTEGRATION

Country: UGA TZA KEN RWA BDI EAC5



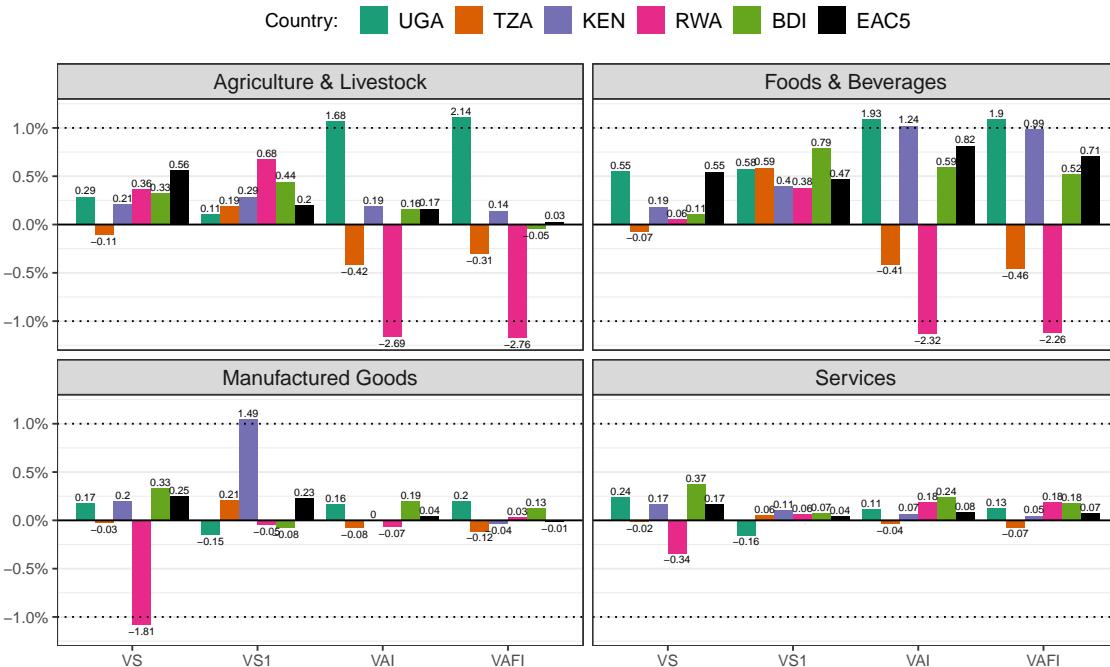
Notes: Figure shows (weighted) linear slopes (dotted lines in Figure 15). In the estimation, all obs. received a weight of  $w = 1$ , except for EM 2010 ( $w = 2$ ) and EORA > 2015 ( $w = 0.1$ ). These weights reflect data availability and quality.

<sup>9</sup>All observations receive a weight of 1, except for EM 2010 obs. which receive a weight of 2 because no further data is observed until 2015, and EORA 2016-21 obs. receive a weight of 0.1 due to the stark structural break.

The more reliable EM database suggests that, with few exceptions, EAC regional integration in VA terms is increasing in most countries. Considering the EAC5 as a whole, the coefficients suggest that  $VS^{EAC}$  is increasing by 0.5 percentage points (pp.) per year, and the EAC shares in EAC VS1, VAI, and VAFI are increasing at a slower rate of around 0.15 pp. per year. These trends mildly contrast those in gross trade (Figure 3).

As with gross trade, the weak aggregate signal indicates that there may be more substantial sectoral developments. I thus recompute all 4 indicators at the sector level using the EM database with full country dimension but only 5 broad sectors. Figure 17 shows weighted linear slope estimates at the sector level (excluding mining), using again a weight of 2 for 2010 estimates.

Figure 17: WEIGHTED SLOPE ESTIMATES OF MEMBERS SECTORAL INTEGRATION SPEED



*Notes:* Figure shows (weighted) linear slopes estimating the speed of regional integration (pp. per year) at the 5-sector level (excl. mining) based on EM (2010-2019). All obs. received a weight of  $w = 1$ , except for EM 2010 ( $w = 2$ ).

Regional integration in goods-producing sectors proceeds substantially faster than in services. The aggregate pattern from Figure 16, with faster integration through  $VS^{EAC}$ , is reflected in agriculture, manufacturing, and services. The FBE sector, on the other hand, experienced stronger integration through forward linkages ( $VS1^{EAC}$ ) and imports ( $VAI^{EAC}$ ,  $VAFI^{EAC}$ ) at greater speeds ( $\geq 0.5$  pp. per year on all metrics). Uganda and Kenya are driving these developments. The regional integration in FBE through  $VS1^{EAC}$  is driven by all 5 members at almost equal shares, whereas the manufacturing expansion through  $VS1^{EAC}$ , proceeding at about half the speed as FBE, is driven almost completely by Kenya, with Tanzania contributing a little bit, and other members experiencing declining  $VS1^{EAC}$ . This analysis of regional integration in VA terms thus complements Figures 5 and 6, indicating that there is some momentum in regional integration through agriculture and FBE, but equitable integration in manufacturing is difficult, and Kenya is strengthening its already favourable trading position through forward GVC linkages.

#### 4.4 EAC Positioning in GVCs

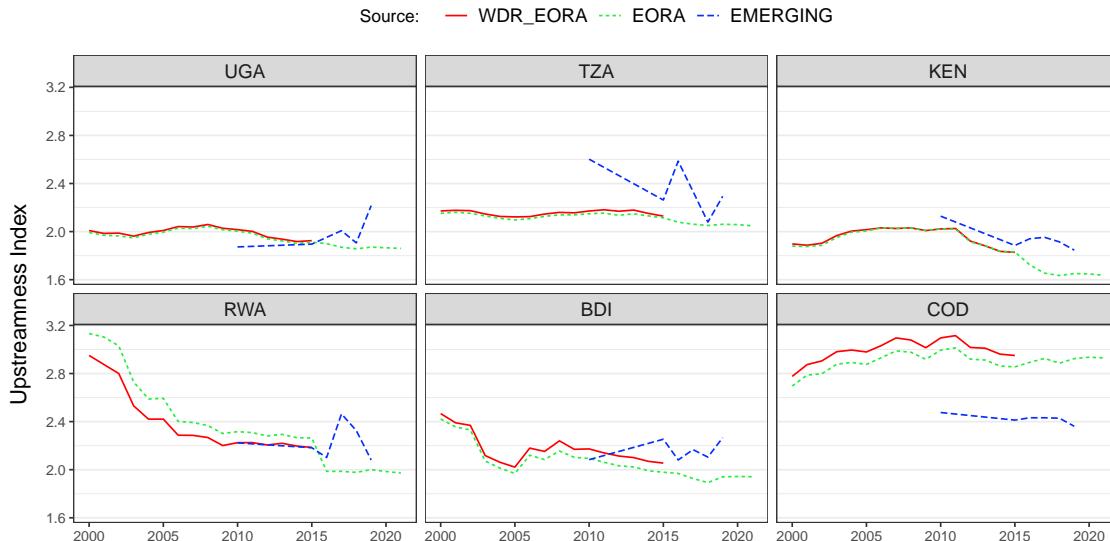
Following Antràs et al. (2012); Antràs & Chor (2022), a common measure of upstreamness  $U_{oi} \in \mathbf{u}$  is obtained by iterating forward the IO model in Eq. 1, multiplying terms by the number of production stages needed to obtain them, and normalizing by gross output. In matrix notation:

$$\mathbf{u}\mathbf{x} = \mathbf{d} + 2\mathbf{Ad} + 3\mathbf{AAD} + 4\mathbf{AAAD} + \dots = (\mathbf{I} - \mathbf{A})^{-2}\mathbf{d}. \quad (12)$$

The index is, by definition, greater than 1, and Antràs et al. (2012) state that it can be interpreted as the dollar amount by which the output of all country-sectors combined increases following a one-dollar increase in the VA of sector  $i$  in country  $o$ . Intuitively, it measures the distance of

the production stage performed by sector  $i$  in country  $o$  to the finally demanded product ( $\mathbf{d}$ ).<sup>10</sup> Antràs et al. (2012) further find that  $U$  is positively correlated with physical capital intensity and negatively correlated with skill intensity across US industries, and negatively correlated with rule of law, private credit to GDP, and education across a sample of OECD countries. Figure 18 shows aggregate upstreamness for the EAC, calculated using the regional MRIO tables with the full sector dimension, where sector-level  $U_{oi}$  estimates were averaged using gross export weights.

Figure 18: UPSTREAMNESS INDEX FOR EAC COUNTRIES



*Notes:* Figure shows upstreamness index following Antràs et al. (2012), computed at the sector level and averaged across sectors using sectoral gross exports as weights. The EORA and EM MRIOs have EAC + 11 regions and all sectors.

Figure 18 shows that the computed  $U$  index closely tracks the version computed by Mancini et al. (2024) using the full EORA 26 database and also performing an inventories adjustment following Antràs & Chor (2018). The data suggest that Congo, as a large commodity exporter, is very upstream, while other members apart from Uganda and Burundi, where EM suggests a slight increase, have moved more downstream since 2010. Rwanda also shows an impressive downstream shift in 2000-2010, but this trend must be scrutinized as EORA lacks a Rwandan IO table.

To investigate developments at the sector level, I aggregate  $U$  to broad sectors using export weights, combining all manufacturing sectors apart from FBE and all service sectors into broad categories. I also aggregate the time dimension over two intervals, 2010-2014 and 2015-2019, using the median to obtain a robust estimate. Table 6 reports the results, including an estimate of the growth rate of  $U$  between the two intervals and an export-weighted EAC5 average.

The most upstream sector, according to both databases, is MIN, where  $U > 3$  implies more than 3 production stages on average before final use. This is followed by MAN with  $2 < U < 3$  in most members, FBE and primary AFF with  $1.5 < U < 2.5$ , and SRV with  $1 < U < 2$ . Except for SRV, this is broadly in line with the world average sectoral upstreamness pattern of, according to EM 2015-19, 3.32 (MIN), 2.86 (MAN), 2.62 (AFF), 2.25 (SRV) and 2.18 (FBE). The  $U$  values of around 2/2.5 for EAC FBE/MAN indicate that these sectors are located at least one step before final use. For FBE, where more than 90% of VS1 is through non-EAC GVC partners (Figure 14), this implies that more processing steps could still be undertaken regionally to export products closer to FD. The change between the two intervals indicates a downstream shift in almost all country-sectors. It is particularly pronounced in AFF, MAN, and FBE, but also in SRV. The shift suggests that all production processes are moving closer to FD. The world average growth rate in  $U$  between these intervals, according to EM, was 3.3% for AFF, 2.1% for MIN, 0.85/0.82% for

<sup>10</sup> An equivalent measure of downstreamness ( $\mathbf{d}$ ) can be computed measuring the distance to VA instead of FD (Antràs & Chor, 2022; Miller & Temurshoev, 2017; Mancini et al., 2024), but, for the sake of brevity, this is omitted. The simplest way of computing this index is as  $\mathbf{d} = \mathbf{1}'\mathbf{B}$ , i.e., it is the column-sum of the Leontief inverse matrix (Miller & Temurshoev, 2017; Antràs & Chor, 2022). It can be interpreted as the total increase in gross output in the world economy that a unit increase in FD in the respective country-sector would generate. At the world level  $\mathbf{u}$  and  $\mathbf{d}$  are identical and measure the length of GVCs (Mancini et al., 2024).

FBE/MAN, and -0.5% for SRV, revealing that, except for SRV, EAC developments run against a global trend towards longer manufacturing GVCs ([Antràs & Chor, 2018](#)).

Table 6: EAC5 Trends in Upstreamness by Broad Sector (Aggregated)

Sector	Year	WDR Estimates using EORA 2015						EMERGING Estimates					
		UGA	TZA	KEN	RWA	BDI	EAC5	UGA	TZA	KEN	RWA	BDI	EAC5
AFF	2010-2014	2.17	2.38	1.56	2.36	3.49	1.81	1.86	1.88	2.30	1.80	1.98	2.09
AFF	2015-2019	2.13	2.36	1.49	2.35	3.33	1.76	1.91	1.84	1.83	1.28	1.77	1.84
AFF	Growth Rate	-1.72	-0.94	-4.11	-0.49	-4.75	-2.68	2.85	-2.08	-20.64	-29.25	-10.73	-12.06
MIN	2010-2014	3.07	3.53	3.49	3.39	2.89	3.48	2.13	3.23	3.28	3.29		3.25
MIN	2015-2019	3.02	3.37	3.45	3.32	2.81	3.41	3.22	3.28	3.12	1.54	1.63	3.17
MIN	Growth Rate	-1.62	-4.46	-1.40	-2.23	-2.81	-2.08	51.20	1.56	-5.02	-53.27		-2.44
FBE	2010-2014	1.47	1.58	1.41	1.39	1.50	1.45	2.23	2.04	2.34	2.57	2.32	2.26
FBE	2015-2019	1.44	1.57	1.34	1.36	1.45	1.38	2.15	2.12	2.18	2.35	2.41	2.13
FBE	Growth Rate	-2.26	-0.88	-5.32	-1.87	-2.81	-4.37	-3.21	3.81	-6.57	-8.54	4.25	-5.59
MAN	2010-2014	2.20	2.09	2.30	2.19	2.03	2.25	2.38	3.28	2.41	3.84	3.17	2.89
MAN	2015-2019	2.14	2.06	2.18	2.15	1.97	2.15	2.47	2.89	2.24	3.16	2.73	2.63
MAN	Growth Rate	-2.62	-1.57	-5.34	-1.68	-3.15	-4.42	3.81	-11.79	-7.11	-17.74	-13.96	-9.02
SRV	2010-2014	1.73	1.89	1.77	1.70	1.79	1.78	1.35	2.19	1.83	1.40	1.47	1.82
SRV	2015-2019	1.70	1.85	1.64	1.67	1.74	1.69	1.52	2.09	1.65	1.61	1.41	1.77
SRV	Growth Rate	-1.40	-1.89	-7.02	-1.78	-2.85	-4.82	12.27	-4.84	-9.75	15.11	-4.01	-2.66

*Notes:* Table shows median upstreamness ( $U$ ) following [Antràs et al. \(2012\)](#) across 2010-14 and 2015-19, and the growth rate in percentage terms between these medians. MAN and SRV are broad categories combining sectors TEX-MAN and EGW-PAO in Table 2, respectively, via an export-weighted average. The EAC5 is an export-weighted average across the 5 countries (excl. COD) computed annually before taking the median.

This appears to be good news for all sectors apart from MAN and SRV, indicating that exports are closer to FD and more local value is added. For MAN, it suggests a shift towards processing trade, which is generally not associated with industrial upgrading. The effect of SRV moving downstream is more ambiguous and depends very much on the type of service.<sup>11</sup>

## 5 (New) Revealed Comparative Advantage

In international trade, including GVC-related trade, competitiveness is closely related to the concept of comparative advantage<sup>12</sup>. A popular way to quantify Ricardo's concept of comparative advantage is [Balassa \(1965\)](#)'s measure of revealed comparative advantage, defined as the share of a sector in gross country exports divided by the share of that sector in gross world exports

$$RCA_{oi} = \frac{E_{oi}}{\sum_i E_{oi}} \left/ \frac{\sum_j E_{ji}}{\sum_{ji} E_{ji}} \right.. \quad (13)$$

$RCA_{oi} > 1$  signifies a revealed comparative advantage of country  $o$  in sector  $i$ . The traditional index based on gross exports, however, does not account for GVCs and double counting in exports. [Koopman et al. \(2014\)](#), therefore, propose a new index based on the DVA in gross exports. [Borin & Mancini \(2019\)](#) show that the decomposition of [Koopman et al. \(2014\)](#) is inexact in allocating DVA and foreign double-counted items and propose refinements. Appendix Figure B3 shows the refined breakdown following [Borin & Mancini \(2019\)](#), and Appendix Figure B4 plots the decomposition of gross exports for each of the EAC members.<sup>13</sup> DVA is the sum of DAVAX, NDAVAX, and REF. According to EM 2015-19, in the average EAC member, DAVAX accounts for 71% of gross exports, NDAVAX for 12%, FVA for 14%, and FDC for 3%. REF and DDC are close to 0 in all EAC countries, implying that the GVCs these countries engage in are relatively short. DVA =  $E - (DDC + FVA + FDC)$ , yielding an average 17% downward adjustment of gross exports. This may appear small, but, as Table 3 shows, manufacturing sectors have higher VS of up to 50%.

<sup>11</sup>For transport/tourism (TRA), a downstream shift could indicate more local value addition, whereas for telecommunications (PTE) and financial intermediation (FIB), downstream shifts might signify the insufficient quality of these services to be used as intermediates in more complicated production processes. EAC data on these sectors are likely of questionable quality, yet a brief disaggregated appraisal using EM yields, notably, 28% upstream/downstream shifts in FIB in Rwanda/Tanzania (EAC5 average is 5.3% downstream), a 9.5/5.9% downstream shift in PTE in Kenya/Rwanda (EAC5 average is 3% downstream) and a 10.2/7.4% downstream shift in TRA in Kenya/Burundi, while other members saw a slight upstream shift (EAC5 average is 1.4% downstream).

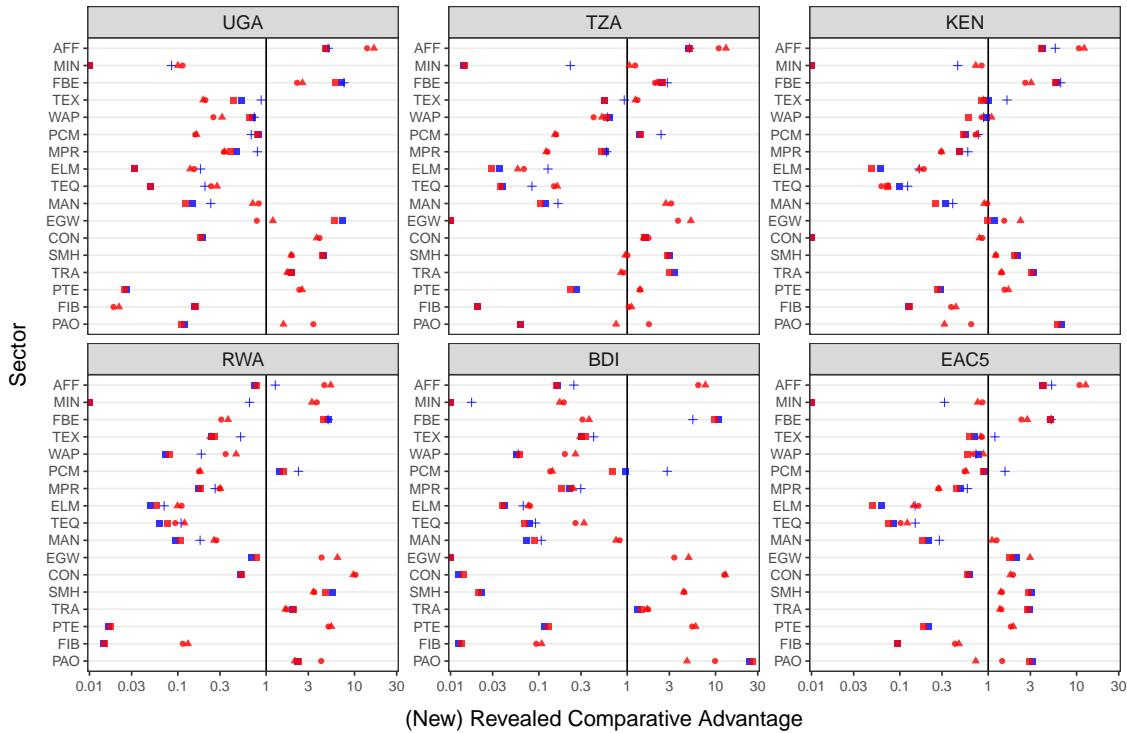
<sup>12</sup>A widely accepted theory of international trade developed by David Ricardo in 1817 stipulating that countries specialize in sectors where their productivity relative to the international average is greatest.

<sup>13</sup>To connect this to the aggregate measures of GVC integration VS and VS1 discussed so far: VS is  $(DVA - GX)/GX$ , or  $(DDC + FVA + FDC)/GX$ , VS1 is  $(DVA - DAVAX)/GX$  or  $(NDAVAX + REF)/GX$ .

For comparison, I compute both classical RCA using gross exports (GX) and NRCA using DVA in exports (VAX) based on all available databases, including BACI (only available for goods-producing sectors) and the WDR GVC indicators based on EORA 2015. Figure 19 shows median (N)RCA estimates across years 2010-19 for the EAC5. Appendix Tables B2 and B3 contain the corresponding values and correlations among different estimates, respectively. Whereas EORA-based estimates correlate around 0.57 with the BACI estimates, EM estimates have a strong correlation of 0.93, confirming that these IO tables are very close to official trade data. In all IO tables, RCA and NRCA estimates are also highly correlated ( $r > 0.96$ ), suggesting that the foreign content shares in exported goods within a sector are quite similar across different countries.

Figure 19: (NEW) REVEALED COMPARATIVE ADVANTAGE: 2010-19 MEDIAN

Flow: • GX • VAX      Source: • WDR\_EORA ▲ EORA □ EMERGING + BACI



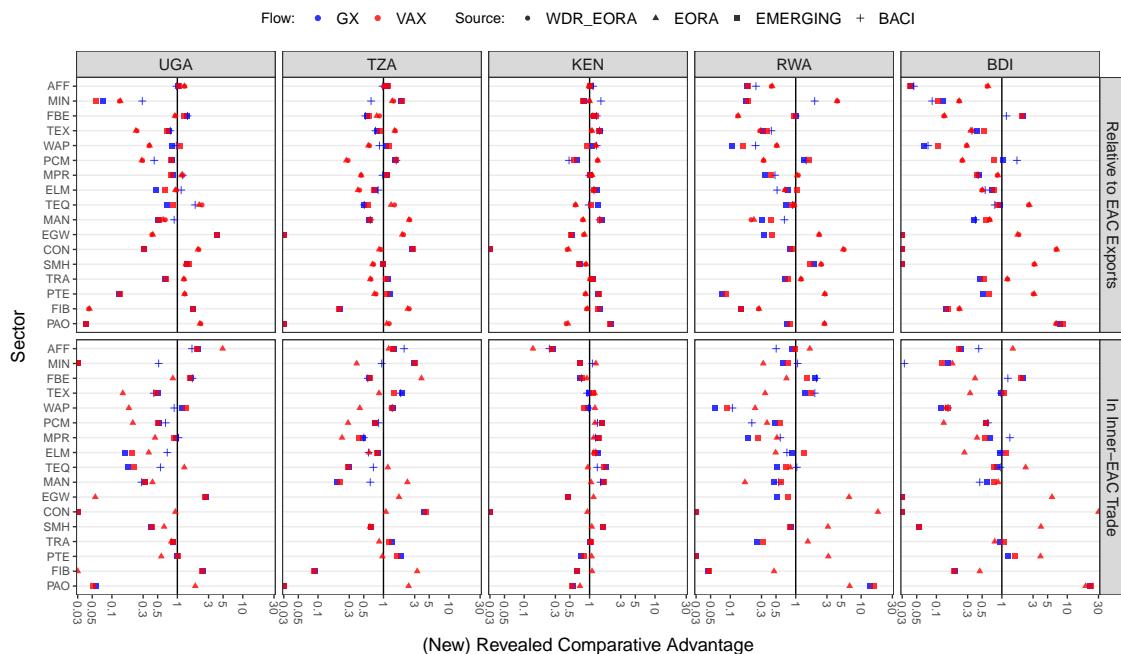
*Notes:* Figure shows median 2010-19 (N)RCA indices based on DVA/gross exports according to different databases. DVA is computed following [Borin & Mancini \(2019\)](#). Appendix Table B2 contains the values and Table B3 their correlations. To not overcrowd the figure, GX-based estimates using (WDR-)EORA are not shown.

All estimates show that the EAC5 as a whole, and, according to EM/BACI, also all the 5 countries individually, have a succinct (N)RCA in agriculture and food processing of, according to the EM NCRA estimates, 4.17 (AFF) and 5.01 (FBE). Similarly, all countries have a sizeable disadvantage in all core manufacturing sectors except for textiles and petrochemicals. Especially ELM (0.05) and TEQ (0.08), core drivers of GVC expansion according to the WDR, have a strong revealed disadvantage. On the services side, all members have a (N)RCA in TRA (2.77, incl. tourism), particularly Kenya (3.12) and Tanzania (3.02), and all members apart from Burundi have a (N)RCA in SMH (2.84), particularly Uganda (4.36) and Rwanda (4.64). Furthermore, with its powerful dams, Uganda has a large (N)RCA in EGW (5.87). EAC members thus exhibit similar patterns of (N)RCA in agriculture, food processing, and tourism, and a disadvantage in core manufacturing sectors. This is constitutive to forming a common trade block, supported by a monetary union as planned, and deepening regional tourism and food processing value chains. Yet, comparing the EAC with ROW masks rivalries and differences in (N)RCA between members.

## 5.1 NRCA Relative to the EAC and in Inner-EAC Trade

To uncover these differences, I compute (N)RCA relative to the EAC5 as the share of a sector in country exports to its share in EAC5 exports. Furthermore, regional trading reveals comparative advantages that can foster or block deeper RVCs. Thus, I also compute (N)RCA w.r.t. intraregional exports. Figure 20 presents both estimates and Appendix Table B4 the corresponding values.

Figure 20: NRCA RELATIVE TO EAC5



*Notes:* Figure shows median 2010-19 (N)RCA indices based on DVA/gross exports according to different databases, calculated w.r.t. EAC5 exports (top panel) and w.r.t. total inner EAC5 trade (bottom panel). DVA is computed following [Borin & Mancini \(2019\)](#). Appendix Table B4 contains the corresponding values. To not overcrowd the figure, GX-based estimates using (WDR-)EORA are not shown.

The estimates unveil that relative to other EAC5 members, Uganda, Tanzania, and Kenya have a slight (1-1.2) (N)RCA in agriculture, and Uganda, Kenya, and Burundi have a (N)RCA in FBE of (1.2-1.5). In inner-EAC trade, Kenya's (N)RCA drops to 0.26/0.78 in AFF/FBE in VAX terms, whereas Uganda's rises to 2/1.5, reflecting its stronger regional supplier role. Rwanda has a (N)RCA in mining, but this is not reflected in inner-EAC5 trade. Kenya has a slight comparative advantage in manufacturing sectors, including TEX, MPR, ELM, TEQ, and other manufactures (MAN). Exempting TEX, including PCM, these estimates are even higher in inner-EAC5 trade, but all are in the range between 1 and 2 and thus significantly lower than with ROW. Tanzania also has a (N)RCA in PCM according to both denominations. As mentioned earlier, Kenya and, to a lesser extent, Tanzania also have a slight (N)RCA in TRA (1-1.2), both relative to the EAC and revealed in inner-EAC trade, and Uganda has a large (N)RCA in EGW (4-5).

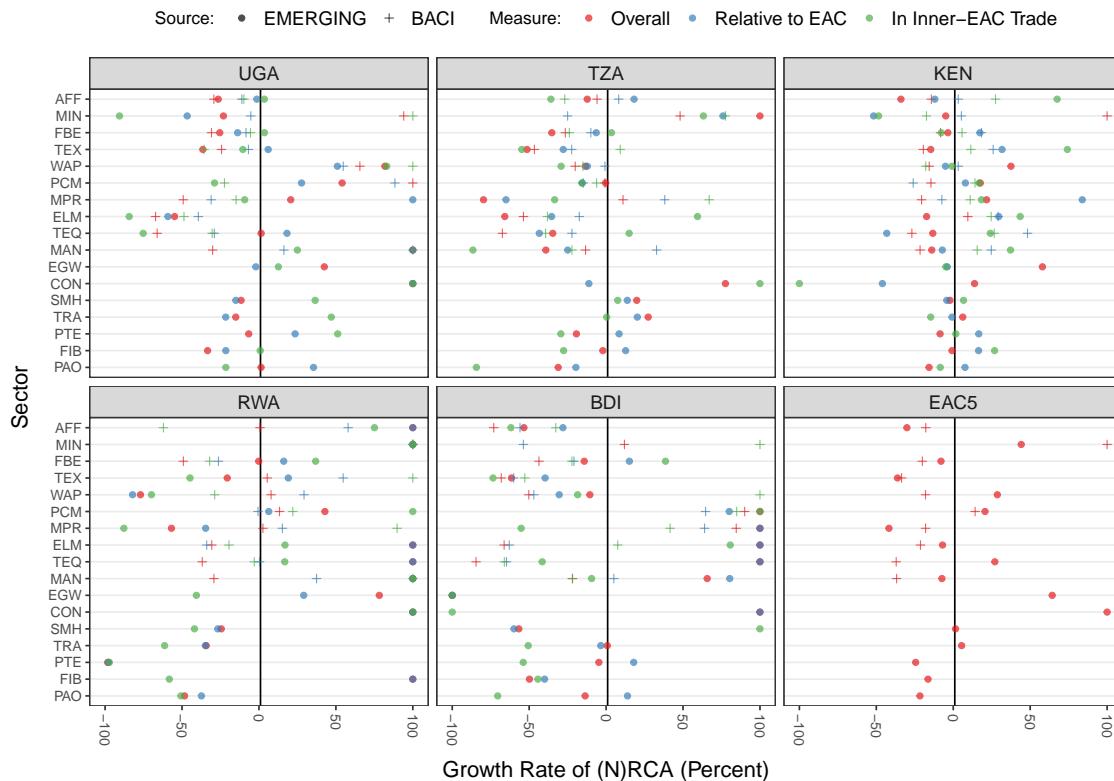
The trading patterns of different members in both gross and VA terms thus reveal differences in comparative advantage, but these are, with few exceptions, such as Ugandan EGW, between 0.5 and 2, and thus moderate in size. It may, however, still require policy action to overcome these differences and foster more horizontal RVCs in critical sectors such as FBE and tourism (TRA).

## 5.2 Trends in (N)RCA

A final question regards the direction and speed of shifts in RCA, both overall and inside the EAC. To measure this, I only use gross trade from BACI and DVA from EM to compute (N)RCA medians over two periods: 2006-2010 and 2015-2019. I then compute the growth rate and report it in Figure 21. Appendix Figure B5 shows estimates for both periods and Table B5 holds all values.

The bottom right panel of Figure 21 shows that the EAC5 has lost some (N)RCA in AFF, FBE, and, in GX terms, in all manufacturing sectors apart from PCM. On the services side, there are overall gains in CON, EGW, TRA, and losses in PTE and financial and FIB. Kenya gained a bit in FBE, TEX, MPR, and ELM relative to the EAC. For core manufacturing sectors, this is accentuated in inner-EAC5 trade, where Kenya's share of EAC5 manufacturing trade increased, as already noted in Section 4.3. These patterns highlight that policy efforts might be needed to strengthen the region's comparative advantage in food processing and tourism and to reverse the trend towards unidirectional regional manufacturing trade and value chains that further strengthen Kenya's role as a supplier of intermediates and regional hegemon.

Figure 21: GROWTH OF (N)RCA BETWEEN 2006-2010 AND 2015-2019 (MEDIAN)



*Notes:* Figure shows the growth rate in percentage terms between the 2006-10 and 2015-19 (N)RCA medians. EM estimates are based on DVA in exports, BACI on gross exports. Appendix Figure B5 and Table B5 show the values.

## 6 Trade, Value Chains, and Economic Development

Having extensively documented the patterns of EAC global and regional integration through both traditional trade (Section 3) and value chains (Section 4) while highlighting salient trends, potentials, imbalances, and policy priorities, a critical remaining policy questing regards the impact of different forms of integration on economic development in the region. This section attempts to provide causal reduced-form evidence on this matter following [Kummritz \(2016\)](#).

### 6.1 Review of the Empirical Literature

The WDR Chapter 3 presents extensive correlational evidence that GVC participation is associated with gains in GDP per capita growth and labor productivity, poverty reduction, skill transfer, and employment creation, often benefiting gender equality, but also with challenges to taxation and higher inequality ([World Bank, 2020](#); [Antràs & Chor, 2022](#)). The report highlights that long-term firm-to-firm links and specialization in GVC-related tasks promote efficient production, technology diffusion, and access to capital. A cross-country dynamic growth regression estimated with System-GMM yields an 11-14% improvement in per-capita GDP following a 10% increase in overall GVC participation, which is contrasted with a 2% gain from increased trade in products fully produced in one country. Developing countries experience the biggest growth spurt upon transitioning from commodities to limited manufacturing, typically reaping 20% income gains within 3 years.

These findings are broadly echoed in much macroeconomic work on GVCs and economic development. Among the first, [Kummritz \(2016\)](#) assesses the effect of GVC participation on labor productivity and DVA using OECD ICIOS for 61 countries and 34 industries from 1995-2011. He develops a novel instrumental variable (IV) for GVC participation - a VA trade resistance index combining third-country trade costs with industry-specific technological variables - and estimates that a 1 percent increase in VS leads to 0.11% higher DVA in the average industry, and a 1 percent increase in VS1 leads to 0.60% higher DVA and 0.33% higher labor productivity. The effects of forward integration (VS1) are greater for high-income countries, whereas low/middle-income countries show stronger returns from backward integration (VS).

[Altomonte et al. \(2018\)](#), using an IV combining the growing size of container ships since 1997 with the ex-ante availability of deep sea ports, also present causal evidence of a positive effect of GVC-related trade (DVA in exports) on growth, which is larger than the effect of traditional trade. Both are three-step IV strategies following [Romer et al. \(1999\)](#) and [Feyrer \(2009, 2019\)](#).

[Constantinescu et al. \(2019\)](#), used the WIOD with 40 countries and 13 sectors over 1995–2009, and find that (backward) GVC participation boosts labor productivity. An increase of 10% yields an average productivity increase of 1.7%. Examining a sample of 24 emerging economies, [Jangam & Rath \(2021\)](#) show that both forward and backward participation significantly improve DVA in exports from 1995–2011. [Altun et al. \(2023\)](#) examine the role of GVC participation in high-technology exports for 120 countries during 1995–2019 and find that GVC participation correlates strongly with high-tech exports. [Kummritz et al. \(2017\)](#) find that GVC participation increases VA, especially in upstream stages. [Pahl & Timmer \(2020\)](#) study the effects of GVC participation on VA in 58 countries (of which 38 developing) between 1970 and 2008 and find a robust positive effect on manufacturing productivity growth, especially for less productive countries where the distance to the global frontier is large. However, they find no positive effects on employment and some negative effects for middle-income countries. Thus, they conclude that GVC participation is a mixed blessing, inducing skill-biased technological change, in line with [Rodrik \(2018\)](#). [Kummritz \(2016\)](#) notes that GVCs do not necessarily need to benefit developing countries as there could be adverse terms of trade effects or decreases in productive endowments from heavy engagement in them, which is also shown in some theoretical models such as [Baldwin & Robert-Nicoud \(2014\)](#). An argument by [Kummritz \(2015\)](#) is also that GVCs might substitute foreign for domestic suppliers. However, his empirical research suggests that FVA is a rather a complement to DVA.

[Beverelli et al. \(2019\)](#) provide empirical evidence on the relationship between domestic value chains (DVCs) and GVCs. They find that across countries at different stages of development, higher domestic integration by 1 standard deviation raises GVC integration through backward linkages (VS) by 0.4%. DVC integration explains up to 30% of overall GVC participation. They explain these results with fixed costs of fragmentation and switching suppliers: "High fragmentation costs allow, due to their sunk nature, DVCs to act as stepping stones to GVCs" ([Beverelli et al., 2019](#)).

[Shen et al. \(2021\)](#) construct a simple dynamic model to illustrate the micro-mechanism of industrial upgrading along the GVC. Using the WIOD, they find that more upstream industries correlate with higher profitability and VA, capital intensity, and R&D investment. Their dynamic model explains this through three effects: endogenous sunk costs, decreasing intermediate input price elasticity, and sequential pricing effect uncertainty. They show that the empirical patterns revealed in China are consistent with the model's predictions. [Tian et al. \(2022\)](#) also study the relationship between GVC participation and industrial upgrading (process, product, and skill upgrading) using the WIOD. They find that GVC integration increases industrial upgrading for developing and developed countries. Developing countries benefit more from backward GVC participation through importing more sophisticated inputs and learning through embodied knowledge, whereas developed countries upgrade more through forward GVC participation. They interpret their findings as evidence against more critical voices and models questioning the benefits of developing country participation in GVCs, such as [Baldwin & Robert-Nicoud \(2014\)](#) or [Dalle et al. \(2013\)](#). The macroeconomic study of [Lwesya \(2022\)](#) on GVCs and economic upgrading in the EAC, discussed in the introduction, also finds a significant positive effect of lagged FVA on DVA in EAC5 exports, with coefficients implying an elasticity of 0.49.

Many more microeconomic studies also find positive effects of GVC participation on industrial development. [Piernartini & Rubínová \(2014\)](#), for example, use industry-level R&D and patent data for a sample of 29 countries during 2000–2008 and show that knowledge spillovers increase with the intensity of supply chains linkages between countries and that these spillovers are larger than spillovers from traditional trade flows. Similar evidence is presented by [Benz et al. \(2015\)](#), who use firm-level data to show that offshoring leads to knowledge spillovers and that forward spillovers (from producers to users if intermediate inputs) are stronger than backward spillovers.

Microeconomic studies involving EAC members include [Barrientos et al. \(2016\)](#)'s case study of supermarket expansion within southern and eastern Africa, showing that higher quality and

sourcing requirements by global and regional supermarket chains induced improved processes in Kenyan and Ugandan horticulture, allowing diversification and higher fruits and vegetable exports ([World Bank, 2020](#)). A study of Kenyan horticulture by [Krishnan \(2018\)](#) shows that incomes increased after contract farmers adopted quality standards by their international buyers, and also that opportunistic RVCs emerged when suppliers found their produce rejected due to lack of standards compliance, which gradually led to more organized RVCs with own standards and procurement strategies. [Dihel et al. \(2018\)](#) study the effects of value chain participation on African farmers via a survey of 3,935 farmers, 60 aggregators, and 56 buyers in the maize, cassava, and sorghum value chains in Ghana, Kenya, and Zambia, and show that contracted farmers saw greater structural transformation, higher output, and better access to seeds, fertilizers, pesticides, technology, and extension services than non-contracted farmers. These findings are commensurate with [Daly et al. \(2016\)](#)'s study of Maize value chains in East Africa, which identifies Kenyan processors as the lead firms demanding Ugandan suppliers to provide high-quality maize, and document investments into Ugandan production facilities by South African and German companies. They also document challenges in access to finance for farmers, insufficient commercial scale, and lack of communication of market signals and standards along the value chain.

## 6.2 Empirical Strategy

A natural idea to assess the impact of GVC integration on economic development is to investigate if higher GVC participation is associated with higher domestic VA (GDP). Many authors do this in one form or another, including [Lwesya \(2022\)](#) who use DVA in exports. Following [Kummritz \(2016\)](#) and [Rodriguez & Rodrik \(2000\)](#), I argue that running regressions at the country level is subject to omitted variable bias from many factors affecting GVC integration and economic development. Thus, a sector-level regression framework with country-sector, country-year, and sector-year fixed effects is advantageous to capture many confounding factors such as infrastructure, geography, institutions, and economic policies, or multilateral resistance. A caveat is that the coefficients only capture within-industry effects, and are therefore likely lower bound estimates of the overall economic effects of GVC integration. My baseline specification is

$$\log(\text{VA}_{cst}) = \beta \log(\text{GVC}_{cst}) + \alpha_{cs} + \beta_{ct} + \gamma_{st} + \epsilon_{cst}, \quad (14)$$

with  $\text{GVC}_{cst}$  a GVC indicator (such as VS or VS1), and  $\alpha_{cs}$ ,  $\beta_{ct}$ , and  $\gamma_{st}$  country-sector, country-year, and sector-year fixed effects, respectively. The coefficient  $\beta$  could still be biased by sector-level confounders, measurement error in  $\text{GVC}_{cst}$ , simultaneity, and reverse causality.

To address these issues, [Kummritz \(2016\)](#) develops an instrument for GVC participation combining third-party trade costs and industry distance in the value chain to induce exogenous variation in FVA in exports, which forms the basis for simple GVC indicators. The first step is to predict the elements of the **VBE** (VA exports) matrix using exogenous trade costs and industry structure and then compute VS and VS1 indicators following Equations 5 and 6 using this predicted matrix **V̂BE**. These exogenous components  $\hat{VS}_{uj}$  and  $\hat{VS1}_{oi}$  can then be used to instrument VS and VS1 in Eq. 14. Specifically, for each GVC instrument, a different **V̂BE<sub>t</sub>** matrix is constructed, whose (time-varying) elements  $vbe_{oijt}$  are predicted using equations

$$\log(\hat{vbe}_{oijt}) = \beta^{\text{VS}} \log(\tau_{out} \times \delta_{oij}) + \alpha_{uj} + \beta_{ut} + \gamma_{jt} + \epsilon_{oijt} \quad \forall u \neq o \quad (15)$$

$$\log(\hat{vbe}_{oijt}) = \beta^{\text{VS1}} \log(\tau_{out} \times \delta_{oij}) + \alpha_{oi} + \beta_{ot} + \gamma_{it} + \epsilon_{oijt} \quad \forall o \neq u \quad (16)$$

to construct  $\hat{VS}_{uj}$  and  $\hat{VS1}_{oi}$ , respectively.<sup>14</sup> Thus, all variation in the foreign sources of VA ( $oit$ ) in Eq. 15 and in the usage ( $ujt$ ) in Eq. 16, is due to the exogenous trade cost term:  $\log(\tau_{out} \times \delta_{oij})$ .

The trade cost term has two components:  $\tau_{out}$  is an export-weighted estimate of the bilateral trade costs of the supplier of VA ( $o$ ) with all other trading partners ( $k \neq u$ ) in period  $t$ . This is done to preserve the exogeneity of the trade cost measure to factors affecting the specific bilateral  $ou$  link, which may be correlated with GVC-related trade along this link. Following [Kummritz \(2016\)](#), I use the World Bank ESCAP trade costs database ([Arvis et al., 2016](#)) based on [Novy \(2013\)](#), which provides a holistic, tariff-equivalent measure of total trade costs implied by an inverse gravity model.

<sup>14</sup>E.g.,  $\hat{VS}_{ujt}$  is obtained by summing column  $uj$  of a matrix  $\hat{\mathbf{VBE}}_t^{\text{VS}}$  where domestic elements are 0 and the non-domestic ( $u \neq o$ ) elements are estimated using Eq. 15, which includes fixed effects for the using country-sector ( $uj$ ) and time ( $ut, jt$ ) dimensions (as in the final model). Similarly, Eq. 16 estimates  $\hat{\mathbf{VBE}}_t^{\text{VS1}}$  to obtain  $\hat{VS1}_{oit}$ .

The second term,  $\delta_{oiuj}$ , is a time-invariant measure of the distance between industries  $oi$  and  $uj$  along the GVC. It is defined as  $\delta_{ijt} = 1/(u_{oi} \times d_{uj})$ , where  $u_{oi} = \frac{1}{T} \sum_t u_{oit}$  is the average upstreamness of country-sector  $oi$  as defined in Eq. 12 and  $d_{uj} = \frac{1}{T} \sum_t d_{ujt}$  a corresponding downstreamness index, as described e.g. in Antràs & Chor (2022) and footnote 10. Kummritz (2016) notes that the indirect trade costs ( $\tau_{out}$ ) have a larger effect on VA for industries separated by more stages ( $\delta_{oiuj}$ ). The index  $\delta_{oiuj}$  is inverted since  $u_{oi}$  and  $d_{uj}$  have a positive relationship with the elements of **VBE**, to yield a trade cost index  $\tau_{out} \times \delta_{oiuj}$  negatively related to  $vbe_{oiujt}$ .<sup>15</sup><sup>16</sup>

I also estimate economic returns to gross trade and trade in final goods. To instrument these, I omit the industry distance component and instead construct a sector-level time-varying 3rd-party trade cost measure  $\tau_{oit}$ , obtained as exports-weighted average of sector  $i$  in country  $o$ 's exports to all destination countries  $k \neq u$ . This is then used to predict bilateral sector-level trade in gross and final goods using similar zero-stage equations to 15 and 16, and the predictions are summed across importers to yield appropriate instruments for gross and final goods exports, respectively.

At last, I also consider returns to regional integration in both gross and VA terms using an alternative final stage model of the form

$$\log(VA_{cst}) = \beta_1 \log(GVC_{cst}) + \beta_2 SH_{cst}^{EAC} \times \log(GVC_{cst}) + \alpha_{cs} + \beta_{ct} + \gamma_{st} + \epsilon_{cst}, \quad (17)$$

where  $SH_{cst}^{EAC}$  is the EAC share in  $GVC_{cst}$ . Following Section 4.3, this is  $VS^{EAC}$  and  $VS1^{EAC}$  for GVC indicators and the EAC share in gross/final exports for traditional trade. The coefficient  $\beta_2$  gives the additional impact when  $SH_{cst}^{EAC}$  is increased by one unit (100%), i.e., a 1% increase in regional trade yields a  $\beta_1 + \beta_2\%$  increase in VA, whereas a 1% increase in extra-regional trade has an impact of  $\beta_1\%$ . Since the regional share is a component of  $GVC_{cst}$ , obtaining an instrument for it from the zero-stage predictions is straightforward. The RHS of the first stages thus mirror Eq. 17, with  $SH_{cst}^{EAC}$  and  $GVC_{cst}$  replaced by their zero-stage predicted measures.

My default sample includes the full number of sectors (26 for EORA, 134 for EM) for 5 EAC countries: Uganda, Rwanda, Tanzania, Kenya, and Burundi.<sup>17</sup> Estimations are run using indicators computed on EORA 2021, EORA 2015, and EM. With each database, I run one set of estimations using the full set of sectors and one using only manufacturing sectors (all sectors mapping to broad sectors FBE, TEX, WAP, PCM, MPR, ELM, TEQ, MAN, in Table 2). Unfortunately, with EM, all estimates are statistically insignificant and close to zero. This indicates that the high resolution of 134 sectors in these tables is not suitable for evaluating returns to trade and GVC participation in the EAC5. Aggregating to 17 broad sectors also yields insignificant results due to the short time dimension of 6 years. Thus, I do not report EM results.

---

<sup>15</sup>Unlike Kummritz (2016), I employ an industry distance measure ( $\delta_{oiuj}$ ) at the country-sector level, whereas he uses a measure ( $\delta_{ij}$ ) of pure industry distance that is averaged across countries as well. While this common technology assumption may be appropriate for his sample of mostly OECD economies in the OECD TIVA ICIO tables, the instrument constructed using this formulation lacks some relevance for the EAC5. This suggests that industries in developing countries use different technologies and have different GVC positions than the same industries in advanced economies. While using a bilateral-sector-level industry distance measure may partly compromise the exogeneity of the instrument, this is unlikely because this distance is still time-invariant, and the 2SLS regressions include triple fixed effects. Thus, the identifying variation still comes from time-variation in the trade cost term ( $\tau_{out}$ ), and using a more accurate measure of industry distance merely helps increase the relevance of this term. Empirically, I find that computing two instruments using both  $\delta_{oiuj}$  and  $\delta_{ou}$ , and including them both in the first stage often yields a sizeable improvement in the fit, indicating that the difference of local industry structure to the world average interacted with trade costs has some predictive power for FVA in developing countries. In all cases, however, the instruments are very weak.

<sup>16</sup>Another difference to Kummritz (2016) is that I smooth bilateral trade costs using a centered 3-year MA and impute missing values at the end of the sample using the last MA observation carried forward. This is sensible because trade costs based on an inverted gravity model are endogenous to current trade flows and, therefore, more volatile than pure technological or regulatory changes would warrant. The smoothing step does not compromise the instrument's relevance, confirming that the ESCAP measure is noisy. Appendix Figure B6 shows the raw and smoothed bilateral trade costs among EAC5 members. Interestingly, the ESCAP estimates suggest that trading with Kenya is significantly less costly, and Kenya and Uganda also report trade costs below 100% on each other. These costs are endogenous to observed trade flows, and the strong trade links between Kenya and Uganda have already been highlighted several times. However, this perspective entertains the possibility that high and asymmetric trading costs may be another reason for sluggish and asymmetric EAC integration in supply chain trade. In the framework of Antràs & De Gortari (2020), high trade costs imply greater importance for regional GVC participation. Since the IV trade cost measure ( $\tau_{out}$ ) is a weighted average of origin's ( $o$ ) trade costs with third parties, and EAC members trade much more with ROW than with each other, an accurate and timely representation of regional trade costs is irrelevant for the identification.

<sup>17</sup>South Sudan is omitted because of data quality concerns, Congo because of lacking RVC integration with the EAC, and different trading patterns. The inclusion of Congo does not significantly alter the results.

### 6.3 Results: Gross Trade and Trade in Final Goods

Table 7 reports results for gross trade using the full sample of sectors, and Table 8 shows identical regressions for the subset of manufacturing sectors. In both tables, the instruments are weak, and with one exception, not significantly different from OLS.<sup>18</sup> The OLS results suggest an elasticity of VA to gross trade of 0.13-0.25, in line with the 0.2 reported by the WDR. The results are also congruent to Altomonte et al. (2018), who find larger effects around 0.3 using the WIOD and very similar OLS and IV coefficients, with IV being slightly larger than OLS. The effects of trade in final goods on VA are slightly lower at 0.1-0.2, and the effects of both gross and final goods trade in manufacturing sectors (Table 8) are even lower at  $\leq 0.1$ . This indicates that intermediate trade, i.e., GVC-related trade, is more critical for economic development in the EAC, particularly for manufacturing sectors where intermediates account for a larger fraction of total trade.

Table 7: GROSS TRADE EAC5 REGRESSIONS

Dependent Variable:	log(VA)							
Exports Measure:	Gross				Final Goods			
Data:	EORA21		EORA15		EORA21		EORA15	
Model:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
<i>Variables</i>								
log(E)	0.2454*** (0.0419)	0.1072 (0.0770)	0.1305*** (0.0233)	0.1503* (0.0777)	0.1892*** (0.0353)	0.1931** (0.0818)	0.1036*** (0.0207)	0.1776** (0.0719)
<i>Fixed-effects</i>								
# country-sector	130	130	129	129	130	130	129	129
# country-year	110	110	80	80	110	110	80	80
# sector-year	572	572	416	416	572	572	416	416
<i>Fit statistics</i>								
Observations	2,740	2,740	2,023	2,023	2,740	2,740	2,023	2,023
R <sup>2</sup>	0.9859	0.9856	0.9928	0.9928	0.9855	0.9855	0.9927	0.9927
Within R <sup>2</sup>	0.0485	0.0331	0.0238	0.0232	0.0269	0.0269	0.0143	0.0070
Wu-Hausman, p-value	0.0101		0.5572		0.9703		0.0819	
Kleibergen-Paap (1st stage), F	23.99		28.82		3.086		20.28	
Wald (1st stage), p-value	< 0.001		< 0.001		0.0790		< 0.001	

*Driscoll-Kraay (L=2) standard-errors in parentheses. Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1*

*Notes:* Table shows the elasticity of DVA to gross and final goods exports using EORA with full 26 sector resolution for 5 EAC countries: Uganda, Tanzania, Kenya, Rwanda, and Burundi. Estimations are done using both the first edition of EORA (EORA15: years 2000-2015) and the extended version (EORA21: years 2000-2021). The IV specification uses a sector-level exports weighted average of third-country trade costs as an instrument.

Table 8: GROSS TRADE EAC5 REGRESSIONS: MANUFACTURING SECTORS

Dependent Variable:	log(VA)							
Exports Measure:	Gross				Final Goods			
Data:	EORA21		EORA15		EORA21		EORA15	
Model:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
<i>Variables</i>								
log(E)	0.1015*** (0.0324)	0.9659 (0.6435)	0.1175*** (0.0299)	0.2198 (0.1746)	-0.0566 (0.0920)	-0.1048 (0.5590)	0.1075*** (0.0303)	-0.0211 (0.0400)
<i>Fixed-effects</i>								
# country-sector	40	40	40	40	40	40	40	40
# country-year	110	110	80	80	110	110	80	80
# sector-year	176	176	128	128	176	176	128	128
<i>Fit statistics</i>								
Observations	859	859	640	640	859	859	640	640
R <sup>2</sup>	0.9883	0.9817	0.9951	0.9951	0.9882	0.9882	0.9951	0.9950
Within R <sup>2</sup>	0.0077	-0.5515	0.0216	0.0052	0.0022	0.0006	0.0216	-0.0093
Wu-Hausman, p-value	0.1060		0.7317		0.9675		0.5074	
Kleibergen-Paap (1st stage), F	1.009		5.278		0.4104		19.31	
Wald (1st stage), p-value	0.3152		0.0218		0.5217		< 0.001	

*Driscoll-Kraay (L=2) standard-errors in parentheses. Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1*

*Notes:* Table shows the elasticity of DVA to gross and final goods exports using EORA with a sample of 8 manufacturing sectors (FBE, TEX, WAP, PCM, MPR, ELM, TEQ, and MAN in Table 2), for 5 EAC countries: Uganda, Tanzania, Kenya, Rwanda, and Burundi. Estimations are done using both the first edition of EORA (EORA15: years 2000-2015) and the extended version (EORA21: years 2000-2021). The IV specification uses a sector-level exports weighted average of third-country trade costs as an instrument.

<sup>18</sup> Appendix Table B7 shows the zero-stage regressions, with sizeable coefficients but weak within-R<sup>2</sup>.

## 6.4 Results: Backward and Forward GVC Participation

Appendix Table B7 reports the zero stage regressions predicting the elements  $vbe_{oiujt}$ . As expected, the trade cost measures correlate negatively with the elements of **VBE**. I then run both OLS and 2SLS fixed-effects regressions according to Eq. 14 using VS and VS1 measures in log-levels and instrumenting them with  $\hat{V}S$  and  $\hat{V}S1$ , also in log-levels. I estimate 4 specifications: (1) OLS, (2) IV with a time-invariant ( $\delta_{ij}$ ) industry-distance instrument as in Kummritz (2016) (see Footnote 15), (2) IV with the bilateral ( $\delta_{oiuj}$ ) industry-distance instrument, and (4) IV with both instruments.

Table 9 shows the results on the full sample, and Table 10 for the manufacturing sample. Appendix Tables B8 and B9 report the corresponding first stages. The first stages are generally very weak, with many coefficients insignificant or of the wrong sign. Since the trade cost measure  $\tau_{out} \times \delta_{oiuj}$  is negatively correlated with **VBE**, these negative first-stage coefficients could indicate some overfitting at the zero stages (Table B7) which are also quite weak. In any case, this indicates that the IV/2SLS results in Tables 9 and 10 need to be treated with caution, even in cases where first-stage statistics at the bottom of these tables (such as a sizeable Kleinbergen & Paap F-statistic) suggest that they are sufficiently strong. Also notable is that coefficients from the full EORA 200-2021 sample are generally smaller and more often insignificant than those of the WDR (EORA 2000-2015) sample. This appears to reflect a trend change in the data update from 2016 (the fixed effects absorb the structural break). Thus, I consider the results on the WDR sample more reliable, as its IO tables were created using a consistent methodology.

Table 9: GVC PARTICIPATION EAC5 REGRESSIONS

Dependent Variable:	log(VA)							
	EORA21 (2000-2021)				WDR EORA15 (2000-2015)			
Data:	OLS	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS	OLS	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS
Model:								
<i>Variables</i>								
log(VS)	-0.2193** (0.0841)	0.5786 (0.3895)	0.6015 (0.4202)	0.2378*** (0.0827)	-0.0664 (0.0539)	0.2050*** (0.0418)	0.2096*** (0.0412)	0.1922*** (0.0435)
<i>Fit statistics</i>								
Observations	2,740	2,740	2,740	2,740	2,023	2,023	2,023	2,023
R <sup>2</sup>	0.9858	0.9776	0.9771	0.9831	0.9927	0.9919	0.9918	0.9920
Within R <sup>2</sup>	0.0416	-0.5091	-0.5413	-0.1391	0.0066	-0.1037	-0.1075	-0.0935
Wu-Hausman, p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Kleibergen-Paap (1st stage), F	1.652	1.547	32.23		47.22	47.07	53.78	
Wald (1st stage), p-value	0.1988	0.2136	< 0.001		< 0.001	< 0.001	< 0.001	
<i>Variables</i>								
log(E2R)	0.7351*** (0.0396)	0.1842 (1.129)	-0.1586 (2.588)	-0.6351 (2.591)	0.7432*** (0.0607)	0.5716*** (0.1831)	0.5359** (0.1907)	0.7366*** (0.0597)
<i>Fit statistics</i>								
Observations	2,740	2,734	2,733	2,733	2,023	2,017	2,016	2,016
R <sup>2</sup>	0.9950	0.9894	0.9803	0.9608	0.9976	0.9973	0.9972	0.9976
Within R <sup>2</sup>	0.6633	0.2889	-0.3138	-1.618	0.6683	0.6329	0.6164	0.6724
Wu-Hausman, p-value	0.0678	0.0568	0.0016		0.0638	0.0460	0.8330	
Kleibergen-Paap (1st stage), F	0.4409	0.1569	0.2861		3.421	3.989	4.301	
Wald (1st stage), p-value	0.5067	0.6920	0.7512		0.0645	0.0459	0.0137	
<i>Fixed-effects</i>								
# country-sector	130	130	130	130	129	129	129	129
# country-year	110	110	110	110	80	80	80	80
# sector-year	572	572	572	572	416	416	416	416

Driscoll-Kraay (L=2) standard-errors in parentheses. Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Notes: Table shows the elasticity of DVA to backward (VS) and forward (E2R) GVC participation using EORA with full 26 sector resolution for 5 EAC countries: Uganda, Tanzania, Kenya, Rwanda, and Burundi. Estimations are done using both the first edition of EORA (EORA15: years 2000-2015) and the extended version (EORA21: years 2000-2021) via OLS and IV/2SLS. The IV models use exogenous GVC participation predicted by an exports-weighted average of third country trade costs interacted with bilateral ( $\delta_{ij}$ ) or bilateral-sector level ( $\delta_{oiuj}$ ) industry distance along the value chain as instruments. Appendix Table B7 shows the zero stage, and Table B8 the first stage estimations, including the same set of triple fixed effects.

Overall, the results suggest that GVC participation positively affects VA, and that this effect is larger for forward integration and manufacturing sectors (E2R = VS1 is used here to avoid confusion). Drawing from the IV results in the WDR sample, a 1% increase in the foreign content of exports (VS) implies a 0.2% increase in VA in the full sample, and a 0.45-0.5% increase in manufacturing VA. On the other hand, a 1% increase in the re-exported content of exports (E2R) implies a 0.5-0.7% increase in VA, and a 0.8-1% increase in manufacturing VA.

Table 10: GVC PARTICIPATION EAC5 REGRESSIONS: MANUFACTURING SECTORS

Dependent Variable:		log(VA)							
Data:		EORA21 (2000-2021)				WDR EORA15 (2000-2015)			
Model:		OLS	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS	OLS	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS
<i>Variables</i>									
log(VS)		0.1344*** (0.0227)	1.404 (1.945)	1.638 (2.768)	0.4800** (0.1835)	0.0590* (0.0291)	0.4394*** (0.1178)	0.4562*** (0.1247)	0.2089*** (0.0459)
<i>Fit statistics</i>									
Observations		859	859	859	859	640	640	640	640
R <sup>2</sup>		0.9884	0.9685	0.9605	0.9870	0.9951	0.9940	0.9939	0.9949
Within R <sup>2</sup>		0.0190	-1.674	-2.356	-0.1064	0.0054	-0.2179	-0.2381	-0.0293
Wu-Hausman, p-value			0.0974	0.1160	0.0010		0.0025	0.0028	0.0249
Kleibergen-Paap (1st stage), F		0.4762	0.3118	51.79		61.77	62.66	24.34	
Wald (1st stage), p-value		0.4901	0.5765	< 0.001		< 0.001	< 0.001	< 0.001	
<i>Variables</i>									
log(E2R)		0.6724*** (0.0775)	0.8149*** (0.0377)	0.8204*** (0.0390)	0.8366*** (0.0443)	0.5275*** (0.1523)	1.094*** (0.3019)	1.214*** (0.3362)	0.7602** (0.2842)
<i>Fit statistics</i>									
Observations		859	859	859	859	640	640	640	640
R <sup>2</sup>		0.9966	0.9963	0.9962	0.9961	0.9978	0.9946	0.9932	0.9972
Within R <sup>2</sup>		0.7148	0.6827	0.6801	0.6721	0.5496	-0.0841	-0.3807	0.4427
Wu-Hausman, p-value			0.0002	< 0.001	< 0.001		0.1988	0.1378	0.5741
Kleibergen-Paap (1st stage), F		7.529	9.387	22.39		11.86	9.076	7.092	
Wald (1st stage), p-value		0.0062	0.0022	< 0.001		0.0006	0.0027	0.0009	
<i>Fixed-effects</i>									
# country-sector		40	40	40	40	40	40	40	40
# country-year		110	110	110	110	80	80	80	80
# sector-year		176	176	176	176	128	128	128	128

Driscoll-Kraay ( $L=2$ ) standard-errors in parentheses. Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Notes: Table shows the elasticity of DVA to backward (VS) and forward (E2R) GVC participation using EORA with a sample of 8 manufacturing sectors (FBE, TEX, WAP, PCM, MPR, ELM, TEQ, and MAN in Table 2), for 5 EAC countries: Uganda, Tanzania, Kenya, Rwanda, and Burundi. Estimations are done using both the first edition of EORA (EORA15: years 2000-2015) and the extended version (EORA21: years 2000-2021) via OLS and IV/2SLS. The IV models use exogenous GVC participation predicted by an exports-weighted average of third country trade costs interacted with bilateral ( $\delta_{ij}$ ) or bilateral-sector level ( $\delta_{oiuj}$ ) industry distance along the value chain as instruments. Appendix Table B7 shows the zero stage, and Table B8 the first stage estimations, including the same set of triple fixed effects.

Larger productivity gains from forward integration are also prevalent in the literature. Kummritz (2016) finds robust benefits of GVC backward and forward integration on VA in both developing and developed countries, with a larger benefit of forward integration (E2R) at elasticities of 0.58 for low/middle-income countries and 0.68 for high-income countries. VS elasticities are smaller around 0.09/0.21, respectively. He also estimates labor productivity elasticities to E2R of 0.29 for low/middle-income countries and 0.49 for high-income countries. In a similar exercise, Kummritz (2015) finds that high-income countries benefit relatively more from forward linkages, whereas middle-income countries also benefit from backward linkages (VS). The results presented here broadly align with these findings, suggesting that both backward and forward integration have sizeable returns in low-income countries. In manufacturing sectors, the estimates for these EAC countries are even greater than those of Kummritz (2016), with VA elasticities from forward integration close to 1, tentatively indicating that low-income African economies (not covered by the OECD TIVA ICIO's) can benefit substantially from increasing their supply of high-quality manufacturing intermediates. I note that these estimates, while large, are still smaller than the 1.1-1.4 elasticities to overall GVC participation (VS + VS1) reported by the WDR.

## 6.5 Results: Regional Integration

Tables 11 and 12 show regional integration estimations for gross and GVC-related trade using the full sample of sectors. Appendix Tables B10 and B11 provide equivalent results for manufacturing sectors. In the manufacturing sample, the interaction term is statistically insignificant.

In this more complex specification, the instruments are even weaker, thus, for GVC-related trade, I only report 2SLS specifications employing both sets of instruments. With gross trade (Table 11), the Wu-Hausmann test fails to reject the exogeneity of the regressor in all but the first IV specification. With GVC-related trade (Table 12) this is also the case for forward integration. For backward integration, the IV specifications have a sizeable negative within- $R^2$  and a huge negative interaction effect. This signifies that the instruments are useless in this more complex case. Therefore, I only interpret the OLS estimates.

Table 11 reports positive interaction terms with significant coefficients between 0.066 and 0.079, suggesting that an increase in regional trade in both gross terms and in final goods yields a 6-8 pp. higher VA return than an increase in extra-regional trade, whose VA return to a doubling of exports is estimated between 10% and 24%. The empirical results thus suggest that regional integration through trade is beneficial for economic growth in the region.

Table 11: EAC5 REGIONAL INTEGRATION VIA GROSS TRADE REGRESSIONS

Dependent Variable:	log(VA)							
Exports Measure:	Gross				Final Goods			
Data:	EORA21		EORA15		EORA21		EORA15	
Model:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
<i>Variables</i>								
log(E)	0.2361*** (0.0463)	0.0290 (0.0910)	0.1034*** (0.0202)	0.1051 (0.0988)	0.1718*** (0.0378)	-0.1314 (0.3728)	0.0837*** (0.0195)	0.2198** (0.1026)
log(E) × SH <sup>EAC5</sup>	0.0555 (0.0405)	0.0661*** (0.0174)	0.0749** (0.0279)	0.0468 (0.0309)	0.0787*** (0.0260)	0.1437 (0.0898)	0.0531 (0.0335)	-0.0397 (0.0352)
<i>Fixed-effects</i>								
# country-sector	130	130	129	129	130	130	129	129
# country-year	110	110	80	80	110	110	80	80
# sector-year	572	572	416	416	572	572	416	416
<i>Fit statistics</i>								
Observations	2,740	2,740	2,023	2,023	2,740	2,740	2,023	2,023
R <sup>2</sup>	0.9859	0.9854	0.9929	0.9929	0.9856	0.9846	0.9928	0.9926
Within R <sup>2</sup>	0.0507	0.0166	0.0307	0.0296	0.0325	-0.0340	0.0181	-0.0071
Wu-Hausman, p-value	0.0001		0.4127		0.2302		0.2113	

Driscoll-Kraay ( $L=2$ ) standard-errors in parentheses. Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Notes: Table reports analogous estimations to Table 7, but now including an interaction term of the log of exports with the regional share in exports, which captures the additional returns from a regional expansion in trade.

Table 12 indicates a significant positive OLS interaction term for forward GVC integration of order 0.13-0.14, implying that a 100% increase in forward integration through regional trade yields a 13-14 pp. higher return than the already sizeable return of 73% to extra-regional forward linkages. For backward integration, the terms on the OLS regression are negative of order 0.14-0.19, but the main effects are also negative. Since backward integration is particularly prone to simultaneity, as evident from Tables 9 and 10, a strong instrument is needed for identification, so these OLS coefficients are likely not very meaningful. Further work is required to create stronger instruments for GVC participation in developing countries.

Table 12: EAC5 REGIONAL INTEGRATION VIA RVCs REGRESSIONS

Dependent Variable:	log(VA)							
GVC Indicator:	Backward Integration (VS)				Forward Integration (E2R)			
Data:	EORA21		EORA15		EORA21		EORA15	
Model:	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
<i>Variables</i>								
log(GVC)	-0.1926** (0.0883)	0.1478*** (0.0481)	-0.0605 (0.0629)	0.0832** (0.0373)	0.7335*** (0.0398)	0.7110*** (0.1008)	0.7315*** (0.0646)	0.7467*** (0.1925)
log(GVC) × SH <sup>EAC5</sup>	-0.1869* (0.1015)	-1.116*** (0.2550)	-0.1397 (0.2464)	-1.960* (1.038)	0.1332*** (0.0280)	0.6306 (0.6673)	0.1383* (0.0754)	0.0082 (0.5347)
<i>Fixed-effects</i>								
# country-sector	130	130	129	129	130	130	129	129
# country-year	110	110	80	80	110	110	80	80
# sector-year	572	572	416	416	572	572	416	416
<i>Fit statistics</i>								
Observations	2,740	2,740	2,023	2,023	2,740	2,733	2,023	2,016
R <sup>2</sup>	0.9858	0.9837	0.9927	0.9917	0.9950	0.9948	0.9976	0.9976
Within R <sup>2</sup>	0.0459	-0.0980	0.0074	-0.1300	0.6648	0.6502	0.6717	0.6731
Wu-Hausman, p-value	< 0.001		< 0.001		0.4314		0.7855	

Driscoll-Kraay ( $L=2$ ) standard-errors in parentheses. Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Notes: Table reports analogous estimations to Table 9, but now including an interaction term of the log of GVC participation (VS or E2R) with its regional share, which captures the additional returns from a regional expansion in GVC participation. Due to the weakness of the instruments, only the 2SLS specification, including both instrumental variables and their respective interaction terms, is reported.

## 7 Summary and Conclusion

Using rich and novel data sources, this study rigorously examines the EAC region's global and regional integration through trade and value chains and their effects on economic development. The analysis focusses on five member countries: Uganda, Tanzania, Kenya, Rwanda, and Burundi.

Several salient patterns stand out. The first is that, exempting a small COVID-related rebound in the share of regional trade, the region is not integrating deeper through gross trade. This is particularly the case for imports, where the EAC share with itself has declined from 10% in 2000 to 7.5% in 2020, while the export share remained constant at around 17%. However, this decline is mainly driven by manufacturing and mining increasing regional trade shares in agriculture, forestry and fishing (AFF) and processed foods and beverages (FBE). Considering total trade (exports+imports), AFF trade with ROW was 10x greater than inner-EAC trade in 2020, down from 40x in 2000. In FBE, this ratio declined from 16x (2000) to 11x (2020). In manufacturing, it increased from 15x (2000) to 20x (2020). Manufacturing trade accounts for 64% of EAC5 goods trade, versus 8.1% (AFF), 15% (FBE), and 13% (mining), and thus drives aggregate patterns.

EAC members assume different roles in regional trade. Kenya is a dominant regional exporter, particularly of manufactured products, where 40% of its exports are regional, but only a moderate importer: 18% of Kenyan agricultural imports and less than 1% of its manufacturing imports come from the region. Tanzania also imports little from the region, only 5% of AFF/FBE and 3% of manufacturing imports. It has regional export shares between 20% (AFF) and 11% (FBE). The smaller economies are much more integrated, with regional export and import shares generally above 20%. Particularly Uganda is becoming a significant regional exporter in AFF and FBE, with regional export shares between 35 and 40%. Burundi also recently became a strong agricultural exporter, at a regional share rising from 5% in 2005 to 60% in 2020.

This suggest that regional integration is unequal and follows a pattern where countries first become regional agricultural exporters and then exporters of limited manufactures. However, these manufactures do not significantly cater to a large share of regional demand and thus do not drive regional integration as manufacturers become more foreign-oriented. The FBE sector is intermediate between these two and shows greater promise for regional integration.

In value added (VA) terms, all members have a foreign content share (VS) between 10% (Kenya) and 30% (Congo). The EU and China are the greatest suppliers of EAC foreign content. Only Uganda, Rwanda, and Burundi have a high regional share in VS of 15-30%. The largest regional supplier is Kenya, mostly of manufacturing inputs, followed by Uganda as a regional supplier of mostly primary agriculture. EAC sectors with the highest regional VS shares are FBE and sale and repair of vehicles, fuel trade and hotels (SMH) at 14% each. Petrochemicals (PCM) and textiles (TEX) also have EAC VS shares of 7-9%. Core manufacturing sectors with overall high VS have small regional shares, such as electrical machinery (ELM), where VS in the EAC is at 35%, but the regional share in VS is only 2.8%. EAC regional integration in supply chains thus concentrates on food processing and light manufacturing but at low regional VS shares. This highlights both the great potential and significant challenges in deepening manufacturing RVCs.

For forward GVC participation (re-exported exports or VS1), the EU is the major GVC partner. Regional forward integration is still in its infancy, at regional VS1 shares below 6% in all EAC members. The strongest forward linkages are between Kenya and Uganda, with Uganda accounting for 4.4% of Kenya's VS1 (approx. 80 million USD) and Kenya accounting for 3.2% of Ugandan VS1 (approx. 30 million USD). At the sector level, 21% of agriculture and manufacturing exports are re-exported, followed by 15% of mining exports and 13% of FBE. Bilaterally, 54% of Kenyan VS1 through Uganda are manufacturing inputs, whereas 45% of Ugandan VS1 through Kenya are agricultural inputs, highlighting the different roles of these two countries in RVCs. 21% of Kenyan manufacturing VS1 is via its EAC partners, indicating that the supply of regional manufacturing inputs is quantitatively important for Kenya. Considering overall EAC VS1 by sector, the highest regional shares are in manufacturing and FBE at 6.7% each, followed by AFF at 4%. Regional forward linkages are thus much weaker than backward linkages, which, in FBE, are twice as large.

The region does not seem to be integrating deeper into GVCs. Backward linkages (VS) show some improvements in the 3 smaller economies in recent years, countered by a slight decline in

the larger economies. Forward linkages (VS1) exhibit a very weak decline. Unlike gross trade, the regional VS share is growing at a slow pace of 0.5 pp. per year, and regional shares in VS1 and VA imports (both gross and final) are growing at 0.2 pp./year. Regional integration is advancing in all sectors, but particularly fast in FBE, at rates above 0.5 pp./year on all metrics. Growth in regional VS1 shares in FBE is also particularly equitable, whereas in core manufacturing, which is integrating in VS and VS1 at 0.25 pp./year, the growth in regional VS1 is entirely driven by Kenya, with other countries experiencing losses. The analysis thus highlights the potential of the FBE sector and challenges fostering more horizontal manufacturing RVCs between members.

Examining the evolving position of EAC sectors in GVCs indicates a downstream shift in all members and sectors, implying a move towards production stages closer to final demand running against the global trend towards longer GVCs. For AFF and FBE, this appears to be good news as it implies more local value addition. For manufacturing sectors, on the other hand, it indicates a shift towards processing trade rather than high-quality intermediates. The transport and tourism (TRA) sector in Kenya also saw a downstream shift, suggesting some local upgrading.

Computing (New) Revealed Comparative Advantage indices signifies that all members and the region as a whole have sizeable (N)RCA in AFF (4.2) and FBE (5) and a strong disadvantage in core manufacturing (below 0.1 in ELM and TEQ). The region also has (N)RCA in travel services (TRA) of 2.8, particularly Kenya (3.1) and Tanzania (3). Relative to the region, Kenya has slight (N)RCA in most manufacturing sectors apart from PCM, where Tanzania and Rwanda perform strongly. Uganda, Kenya, and Burundi have regional (N)RCA in FBE, Kenya and Tanzania in tourism (TRA), and Uganda in electricity supply (EGW). Except for the latter, these estimates are below 2 and thus moderate. They may nevertheless require policy attention, particularly in manufacturing where Kenya has gained relatively. Trends also signify a slight overall EAC (N)RCA loss in AFF and FBE, encouraging policy efforts to increase foods production and exports.

OLS and IV estimates imply that EAC integration through trade and GVCs benefits sector-level economic growth at elasticities of VA to gross exports of 0.13-0.25 and 0.1-0.2 to exports of final goods. This suggests that intermediates trade is more important for economic development than trade in final goods. Manufacturing sectors show lower returns to gross trade. Examining GVC participation yields IV estimates of 0.2 (all sectors) and 0.45 (manufacturing sectors) to backward GVC participation (VS) and 0.6 (all sectors) and 0.9 (manufacturing sectors) to forward GVC participation (VS1/E2R). These resonate with other papers and the 2020 World Development Report finding that GVC participation benefits economic development, particularly forward linkages in manufacturing. The paper also investigates the returns of deeper regional integration vis-a-vis global integration. OLS estimates suggest that deeper regional linkages yield additional returns: The elasticity to gross and final goods exports increases by 0.06-0.08 for regional trade, and the elasticity to forward GVC participation (VS1) increases by 0.13-0.14 for regional links.

The paper thus highlights both prospects of and challenges to EAC regional integration. The region demonstrates a modest level of integration through trade and RVCs, which are concentrated in certain sectors. In particular, the FBE sector demonstrates higher levels of regional integration and growth. At the same time, integration in manufacturing is concentrating on Kenya's role as a supplier of inputs, with a limited supplier role of other countries. There is no clear trend towards greater regional integration through gross trade, and the pace of integration in VA trade, while positive, is very slow. Shifts in comparative advantage suggest a loss of (N)RCA in manufacturing, including FBE, alongside a downstream shift. This should prompt policy action to at least increase output and deepen RVCs in the FBE sector. Broader industrial policy coordination may also be necessary to mitigate Kenya's increasing role as a regional manufacturing hegemon. Sector-level estimates suggest that integration through trade and GVCs benefits domestic activity, particularly within RVCs. Thus, any policy action should be considerate not to slow down or reverse the (already sluggish) trend towards increased EAC regional integration through RVCs.

Regarding the AfCFTA, this study shows that establishing a common market among economies with different distributions of comparative advantage may result in vertical GVCs and RVCs, leading to a loss of competitiveness in certain sectors and countries, particularly in smaller manufacturing sectors. Thus coordination of industrial and GVC-related policies should be considered together with the planned protocols to establish and regulate a common African market.

## References

- Abel, G. J. (2023). *migest: Methods for the indirect estimation of bilateral migration* [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=migest> (R package version 2.0.4)
- Aloo, L. O. (2017). Free movement of goods in the EAC. In *East african community law* (pp. 303–325). Brill Nijhoff.
- Altomonte, C., Bonacorsi, L., & Colantobe, I. (2018). *Trade and Growth in the Age of Global Value Chains* (BAFFI CAREFIN Working Papers No. 1897). BAFFI CAREFIN, Centre for Applied Research on International Markets Banking Finance and Regulation, Universita' Bocconi, Milano, Italy. Retrieved from <https://ideas.repec.org/p/baf/cbafwp/cbafwp1897.html>
- Altun, A., Avsar, I. I., Turan, T., & Yanikkaya, H. (2023). Does global value chain participation boost high technology exports? *Journal of International Development*, 35(5), 820–837.
- Antràs, P., & Chor, D. (2013). Organizing the global value chain. *Econometrica*, 81(6), 2127–2204.
- Antràs, P., & Chor, D. (2018). On the measurement of upstreamness and downstreamness in global value chains. *World Trade Evolution: Growth, Productivity and Employment*, 5, 126–194.
- Antràs, P., & Chor, D. (2022). Global value chains. *Handbook of International Economics*, 5, 297–376.
- Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, 102(3), 412–416.
- Antràs, P., & De Gortari, A. (2020). On the geography of global value chains. *Econometrica*, 88(4), 1553–1598.
- Arvis, J.-F., Duval, Y., Shepherd, B., Utoktham, C., & Raj, A. (2016). Trade costs in the developing world: 1996–2010. *World Trade Review*, 15(3), 451–474.
- Balassa, B. (1965). Trade liberalisation and “revealed” comparative advantage 1. *The Manchester School*, 33(2), 99–123.
- Baldwin, R., & Lopez-Gonzalez, J. (2015). Supply-chain trade: A portrait of global patterns and several testable hypotheses. *The World Economy*, 38(11), 1682–1721.
- Baldwin, R., & Robert-Nicoud, F. (2014). Trade-in-goods and trade-in-tasks: An integrating framework. *Journal of International Economics*, 92(1), 51–62.
- Balié, J., Del Prete, D., Magrini, E., Montalbano, P., & Nenci, S. (2019). Does trade policy impact food and agriculture global value chain participation of sub-saharan african countries? *American Journal of Agricultural Economics*, 101(3), 773–789.
- Barrientos, S., Knorringa, P., Evers, B., Visser, M., & Oondo, M. (2016). Shifting regional dynamics of global value chains: Implications for economic and social upgrading in african horticulture. *Environment and Planning A: Economy and Space*, 48(7), 1266–1283.
- Belotti, F., Borin, A., & Mancini, M. (2020). *icio: Economic analysis with inter-country input-output tables in Stata* (Policy Research Working Paper Series No. 9156). The World Bank. Retrieved from <https://ideas.repec.org/p/wbk/wbrwps/9156.html>
- Benz, S., Larch, M., & Zimmer, M. (2015). Trade in ideas: Outsourcing and knowledge spillovers. *International Economics and Economic Policy*, 12(2), 221–237.
- Beverelli, C., Stolzenburg, V., Koopman, R. B., & Neumueller, S. (2019). Domestic value chains as stepping stones to global value chain integration. *The World Economy*, 42(5), 1467–1494.
- Borin, A., & Mancini, M. (2019). *Measuring what matters in global value chains and value-added trade* (Tech. Rep. No. 8804). World Bank Policy Research Working Paper.
- Constantinescu, C., Mattoo, A., & Ruta, M. (2019). Does vertical specialisation increase productivity? *The World Economy*, 42(8), 2385–2402.

- Dalle, D., Fossati, V., & Lavopa, F. (2013). Industrial policy and developmental space: The missing piece in the GVCs debate. *Revista Argentina de Economía Internacional*, 2, 3–14.
- Daly, J., Abdulsamad, A., & Gereffi, G. (2017). *Regional value chains in east africa: Summary report* (Project report). International Growth Center. Retrieved from <https://www.theigc.org/sites/default/files/2017/05/Daly-et-al-2017-summary-paper.pdf>
- Daly, J., Hamrick, D., Gereffi, G., & Guinn, A. (2016). Maize value chains in east africa. *Duke Center on Globalization, Governance, and Competitiveness*.
- Daudin, G., Rifflart, C., & Schweisguth, D. (2011). Who produces for whom in the world economy? *Canadian Journal of Economics/Revue Canadienne d'Économique*, 44(4), 1403–1437.
- Dihel, N. C., Grover, A., Hollweg, C. H., & Slany, A. (2018). How does participation in value chains matter to african farmers? *World Bank Policy Research Working Paper*(8506).
- Engel, J., Winkler, D., & Farole, T. (2016). *SACU in global value chains : Measuring GVC integration, position, and performance of Botswana, Lesotho, Namibia, South Africa, and Swaziland* (Tech. Rep.). Washington, DC, United States of America: World Bank. Retrieved from <https://policycommons.net/artifacts/1291484/sacu-in-global-value-chains/1894428/> (Retrieved on 20 Jul 2024. CID: 20.500.12592/5bdj09)
- Fally, T. (2012). Production staging: measurement and facts. *University of Colorado Boulder*, 155–168.
- Fernandes, A. M., Kee, H. L., & Winkler, D. (2022). Determinants of global value chain participation: cross-country evidence. *The World Bank Economic Review*, 36(2), 329–360.
- Feyrer, J. (2009). *Distance, trade, and income—the 1967 to 1975 closing of the suez canal as a natural experiment* (Working Paper). National Bureau of Economic Research. Retrieved from <https://www.nber.org/papers/w15557> doi: 10.3386/w15557
- Feyrer, J. (2019). Trade and income—exploiting time series in geography. *American Economic Journal: Applied Economics*, 11(4), 1–35.
- Foster-McGregor, N., Kaulich, F., & Stehrer, R. (2015). *Global Value Chains in Africa* (MERIT Working Papers No. 2015-024). United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT). Retrieved from <https://ideas.repec.org/p/unm/unumer/2015024.html>
- Fowler, M., & Rauschendorfer, J. (2019). *Agro-industrialisation in Uganda* (Working Paper No. F-IH-UGA-006-2). International Growth Center. Retrieved from <https://www.theigc.org/sites/default/files/2019/11/Fowler-and-Rauschendorfer-2019-Working-paper-v2.pdf>
- Gaulier, G., & Zignago, S. (2010). *BACI: International trade database at the product-level. The 1994-2007 version* (Working Papers No. 2010-23). CEPII. Retrieved from <http://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=2726>
- Hummels, D., Ishii, J., & Yi, K.-M. (2001). The nature and growth of vertical specialization in world trade. *Journal of International Economics*, 54(1), 75–96.
- Huo, J., Chen, P., Hubacek, K., Zheng, H., Meng, J., & Guan, D. (2022). Full-scale, near real-time multi-regional input–output table for the global emerging economies (EMERGING). *Journal of Industrial Ecology*, 26(4), 1218–1232.
- IMF General Statistics Division. (1993). *Direction of trade statistics*. International Monetary Fund. Retrieved from <https://data.imf.org/?sk=9d6028d4f14a464ca2f259b2cd424b85>
- Jangam, B. P., & Rath, B. N. (2021). Does global value chain participation enhance domestic value-added in exports? evidence from emerging market economies. *International Journal of Finance & Economics*, 26(2), 1681–1694.
- Koopman, R., Wang, Z., & Wei, S.-J. (2014). Tracing value-added and double counting in gross exports. *American Economic Review*, 104(2), 459–94.

- Kowalski, P., Gonzalez, J. L., Ragoussis, A., & Ugarte, C. (2015, April). *Participation of developing countries in global value chains: Implications for trade and trade-related policies* (OECD Trade Policy Papers No. 179). OECD Publishing. Retrieved from <https://ideas.repec.org/p/oec/traaab/179-en.html> doi: 10.1787/5js33lfw0xxn-en
- Krishnan, A. (2018). The origin and expansion of regional value chains: The case of kenyan horticulture. *Global Networks*, 18(2), 238–263.
- Kummritz, V. (2015). *Global value chains: benefiting the domestic economy?* (Tech. Rep.). Graduate Institute of International and Development Studies Working Paper.
- Kummritz, V. (2016). Do global value chains cause industrial development? *CTEI Working Paper No 2016-01*. Retrieved from <http://repec.graduateinstitute.ch/pdfs/cteiwp/CTEI-2016-01.pdf> doi: 10.1002/pd
- Kummritz, V., & Quast, B. (2016). Global value chains in low and middle income countries. *CTEI Working Papers*(10). Retrieved from <https://ideas.repec.org/p/gii/cteiwp/ctei-2016-10.html>
- Kummritz, V., Taglioni, D., & Winkler, D. E. (2017). *Economic upgrading through global value chain participation: Which policies increase the value added gains?* (Tech. Rep. No. 8007). World Bank Policy Research Working Paper.
- Lenzen, M., Kanemoto, K., Moran, D., & Geschke, A. (2012). Mapping the structure of the world economy. *Environmental Science & Technology*, 46(15), 8374–8381.
- Lenzen, M., Moran, D., Kanemoto, K., & Geschke, A. (2013). Building eora: a global multi-region input–output database at high country and sector resolution. *Economic Systems Research*, 25(1), 20–49.
- Leontief, W. W. (1936). Quantitative input and output relations in the economic systems of the united states. *The Review of Economic Statistics*, 105–125.
- Lwesya, F. (2022). Integration into regional or global value chains and economic upgrading prospects: an analysis of the East African Community (EAC) bloc. *Future Business Journal*, 8(1), 33.
- Mancini, M., Montalbano, P., Nenci, S., & Vurchio, D. (2024). Positioning in global value chains: World map and indicators, a new dataset available for GVC analyses. *The World Bank Economic Review*, lhae005.
- Miller, R. E., & Temurshoev, U. (2017). Output upstreamness and input downstreamness of industries/countries in world production. *International Regional Science Review*, 40(5), 443–475.
- MoFPED. (2021). *Beyond recovery: Policies towards resurgent growth in uganda* (A. Musisi & R. Newfarmer, Eds.). Ministry of Finance, Planning and Economic Development of Uganda and International Growth Center. Retrieved from <https://mepd.finance.go.ug/documents/EGF/EGF-Book.pdf>
- Novy, D. (2013). Gravity redux: measuring international trade costs with panel data. *Economic Inquiry*, 51(1), 101–121.
- Obasaju, B. O., Olayiwola, W. K., Okodua, H., Adediran, O. S., & Lawal, A. I. (2021). Regional economic integration and economic upgrading in global value chains: selected cases in africa. *Heliyon*, 7(2).
- Pahl, S., & Timmer, M. P. (2020). Do global value chains enhance economic upgrading? a long view. *The Journal of Development Studies*, 56(9), 1683–1705.
- Piermartini, R., & Rubínová, S. (2014). *Knowledge spillovers through international supply chains* (Tech. Rep.). The Graduate Institute of International and Development Studies.
- Quast, B., & Kummritz, V. (2015). *Decompr: Global value chain decomposition in r* (CTEI Working Papers series No. 01-2015). Centre for Trade and Economic Integration, The Graduate Institute. Retrieved from <https://ideas.repec.org/p/gii/cteiwp/ctei-2015-01.html>

- Rodriguez, F., & Rodrik, D. (2000). Trade policy and economic growth: a skeptic's guide to the cross-national evidence. *NBER Macroeconomics Annual*, 15, 261–325.
- Rodrik, D. (2018). *New technologies, global value chains, and developing economies* (Tech. Rep.). National Bureau of Economic Research.
- Romer, D. H., Frankel, J. A., et al. (1999). Does trade cause growth? *American Economic Review*, 89(3), 379–399.
- Shen, J. H., Zhang, L., Lee, C.-C., Zhang, J., & Shen, L. (2021). Towards a dynamic model of the industrial upgrading with global value chains. *The World Economy*, 44(9), 2683–2702.
- Tian, K., Dietzenbacher, E., & Jong-A-Pin, R. (2022). Global value chain participation and its impact on industrial upgrading. *The World Economy*, 45(5), 1362–1385.
- Timmer, M., Erumban, A. A., Gouma, R., Los, B., Temurshoev, U., de Vries, G. J., ... others (2012). *The world input-output database (WIOD): Contents, sources and methods* (Tech. Rep.). Institute for International and Development Economics.
- Tinta, A. A. (2017). The determinants of participation in global value chains: The case of ECOWAS. *Cogent Economics & Finance*, 5(1), 1389252.
- Van Campenhout, B., Nabwire, L., Minten, B., & Ariong, R. M. (2020). *Institutional and technological innovations to foster agro-industrialization in uganda: Insights from the dairy value chain*. Intl Food Policy Research Institute.
- World Bank. (2020). *Trading for development in the age of global value chains*. The World Bank Washington, DC.

## Appendix

The appendix has two parts: Part A provides additional concordances for countries and sectors to their corresponding aggregates and EORA data quality reports; Part B provides additional tables and figures, many of which are referred to from the main text.

### A. Data Aggregation and EORA Quality Reports

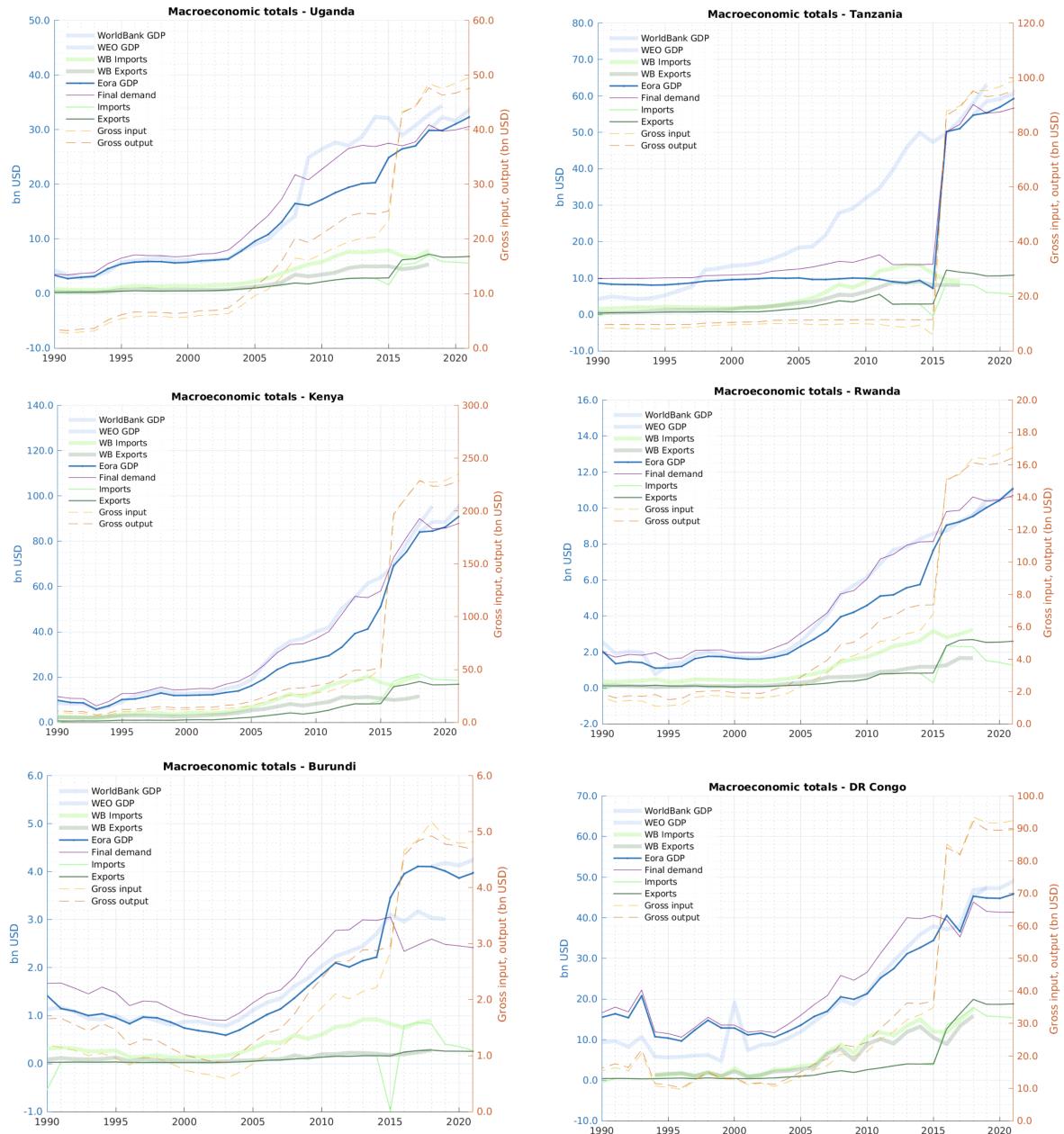
Table A1: COUNTRIES AND REGIONS

<i>Region</i>	<i>Description</i>	<i>Countries</i>
EAC	East African Community	UGA, TZA, KEN, RWA, BDI, COD, SSD
SSA	Sub-Saharan Africa (Excluding EAC)	AGO, BEN, BFA, BWA, CAF, CIV, CMR, COG, COM, CPV, ERI, ETH, GAB, GHA, GIN, GMB, GNB, GNQ, LBR, LSO, MDG, MLI, MOZ, MRT, MUS, MWI, NAM, NER, NGA, SDN, SEN, SLE, SOM, STP, SWZ, SYC, TCD, TGO, ZAF, ZMB, ZWE
EUU	European Union + GBR	AUT, BEL, BGR, CYP, CZE, DEU, DNK, ESP, EST, FIN, FRA, GBR, GRC, HRV, HUN, IRL, ITA, LTU, LUX, LVA, NLD, POL, PRT, ROU, SVK, SVN, SWE, MLT
ECA	Europe and Central Asia (Non-EU)	ALB, AND, ARM, AZE, BIH, BLR, CHE, CHI, FRO, GEO, GIB, GRL, IMN, ISL, KAZ, KGZ, LIE, MCO, MDA, MKD, MNE, NOR, RUS, SMR, SRB, TJK, TKM, TUR, UKR, UZB, XKX
MEA	Middle East and North Africa	ARE, BHR, DJI, DZA, EGY, IRN, IRQ, ISR, JOR, KWT, LBN, LBY, MAR, OMN, PSE, QAT, SAU, SYR, TUN, YEM
NAC	North America and Canada	BMU, CAN, USA
LAC	Latin America and Caribbean	ABW, ARG, ATG, BHS, BLZ, BOL, BRA, BRB, CHL, COL, CRI, CUB, CUW, CYM, DMA, DOM, ECU, GRD, GTM, GUY, HND, HTI, JAM, KNA, LCA, MAF, MEX, NIC, PAN, PER, PRI, PRY, SLV, SUR, SXM, TCA, TTO, URY, VCT, VEN, VGB, VIR
ASE	ASEAN	BRN, IDN, KHM, LAO, MMR, MYS, PHL, SGP, THA, VNM
SAS	South Asia	AFG, BGD, BTN, IND, LKA, MDV, NPL, PAK
CHN	China	CHN, HKG, TWN
ROA	Rest of Asia	ASM, GUM, JPN, KOR, MAC, MNG, MNP, NCL, PRK, PYF, TLS
OCE	Oceania	AUS, FJI, FSM, KIR, MHL, NRU, NZL, PLW, PNG, SLB, TON, TUV, VUT, WSM

Table A2: EMERGING Sectors and Mapping to Broad Sectors

#	EMERGING Sector Definition	BSC	Broad Sector Definition of Huo et al. (2022)
1	Live Animals	AFF	Agriculture, Hunting, Forestry & Fishing
2	Meat and Edible Meat Offal	FBE	Food Production, Beverages & Tobacco
3	Fish, Crustaceans, Molluscs, Aquatic Invertebrates Ne	AFF	Agriculture, Hunting, Forestry & Fishing
4	Dairy Products, Eggs, Honey, Edible Animal Product Ne	FBE	Food Production, Beverages & Tobacco
5	Products of Animal Origin, Nes	AFF	Agriculture, Hunting, Forestry & Fishing
6	Live Trees, Plants, Bulbs, Roots, Cut Flowers Etc	AFF	Agriculture, Hunting, Forestry & Fishing
7	Edible Vegetables and Certain Roots and Tubers	AFF	Agriculture, Hunting, Forestry & Fishing
8	Edible Fruit, Nut, Peel of Citrus Fruit, Melons	AFF	Agriculture, Hunting, Forestry & Fishing
9	Coffee, Tea, Mate and Spices	FBE	Food Production, Beverages & Tobacco
10	Cereals	AFF	Agriculture, Hunting, Forestry & Fishing
11	Milling Products, Malt, Starches, Inulin, Wheat Glute	AFF	Agriculture, Hunting, Forestry & Fishing
12	Oil Seed, Oleaginous Fruits, Grain, Seed, Fruit, Etc, Ne	FBE	Food Production, Beverages & Tobacco
13	Lac, Gums, Resins, Vegetable Saps and Extracts Nes	FBE	Food Production, Beverages & Tobacco
14	Vegetable Plaiting Materials, Vegetable Products Nes	FBE	Food Production, Beverages & Tobacco
15	Animal, vegetable Fats and Oils, Cleavage Products, et	FBE	Food Production, Beverages & Tobacco
16	Meat, Fish and Seafood Food Preparations Ne	FBE	Food Production, Beverages & Tobacco
17	Sugars and Sugar Confectionery	FBE	Food Production, Beverages & Tobacco
18	Cocoa and Cocoa Preparations	FBE	Food Production, Beverages & Tobacco
19	Cereal, Flour, Starch, Milk Preparations and Products	FBE	Food Production, Beverages & Tobacco
20	Vegetable, Fruit, Nut, Etc Food Preparations	FBE	Food Production, Beverages & Tobacco
21	Miscellaneous Edible Preparations	FBE	Food Production, Beverages & Tobacco
22	Beverages, Spirits and Vinegar	FBE	Food Production, Beverages & Tobacco
23	Residues, Wastes of Food Industry, Animal Fodder	FBE	Food Production, Beverages & Tobacco
24	Tobacco and Manufactured Tobacco Substitutes	FBE	Food Production, Beverages & Tobacco
25	Salt, Sulphur, Earth, Stone, Plaster, Lime and Cement	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
26	Ores, Slag and Ash	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
27	Mineral Oils, Oils, Distillation Products, Etc	MIN	Mining & Quarrying
28	Inorganic Chemicals, Precious Metal Compound, Isotope	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
29	Organic Chemicals	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
30	Pharmaceutical Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
31	Fertilizers	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
32	Tanning, Dyeing Extracts, Tannins, Derivatives, Pigments et	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
33	Essential Oils, Perfumes, Cosmetics, Toilettries	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
34	Sooty, Lubricants, Waxes, Candles, Modelling Pastes	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
35	Allotropoids, Modified Starches, Glue, Enzymes	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
36	Explosives, Pyrotechnic Matches, Pyromorphics, Etc	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
37	Photographic or Cinematographic Goods	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
38	Miscellaneous Chemical Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
39	Plastics and Articles Thereof	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
40	Rubber and Articles Thereof	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
41	Row Hides and Skins (Other than Furskins) and Leather	TEX	Textiles, Leather & Wearing Apparel
42	Articles of Leather, Animal Gut, Harness, Travel Good	TEX	Textiles, Leather & Wearing Apparel
43	Furskins and Artificial Fur, Manufactures Thereof	TEX	Textiles, Leather & Wearing Apparel
44	Wood and Articles of Wood, Wood Charcoal	WAP	Wood, Paper & Publishing
45	Cork and Articles of Cork	WAP	Wood, Paper & Publishing
46	Manufactures of Plaiting Material, Basketwork, Etc.	WAP	Wood, Paper & Publishing
47	Pulp of Wood, Fibrous Cellulosic Material, Waste Etc	WAP	Wood, Paper & Publishing
48	Paper & Paperboard, Articles of Pulp, Paper and Board	WAP	Wood, Paper & Publishing
49	Printed Books, Newspapers, Pictures Etc	WAP	Wood, Paper & Publishing
50	Silk	TEX	Textiles, Leather & Wearing Apparel
51	Wool, Animal Hair, Horsehair Yarn and Fabric Thereof	TEX	Textiles, Leather & Wearing Apparel
52	Cotton	TEX	Textiles, Leather & Wearing Apparel
53	Vegetable Textile Fibres Nes, Paper Yarn, Woven Fabri	TEX	Textiles, Leather & Wearing Apparel
54	Mannmade Filaments	TEX	Textiles, Leather & Wearing Apparel
55	Mannmade Staple Fibres	TEX	Textiles, Leather & Wearing Apparel
56	Wadding, Felt, Nonwovens, Yarns, Twine, Cordage, Etc	TEX	Textiles, Leather & Wearing Apparel
57	Carpets and Other Textile Floor Coverings	TEX	Textiles, Leather & Wearing Apparel
58	Special Woven or Tufted Fabric, Lace, Tapestry Etc	TEX	Textiles, Leather & Wearing Apparel
59	Impregnated, Coated or Laminated Textile Fabric	TEX	Textiles, Leather & Wearing Apparel
60	Knitted or Crocheted Fabric	TEX	Textiles, Leather & Wearing Apparel
61	Articles of Apparel, Accessories Knit or Crochet	TEX	Textiles, Leather & Wearing Apparel
62	Articles of Apparel, Accessories, not Knit or Crochet	TEX	Textiles, Leather & Wearing Apparel
63	Other Made Textile Articles, Sets, Worn Clothing Etc	TEX	Textiles, Leather & Wearing Apparel
64	Footwear, Gaiters and the Like, Parts Thereof	TEX	Textiles, Leather & Wearing Apparel
65	Headgear and Parts Thereof	TEX	Textiles, Leather & Wearing Apparel
66	Umbrellas, Walking-Sticks, Seat-Sticks, Whips, Etc	TEX	Textiles, Leather & Wearing Apparel
67	Bird Skin, Feathers, Artificial Flowers, Human Hair	TEX	Textiles, Leather & Wearing Apparel
68	Stone, Plaster, Cement, Asbestos, Mica, Etc Articles	TEX	Textiles, Leather & Wearing Apparel
69	Ceramic Products Undata	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
70	Glass and Glassware	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
71	Pearls, Precious Stones, Metals, Coins, Etc	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
72	Iron and Steel	MPR	Metal & Metal Products
73	Articles of Iron or Steel	MPR	Metal & Metal Products
74	Copper and Articles Thereof	MPR	Metal & Metal Products
75	Nickel and Articles Thereof	MPR	Metal & Metal Products
76	Aluminium and Articles Thereof	MPR	Metal & Metal Products
77	Lead and Articles Thereof	MPR	Metal & Metal Products
79	Zinc and Articles Thereof	MPR	Metal & Metal Products
80	Tin and Articles Thereof	MPR	Metal & Metal Products
81	Other Base Metals, Cermets, Articles Thereof	MPR	Metal & Metal Products
82	Tools, Implements, Cutlery, Etc of Base Metal	MPR	Metal & Metal Products
83	Miscellaneous Articles of Base Metal	MPR	Metal & Metal Products
84	Nucleus, Reactors, Boilers, Machinery, Etc	ELM	Electrical & Machinery
85	Electrical, Electronic Equipment	ELM	Electrical & Machinery
86	Railway, Tramway Locomotives, Rolling Stock, Equipment	TEQ	Transport Equipment
87	Vehicles Other than Railway, Tramway	TEQ	Transport Equipment
88	Aircraft, Spacecraft, and Parts Thereof	TEQ	Transport Equipment
89	Ships, Boats and Other Floating Structures	TEQ	Transport Equipment
90	Optical, Photo, Technical, Medical, Etc Apparatus	ELM	Electrical & Machinery
91	Clocks and Watches and Parts Thereof	ELM	Electrical & Machinery
92	Musical Instruments, Parts and Accessories	ELM	Electrical & Machinery
93	Arms and Ammunition, Parts and Accessories Thereof	ELM	Electrical & Machinery
94	Furniture, Lighting, Signs, Prefabricated Buildings	MAN	Manufacturing & Recycling
95	Toys, Games, Sports Requisites	MAN	Manufacturing & Recycling
96	Miscellaneous Manufactured Articles	MAN	Manufacturing & Recycling
97	Works of Art, Collectors Pieces and Antiques	MAN	Manufacturing & Recycling
98	Commodities not Specified According to Kind	MAN	Manufacturing & Recycling
99	Electricity	EGW	Electricity, Gas & Water
100	Gas Manufacture, Distribution	EGW	Electricity, Gas & Water
101	Water Collection, Purification, and Distribution	EGW	Electricity, Gas & Water
102	Coal	MIN	Mining & Quarrying
103	Oil	MIN	Mining & Quarrying
104	Gas	MIN	Mining & Quarrying
105	Petroleum, Coal Products	PCM	Petroleum, Chemicals & Non-Metallic Mineral Products
106	Manufacturing Services on Physical Inputs Owned by Others	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
107	Maintenance and Repair Services N.i.e.	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
108	Sea Transport	TRA	Transport
109	Air Transport	TRA	Transport
110	Other Modes of Transport	TRA	Transport
111	Postal and Courier Services	PTE	Post & Telecommunications
112	Goods (Travel)	TRA	Transport
113	Local Transport Services	TRA	Transport
114	Accommodation Services	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
115	Food-Serving Services	SMH	Sale, Maintenance & Repair of Vehicles; Fuel; Trade; Hotels & Restaurants
116	Construction	CON	Construction
117	Direct Insurance	FIB	Financial Intermediation & Business Activity
118	Pension and Standardized Guaranteed Services	FIB	Financial Intermediation & Business Activity
119	Financial Services	FIB	Financial Intermediation & Business Activity
120	Real Estate	FIB	Financial Intermediation & Business Activity
121	Charges for the Use of Intellectual Property N.i.e.	FIB	Financial Intermediation & Business Activity
122	Telecommunications Services	PTE	Post & Telecommunications
123	Computer Services	PTE	Post & Telecommunications
124	Information Services	FIB	Financial Intermediation & Business Activity
125	Research and Development Services	FIB	Financial Intermediation & Business Activity
126	Professional and Management Consulting Services	FIB	Financial Intermediation & Business Activity
127	Engineering Services	FIB	Financial Intermediation & Business Activity
128	Waste Treatment and De-Pollution Agricultural and Mining Services	PAO	Public Administration, Education, Health; Recreation; Other Services
129	Operating Leasing Services	FIB	Financial Intermediation & Business Activity
130	Other Business Services N.i.e.	PAO	Public Administration; Education; Health; Recreation; Other Services
131	Audiovisual and Related Services	FIB	Financial Intermediation & Business Activity
132	Health Services	PAO	Public Administration; Education; Health; Recreation; Other Services
133	Education Services	PAO	Public Administration; Education; Health; Recreation; Other Services
134	Recreation & Other Services	PAO	Public Administration; Education; Health; Recreation; Other Services
135	Government Goods and Services N.i.e.	PAO	Public Administration; Education; Health; Recreation; Other Services

Figure A1: EORA Data Quality Reports: EAC Macroeconomic Totals



## B. Additional Tables and Figures

Figure B1: AVERAGE TRADE FLOWS BY BROAD SECTOR, 2010-2015: EORA: USD BILLIONS

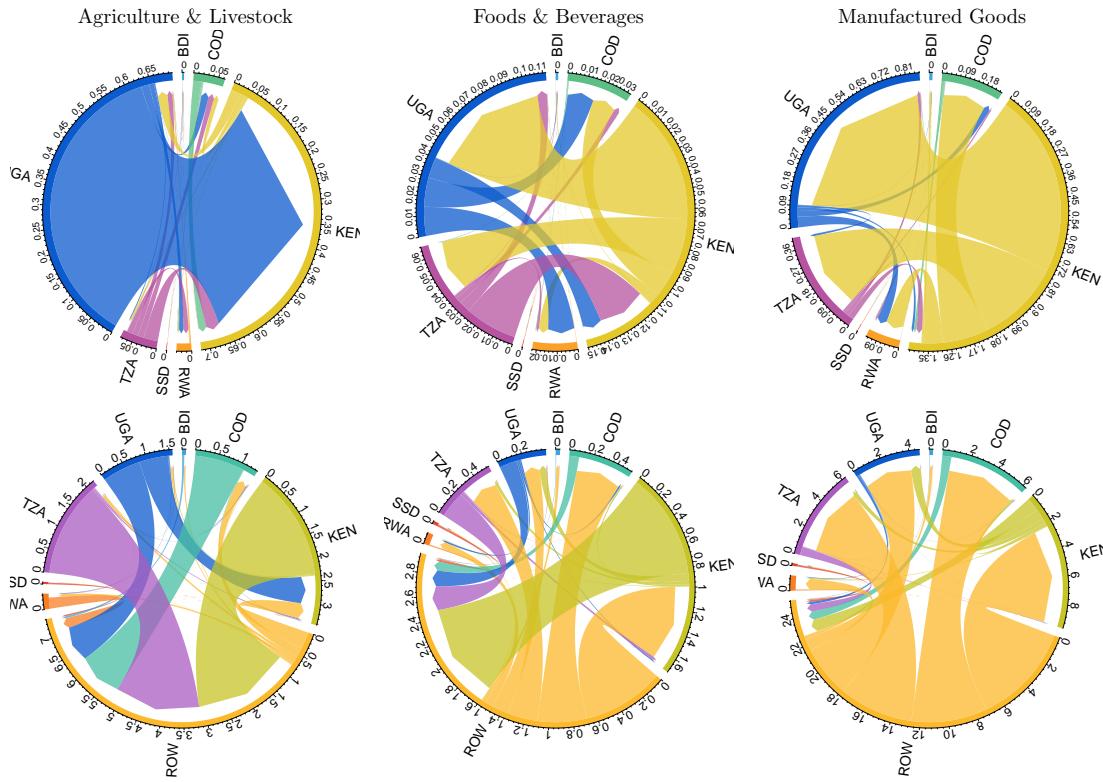


Figure B2: AVERAGE TRADE FLOWS BY BROAD SECTOR, 2010-2015: EMERGING: USD BILLIONS

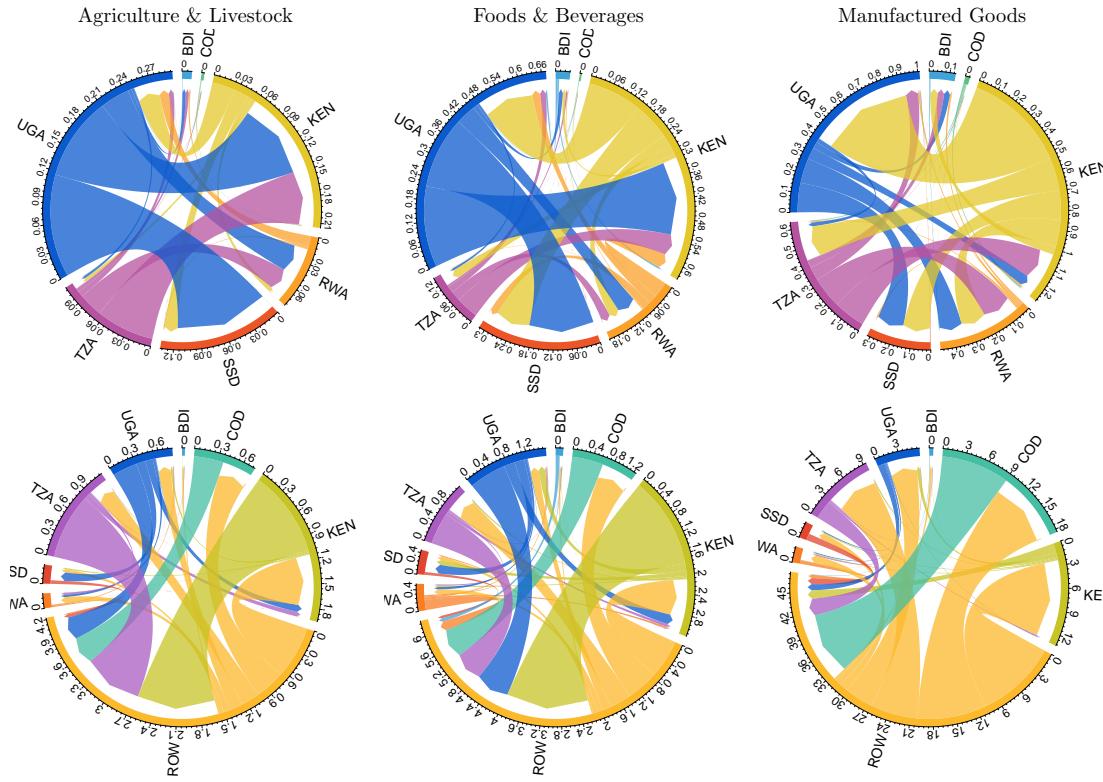
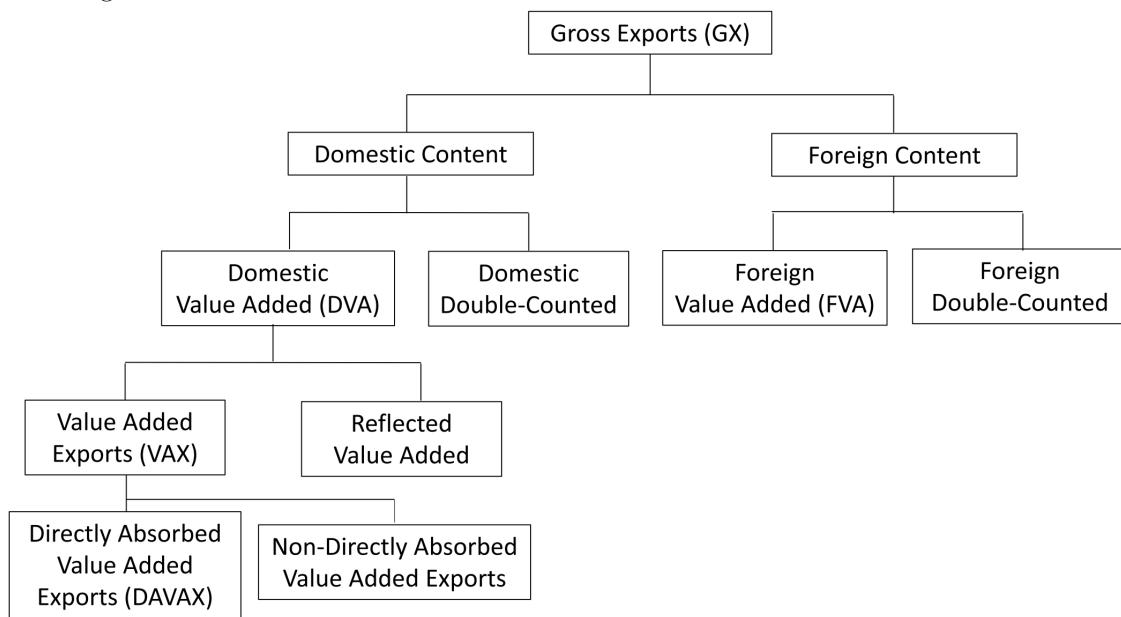


Table B1: LARGEST 50 INTERMEDIATE EAC TRADE FLOWS: EMERGING 2015-19 AVERAGE  
*Millions of Current USD at Basic Prices*

	Overall			Inner-EAC		
	From	To	Value	From	To	Value
2	UGA.PCM	MEA.MIN	254.72	KEN.PCM	UGA.CON	65.28
3	MEA.PCM	KEN.FBE	245.28	KEN.PCM	RWA.AFF	41.38
4	SAS.PCM	TZA.TRA	244.89	KEN.PCM	UGA.FBE	40.42
5	SAS.PCM	KEN.FBE	233.83	UGA.FBE	KEN.TRA	40.12
6	CHN.TEX	TZA.TEX	222.33	UGA.FBE	KEN.FBE	38.88
7	CHN.TEX	KEN.TEX	213.58	TZA.PCM	RWA.AFF	38.36
8	CHN.PCM	KEN.PCM	213.41	UGA.PCM	RWA.AFF	31.44
9	CHN.ELM	TZA.ELM	211.46	KEN.MPR	UGA.CON	31.15
10	CHN.TEX	KEN.TRA	211.10	TZA.AFF	KEN.FBE	30.07
11	KEN.TRA	EUU.TRA	207.65	KEN.FBE	UGA.FBE	29.40
12	TZA.PCM	ECA.PCM	196.53	KEN.PCM	UGA.AFF	29.18
13	KEN.FBE	SAS.FBE	196.09	UGA.AFF	KEN.FBE	28.80
14	KEN.AFF	EUU.FBE	191.72	KEN.PCM	TZA.PCM	27.12
15	MEA.PCM	TZA.TRA	191.05	KEN.MPR	UGA.MPR	22.62
16	CHN.MPR	KEN.EGW	186.79	RWA.FBE	KEN.FBE	21.90
17	UGA.PCM	MEA.CON	180.61	TZA.TEX	KEN.TEX	19.46
18	CHN.TEX	KEN.WAP	175.45	RWA.FBE	KEN.TRA	17.37
19	SAS.PCM	KEN.PCM	172.61	UGA.EGW	KEN.CON	17.14
20	TZA.AFF	SAS.AFF	163.76	UGA.PCM	RWA.TRA	16.29
21	KEN.AFF	EUU.AFF	160.71	KEN.FBE	UGA.SMH	14.99
22	CHN.PCM	KEN.FBE	158.19	UGA.AFF	KEN.AFF	14.28
23	SAS.PCM	KEN.EGW	156.87	KEN.MPR	UGA.PTE	13.76
24	TZA.PCM	SSA.MPR	155.21	TZA.WAP	KEN.WAP	13.70
25	MEA.PCM	KEN.PCM	142.64	TZA.FBE	KEN.TEX	13.64
26	MEA.PCM	KEN.EGW	142.53	KEN.PCM	TZA.FBE	13.10
27	CHN.PCM	TZA.CON	136.05	UGA.FBE	RWA.TRA	13.07
28	CHN.PCM	TZA.PCM	135.90	KEN.AFF	UGA.FBE	12.36
29	TZA.TRA	EUU.TRA	131.02	UGA.FBE	KEN.TEX	12.33
30	KEN.FBE	SAS.PCM	129.12	KEN.PCM	UGA.TRA	12.22
31	UGA.PCM	MEA.PCM	129.11	KEN.PCM	UGA.EGW	11.55
32	MEA.PCM	KEN.CON	128.95	UGA.WAP	KEN.WAP	11.23
33	EUU.PCM	KEN.PCM	114.90	KEN.TEX	UGA.TEX	10.58
34	CHN.PCM	KEN.AFF	114.73	TZA.FBE	KEN.FBE	10.45
35	SAS.PCM	KEN.AFF	113.57	TZA.FBE	KEN.PCM	10.21
36	EUU.PCM	KEN.FBE	113.21	TZA.WAP	KEN.FBE	9.72
37	CHN.ELM	TZA.CON	113.02	UGA.FBE	KEN.PCM	8.79
38	CHN.ELM	UGA.EGW	112.73	UGA.WAP	KEN.FBE	8.42
39	MEA.PCM	TZA.CON	106.08	KEN.PCM	RWA.FBE	8.40
40	KEN.FBE	MEA.FBE	104.53	TZA.TEX	KEN.TRA	8.00
41	CHN.MAN	KEN.MAN	104.29	KEN.PCM	BDI.AFF	8.00
42	UGA.FBE	EUU.FBE	102.62	KEN.FBE	UGA.TRA	7.66
43	TZA.AFF	ASE.AFF	102.24	TZA.FBE	KEN.TRA	7.18
44	CHN.ELM	KEN.CON	100.89	TZA.AFF	KEN.AFF	6.96
45	CHN.MPR	TZA.ELM	98.29	KEN.PCM	TZA.CON	6.93
46	KEN.FBE	EUU.FBE	96.18	KEN.PCM	TZA.TRA	6.83
47	TZA.PCM	SSA.PCM	95.74	TZA.PCM	BDI.AFF	6.80
48	CHN.ELM	KEN.FBE	94.59	KEN.FBE	UGA.PTE	6.49
49	SAS.PCM	KEN.CON	93.59	KEN.PCM	TZA.AFF	6.48
50	TZA.TRA	SAS.TRA	91.13	KEN.WAP	RWA.CON	6.43

Figure B3: REFINED KOOPMAN WANG WEI DECOMPOSITION OF GROSS EXPORTS



Source: Antràs &amp; Chor (2022)

Figure B4: KWW DECOMPOSITION OF GROSS EXPORTS

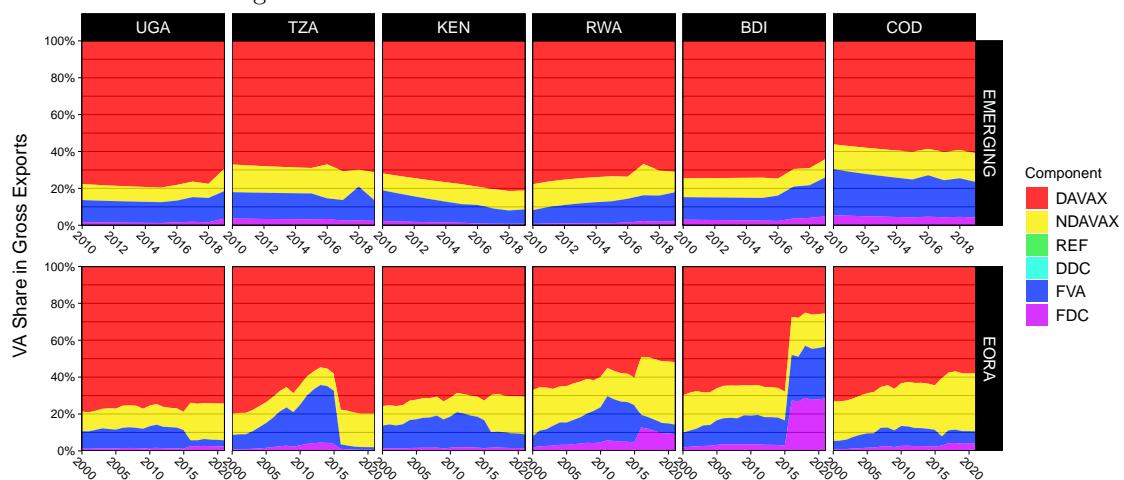


Table B2: (N)RCA Estimates from Figure 19

Country	Source	Flow	AFF	MIN	FBE	TEX	WAP	PCM	MPR	ELM	TEQ	MAN	EGW	CON	SMH	TRA	PTE	FIB	PAO
UGA	WDR_EORA	GX	16.07	0.14	2.49	0.24	0.29	0.18	0.39	0.18	0.28	0.95	0.93	4.57	2.33	2.04	2.76	0.02	1.09
TZA	WDR_EORA	GX	10.48	1.48	2.42	1.63	0.57	0.21	0.20	0.12	0.29	4.58	3.99	2.03	1.15	1.06	1.43	1.02	0.58
KEN	WDR_EORA	GX	11.57	1.09	2.95	0.94	0.94	0.73	0.44	0.26	0.08	1.04	1.61	1.11	1.41	1.70	1.73	0.44	0.19
RWA	WDR_EORA	GX	4.40	4.49	0.29	0.29	0.35	0.16	0.32	0.11	0.10	0.67	4.86	9.81	4.05	1.69	5.12	0.12	2.55
BDI	WDR_EORA	GX	6.83	0.32	0.31	0.38	0.20	0.13	0.24	0.07	0.25	1.18	4.23	12.99	5.53	1.82	6.10	0.10	3.50
EAC5	WDR_EORA	GX	11.54	1.10	2.68	0.96	0.77	0.55	0.38	0.22	0.14	1.64	2.09	2.17	1.64	1.63	2.00	0.48	0.50
UGA	EORA	GX	17.33	0.11	2.83	0.22	0.36	0.19	0.39	0.16	0.35	0.79	1.25	3.98	2.05	1.81	2.69	0.02	1.35
TZA	EORA	GX	11.30	1.08	2.69	1.56	0.71	0.22	0.20	0.10	0.38	3.97	4.98	1.72	0.98	0.93	1.34	1.00	0.67
KEN	EORA	GX	11.95	0.79	3.41	0.89	1.20	0.77	0.44	0.23	0.11	0.92	2.20	0.95	1.24	1.56	1.73	0.44	0.25
RWA	EORA	GX	4.82	3.38	0.34	0.28	0.45	0.17	0.32	0.10	0.14	0.58	6.55	8.74	3.56	1.50	5.07	0.12	3.23
BDI	EORA	GX	7.43	0.24	0.36	0.35	0.26	0.13	0.25	0.07	0.34	0.98	5.49	11.58	4.86	1.62	6.05	0.10	4.40
EAC5	EORA	GX	12.28	0.83	3.07	0.90	0.98	0.58	0.38	0.20	0.19	1.39	2.92	1.91	1.43	1.45	1.94	0.47	0.65
UGA	EMERGING	GX	4.76	0.00	7.16	0.52	0.70	0.82	0.46	0.03	0.05	0.14	7.29	0.19	4.42	1.92	0.03	0.16	0.12
TZA	EMERGING	GX	4.93	0.01	2.46	0.55	0.64	1.38	0.58	0.04	0.04	0.12	0.00	1.61	3.04	3.41	0.27	0.02	0.06
KEN	EMERGING	GX	4.16	0.01	5.98	1.01	0.94	0.55	0.47	0.06	0.10	0.33	1.17	0.00	2.17	3.28	0.29	0.13	6.78
RWA	EMERGING	GX	0.74	0.00	5.11	0.24	0.07	1.41	0.17	0.05	0.06	0.10	0.69	0.51	5.67	2.00	0.02	0.01	2.26
BDI	EMERGING	GX	0.16	0.00	10.83	0.31	0.06	0.95	0.22	0.04	0.08	0.07	0.00	0.01	0.02	1.30	0.12	0.01	24.16
EAC5	EMERGING	GX	4.12	0.01	5.12	0.70	0.77	0.90	0.49	0.06	0.08	0.21	2.09	0.61	3.08	2.90	0.21	0.09	3.16
UGA	BACI	GX	5.03	0.08	7.60	0.88	0.73	0.68	0.80	0.18	0.20	0.24							
TZA	BACI	GX	5.08	0.23	2.85	0.93	0.60	2.42	0.58	0.13	0.08	0.16							
KEN	BACI	GX	5.74	0.45	6.61	1.63	0.89	0.77	0.59	0.17	0.12	0.40							
RWA	BACI	GX	1.27	0.65	5.02	0.51	0.18	2.32	0.26	0.07	0.11	0.18							
BDI	BACI	GX	0.25	0.02	5.53	0.42	0.06	2.84	0.30	0.07	0.09	0.11							
EAC5	BACI	GX	5.21	0.32	5.16	1.19	0.73	1.55	0.58	0.15	0.15	0.28							
UGA	WDR_EORA	VAX	13.87	0.11	2.24	0.20	0.25	0.16	0.33	0.15	0.24	0.82	0.78	4.02	1.94	1.79	2.37	0.02	3.42
TZA	WDR_EORA	VAX	10.87	1.23	2.07	1.31	0.42	0.16	0.13	0.07	0.15	3.15	3.76	1.75	1.01	0.91	1.39	1.04	1.76
KEN	WDR_EORA	VAX	10.62	0.84	2.63	0.91	0.84	0.72	0.29	0.19	0.06	0.98	1.52	0.86	1.22	1.42	1.53	0.38	0.64
RWA	WDR_EORA	VAX	4.55	3.74	0.31	0.24	0.35	0.17	0.30	0.11	0.09	0.27	4.23	10.26	3.49	1.70	5.07	0.11	4.19
BDI	WDR_EORA	VAX	6.36	0.19	0.31	0.30	0.20	0.13	0.23	0.08	0.26	0.82	3.41	12.58	4.41	1.72	5.43	0.09	9.87
EAC5	WDR_EORA	VAX	10.80	0.86	2.38	0.85	0.68	0.54	0.27	0.16	0.10	1.24	1.87	1.91	1.42	1.41	1.80	0.42	1.44
UGA	EORA	VAX	16.59	0.10	2.57	0.19	0.32	0.16	0.34	0.14	0.28	0.71	1.19	3.72	1.92	1.73	2.55	0.02	1.57
TZA	EORA	VAX	13.12	1.06	2.19	1.24	0.52	0.15	0.12	0.06	0.16	2.74	5.25	1.53	0.95	0.85	1.40	1.12	0.75
KEN	EORA	VAX	12.20	0.72	3.07	0.89	1.09	0.75	0.30	0.17	0.07	0.90	2.32	0.80	1.21	1.41	1.69	0.43	0.32
RWA	EORA	VAX	5.37	3.29	0.37	0.23	0.46	0.18	0.30	0.10	0.12	0.26	6.42	9.69	3.43	1.65	5.46	0.13	2.10
BDI	EORA	VAX	7.67	0.17	0.37	0.29	0.26	0.14	0.25	0.08	0.32	0.75	4.96	12.92	4.37	1.68	5.95	0.11	4.75
EAC5	EORA	VAX	12.68	0.76	2.76	0.83	0.88	0.56	0.28	0.14	0.12	1.10	2.97	1.78	1.39	1.35	1.93	0.47	0.72
UGA	EMERGING	VAX	4.67	0.00	6.15	0.42	0.65	0.79	0.40	0.03	0.05	0.12	5.87	0.18	4.36	1.92	0.02	0.16	0.11
TZA	EMERGING	VAX	5.08	0.01	2.53	0.55	0.58	1.43	0.51	0.03	0.04	0.10	0.00	1.58	2.83	3.02	0.23	0.02	0.06
KEN	EMERGING	VAX	4.05	0.01	5.79	0.84	0.60	0.52	0.47	0.05	0.07	0.26	0.99	0.00	2.00	3.13	0.27	0.13	6.02
RWA	EMERGING	VAX	0.77	0.00	4.50	0.26	0.08	1.56	0.18	0.06	0.08	0.11	0.79	0.52	4.64	2.06	0.02	0.01	2.30
BDI	EMERGING	VAX	0.16	0.00	9.67	0.34	0.06	0.69	0.18	0.04	0.07	0.09	0.00	0.01	0.02	1.47	0.13	0.01	26.19
EAC5	EMERGING	VAX	4.17	0.01	5.01	0.61	0.58	0.89	0.43	0.05	0.08	0.18	1.73	0.58	2.85	2.77	0.19	0.09	2.91

Table B3: (NEW) REVEALED COMPARATIVE ADVANTAGE: CORRELATIONS OF 2010-19 MEDIAN

	WDR_EORA_GX	EORA_GX	EORA_VAX	EMERGING_GX	BACI_GX
WDR_EORA_GX	1	.962	.990	.987	.179
WDR_EORA_VAX	.962	1	.960	.967	.341
EORA_GX	.990	.960	1	.995	.215
EORA_VAX	.987	.967	.995	1	.217
EMERGING_GX	.179	.341	.215	.217	1
EMERGING_VAX	.186	.362	.223	.227	.994
BACI_GX	.550	.563	.571	.570	.925
					.934
					1

Notes: WDR EORA is only available for 2010-15, EMERGING misses years 2011-14.

Figure B5: (N)RCA IN 2006-2010 AND 2015-2019 (MEDIAN)

Source: • EMERGING + BACI Flow: ● 2006–2010 ○ 2015–2019

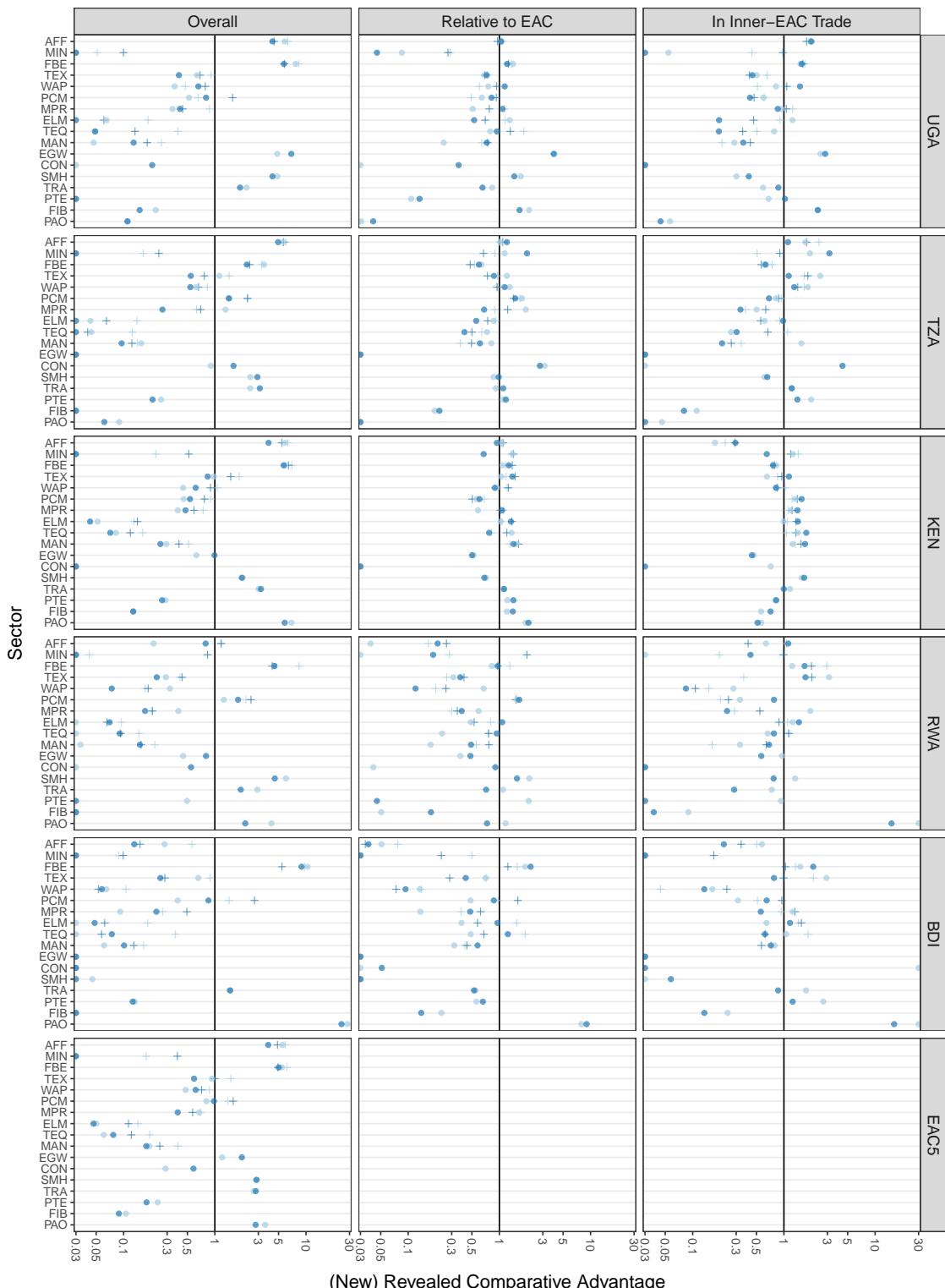


Table B4: (N)RCA Estimates from Figure 20

Country	Source	Flow	AFF	MIN	FBE	TEX	WAP	PCM	MPR	ELM	TEQ	MAN	EGW	CON	SMH	TRA	PTE	FIB	PAO
Relative to the EAC																			
UGA	WDR_EORA	GX	1.39	0.13	0.93	0.26	0.38	0.33	1.02	0.81	1.95	0.58	0.43	2.12	1.41	1.25	1.37	0.05	2.17
TZA	WDR_EORA	GX	0.91	1.32	0.91	1.71	0.74	0.38	0.52	0.54	2.04	2.80	1.90	0.93	0.70	0.65	0.71	2.12	1.16
KEN	WDR_EORA	GX	1.00	0.98	1.10	0.99	1.22	1.32	1.14	1.19	0.56	0.64	0.77	0.51	0.87	1.04	0.88	0.92	0.38
RWA	WDR_EORA	GX	0.39	4.06	0.11	0.30	0.45	0.29	0.84	0.50	0.69	0.40	2.35	4.57	2.47	1.04	2.56	0.24	5.20
BDI	WDR_EORA	GX	0.60	0.29	0.12	0.39	0.26	0.23	0.64	0.34	1.73	0.71	2.01	6.05	3.38	1.12	3.07	0.21	7.14
UGA	EORA	GX	1.34	0.13	0.91	0.26	0.39	0.32	1.03	0.80	1.88	0.57	0.45	2.06	1.40	1.24	1.36	0.05	2.10
TZA	EORA	GX	0.92	1.25	0.87	1.68	0.72	0.37	0.50	0.52	2.02	2.76	1.83	0.90	0.68	0.64	0.69	2.12	1.11
KEN	EORA	GX	0.99	1.00	1.12	0.99	1.23	1.33	1.15	1.20	0.58	0.64	0.79	0.54	0.88	1.05	0.89	0.94	0.40
RWA	EORA	GX	0.40	3.96	0.11	0.31	0.46	0.30	0.86	0.50	0.72	0.41	2.24	4.48	2.46	1.01	2.55	0.24	5.02
BDI	EORA	GX	0.59	0.28	0.12	0.40	0.27	0.23	0.65	0.34	1.78	0.73	1.94	6.26	3.42	1.09	3.03	0.22	6.88
UGA	EMERGING	GX	1.02	0.07	1.43	0.74	0.82	0.82	0.85	0.47	0.70	0.51	4.06	0.31	1.40	0.65	0.13	1.71	0.04
TZA	EMERGING	GX	1.12	1.83	0.56	0.80	1.08	1.48	1.14	0.73	0.50	0.59	0.00	2.71	0.97	1.17	1.27	0.22	0.02
KEN	EMERGING	GX	1.04	0.81	1.18	1.46	1.08	0.63	1.02	1.30	1.34	1.52	0.54	0.00	0.71	1.11	1.36	1.42	2.14
RWA	EMERGING	GX	0.18	0.17	1.02	0.31	0.10	1.34	0.34	0.77	0.70	0.30	0.33	0.81	1.88	0.69	0.07	0.14	0.74
BDI	EMERGING	GX	0.04	0.13	2.08	0.42	0.07	1.03	0.44	0.72	0.95	0.38	0.00	0.02	0.01	0.46	0.51	0.14	7.64
UGA	BACI	GX	0.97	0.29	1.43	0.78	0.96	0.44	1.21	1.14	1.87	0.89							
TZA	BACI	GX	0.97	0.65	0.53	0.75	0.87	1.57	0.97	0.83	0.51	0.63							
KEN	BACI	GX	1.13	1.49	1.28	1.42	1.28	0.48	0.99	1.16	0.98	1.43							
RWA	BACI	GX	0.25	1.95	1.00	0.42	0.24	1.44	0.48	0.52	0.89	0.67							
BDI	BACI	GX	0.05	0.09	1.17	0.35	0.07	1.70	0.43	0.56	0.78	0.39							
UGA	WDR_EORA	VAX	1.28	0.13	0.94	0.24	0.37	0.29	1.22	0.97	2.38	0.65	0.41	2.13	1.35	1.27	1.30	0.04	2.28
TZA	WDR_EORA	VAX	1.01	1.43	0.87	1.51	0.61	0.29	0.45	0.43	1.48	2.51	2.02	0.92	0.70	0.64	0.77	2.48	1.23
KEN	WDR_EORA	VAX	0.98	0.98	1.10	1.07	1.24	1.32	1.08	1.16	0.60	0.78	0.81	0.46	0.87	1.02	0.86	0.90	0.44
RWA	WDR_EORA	VAX	0.42	4.33	0.13	0.28	0.51	0.32	1.10	0.69	0.93	0.21	2.29	5.48	2.46	1.21	2.81	0.27	2.75
BDI	WDR_EORA	VAX	0.60	0.22	0.13	0.35	0.29	0.25	0.85	0.50	2.57	0.65	1.82	6.74	3.11	1.22	3.03	0.22	6.76
UGA	EORA	VAX	1.28	0.13	0.92	0.23	0.38	0.29	1.16	0.93	2.16	0.61	0.42	2.06	1.34	1.26	1.29	0.05	2.21
TZA	EORA	VAX	1.03	1.36	0.79	1.49	0.59	0.27	0.45	0.41	1.31	2.46	1.93	0.86	0.68	0.63	0.73	2.35	1.12
KEN	EORA	VAX	0.97	1.00	1.12	1.09	1.25	1.33	1.10	1.17	0.61	0.80	0.84	0.48	0.89	1.02	0.87	0.92	0.47
RWA	EORA	VAX	0.44	4.28	0.13	0.29	0.52	0.32	1.05	0.68	0.95	0.23	2.27	5.30	2.43	1.20	2.77	0.27	2.75
BDI	EORA	VAX	0.62	0.22	0.13	0.33	0.29	0.25	0.87	0.51	2.64	0.65	1.74	6.94	3.19	1.21	3.13	0.23	6.76
UGA	EMERGING	VAX	1.05	0.06	1.26	0.71	1.09	0.80	0.80	0.66	0.87	0.53	3.98	0.31	1.50	0.67	0.13	1.73	0.04
TZA	EMERGING	VAX	1.15	1.89	0.60	0.87	1.21	1.53	1.12	0.71	0.57	0.62	0.00	2.86	0.97	1.09	1.15	0.21	0.02
KEN	EMERGING	VAX	1.00	0.80	1.24	1.37	0.91	0.58	1.00	1.19	1.07	1.43	0.52	0.00	0.70	1.13	1.41	1.37	2.05
RWA	EMERGING	VAX	0.19	0.19	0.93	0.37	0.15	1.60	0.41	1.03	0.88	0.42	0.44	0.88	1.63	0.75	0.09	0.15	0.80
BDI	EMERGING	VAX	0.04	0.11	2.06	0.53	0.10	0.75	0.41	0.76	0.86	0.57	0.00	0.03	0.01	0.54	0.63	0.15	8.59
In Inner-EAC Trade																			
UGA	EORA	GX	5.74	0.02	0.89	0.18	0.21	0.25	0.41	0.33	1.26	0.47	0.07	0.94	0.71	0.85	0.64	0.00	2.13
TZA	EORA	GX	1.14	0.37	3.74	0.92	0.49	0.38	0.24	0.71	1.49	2.81	1.62	0.98	0.63	0.85	0.90	2.88	2.64
KEN	EORA	GX	0.13	1.22	0.86	1.15	1.18	1.18	1.15	1.14	0.93	0.98	1.11	0.94	1.06	1.03	1.06	1.07	0.66
RWA	EORA	GX	1.64	0.33	0.66	0.43	0.24	0.36	0.43	0.38	0.75	0.54	7.71	15.20	3.40	1.41	3.10	0.45	15.32
BDI	EORA	GX	1.09	0.23	0.35	0.42	0.13	0.12	0.33	0.18	1.83	1.13	7.06	27.72	4.85	0.73	4.00	0.44	24.83
UGA	EMERGING	GX	2.04	0.01	1.61	0.51	1.18	0.53	0.95	0.16	0.17	0.33	0.00	0.40	0.85	0.98	2.47	0.06	2.06
TZA	EMERGING	GX	1.47	2.92	0.61	1.90	1.38	0.75	0.48	0.84	0.29	0.19	4.11	0.64	1.34	1.87	0.09	0.01	
KEN	EMERGING	GX	0.28	0.70	0.72	0.99	0.93	1.52	1.34	1.37	1.79	1.63	0.47	0.00	1.60	1.01	0.74	0.64	0.56
RWA	EMERGING	GX	0.87	0.64	2.00	1.38	0.06	0.49	0.19	0.88	0.53	0.46	0.51	0.00	0.84	0.25	0.02	0.05	13.67
BDI	EMERGING	GX	0.24	0.15	2.11	0.99	0.12	0.57	0.65	0.95	0.84	0.60	0.00	0.00	0.05	0.94	1.26	0.20	21.95
UGA	BACI	GX	1.67	0.52	1.70	0.44	0.89	0.66	1.04	0.70	0.55	0.28							
TZA	BACI	GX	2.07	0.94	0.58	1.83	1.36	0.84	0.50	0.59	0.70	0.63							
KEN	BACI	GX	0.24	1.10	0.76	0.92	0.97	1.32	1.23	1.25	1.31	1.48							
RWA	BACI	GX	0.50	1.07	2.08	1.96	0.11	0.21	0.58	0.73	1.04	0.55							
BDI	BACI	GX	0.44	0.03	1.24	0.97	0.15	0.61	1.33	0.98	0.95	0.46							
UGA	EORA	VAX	4.92	0.02	0.85	0.15	0.18	0.21	0.46	0.37	1.28	0.42	0.06	0.93	0.63	0.81	0.57	0.00	1.88
TZA	EORA	VAX	1.20	0.39	3.80	0.86	0.43	0.29	0.23	0.60	1.17	2.32	1.72	1.09	0.62	0.86	0.96	3.27	2.42
KEN	EORA	VAX	0.14	1.25	0.91	1.19	1.21	1.21	1.16	1.16	0.94	1.05	1.15	0.93	1.08	1.04	1.08	1.09	0.71
RWA	EORA	VAX	1.64	0.32	0.72	0.34	0.24	0.36	0.52	0.49	0.83	0.17	6.53	18.05	3.11	1.53	3.15	0.47	6.67
BDI	EORA	VAX	1.47	0.18	0.39	0.33	0.14	0.13	0.42	0.27	2.31	0.88	5.84	31.37	3.95	0.78	3.89	0.46	19.08
UGA	EMERGING	VAX	1.97	0.01	1.54	0.48	1.36	0.51	0.90	0.20	0.22	0.32	2.68	0.00	0.41	0.85	1.01	2.35	0.05
TZA	EMERGING	VAX	1.42	2.99	0.62	1.43	1.33	0.75	0.42	0.80	0.28	0.22	4.39	0.63	1.22	1.62	0.09	0.01	
KEN	EMERGING	VAX	0.26	0.70	0.78	1.12	0.82	1.54	1.39	1.21	1.68	1.63	0.47	0.00	1.62	1.04	0.82	0.65	0.53
RWA	EMERGING	VAX	0.96	0.77	1.50	1.73	0.09	0.57	0.26	1.36	0.72	0.60	0.76	0.00	0.81	0.31	0.03	0.05	15.93
BDI	EMERGING	VAX	0.22	0.12	1.96	1.06	0.15	0.57	0.56	1.16	0.77	0.66	0.00	0.00	0.05	1.09	1.59		

Table B5: (N)RCA Estimates from Figures B5 and 21

Country	Source	Flow	Period	AFF	MIN	FBE	TEX	WAP	PCM	MPR	ELM	TEQ	MAN	EGW	CON	SMH	TRA	PTE	FIB	PAO
Relative to the EAC																				
UGA	EM	VAX	2006-2010	1.06	0.09	1.41	0.68	0.76	0.64	0.51	1.30	0.79	0.25	4.02	0.00	1.71	0.83	0.11	2.12	0.03
UGA	EM	VAX	2015-2019	1.04	0.05	1.22	0.73	1.14	0.82	1.09	0.53	0.94	0.73	3.94	0.36	1.46	0.65	0.13	1.66	0.04
UGA	EM	VAX	Growth Rate	-1.39	-46.65	-13.86	6.01	50.98	27.69	112.59	-59.19	18.23	196.66	-2.04	Inf	-15.04	-21.60	23.52	-21.53	35.46
TZA	EM	VAX	2006-2010	1.02	1.14	0.64	1.21	1.30	1.77	1.95	0.86	0.73	0.82	0.00	3.14	0.86	0.92	1.09	0.20	0.02
TZA	EM	VAX	2015-2019	1.21	2.01	0.60	0.87	1.14	1.50	0.68	0.56	0.41	0.61	0.00	2.79	0.98	1.10	1.18	0.22	0.02
TZA	EM	VAX	Growth Rate	18.19	75.99	-6.40	-27.84	-12.24	-15.45	-65.06	-35.41	-43.39	-24.94		-11.18	13.80	20.31	8.45	12.72	-19.63
KEN	EM	VAX	2006-2010	1.07	1.39	1.08	1.05	0.94	0.56	0.58	1.03	1.36	1.54	0.52	0.00	0.71	1.13	1.22	1.21	1.94
KEN	EM	VAX	2015-2019	0.94	0.67	1.27	1.39	0.88	0.61	1.07	1.34	0.77	1.43	0.50	0.00	0.68	1.12	1.42	1.41	2.09
KEN	EM	VAX	Growth Rate	-11.93	-51.73	17.33	31.73	-5.04	7.90	83.87	29.57	-43.19	-7.11	-3.87	-46.04	-4.20	-0.80	16.61	16.47	7.66
RWA	EM	VAX	2006-2010	0.04	0.01	0.83	0.31	0.67	1.55	0.59	0.49	0.24	0.18	0.37	0.04	2.13	1.10	2.10	0.05	1.17
RWA	EM	VAX	2015-2019	0.21	0.19	0.96	0.37	0.12	1.65	0.39	1.07	0.93	0.49	0.48	0.90	1.56	0.72	0.05	0.18	0.73
RWA	EM	VAX	Growth Rate	443.79	1341.66	16.12	19.09	-82.12	6.35	-34.65	120.96	296.51	177.25	29.12	2082.03	-26.75	-34.81	-97.83	251.07	-37.38
BDI	EM	VAX	2006-2010	0.05	0.00	1.91	0.71	0.13	0.48	0.14	0.38	0.49	0.32	0.00	0.00	0.02	0.55	0.56	0.23	7.95
BDI	EM	VAX	2015-2019	0.04	0.00	2.20	0.43	0.08	0.87	0.48	0.95	1.24	0.58	0.00	0.05	0.01	0.53	0.66	0.14	9.05
BDI	EM	VAX	Growth Rate	-28.04	15.16	-39.62	-30.48	80.02	252.01	148.31	155.41	80.37	-100.00	1936.67	-59.84	-3.48	17.94	-39.97	13.92	
UGA	BACI	GX	2006-2010	1.07	0.29	1.37	0.74	0.60	0.49	1.13	1.16	1.86	0.64							
UGA	BACI	GX	2015-2019	0.95	0.27	1.26	0.69	0.93	0.78	0.70	1.32	0.74								
UGA	BACI	GX	Growth Rate	-11.13	-5.31	-8.41	-6.84	54.78	88.34	-31.09	-39.41	-28.98	16.22							
TZA	BACI	GX	2006-2010	1.01	0.89	0.53	0.95	0.95	1.68	0.89	0.90	0.64	0.37							
TZA	BACI	GX	2015-2019	1.09	0.67	0.48	0.74	0.94	1.43	1.23	0.74	0.50	0.49							
TZA	BACI	GX	Growth Rate	8.17	-25.00	-9.91	-22.37	-0.75	-14.85	38.13	-17.42	-22.21	32.79							
KEN	BACI	GX	2006-2010	1.06	1.35	1.17	1.18	1.22	0.68	1.11	1.06	0.82	1.30							
KEN	BACI	GX	2015-2019	1.10	1.42	1.38	1.49	1.26	0.51	1.03	1.37	1.21	1.62							
KEN	BACI	GX	Growth Rate	3.24	5.28	18.32	25.93	3.24	-26.02	-7.38	29.15	48.27	24.73							
RWA	BACI	GX	2006-2010	0.17	0.28	1.30	0.26	0.24	1.53	0.30	0.80	0.76	0.56							
RWA	BACI	GX	2015-2019	0.26	2.00	0.96	0.41	0.26	1.52	0.35	0.53	0.76	0.77							
RWA	BACI	GX	Growth Rate	57.96	607.63	-26.33	54.71	29.28	-0.60	15.23	-33.97	0.35	37.42							
BDI	BACI	GX	2006-2010	0.08	0.50	1.57	0.71	0.14	0.97	0.38	1.55	1.92	0.42							
BDI	BACI	GX	2015-2019	0.03	0.23	1.24	0.28	0.07	1.59	0.62	0.58	0.68	0.44							
BDI	BACI	GX	Growth Rate	-55.95	-53.77	-20.90	-60.08	-46.96	64.66	63.99	-62.93	-64.74	5.04							
In Inner-EAC Trade																				
UGA	EM	VAX	2006-2010	1.94	0.05	1.52	0.50	0.82	0.60	0.94	1.24	0.78	0.29	2.52	0.00	0.30	0.59	0.68	2.34	0.06
UGA	EM	VAX	2015-2019	2.00	0.01	1.57	0.45	1.50	0.43	0.86	0.19	0.19	0.36	2.84	0.00	0.41	0.87	1.03	2.36	0.04
UGA	EM	VAX	Growth Rate	3.49	-90.65	3.57	-10.48	83.04	-28.89	-9.28	-84.33	-75.23	24.96	12.66	Inf	36.54	47.02	51.20	0.86	-21.50
TZA	EM	VAX	2006-2010	1.73	1.93	0.61	2.50	1.83	0.81	0.50	0.62	0.26	1.56	0.00	0.00	0.61	1.21	1.99	0.11	0.05
TZA	EM	VAX	2015-2019	1.11	3.15	0.63	1.13	1.30	0.69	0.33	0.98	0.30	0.21	0.00	4.39	0.66	1.22	1.41	0.08	0.01
TZA	EM	VAX	Growth Rate	-35.87	63.25	3.48	-54.80	-29.22	-15.67	-33.43	59.41	14.99	-86.53	Inf	7.54	0.41	-29.39	-27.59	-84.28	
KEN	EM	VAX	2006-2010	0.18	1.26	0.82	0.65	0.83	1.34	1.19	0.99	1.42	1.24	0.47	0.72	1.56	1.17	0.81	0.56	0.56
KEN	EM	VAX	2015-2019	0.29	0.65	0.76	1.14	0.82	1.56	1.41	1.42	1.76	1.71	0.45	0.00	1.67	1.00	0.82	0.71	0.51
KEN	EM	VAX	Growth Rate	67.56	-48.48	-7.78	74.19	-0.82	16.84	18.37	43.50	24.12	37.23	-4.84	-100.00	6.69	-14.63	1.69	26.88	-8.42
RWA	EM	VAX	2006-2010	0.64	0.00	1.23	3.12	0.28	0.33	1.95	1.25	0.67	0.33	0.95	0.00	1.33	0.74	0.93	0.09	30.72
RWA	EM	VAX	2015-2019	1.11	0.43	1.69	1.72	0.08	0.78	0.24	1.46	0.78	0.69	0.56	0.00	0.77	0.28	0.03	0.04	15.14
RWA	EM	VAX	Growth Rate	75.06	Inf	36.84	-44.80	-69.86	136.03	-87.79	16.94	16.78	108.43	-40.66	Inf	-41.92	-61.31	-97.23	-58.24	-50.72
BDI	EM	VAX	2006-2010	0.57	0.00	1.51	2.93	0.16	0.31	1.25	0.65	1.06	0.79	0.00	167.36	0.02	1.75	2.70	0.24	54.71
BDI	EM	VAX	2015-2019	0.22	0.00	2.10	0.78	0.13	0.65	0.56	1.17	0.62	0.72	0.00	0.06	0.86	1.25	0.13	16.18	
BDI	EM	VAX	Growth Rate	-61.82	38.60	-73.45	-18.47	106.52	-55.19	80.69	-41.52	-9.43	-100.00	150.84	-50.61	-53.85	-44.31	-70.42		
UGA	BACI	GX	2006-2010	1.97	0.45	1.70	0.65	0.51	0.61	1.25	0.91	0.51	0.21							
UGA	BACI	GX	2015-2019	1.77	0.98	1.61	0.42	1.07	0.47	1.06	0.47	0.35	0.43							
UGA	BACI	GX	Growth Rate	-9.81	118.48	-5.50	-35.71	109.45	-22.41	-14.85	-48.72	-30.42	104.38							
TZA	BACI	GX	2006-2010	2.42	0.51	0.74	1.68	1.67	0.93	0.38	0.90	1.10	0.34							
TZA	BACI	GX	2015-2019	1.77	0.90	0.56	1.83	1.42	0.88	0.63	0.56	0.67	0.26							
TZA	BACI	GX	Growth Rate	-26.89	77.56	-23.87	9.19	-15.04	-6.17	67.01	-38.13	-39.34	-22.19							
KEN	BACI	GX	2006-2010	0.23	1.44	0.74	0.85	1.03	1.23	1.11	1.09	1.07	1.33							
KEN	BACI	GX	2015-2019	0.29	1.19	0.79	0.95	0.85	1.40	1.23	1.36	1.35	1.54							
KEN	BACI	GX	Growth Rate	27.48	-17.40	5.71	11.44	-17.87	14.15	10.99	24.65	26.66	15.57							
RWA	BACI	GX	2006-2010	1.07	0.20	2.95	0.36	0.15	0.20	0.29	1.10	1.16	0.16							
RWA	BACI	GX	2015-2019	0.41	0.99	2.01	2.02	0.11	0.25	0.55	0.88	1.12	0.64							
RWA	BACI	GX	Growth Rate	-62.09	405.49	-32.05	456.91	-28.76	22.00	89.77	-19.51	-3.06	293.22							
BDI	BACI	GX	2006-2010	0.51	0.00	1.34	2.10	0.04	0.51	0.93	1.45	1.83	0.73							
BDI	BACI	GX	2015-2019	0.34	0.17	1.04	0.99	0.24	0.95	1.32	1.55	0.62	0.57	</						

Figure B6: ESCAP BILATERAL TRADE COST MEASURE FOR THE EAC5

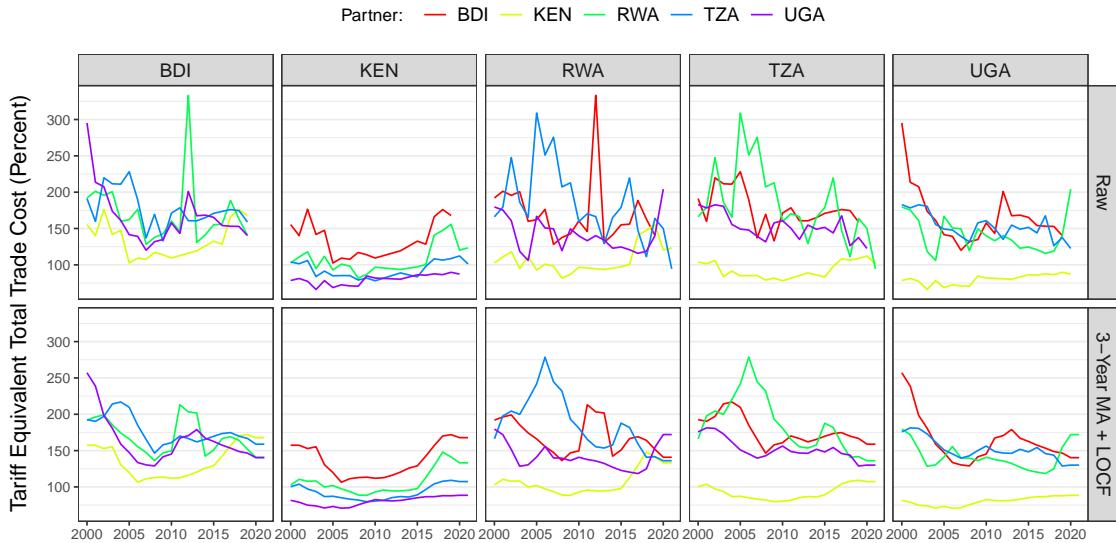


Table B6: ZERO-STAGE REGRESSIONS: GROSS TRADE

Dependent Variable:	Gross Exports		Final Goods Exports	
	EORA	EMERGING	EORA	EMERGING
$\log(\tau_{oiut})$	6.454*** (0.3439)	1.245*** (0.0792)	5.312*** (0.3959)	1.023*** (0.0695)
Observations	155,584	208,635	155,584	208,519
R <sup>2</sup>	0.5971	0.4951	0.6306	0.4915
Within R <sup>2</sup>	0.0225	0.0072	0.0188	0.0073
<i>Fixed-effects</i>				
# country-sector	442	2,202	442	2,201
# country-year	374	102	374	102
# sector-year	572	780	572	780

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Table B7: ZERO-STAGE REGRESSIONS: VALUE ADDED TRADE

Dependent Variable:	$\log(vbe_{oiujt}+1)$			
	EORA	EMERGING	EORA	EMERGING
$\log(\tau_{out} \times \delta_{oiuj})$	-2.774*** (0.0717)	-0.1234*** (0.0049)	-1.820*** (0.0300)	-0.1959*** (0.0062)
R <sup>2</sup>	0.6279	0.2629	0.4301	0.1906
Within R <sup>2</sup>	0.3419	0.0395	0.0303	0.0451
$\log(\tau_{out} \times \delta_{ij})$	-3.018*** (0.0772)	-0.1484*** (0.0060)	-3.112*** (0.0560)	-0.2373*** (0.0072)
R <sup>2</sup>	0.6241	0.2573	0.4402	0.1636
Within R <sup>2</sup>	0.3352	0.0322	0.0475	0.0132
<i>Fixed-effects</i>				
Using Country		Source Country		
# country-sector	441	2,202	440	2,202
# country-year	374	102	374	102
# sector-year	572	780	572	780
Observations	3,928,956	27,381,322	3,928,956	27,381,322

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Table B8: GVC PARTICIPATION EAC5 REGRESSIONS: FIRST STAGES

Data:	EORA21 (2000-2021)			WDR EORA15 (2000-2015)		
Model:	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS
Dependent Variable:				log(VS)		
<i>Variables</i>						
log( $\hat{VS}^{\delta_{ij}}$ )	-0.9737 (0.7575)		-24.23* (12.41)	-2.472*** (0.3597)		-9.525 (6.322)
log( $\hat{VS}^{\delta_{oiuj}}$ )		-1.054 (0.8474)	25.82* (14.56)		-2.735*** (0.3985)	7.846 (7.331)
<i>Fit statistics</i>						
Observations	2,740	2,740	2,740	2,023	2,023	2,023
R <sup>2</sup>	0.9868	0.9868	0.9870	0.9914	0.9914	0.9914
Within R <sup>2</sup>	0.0314	0.0299	0.0493	0.1952	0.1932	0.1976
Kleibergen-Paap, F-stat.	1.652	1.547	32.23	47.22	47.07	53.78
Wald, p-value	0.1988	0.2136	< 0.001	< 0.001	< 0.001	< 0.001
Dependent Variable:				log(E2R)		
<i>Variables</i>						
log( $\hat{E2R}^{\delta_{ij}}$ )	-0.1017 (0.1531)		0.1118 (0.5633)	-0.2892* (0.1564)		-0.7687** (0.3569)
log( $\hat{E2R}^{\delta_{oiuj}}$ )		-0.0529 (0.1334)	-0.1306 (0.3157)		-0.2005* (0.1003)	0.2822 (0.1743)
<i>Fit statistics</i>						
Observations	2,734	2,733	2,733	2,017	2,016	2,016
R <sup>2</sup>	0.9768	0.9767	0.9767	0.9879	0.9878	0.9879
Within R <sup>2</sup>	0.0015	0.0006	0.0007	0.0202	0.0163	0.0278
Kleibergen-Paap, F-stat.	0.4409	0.1569	0.2861	3.421	3.989	4.301
Wald, p-value	0.5067	0.6920	0.7512	0.0645	0.0459	0.0137
<i>Fixed-effects</i>						
# country-sector	130	130	130	129	129	129
# country-year	110	110	110	80	80	80
# sector-year	572	572	572	416	416	416

Driscoll-Kraay (L=2) standard-errors in parentheses

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Table B9: GVC PARTICIPATION EAC5 REGRESSIONS: MANUFACTURING SECTORS: FIRST STAGES

Data:	EORA21 (2000-2021)			WDR EORA15 (2000-2015)		
Model:	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS	IV- $\delta_{ij}$	IV- $\delta_{oiuj}$	2SLS
Dependent Variable:	log(VS)			log(E2R)		
<i>Variables</i>						
$\log(\hat{VS}^{\delta_{ij}})$	-0.5992 (0.8678)		-135.1*** (15.26)	-2.674*** (0.3400)		-113.5*** (16.48)
$\log(\hat{VS}^{\delta_{oiuj}})$		-0.5249 (0.9394)	147.1*** (17.37)		-2.785*** (0.3516)	121.1*** (17.76)
<i>Fit statistics</i>						
Observations	859	859	859	640	640	640
R <sup>2</sup>	0.9928	0.9928	0.9938	0.9968	0.9967	0.9974
Within R <sup>2</sup>	0.0030	0.0019	0.1354	0.0929	0.0846	0.2681
Kleibergen-Paap, F-stat.	0.4762	0.3118	51.79	61.77	62.66	24.34
Wald, p-value	0.4901	0.5765	< 0.001	< 0.001	< 0.001	< 0.001
Dependent Variable:	log(E2R)			log(VS)		
<i>Variables</i>						
$\log(\hat{E2R}^{\delta_{ij}})$	1.251** (0.4556)		-4.393*** (1.140)	-0.1504*** (0.0437)		-0.6463 (1.044)
$\log(\hat{E2R}^{\delta_{oiuj}})$		1.058*** (0.3451)	4.618*** (0.8651)		-0.1237*** (0.0410)	0.4307 (0.9137)
<i>Fit statistics</i>						
Observations	859	859	859	640	640	640
R <sup>2</sup>	0.9873	0.9875	0.9878	0.9921	0.9921	0.9921
Within R <sup>2</sup>	0.1821	0.1965	0.2180	0.0030	0.0027	0.0034
Kleibergen-Paap, F-stat.	7.529	9.387	22.39	11.86	9.076	7.092
Wald, p-value	0.0062	0.0022	< 0.001	0.0006	0.0027	0.0009
<i>Fixed-effects</i>						
# country-sector	40	40	40	40	40	40
# country-year	110	110	110	80	80	80
# sector-year	176	176	176	128	128	128

Driscoll-Kraay (L=2) standard-errors in parentheses

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Table B10: EAC5 REGIONAL INTEGRATION VIA GROSS TRADE REGRESSIONS: MANUFACTURING SECTORS

Dependent Variable:		log(VA)							
Exports Measure:		Gross				Final Goods			
Data:		EORA21		EORA15		EORA21		EORA15	
Model:		OLS	IV	OLS	IV	OLS	IV	OLS	IV
<i>Variables</i>									
log(E)		0.0923*** (0.0263)	-0.4191 (0.3939)	0.1154*** (0.0295)	0.2048 (0.1446)	-0.0556 (0.0892)	0.0411 (0.1168)	0.1087*** (0.0302)	0.0246 (0.1038)
log(E) × SH <sup>EAC5</sup>		-0.0283 (0.0284)	-0.1999 (0.1367)	-0.0481 (0.0360)	0.0073 (0.0362)	0.0037 (0.0196)	-0.0091 (0.0300)	-0.0360 (0.0338)	-0.0287 (0.0418)
<i>Fixed-effects</i>									
# country-sector		40	40	40	40	40	40	40	40
# country-year		110	110	80	80	110	110	80	80
# sector-year		176	176	128	128	176	176	128	128
<i>Fit statistics</i>									
Observations		859	859	640	640	859	859	640	640
R <sup>2</sup>		0.9883	0.9860	0.9952	0.9951	0.9883	0.9882	0.9952	0.9951
Within R <sup>2</sup>		0.0088	-0.1884	0.0276	0.0078	0.0022	-0.0050	0.0252	0.0119
Wu-Hausman, p-value			0.2981		0.7154		0.8432		0.8630

Driscoll-Kraay (L=2) standard-errors in parentheses

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Table B11: EAC5 REGIONAL INTEGRATION VIA RVCs REGRESSIONS: MANUFACTURING SECTORS

Dependent Variable:		log(VA)							
GVC Indicator:		Backward Integration (VS)				Forward Integration (E2R)			
Data:		EORA21		EORA15		EORA21		EORA15	
Model:		OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
<i>Variables</i>									
log(GVC)		0.1991*** (0.0471)	0.4829*** (0.1146)	0.0742** (0.0271)	0.3588*** (0.0814)	0.6728*** (0.0785)	0.8804*** (0.0284)	0.5249*** (0.1524)	0.7451*** (0.1848)
log(GVC) × SH <sup>EAC5</sup>		-0.1954* (0.0967)	-0.4012* (0.1929)	-0.1208 (0.0914)	-1.314** (0.4490)	0.0553 (0.0582)	-0.0546 (0.4691)	0.0368 (0.0386)	-0.3621* (0.1951)
<i>Fixed-effects</i>									
# country-sector		40	40	40	40	40	40	40	40
# country-year		110	110	80	80	110	110	80	80
# sector-year		176	176	128	128	176	176	128	128
<i>Fit statistics</i>									
Observations		859	859	640	640	859	859	640	640
R <sup>2</sup>		0.9886	0.9879	0.9951	0.9944	0.9967	0.9958	0.9978	0.9969
Within R <sup>2</sup>		0.0318	-0.0313	0.0064	-0.1233	0.7155	0.6439	0.5504	0.3820
Wu-Hausman, p-value			0.0052		0.0082		< 0.001		0.0541

Driscoll-Kraay (L=2) standard-errors in parentheses

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

## **Chapter 4**

# **Infrastructure for Trade and Structural Transformation**

# Mapping Africa's Infrastructure Potential with Geospatial Big Data and Causal ML

Sebastian Krantz\*

August 21, 2024

## Abstract

Using rich geospatial data and causal machine learning, this paper maps potential economic benefits of incremental investments in all major types of public and economic infrastructure across Africa. These 'infrastructure potential maps' cover all populated areas in Africa, at a spatial resolution of  $9.7\text{km}$  ( $96\text{km}^2$  hexagons). They show that the local benefits of additional infrastructure are highly variable and context-specific. The results broadly suggest that 'hard infrastructure' such as paved roads, power, transport, and communications is more beneficial in cities, whereas 'social infrastructure' such as education, health, public services and utilities, is more important in rural areas. Market access and agglomeration effects are important forces governing these returns. Descriptive analysis further reveals that infrastructure in Africa is concentrated in urban areas and often inefficiently allocated. African cities exhibit marked heterogeneity in infrastructure, public services, and economic activities.

**Keywords:** Africa, infrastructure, investment potential, geospatial big data, causal ML, XAI

**JEL Classification:** O18; R11; R40; C14

## 1 Introduction

It is well recognized in economic literature and policy discourse that Africa has great public infrastructure deficits. Estimates by the African Development Bank (ADB) suggest that Africa's infrastructure needs amount to \$130-170 billion a year, with a financing gap in the range of \$68-108 billion (ADB, 2018). A World Bank report from 2010 states that Sub-Saharan Africa (SSA) has 31 paved roads  $\text{km}$  per  $100\text{km}^2$  of land, compared to  $134\text{km}$  in other low-income countries, estimates the annual infrastructure gap at \$93 billion, and urges SSA countries to spend one percent of GDP on roads (Foster & Briceño-Garmendia, 2010). Large gaps also remain in other areas; for example, according to World Bank statistics, by 2022, 51% of Sub-Saharan Africans had access to electricity, 36% were using the internet (although 84% had mobile phones), and 34% were using basic sanitation services. These deficits suggest that public infrastructure investments in Africa may have high returns for economic activity and wealth generation. At least the World Bank spends around a third of its budget on infrastructure projects (Akpanjar & Kitchens, 2017), and much academic research of recent years suggests sizeable economic returns - such as historical railway access increasing real agricultural incomes by 16% (Donaldson, 2018) or land values by 60% (Donaldson & Hornbeck, 2016), power cuts reducing firm revenues and producer surplus by 5-10% (Allcott et al., 2016), internet availability inducing 2% higher growth and structural change (Goldbeck & Lindlacher, 2021), joint roads/power investments yielding an 11% increase in welfare (Moneke, 2020), road network inefficiencies implying a 1.3% welfare loss (Graff, 2024), and changes in transport costs invoking large reshufflings of population, wealth, and economic activity (Storeygard, 2016; Jedwab & Storeygard, 2022; Faber, 2014; Baum-Snow et al., 2020).

This paper contributes to the debate on the returns to infrastructure investments in Africa by utilizing the wealth of granular geospatial data that has become available in recent years to observationally estimate local returns to different types of public and economic infrastructures across Africa. It utilizes recently developed causal machine learning methods (e.g., Chernozhukov, Chetverikov, et al. (2018); Athey et al. (2019); Nie & Wager (2021)), enabling more credible and

---

\*Kiel Institute for the World Economy  
Address: Haus Welt-Club, Kiellinie 66, D-24105 Kiel  
E-mail: sebastian.krantz@ifw-kiel.de

*Disclaimer:* This is a working paper circulated for discussion and comment purposes. It has not been peer reviewed yet, and any views expressed in it are those of the author and not of the Kiel Institute for the World Economy.

specific inferences from observational datasets. These methods, also known as 'double' or 'debiased' ML, employ predictive ML models to remove factors confounding the relationship between a treatment (infrastructure), and an outcome (economic activity or wealth). Following this first-stage 'debiasing' step, a second-stage 'causal model' estimates a heterogeneous treatment effect (of infrastructure). With only a large spatial cross-section, the causal model may still suffer from reverse causality, but is plausibly free from confounding, which is a large improvement over single-stage predictive models typically employed in geoAI or other ML contexts.

Causal ML thus does not supersede causal econometrics, which, once a convincing source of quasi-random variation has been identified, estimates the causal effect of a specific infrastructure. However, convincing quasi-random variation is hard to find, particularly with infrastructure, and causal studies of infrastructure seldom extend beyond a single country or region. Few econometric studies are also able to provide local heterogeneity in treatment effects. In other words, limited external validity and differences in data and methodology complicate drawing broad inferences from such studies for policy purposes. Conflicting studies of the same intervention in different settings also produce conundrums in academic knowledge, as exemplified by the academic literature on returns to household electrification (Lee et al., 2020; Bayer et al., 2020).

In contrast, granular geospatial data has recently become increasingly rich and uniform and is available at high resolution for vast geographic areas. By analyzing this data with causal ML, I am able to make a contribution that is quite different from most causal econometric research on infrastructure. In particular, I provide localized (1) marginal effects and counterfactual predictions for (2) 14 different types of public and economic infrastructure as well as local and overseas market access, at (3)  $9.7\text{km}$  ( $96\text{km}^2$ ) spatial resolution,<sup>1</sup> for (4)  $>100,000$  populated locations across the entire African continent. My results thus exhibit comparability between different types of infrastructure and across space, and suggest that heterogeneity in returns to infrastructure is indeed eminent. The high-resolution 'infrastructure potential maps' produced in this paper thus present an advance toward policymakers asking about the returns to an additional road, generator, school, or hospital in a specific city, village, or suburb. These maps are necessarily limited by the quality of open geospatial data for Africa, ground-truth measures for wealth and economic activity, and by the identifying assumptions of causal ML.<sup>2</sup> They also only provide partial equilibrium estimates ignoring migration and reshuffling of economic activity following infrastructure investments.

Analyzing granular geospatial data is also useful for gaining insights and uncovering patterns in Africa's spatial economy. Explainable AI (XAI) and other nonparametric methods, such PCA and clustering, can help understand the concentration and the local and global economic significance of different types of infrastructure, as demonstrated in Section 3 of this paper.

The remainder of this paper is structured as follows: Section 2 introduces the detailed geospatial data and describes how it is processed into analyzable form. Section 3 studies the processed data and expounds several stylized facts about Africa's spatial economy. Section 4 introduces the causal ML framework and presents average and heterogeneous partial infrastructure effects. Section 5 does the same for counterfactual predictions. Section 6 summarizes the findings and concludes.

## 2 Data and Preprocessing

The database of African infrastructure constructed for this project is large and built up from granular data sources. It combines detailed geospatial point and vector data with raster data on population, accessibility, and various economic outcome measures.

### 2.1 Geospatial Data on Infrastructure

The main source of open geospatial data on infrastructure is Open Street Map (OSM). The Africa OSM has grown rapidly over the past years. As of November 2022, it contains  $>130$  million spatial

---

<sup>1</sup>Which is roughly the size of a medium-sized city, or of the center of a smaller capital city such as Kigali.

<sup>2</sup>Well known in the literature as the *unconfoundedness assumption*, stipulating the conditional independence of treatment and outcome based on observables. I will present evidence that, in the face of rich geospatial data, this assumption is not easily dismissed via classical arguments such as political favoritism (Dreher et al., 2019).

features covering all areas of public and economic life. More than 50 local OSM groups spawn nearly every African country<sup>3</sup>. The growing reliability of OSM is also reflected in the increased research use of the map. For example, Peng & Chen (2021) use the Zambia OSM in 2019 to train their image segmentation model, and Graff (2024) uses the OSM routing service to generate network connectivity data for his spatial model. As a statistical point of reference, Microsoft Research released a global dataset of segmented roads in 2022 indicating that globally only 2.3% of roads are missing from OSM, and in Africa around 2.6%. To optimally utilize the map, I develop a basic functional classification of OSM features and apply it to the entire Africa OSM, yielding a database of  $\sim$ 12.6 million points/buildings and  $\sim$ 3.8 million km of lines of economic interest. An R package *osmclass* (Krantz, 2023) to classify OSM features is released together with this paper.

Apart from OSM, the Overture Maps Foundation aims to create comprehensive open maps data for developers, with steering members Amazon, Meta, Microsoft, and TomTom. In July 2023, the first production-grade maps dataset was released. With all transport-related data taken from OSM and buildings untagged, the biggest addition is a global dataset of 57 million places of interest (POIs) combined from OSM, Meta, and Microsoft. Within Africa, this adds an additional 1.3 million places mostly from Meta, of which 824k have a minimum confidence score above 0.4.

In addition to OSM and Overture, the Alltheplaces open-source project provides web-scraped POI data, adding 114k POIs mostly from established brands like Starbucks or KFC in Africa.

These sources do not reliably map all infrastructures of economic significance, thus I complement them with several curated datasets covering specific infrastructures. In particular, I assemble data on cell towers (OpenCellid), health facilities managed by the public sector (Maina et al., 2019), power plants (Global Integrated Power Tracker and WRI database (Byers et al., 2018)), steel plants (Global Steel Plant Tracker), special economic zones (SEZs) (Open Zone Map), and ports (2015 World Port Index (MSI, 2019) and World Bank). Table 1 summarizes the places dataset by source.

Table 1: Africa Infrastructure Database: Places Dataset by Source

Source	Count	of which Polygons	Categories
Open Street Map (OSM)	12,221,198	9,038,206	45
OpenCellid (Cell Towers)	1,894,356	0	1
Overture Maps Places (confidence > 0.4)	823,786	0	44
All The Places (Open Web-Scraped POIs)	114,382	0	12
Health Facilities in SSA (Nature Scientific Data)	96,290	0	1
Global Integrated Power Tracker	871	0	1
WRI Global Power Plants Database	363	0	1
Global Steel Plant Tracker	41	0	1
Open Zone Map (Special Economic Zones)	387	0	1
World Port Index	235	0	1
WorldBank Global International Ports	9	0	1
SUM	15,151,918	9,038,206	47

*Notes:* Table shows places of interest (POIs) data collected from different sources. In OSM POIs may be tagged buildings/have geometries.

While the *osmclass* package provides a basic classification of features closely aligned with the OSM tagging system, and curated datasets have a limited number of functionally distinct features, the Overture places (OVP) data follows a different categorization scheme with  $>$  1000 primary place categories. I thus perform a second, more rigorous, classification step to categorize features across sources into 47 specific economic categories and 26 simplified ones. The full classification process is detailed in Appendix A. Table A3 shows the final harmonized classification.

Following harmonization, I also deduplicate the data across sources by allowing only features of the same category from one source within a 10m radius.<sup>4</sup> Curated datasets thereby take precedence over OSM, which in turn supersedes OVP. Tables 1 and A3 summarize already deduplicated data.

<sup>3</sup><https://medium.com/@katereggal/the-state-of-openstreetmap-in-africa-223ecadd5556>

<sup>4</sup>This is done by shifting a 10m grid over the POI features in steps of 1m and deduplicating features within each 10m  $\times$  10m square, thus there is some path dependence in terms of which POIs are compared first.

In addition to POIs, I extract linestring features such as (non-residential) roads, larger waterways, power lines, railways, aeroways, pipelines, and telecommunication lines from OSM. I complement OSM power lines with electricity grid maps from the European Commission ([Kakoulaki & Moner-Girona, 2020](#)) and the [World Bank](#). Table 2 provides a breakdown. In Total, I collect  $\sim 4.4$  million km of network infrastructure, of which  $\sim 1.6$  million km are roads,  $\sim 1.5$  million km waterways, and 967 thousand km of power lines. The other categories sum to 272 thousand km.

Table 2: Africa Infrastructure Database: Lines Data by Category

Category	Count	Length (Km)
road	763,912	1,621,144
waterway	359,756	1,507,112
power <sup>1</sup>	1,013,150	967,317
railway	84,707	128,408
aeroway	27,360	11,019
pipeline	9,453	55,394
storage	8,551	389
ferry	2,412	48,259
aerialway	171	175
telecom	87	28,682
SUM	2,269,903	4,379,392

<sup>1</sup> OSM power lines were combined with datasets from the EC's Joint Research Centre (JRC) and the World Bank.

Finally, I also obtain 2022 fixed and mobile download and upload speeds from [OOKLA](#) via the [EU Africa Knowledge Platform](#) - within map tiles at a resolution of around 610m at the equator.

## 2.2 Spatial Measures of Wealth and Economic Activity

To study the returns to infrastructure, accurate spatial measures of quantities of interest such as economic activity/value-added or household wealth are needed. A popular spatial proxy for economic activity, following the seminal work of [V. Henderson et al. \(2011, 2012\)](#), is remotely sensed nightlight luminosity ([Donaldson & Storeygard, 2016](#)). Since 2011, nightlights data is available at high resolution (15 arc second or  $\sim 500$ m at the equator) from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi satellite ([Gibson et al., 2020](#)). Recently, NASA released a more processed version of the VIIRS imagery for monitoring human activities across space and time under the label 'NASA Black Marble' ([Román et al., 2018](#)). Among the first research users in economics, [Peng & Chen \(2021\)](#) show that this data is substantially more accurate than the conventional VIIRS data in tracking Zambian GDP over time.

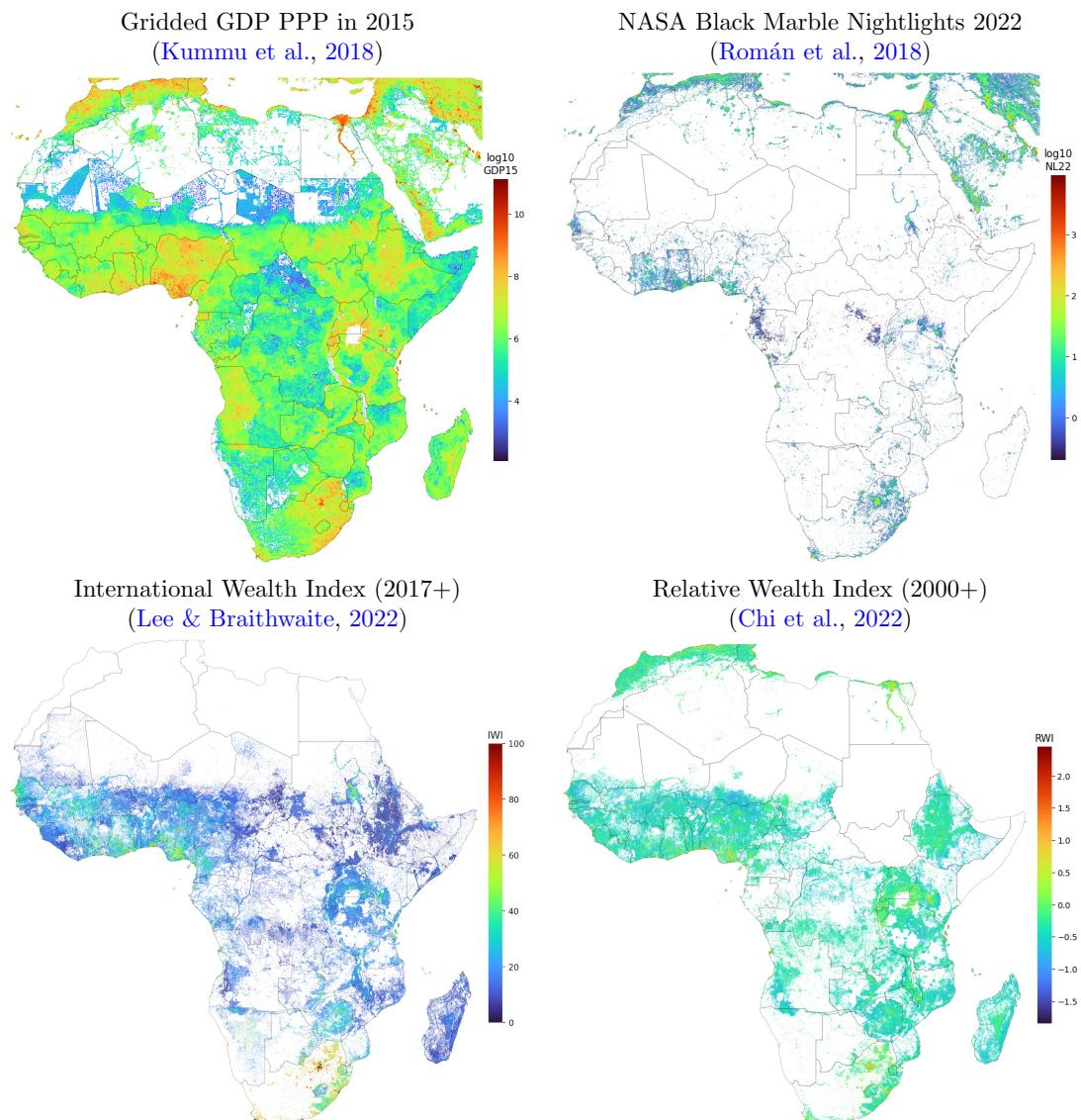
Several contributions also spatially distribute GDP (value-added) from national and/or regional accounts via high-resolution data on population, geophysical features, and nightlights. A widely used dataset is the G-Econ database ([Nordhaus et al., 2006](#)), providing GDP estimates for 1-degree grid cells from 1990 to 2005. A more recent attempt by [Kummu et al. \(2018\)](#) distributes national GDP estimates from the World Development Indicators and CIA World Factbook in constant 2015 PPP dollars at 5 arc-min (0.0833 degrees or 9.3km at the equator) resolution, using subnational value-added estimates from [Gennaioli et al. \(2013\)](#) and population from the HYDE 3.2 database.

Recent efforts, starting with [Jean et al. \(2016\)](#), use richer data sources to predict wealth/poverty at high spatial resolutions. Due to the requirements of high-resolution ground-truth data, this literature has focused on measures of wealth computed from standardized surveys such as DHS and LSMS, rather than coarse measures such as GDP. A popular contribution by [Chi et al. \(2022\)](#) combines vast and heterogeneous data from satellites, mobile phone networks, topographic maps, as well as aggregated and de-identified connectivity data from Facebook to estimate nationally comparable estimates of wealth - a Relative Wealth Index (RWI) - for all low and middle-income countries at 2.4km resolution. Another recent contribution by [Lee & Braithwaite \(2022\)](#) employs a cross-country prediction methodology that combines day- and nighttime satellite imagery, high-resolution population estimates, and OSM to predict the International Wealth Index (IWI) -

a comparable asset-based wealth index calculated from DHS Surveys for 25 countries in SSA conducted since 2017 - for 929,295 populated places in 44 SSA countries at 1-square mile (1.6km) resolution.<sup>5</sup> They obtain a cross-country  $R^2$  of 91.7% for the IWI, outperforming previous results.

Figure 1 shows four of the latest estimates of wealth/economic activity in Africa discussed above. None of these measures is ideal to study the effects of infrastructure. Nightlights are, by definition, correlated with power infrastructure and also relatively sparse since very low-light areas are set to zero in the Black Marble product. Gridded GDP is, by definition, highly correlated with population and may thus be biased towards residential areas. The RWI is not constructed to be comparable across countries and is not available for (South-)Sudan, whereas the IWI is not available for North Africa and uses parts of OSM and population next to daylight satellite imagery in its construction. In the following, I use all 4 estimates shown in Figure 1 to determine weights applied during the aggregation of granular data, but focus on the IWI for final estimation since it is an accurate high-resolution and cross-country comparable estimate. I also conduct robustness exercises with nightlights and ground truth IWI estimates from DHS surveys conducted since 2010 to ensure that key results are not driven by the ML model of [Lee & Braithwaite \(2022\)](#).

Figure 1: Spatial Measures of Wealth and Economic Activity



*Notes:* Figure shows different measures of wealth and economic activity at their original resolution, with blank missing or zero values. The IWI is the main outcome measure, but all measures assist data aggregation (Section 2.5, Table A6).

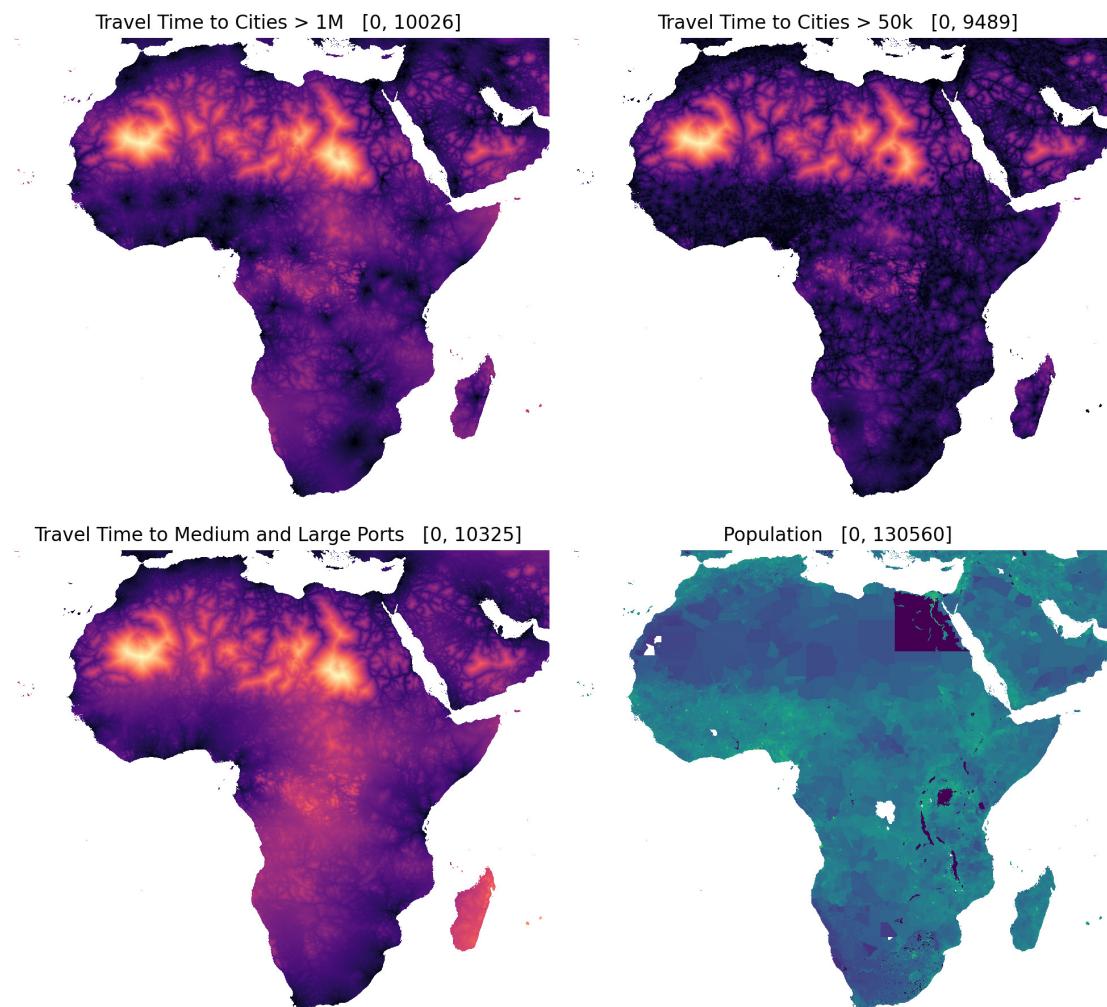
<sup>5</sup>From OSM, [Lee & Braithwaite \(2022\)](#) employ the total length of roads, distance to the closest road, number of junctions, distance to the closest junction, total building area, and the number of buildings for each 1 square-mile populated area, and the number of and distance to 24 locations of interest such as schools, hospitals, and markets.

### 2.3 Covariate Raster Layers

Many further sources of high-resolution data layers about geophysical features, agriculture, climate, and conflict could be included as covariates in an analysis of infrastructure and wealth/economic activity. Economic research such as [Storeygard \(2016\)](#), [Jedwab & Storeygard \(2022\)](#), [Donaldson \(2018\)](#) and [Peng & Chen \(2021\)](#) has, however, focussed on two particularly important dimensions of spatial variation affecting economic outcomes: population (urbanization) and market access.

I obtain population estimates for 2020 from the Gridded Population of the World Version 4 (GPW4) project ([CIESIN, 2016](#)), which is based on administrative data. I broadly distinguish between the infant (0-14 years) and working-age (15-49 years) population to allow for local variation in demographic characteristics. To approximate market access, I consider global accessibility indicators from [Weiss et al. \(2018\)](#), who developed a global map of travel time (in minutes) to cities with more than 50,000 people in the year 2015 at  $1 \text{ km}^2$  resolution. The map is based on a global friction surface constructed from detailed spatial data on transport networks and geophysical features. [Nelson et al. \(2019\)](#) expands this work to settlements of 9 different sizes, from towns of 5000 inhabitants to megacities with more than 5 million inhabitants. [Nelson \(2022\)](#) further computes travel time to ports of 4 different sizes (very small, small, medium, and large) using data from the 2015 (26th) edition of the WPI. From these 12 accessibility maps, I compute 4 which appear most relevant in Africa: (1) travel time to cities  $>50,000$  as in [Weiss et al. \(2018\)](#); (2) travel time to cities  $>1 \text{ million}$ ; (3) travel time to the nearest port, and (4) travel time to one of 43 medium or large African ports. Figure 2 shows 3 of these accessibility maps and total GPW4 population.

Figure 2: Raster Covariate Layers: GPW4 Population (2020) and Market Access (2015)



*Notes:* Figure shows accessibility maps from [Nelson et al. \(2019\)](#); [Nelson \(2022\)](#), and GPW4 population ([CIESIN, 2016](#)). The population layer has some missing data. These gaps are imputed in the final dataset following aggregation over a  $96\text{km}^2$  grid (in Section 2.4) using the 'missForest' algorithm by [Stekhoven & Bühlmann \(2012\)](#), yielding a 97% OOB- $R^2$ .

This choice of accessibility maps is sensible as diversified economic activity in SSA tends to take place in the largest urban centers, of which most countries have one or two. Towns of 50,000 people often function as important hubs to gather agricultural produce from the region for sale on local markets or transport to larger cities and ports (Jedwab & Storeygard, 2022). The vast majority of exporting or importing in Africa also happens through medium and large-sized ports, with modern container terminals and often intersecting international shipping routes.

## 2.4 The Ideal Spatial Grid

Jointly analyzing infrastructure and wealth/activity will require some form of spatial binning and data aggregation. For accurate spatial analysis, an equal area grid is desirable. Discrete Global Grid Systems (DGGS) use regular polyhedra for hierarchical tessellation of cells partitioning the globe. Sahr et al. (2003) discuss different DGGS design choices and implementations, and propose the Icosahedral Snyder Equal Area Aperture 3 Hexagon (ISEA3H) as a good general-purpose geodesic DGGS.<sup>6</sup> An aperture of 3 implies that lower-resolution hexagons have an area 3x larger than the next higher-resolution hexagons. The grid is implemented as part of the DGGRID C++ library (Sahr, 2022), and accessed in R through the 'dggridR' package (Barnes, 2020).

ISEA3H is available at 31 different resolutions, from 12 global cells spaced 7054km apart to 2059 trillion cells spaced 0.5m apart. High spatial resolutions are desirable but increase the computational burden and reduce the number of features in each cell, limiting statistical models' ability to learn about the spatial economy. I thus empirically gauge the highest resolution grid that still yields acceptable predictions for wealth/economic activity by counting POIs in each cell and category and computing the average correlation of these counts with the 4 indicators in Figure 1. I also compute the average  $R^2$  of linear models predicting the wealth/activity indicator from all category counts. Table 3 reports the results for ISEA3H grids at 7 different resolutions, ranging from 3,901 cells 87km apart down to resolution 557,766 cells 3.2km apart. Both individual and joint predictions decrease in strength as the grid resolution increases. The largest drop is experienced when moving from a resolution 11 grid (16.8km) to a resolution 12 grid (9.7km). The resolution 12 grid cells have a size comparable to the city center of Kigali (a larger city like Kampala being covered by 3-4 cells) and contain around 94 POIs on average and 10 in the median cell.

Table 3: POIs in ISEA3H Grids of Different Resolutions

Res.	Area ( $km^2$ )	Spacing (km)	N. Cells (N)	N. Features		Mean[Feat./Cat.]		Corr. ( $r$ )	LM $R^2$
				Mean	Median	Mean	Median		
8	7774.21	87.08	4,091	3,703.720	509	116.978	33.235	0.347	0.592
9	2591.40	50.28	11,243	1,347.676	144	58.674	16.200	0.336	0.552
10	863.80	29.03	29,247	518.067	48	33.677	9.250	0.317	0.504
11	287.93	16.76	71,192	212.832	20	21.489	6.000	0.290	0.425
<b>12</b>	<b>95.98</b>	<b>9.68</b>	<b>160,719</b>	<b>94.276</b>	<b>10</b>	<b>14.657</b>	<b>4.000</b>	<b>0.255</b>	<b>0.358</b>
13	31.99	5.59	336,587	45.016	6	10.154	3.000	0.224	0.321
14	10.66	3.23	646,959	23.420	4	7.121	2.000	0.195	0.278

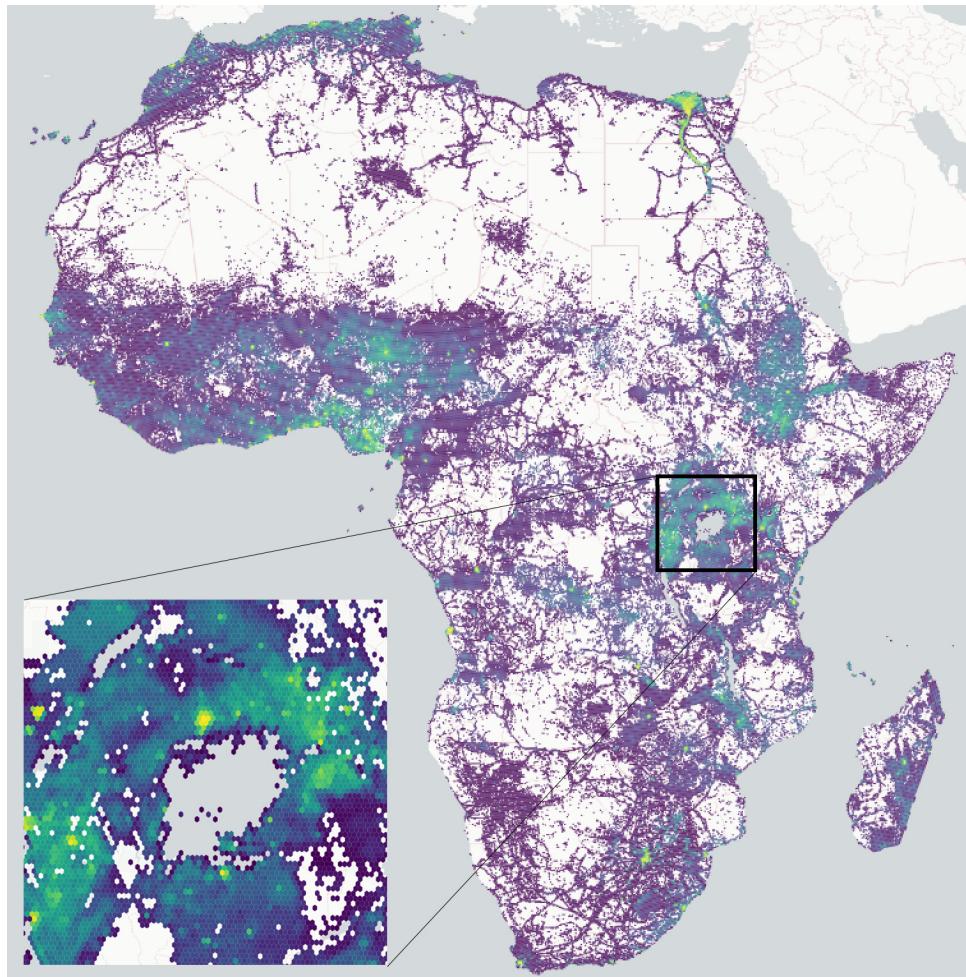
*Notes:* Res. is the grid resolution, also given in terms of cell area and spacing between the centroids of adjacent cells. N. Cells is the number of grid cells containing any POI in Africa. N. Features is average/median number of POIs in a cell. Mean[Feat./Cat.] is the average number of POIs per detailed category (47) in Table A3, computed for each cell and aggregated across cells using the mean or median. Corr. gives the average Pearson's correlation of the feature count within each category with each of the 4 spatial wealth/activity measures in Figure 1. Thus, it is an average of  $47 \times 4 = 188$  correlation coefficients. LM gives the average  $R^2$  of a linear model of the same 4 outcome measures against the features counted in 47 categories, thus it is the average of 4  $R^2$  estimates. The chosen resolution is highlighted.

To enable high-resolution estimation for city centers and suburban regions, I opt for the

<sup>6</sup>Sahr et al. (2003) summarize the design choices for ISEA3H as: "First, due to its lower distortion characteristics we choose the icosahedron for our base platonic solid. We orient it with the north and south poles lying on edge midpoints such that the resulting DGGS will be symmetrical about the equator. Next, we select a suitable partition. The hexagon partition has numerous advantages, and we choose aperture 3, the smallest possible aligned hexagon aperture. Because equal-area cells are advantageous for many applications, we choose the inverse ISEA projection to transform the hexagon grid to the sphere, and we specify that each DGGS point lies at the center of the corresponding planar cell region". Kmoch et al. (2022) further compare 5 open source DGGS implementations (Uber H3, Google S2, rHEALPix, OpenEAGGR, DGGRID) and constate that DGGRID is excellent for constructing grid cells with desirable properties, i.e., low apertures, convenient shapes, and equal area representation.

resolution 12 grid and mitigate the drop in predictive performance and effects of hard cell borders by allowing spatial spillovers from up to 2nd-order neighbours. The implementation of these spillovers is described below. Figure 3 visualizes the grid with GPW4 2020 total population estimates.

Figure 3: GPW4 2020 Population in 160,719 96km<sup>2</sup> ISEA3H Cells Covering Areas with any POI



*Notes:* Figure shows GPW4 population summed over an ISEA3H 96km<sup>2</sup> discrete global grid covering areas with any POI.

## 2.5 Data Aggregation

I aggregate the raster data over the 96km<sup>2</sup> equal area ISEA3H grid by taking the mean of wealth indices, travel times, and internet speed, and the sum of GDP, nightlights, and population within each cell. I also compute the total length in  $m$  of network features (Table 2) per cell, distinguishing paved from unpaved roads and combining 3 types of waterways using average harmonized coefficients from a Ridge Regression against the outcomes in Figure 1.<sup>7</sup> I count POIs in each cell and category (Table A3), but also consider a weighted approach where I compute quartiles of the building-areas of all OSM features tagged to buildings, and use counts of 2/3/4 if the area is within the 2nd/3rd/4th quartile. In this way, large features such as large school buildings receive up to 4 times the weight of small school buildings or schools that are just points. Similarly, I also apply these quartiles counts to the areas of SEZ's, the capacity of power and steel plants, and the outflows of ports. The quartile method thus takes into account the intensive margin.

<sup>7</sup>Namely rivers, man-made canals, and man-made waterways relating to utilities or agricultural activities such as drains and ditches, which are initially summed into a variable called 'waterway\_other'. I then use Ridge regression to estimate an equation of the form  $y = \beta_0 + \beta_1 + \text{river} + \beta_2 + \text{canal} + \beta_3 + \text{waterway\_other} + \epsilon$ , where  $y$  is the log of GDP, the IWI or the RWI (nightlights is too sparse). The coefficients are restricted to be greater than zero, and the optimal Ridge penalty is chosen by 10-fold cross-validation. I then obtain relative coefficients by dividing through by  $\beta_1$ , and averaging them across the 3 outcomes. The result is  $\beta_1 = 1$ ,  $\beta_2 = 9.035$ , and  $\beta_3 = 6.13$ , indicating that man-made features are much more important for spatial activity. Their length is thus increased by a factor  $\beta_j$ .

To aggregate the 47 detailed categories in Table A3 to the 26 simplified ones, which are more independent and thus more useful for analysis, I employ weights derived from penalized regressions on the 4 outcomes from Figure 1. Formally, let there be  $P_j$  detailed categories for simplified category  $j$  indexed by  $p$ . For example the 'communications' category ( $j$ ) combines  $P_j = 3$  detailed categories: 'communications\_network' (cell towers, antennas), 'communications\_other' (TV or radio station, newspaper, publisher, internet cafe), and 'telecom\_len' which is the length (in  $m$ ) of OSM telecommunications lines in each cell. I combine these detailed categories into a simplified one ( $\mathbf{x}_j$ ) using appropriate linear weights  $\beta_{jp}$  maximizing the correlation with the outcomes ( $\mathbf{y}$ )

$$\max_{\beta_j} \text{cor}(\mathbf{x}_j, \mathbf{y}) \quad \forall j \quad \text{s.t.} \quad \beta_{pj} \geq 0 \quad \text{where} \quad \mathbf{x}_j = \sum_{p=1}^{P_j} \beta_{jp} \mathbf{x}_{jp}. \quad (1)$$

To prevent overfitting and negative coefficients  $\beta_{jp}$ , this problem is solved using a Ridge regression, restricting  $\beta_{jp}$  to be positive, and choosing the optimal penalty parameter ( $\lambda^*$ ) via 10-fold cross-validation. The resulting coefficients  $\beta_{jp}$ , are normalized by the coefficient of the most populous category ('communications\_network'), and surprisingly consistent across outcomes. I thus average them to compute final weights. Appendix Table A6 provides three examples. For 'communications', the coefficient on 'communications\_other' is 11.7 and the coefficient on 'telecom\_len' is 0.1, which are sensible in relation to a cell tower ('communications\_network') having a weight of 1.

Finally, I account for economic geography and mitigate cell-border effects by creating additional 'spillover' variables (a.k.a. spatial lags) as inverse-distance-weighted average of neighbouring cells

$$\mathbf{x}_j^{\text{neigh}} = \sum_i \frac{\mathbf{x}_i}{\delta_{ij}} / \sum_i \frac{1}{\delta_{ij}} \quad \forall i \quad \text{where} \quad \delta_{ij} < \tau. \quad (2)$$

I choose  $\tau = 24.2\text{km}$ , which includes all second-order neighbours. The results are robust to the absence of spillover variables and also to using simple counts instead of quantile counts, but richer processing results in increased predictive power and spatial lags limit confounding influences.

Appendix Table A4 shows summary statistics for the final gridded dataset (simple counts). Appendix Figure A1 additionally shows histograms of the aggregated wealth/activity measures, and Table A5 shows pairwise Pearson's correlations of these measures aggregated over the grid. All measures are moderately correlated but follow slightly different distributions.

### 3 Africa's Spatial Economy

With rich geospatial data about infrastructure, population, and wealth/activity in Africa at hand, I investigate Africa's spatial economy and the current allocation of infrastructure using a graduated approach. I first analyze the spatial concentration of infrastructure and visualize its allocation. This yields a characterization of locations relatively lacking infrastructure, further illustrated by a case study of 5 African capital cities. I then examine the spatial clustering of different infrastructures and create an index of spatial efficiency measuring the proximity of residential areas, core infrastructure, and economic clusters. I find that this index is correlated with development indicators, logistic performance, and Graff (2024)'s road network inefficiency measure. Finally, I predict the IWI from infrastructure using ML models and interpret them with XAI methods, yielding a global and local characterization of important infrastructure predictors of wealth.

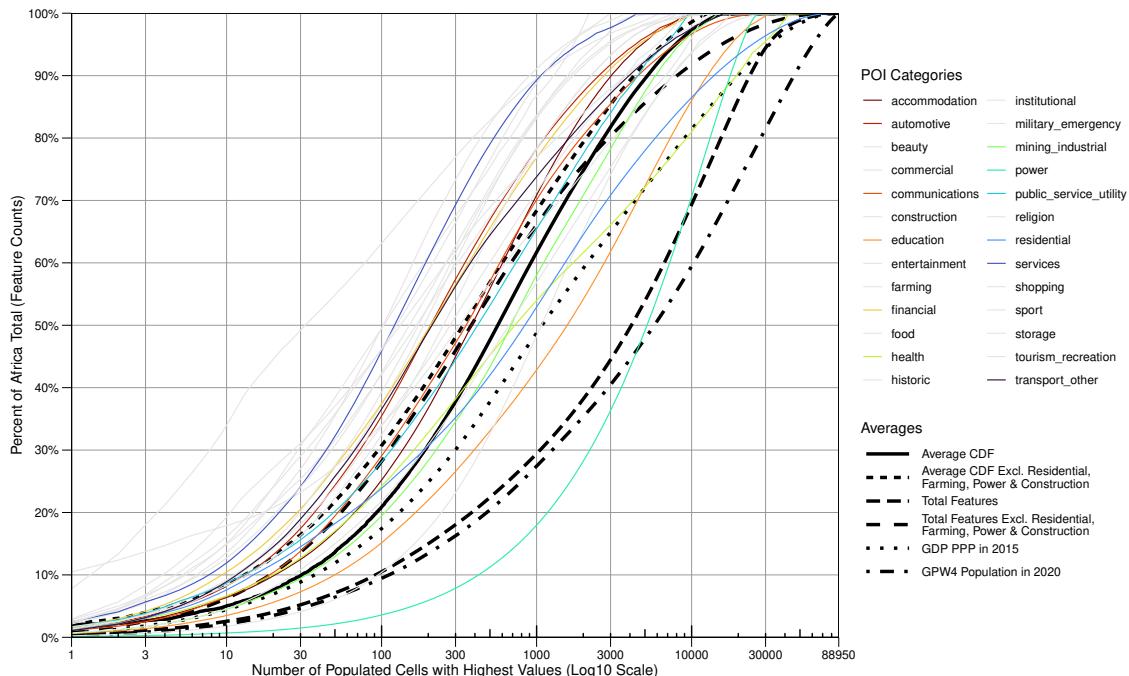
#### 3.1 Infrastructure Concentration

To study infrastructure concentration, I only consider grid cells with more than 960 inhabitants according to both the GPW4 and WorldPop 2020<sup>8</sup> estimates, i.e., cells with more than 10 persons/ $\text{km}^2$  and that have any POI. There are 88,960 such cells. I then count the POIs in simplified categories and compute proportions across cells. Sorting these proportions in descending order and cumulatively summing them yields an empirical CDF measuring infrastructure concentration, which I plot against the cell index. Figure 4 shows the result, both for individual feature categories, some of which are highlighted in color, and for averages across categories, computed before or after the CDF calculation. GDP and GPW4 population are also included as a reference.

<sup>8</sup>WorldPop (<https://www.worldpop.org/>) uses rich geospatial data to estimate population at 1km<sup>2</sup> resolution.

Figure 4 reveals that infrastructure in Africa is highly concentrated - more than population and, for most categories, GDP. The top 1000 cells (1.12% of 88,960) account for 62% of infrastructure in the average feature category, but only 49% of total GDP and 27% of total population. Only education (schools) and power infrastructure (generators) are less concentrated than GDP. The average CDF suggests that close to 100% of any given infrastructure is allocated in  $\leq 10,000$  cells. The only widely mapped feature present in nearly all cells is residential buildings. When infrastructure is pooled across categories, it is less concentrated, and the top 1000 cells only account for 30% of infrastructure. If residential buildings, farmland, power, and construction are excluded, this share rises to 66%. Excluding these four categories also yields 21,891 populated cells (24.6%) that have no other POI. Infrastructure is only moderately correlated with GDP and population, and weakly when aggregated to the country level (see Appendix Table A7), indicating spatial disparities between the allocation of infrastructure, wealth, and economic activities.

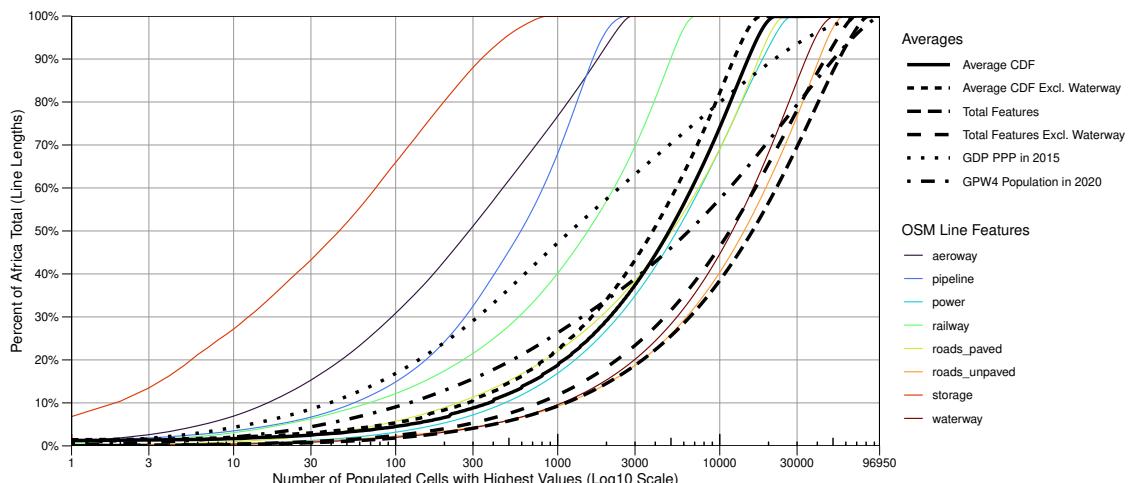
Figure 4: POI Feature Concentration - Empirical CDF's



*Notes:* Figure shows empirical CDF's (with cell-counts on the x-axis) of POIs in different categories.

Figure 5 shows analogous results for line features, indicating that frequent features such as roads, waterways, power, and railways (see Table 2) are less concentrated than GDP, and, in the case of unpaved roads and waterways, also than population.

Figure 5: OSM Line Feature Concentration - Empirical CDF's



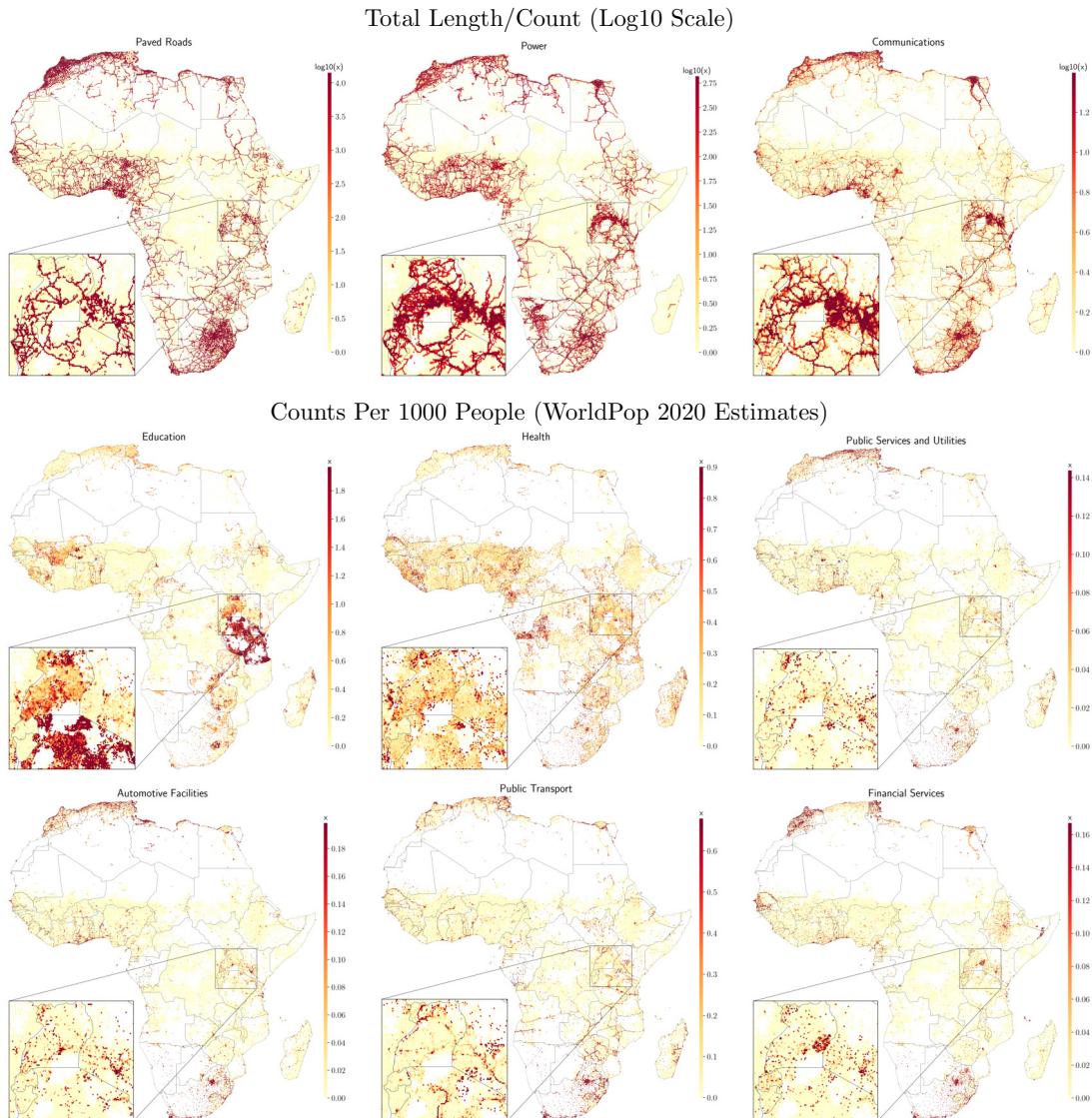
*Notes:* Figure shows empirical CDF's (with cell-counts on the x-axis) of line lengths in different categories.

With lines, the top 1000 cells only account for 10% of waterways and unpaved roads, 17% of power lines, 23% of paved roads, and 40% of railways. Concentration also proceeds gradually: the top 3000 cells account for 20% of unpaved roads and waterways, 39% of paved roads and 70% of railways, and the top 10,000 cells account for 40% of unpaved roads and 70% of paved roads. At the structurally weak end of the spectrum, excluding waterways yields 27.3% of populated cells that have no other line feature. Appendix Table A7 shows that the total line length in each cell correlates slightly stronger with household wealth than the total POI feature count.

### 3.2 The Spatial Allocation of Infrastructure

Figure 6 plots the spatial allocation of 9 critical infrastructures. The top panel shows paved roads, power, and communications, indicating absolute quantities in each cell on a log10 scale. Evidently, there are large gaps in central Africa, the Sahel, and the Horn of Africa regions. The great lakes region zoomed in in these plots is well connected in the central populated areas, but still lacks connectivity in sparsely populated peripheries such as eastern Congo, and Northern Kenya.

Figure 6: Spatial Distribution of Selected Infrastructure Features



*Notes:* Figure shows total length/count (top) or count per 1000 people (middle, bottom) per cell for 9 key infrastructures.

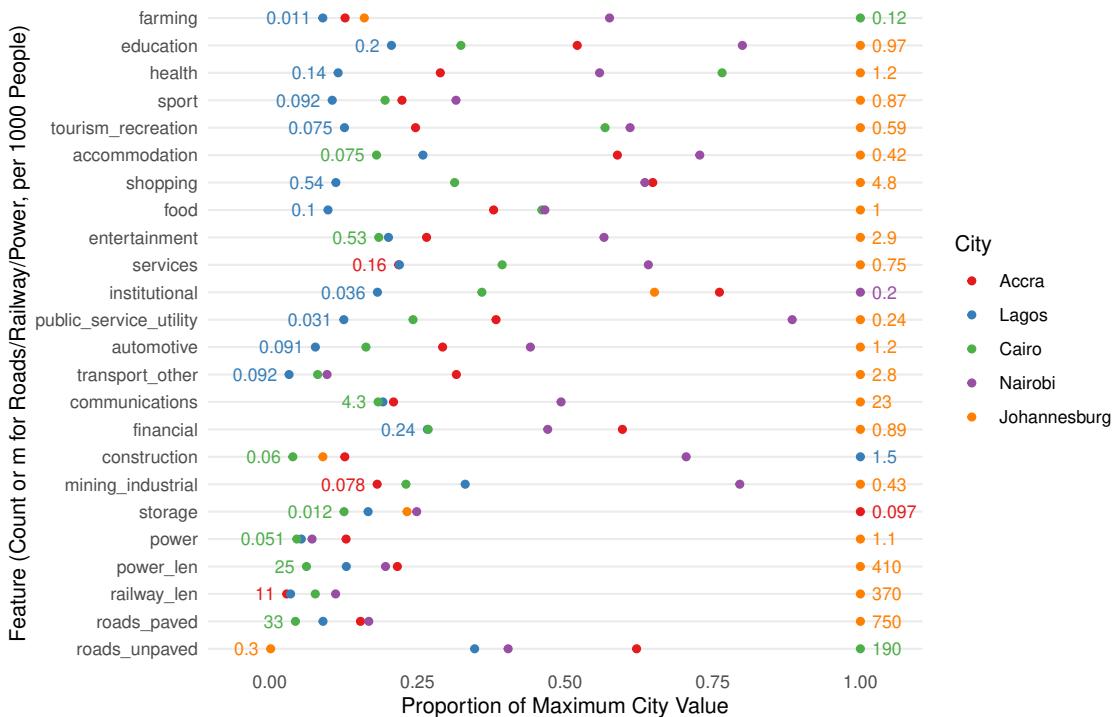
The middle and bottom panels of Figure 6 show education and health facilities, public services and utilities (excl. power), automotive facilities, public transport and financial services per 1000 inhabitants - for cells with more than 10 people per  $km^2$ . Some of these plots evidence differences in data coverage; for example, Tanzania and Uganda both have a very large amount of educational

facilities, averaging around 1 facility per 1000 people. Health seems more balanced across countries partly due to the data contribution by Maina et al. (2019), but also shows large differences in access across space. Automotive and public transport facilities as well as financial services are more concentrated in towns. Curiously, there appears to be a high concentration of financial services in eastern Uganda, but this is also an administratively very fragmented region. In all three categories, South Africa has significantly higher concentrations per 1000 people than other parts of SSA.

### 3.3 Comparing Major African Cities

To engage more specifically with the data, I undertake a case study comparing 5 of the largest African cities: Accra, Cairo, Lagos, Nairobi, and Johannesburg. For each city, I consider 7 hexagons covering the central parts of the city, jointly spanning an area of  $692\text{km}^2$ . I then compute the counts of POIs and the length of lines within these areas and divide them by the WorldPop 2020 population estimate. Figure 7 shows the feature intensity per 1000 people in each city.

Figure 7: Feature Density in 5 Major African Capital Cities per 1000 People



*Notes:* Figure shows feature counts per 1000 inhabitants (WorldPop 2020 estimates) for 5 significant African cities. For each city, 7  $96\text{km}^2$  hexagons covering essential parts of the city are considered. Features are counted in detailed categories and then combined into simplified categories using the weighted aggregation procedure described in Section 2.5.

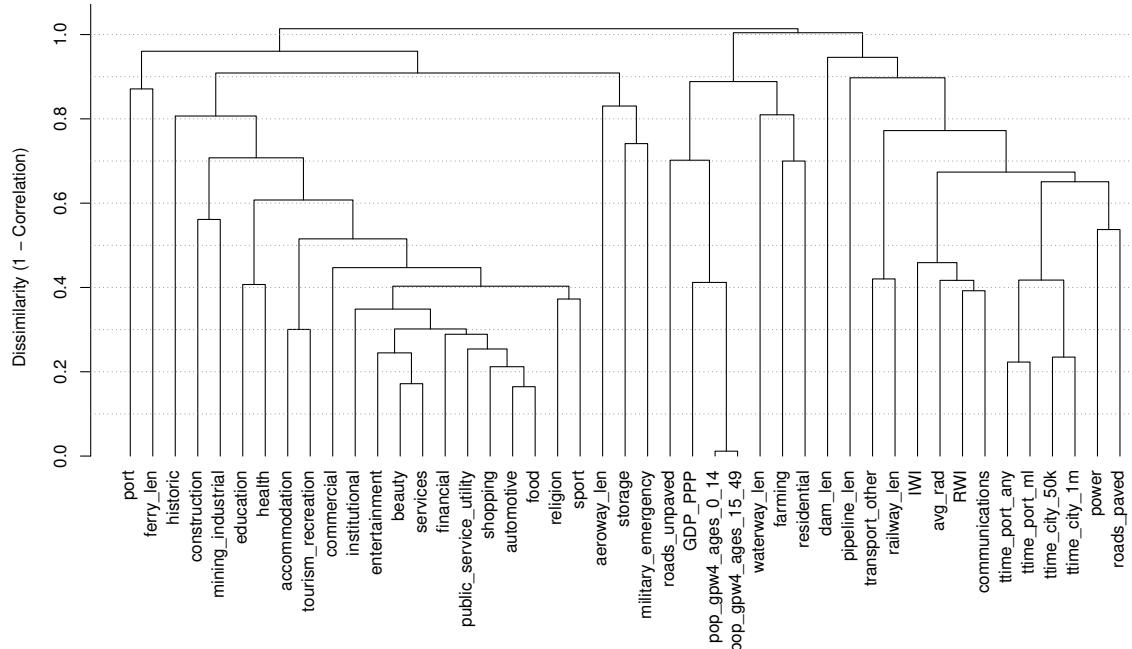
It is apparent that the cities are very heterogeneous in terms of spatial features, with Johannesburg topping the ranks for most feature categories. Notably, Johannesburg has 750m of paved transport roads, 370m of railway, and 410m of power lines per 1000 inhabitants, as well as significantly higher automotive, other (public) transport, communications, education, and health infrastructure than other cities. Conversely, Lagos ranks at the bottom in many categories, providing less than a quarter of the services per capita than Johannesburg. Between these two, Nairobi and Accra are doing well, with Nairobi having the most institutions alongside high densities of education, public services, industrial facilities, construction, accommodation (hotels), services, tourism and recreation. Accra also has a high level of institutions, shopping, financial services, accommodation, and education. It is remarkable that, apart from Johannesburg, the level of infrastructure in these cities does not align with their IWI (Accra: 73.5, Lagos: 67.4, Nairobi: 64.2, Cairo: 70.7 [RF prediction], Johannesburg: 81.6) or GDP/Capita<sup>9</sup> (Accra: 3953, Lagos: 7872, Cairo: 9672, Nairobi: 10439, Johannesburg: 14317 in 2015 USD PPP) estimates, with Lagos and Cairo in particular providing less services than their wealth levels would suggest.

<sup>9</sup>These estimates are based on coarse admin. data by Gennaioli et al. (2013), scaled by Kummu et al. (2018).

### 3.4 Correlations and Clustering

It is also informative to examine the spatial correlations among different features to determine which types of infrastructure often appear together and which are found in populated and/or high-income locations. To study this, I again use the counts of features in the simplified classification. I take the natural log, compute Pearson's correlations among all variables, and use  $1 - \text{correlation}$  as a distance matrix for hierarchical clustering with complete linkage. Figure 8 shows a dendrogram, and Appendix Figure A2 the corresponding correlation matrix.

Figure 8: Hierarchical Clustering of Variables using Correlation and Complete Linkage



*Notes:* Figure shows dendrogram from hierarchical clustering with complete linkage using a correlation-based distance metric, i.e.,  $1 - \text{pairwise\_Pearson's\_correlations}$ . Variables in simple counts correspond to Appendix Table A4 and are cast in logs before computing correlations, except for the IWI and RWI which are included in levels. Also, travel time estimates (in minutes) are first logged and then negated to yield positive correlations with other variables.

Figure 8 presents a detailed view of spatial activity in Africa. Different dissimilarity cutoffs reveal distinct groups of correlated variables. Notably, setting the cutoff around  $h = 0.98$  reveals three prominent groups in the dendrogram, corroborated by the correlation matrix (Figure A2). The group on the RHS includes travel times, household wealth, nightlights, power infrastructure, paved roads, and communications. These variables seem to be a proxy for physical infrastructure, which in turn correlates highly with nightlights and wealth. The second group, comprising most variables on the LHS of the dendrogram, includes economic activities in the broadest sense. Finally, the middle cluster includes population, residential areas, farming, and GDP, which is interpolated across space using population data. The dendrogram and correlation matrix (Figure A2) thus suggest that there is a disparity in space between where people live, where they work, and where most physical infrastructure is located. Presumably, an efficient spatial organization would imply a stronger spatial correlation among these three constructs/groups.

### 3.5 An Index of Spatial Efficiency

Informed by these observations, I compute an index of spatial efficiency (ISE) along 3 dimensions: the GPW4 working age (15-49) population, the first principal component (PC1) of paved roads, power, and communications as a compound measure of hard infrastructure, and the PC1 of education, institutional, health, religion, public\_service\_utility, food, shopping, beauty, services, commercial, mining\_industrial, tourism\_recreation, sport, construction, farming, entertainment, financial, and accommodation as a compound measure of economic activity. I use simple feature counts without spatial spillovers to keep the index simple. The PC1 of roads, power, and communications accounts for 63% of their joint variance, and the PC1 of the activity variables captures

56% of their joint variance. Table 4 shows Pearson's correlations among these components. The ISE is then obtained simply as the geometric mean of these correlations:

$$\text{ISE} = \text{cor}(P, I)^{\frac{1}{3}} \text{cor}(P, A)^{\frac{1}{3}} \text{cor}(I, A)^{\frac{1}{3}} = 0.642. \quad (3)$$

Table 4: Correlations of ISE Dimensions

$N = 160,499$	P	I	A
GPW4 POP 15-49 (P)	1.000		
Roads & Power PC1 (I)	0.589	1.000	
Economic Activity PC1 (A)	0.634	0.708	1.000

*Notes:* Table reports Pearson's correlations among dimension indices (PC1) and population. Their geometric mean is the ISE (Eq. 3).

It is thus an index on the range [0, 1], a value of 1 indicating perfect spatial efficiency with all productive resources concentrated in the same location. Perfect spatial efficiency is unachievable in any real-world setting, but the all-Africa ISE of 0.64 suggests much room for improvement. Table 5 shows ISE estimates at the country level, and Appendix Figure A3 a corresponding map.

Table 5: Country-Level ISE Estimates

#	ISO3	ISE								
1	SYC	0.950	TUN	0.788	ZAF	0.719	COD	0.624	BWA	0.508
2	MLI	0.908	MAR	0.783	BFA	0.717	MDG	0.621	LBY	0.486
3	MUS	0.889	TGO	0.779	SWZ	0.715	DJI	0.619	LBR	0.440
4	UGA	0.868	EGY	0.779	GMB	0.712	BDI	0.617	NAM	0.429
5	RWA	0.860	BEN	0.774	GAB	0.694	CAF	0.612	ESH	0.410
6	KEN	0.849	MWI	0.763	NGA	0.693	MOZ	0.593	ERI	0.360
7	GHA	0.830	LSO	0.750	ZWE	0.688	AGO	0.573	SOM	0.284
8	CIV	0.822	TZA	0.745	GIN	0.686	SDN	0.568	CPV	0.252
9	SLE	0.813	ETH	0.744	MRT	0.675	CMR	0.564	COM	0.214
10	STP	0.812	DZA	0.741	COG	0.672	TCD	0.543	GNQ	0.153
11	SEN	0.809	NER	0.729	ZMB	0.651	GNB	0.542	SSD	0.105

*Notes:* Table reports sorted country-level ISE estimates (Eq. 3) computed from cells within each country.

The average ISE across countries is 0.645. The country-level ISE is mildly correlated with key development indicators in 2020, such as GDP per Capita PPP ( $r = 0.159$ ), Life-Expectancy at Birth ( $r = 0.215$ ), and the Human Development Index ( $r = 0.185$ ). Interestingly, it shows stronger correlations with the 2018 Logistics Performance Index ( $r = 0.396$ ) and the 2020 Doing Business Index ( $r = 0.592$ ). It is also negatively correlated to the hypothetical welfare gains (in percent) from an optimal reallocation of the road network in each country as calculated by Graff (2024) ( $r = -0.360$ ). The stronger association of the ISE with these indicators vis-a-vis development outcomes suggests that it indeed measures spatial (in)efficiency. The index is uncorrelated with total land area ( $r = -0.033$ ), although some smaller states like Seychelles score particularly high.

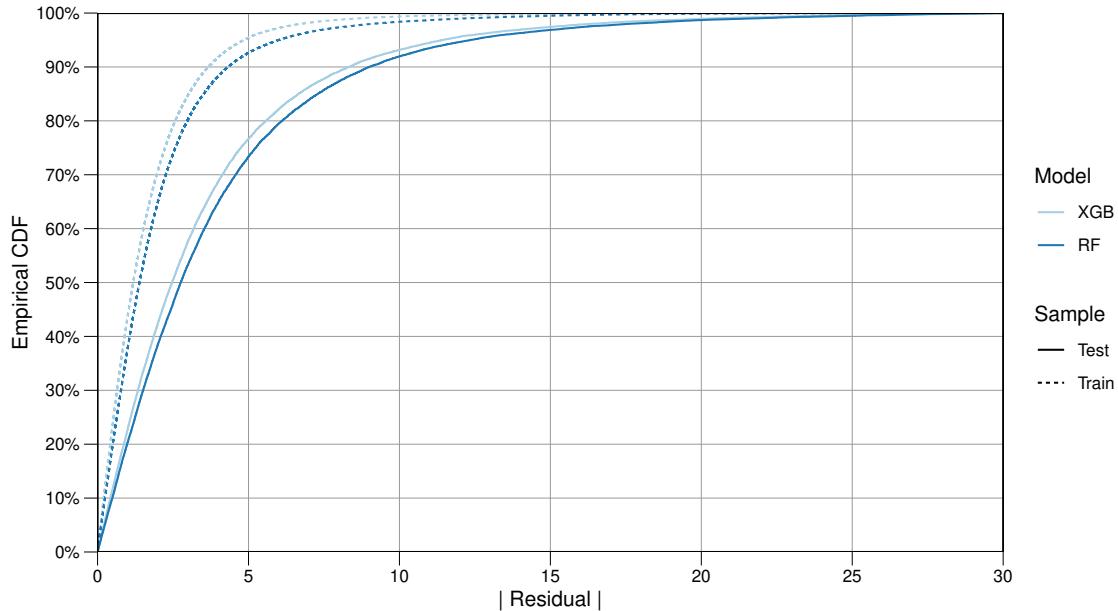
### 3.6 ML Prediction and Interpretation

Given the rich and diverse nature of Africa's spatial economy, a modern ML approach predicting wealth/activity from infrastructure that is able to capture non-linear associations in the data can yield further insights. The gridded data is in tabular form, and a recent empirical assessment by Borisov et al. (2021) shows that gradient-boosting machines (GBMs) (J. H. Friedman, 2001) still outperform deep learning methods on tabular data. Thus, I employ the competition-winning XGBoost algorithm (Chen & Guestrin, 2016) and tune its hyperparameters with Optuna (Akiba et al., 2019) on a test set containing 25% of the data.<sup>10</sup> The IWI model trained with early stopping

<sup>10</sup>The optimal hyperparameters generally feature a low learning rate ( $\eta = 0.01 - 0.02$ ), deep trees ( $\max\_depth = 9 - 15$ ), significant randomization over samples ( $\text{subsample} = 0.5 - 0.8$ ) and significant regularization, especially through constraints on the minimal size of final nodes ( $\min\_child\_weight = 6 - 20$ ).

achieves a test set  $R^2$  of 78.8%. For comparison, I also train a Random Forest (RF) model with default parameters and 1000 trees, which achieves a test set  $R^2$  of 73.3%. Figure 9 shows empirical CDFs of the absolute values of the residuals, indicating that the XGB model predicts 76% of test-set observations with an error of <5 IWI points. Appendix Figure A4 predicts nightlights.

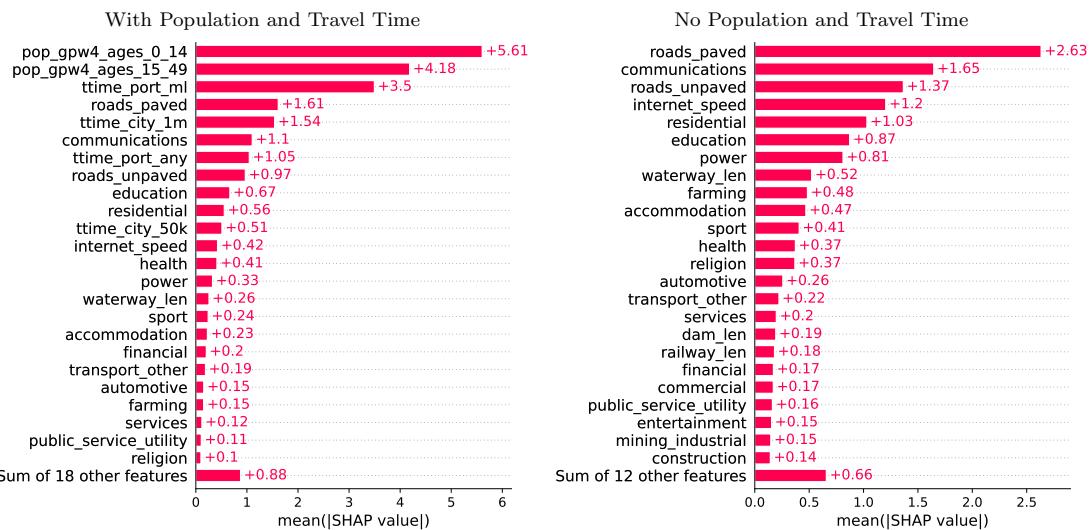
Figure 9: Empirical CDF's of Residuals from ML Models Predicting the IWI [0, 100]



*Notes:* Dataset has 42 predictors, not including spatial spillover variables. The training set has 86,354 observations, the test set 28,782. Training set  $R^2$  is 96.0% for XGB and 93.9% for RF. Test set  $R^2$  is 78.8% for XGB and 73.3% for RF.

To accurately attribute predictions across the different variables, I compute Shapely Values, a game theoretic approach to fairly attribute the contribution of variables to a single prediction. In particular, SHAP values following Lundberg & Lee (2017); Lundberg et al. (2020) give additive variable contributions for each instance that sum to the difference of the prediction from the average model prediction. To gauge the relative importance of different infrastructure categories, I compute (interventional) SHAP values following Lundberg et al. (2020) for two different XGBoost models: the model trained on the full dataset evaluated above, and a model excluding population and travel time estimates. The latter removes strong correlations of infrastructure with these contextual variables. Figure 10 shows overall variable importance for the IWI based on the average absolute SHAP value across all instances. Appendix Figure A5 does the same for nightlights.

Figure 10: Average SHAP Variable Importance for XGBoost Models Predicting the IWI

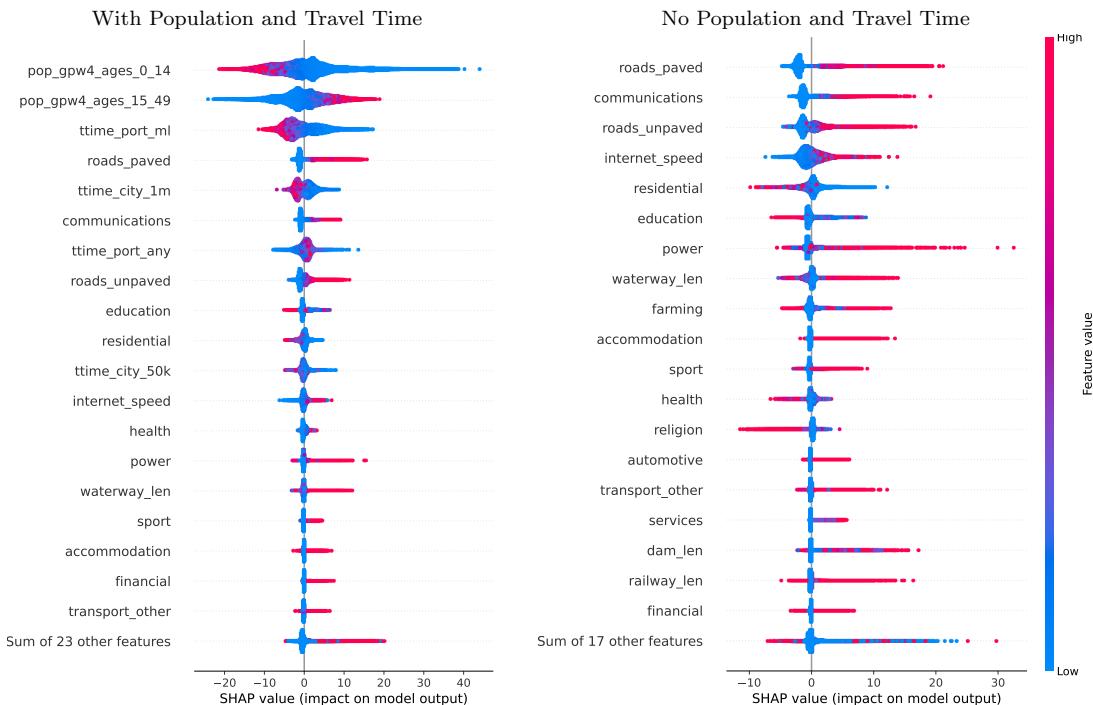


*Notes:* Figure shows average absolute interventional SHAP values for XGBoost models following Lundberg et al. (2020).

Figure 10 suggests that apart from population and travel time to major cities and ports, roads, communications, education, power infrastructure, and residential areas are the most important predictors of wealth. Many further features such as accommodation (hotels), sport, health, automotive and public transport facilities are also important. With nightlights (Figure A5), population, travel time, communications, roads, and power remain important, but sports, automotive facilities, and industrial areas are also large emitters of light. Appendix Figure A7 shows the same plots using SHAP values derived from feature counts for both IWI and nightlights, with similar results.

To gauge the direction of the effects on the predictions, which may be heterogeneous as SHAP values are computed at the instance level and tree ensembles may be highly non-linear, Figure 11 provides a beeswarm plot of the SHAP values coloured by feature intensity. The plot suggests that most variables have the intended effect, with larger values translating into higher SHAP values. Defying the intuition, residential buildings, education, health, and religion have largely negative effects. This spatial pattern needs to be cautiously interpreted *ceteris paribus*, i.e., in the presence of other correlated features often found in wealthy urban locations (e.g., paved roads, power, communications), more education and health facilities may decrease model predictions. Appendix Figure A6 shows the same plot for nightlights, with similar results. Appendix Figure A8 further provides the same plots derived from feature counts data, with similar outcomes.

Figure 11: SHAP Values for XGBoost Models Predicting the IWI



*Notes:* Figure shows interventional SHAP values for XGBoost models following Lundberg et al. (2020).

The relationship between predictors levels and SHAP values also resembles Accumulated Local Effects (ALE) plots (Apley & Zhu, 2016; Lundberg et al., 2020). Appendix Figures A9 and A10 provide ALE scatterplots corresponding to the beeswarm plots. They are particularly useful for detecting thresholds where a feature's effect on the prediction begins to change. For example, the SHAP value of paved roads strongly increases above  $10^4 = 10\text{km}$  in a cell, suggesting that roads are more important for prediction in urban areas. The same applies to power beyond a threshold of  $10^{2.5} \approx 300$  facilities (e.g., transformers, generators) per cell. With education, the opposite is the case: up to 10 facilities per cell, the SHAP value is high, but beyond that, it reduces, indicating that in urban spaces, schools become less critical for wealth prediction. Health and religious facilities exhibit similar but weaker dynamics. In contrast, communications (mainly cell towers) have an almost log-linear positive effect on model predictions. In models with population, the infant (0-14) population has a strong negative effect on predictions, whereas the adult population (15-49) has a strong positive effect, implying Malthusian dynamics in the data.

## 4 Estimating Marginal Infrastructure Benefits

Having explored the data in some depth, this section advances by asking about the marginal effects of infrastructure on wealth and economic activity. If this were an applied economics paper, I would now present an identification strategy (such as RCT, IV, RDD, DID) to causally identify the effects of infrastructure. However, I consider it impossible to causally identify the effect of every infrastructure in every location in Africa. Thus, I use observational causal inference a.k.a. causal ML.

The traditional econometric view, still held to some extent in other disciplines, is that if one can control for the most important confounding factors in an observational setting, a careful *ceteris paribus* interpretation of the partial effect is possible. So far, this premise has been applied almost exclusively in the context of linear regression, meaning that if all confounders are observed and the population model is linear-additive, a careful *ceteris paribus* interpretation is possible.

However, the *ceteris paribus* statement need not be that strong, as one can relax the assumptions of linearity and additivity. The premise of causal ML (also known as 'double' or 'debiased' ML) is that if one observes all factors that confound or proxy for confounding influences, the relationship between treatment and outcome can be specified conditional on an optimal ML prediction of both from observables. In the setting at hand, this means that if Africa's spatial economy is sufficiently observed through the available granular data on infrastructure, POIs, population, and market access, and if appropriate ML models are deployed, it may be possible to identify marginal partial-equilibrium effects of individual infrastructures on wealth/economic activity. Before examining these identification assumptions in more detail, I introduce this estimation strategy more formally.

I follow Nie & Wager (2021) and Chernozhukov, Chetverikov, et al. (2018); Chernozhukov et al. (2017), and adopt some notation from Hirano & Imbens (2004). Let  $Y$  be an outcome of interest,  $W$  a continuous treatment of interest with possible values  $\omega \in \Omega$ , and  $\mathbf{X} = [\mathbf{X}_H', \mathbf{X}_C']'$  be a vector of observed confounders, where  $\mathbf{X}_H$  includes covariates that also affect treatment effect heterogeneity. The *unconfoundedness assumption* is then formally stated as

$$Y(\omega) \perp W \mid \mathbf{X} \quad \forall \omega \in \Omega, \quad (4)$$

meaning that the potential outcome  $Y(\omega)$  is independent of the treatment realization  $W = \omega$  conditional on characteristics  $\mathbf{X}$ . The quantity of interest is the Conditional Average Partial Effect (CAPE)  $\tau(\mathbf{X}_H) = E[\partial Y / \partial W | \mathbf{X}_H]$ . I furthermore make the *strong* assumption that the CAPE can be additively separated from the effect of  $\mathbf{X}$  on  $Y$ , such that the following partial linear specification holds<sup>11</sup>

$$Y = \tau(\mathbf{X}_H) \cdot W + g(\mathbf{X}) + \epsilon \quad (5)$$

$$W = f(\mathbf{X}) + \eta \quad (6)$$

$$E[\epsilon \mid \mathbf{X}] = E[\eta \mid \mathbf{X}] = E[\eta \cdot \epsilon \mid \mathbf{X}] = 0. \quad (7)$$

The *nuisance functions*  $g()$  and  $f()$ , and also the treatment effect function  $\tau()$ , are assumed to be general functions that can be approximated by appropriate ML estimators. Taking the expectations of Eq. 5 conditional on  $\mathbf{X}$  and subtracting it yields

$$Y - E[Y \mid \mathbf{X}] = \tau(\mathbf{X}_H) \cdot (W - E[W \mid \mathbf{X}]) + \epsilon. \quad (8)$$

This is the specification of Robinson (1988). Assuming one can estimate  $m(\mathbf{X}) = E[Y \mid \mathbf{X}]$  and  $f(\mathbf{X}) = E[W \mid \mathbf{X}]$ , and denoting the 'debiased' variables by  $\tilde{Y} = Y - m(\mathbf{X})$  and  $\tilde{W} = W - f(\mathbf{X})$ , following Nie & Wager (2021), one can then estimate the CAPE  $\hat{\tau}(\mathbf{X}_H)$  by minimizing the R-Loss

$$\tau^*(\cdot) = \operatorname{argmin}_{\tau} \left\{ E[(\tilde{Y} - \tau(\mathbf{X}_H)\tilde{W})^2] \right\} = \operatorname{argmin}_{\tau} \left\{ E[\tilde{W}^2 (\tilde{Y}/\tilde{W} - \tau(\mathbf{X}_H))^2] \right\}. \quad (9)$$

The RHS of Eq. 9 demonstrates the so-called 'weight trick', i.e., minimizing the R-Loss with a nonparametric (ML) estimator  $\tau(\mathbf{X}_H)$  amounts to predicting the target  $\tilde{Y}/\tilde{W}$  using sampling weights  $\tilde{W}^2$ . For the practical implementation of Eq. 9, Chernozhukov, Chetverikov, et al. (2018) show that it is important to estimate the first-stage *nuisance functions*  $\hat{m}, \hat{f}$  via cross-fitting.

---

<sup>11</sup>This assumption will be relaxed in section 5 where I estimate causal dose-response functions.

In that case, a Neyman Orthogonality condition exists to ensure that the estimate  $\hat{\tau}(\mathbf{X}_H)$  from minimizing the R-Loss is insensitive to biases (e.g., regularization bias) in  $\hat{m}, \hat{f}$ .<sup>12</sup> The final stage estimate  $\hat{\tau}(\mathbf{X}_H)$  should also be cross-validated to ensure credible out-of-sample CAPE predictions.

Once  $\hat{\tau}(\mathbf{X}_H)$  is obtained, further quantities such as the Average Partial Effect (APE) or Group Average Partial Effects (GAPE) can be obtained by doubly-robust methods analogous to Augmented Inverse Probability Weighting (AIPW) (Robins et al., 1994) in the binary treatment case. Following Chernozhukov et al. (2022), Athey & Wager (2021), and *grf* R package (Athey et al., 2019; Tibshirani et al., 2023), the general form of such an AIPW estimator is

$$\bar{\tau} = \frac{1}{n} \sum_{i=1}^n \Gamma_i, \quad \Gamma_i = \tau_i(\mathbf{X}_{Hi}) + h_i(\mathbf{X}_i, W_i)(\tilde{Y}_i - \tau_i(\mathbf{X}_{Hi})\tilde{W}_i). \quad (10)$$

In the binary  $W$  case the debiasing weight  $h(\mathbf{X}, W)$  amounts to the so-called Horvitz-Thompson transformation.<sup>13</sup> In the continuous  $W$  case, Tibshirani et al. (2023) propose  $h(\mathbf{X}, W) = \tilde{W}/\hat{W}^2$ , where  $\hat{W}^2$  denotes a cross-fitted estimate of  $\tilde{W}^2$  from  $\mathbf{X}$ , which Tibshirani et al. (2023) obtain via a honest Random Forest (Athey & Imbens, 2016; Breiman, 2001) OOB predictions. To assess treatment heterogeneity, Athey & Wager (2019) average the doubly robust scores  $\Gamma_i$  in high and low regions of  $\hat{\tau}(\mathbf{X}_H)$  and test the difference in means. They also compute the Best Linear Prediction (BLP) of the CATE following Chernozhukov, Demirer, et al. (2018) via Eq. 9 as a calibration test.

To satisfy the unconfoundedness assumption, it is critical to choose appropriate estimators for  $m()$ ,  $f()$  and  $\tau()$ , and a suitable cross-fitting strategy. Because I have no strong prior beliefs on how infrastructure ( $\mathbf{X}$ ) affects wealth/economic activity ( $Y$ ) and other infrastructure ( $W$ ), I remain as model-agnostic as possible about these functional forms by employing a *Super Learner* approach (Van der Laan et al., 2007) to create an optimal weighted combination of several candidate learners via cross-validation. The most commonly used functional forms in the heterogeneous treatment effect literature are the LASSO, Random Forests, and Gradient Boosting. Recently, Friedberg et al. (2020) have proposed Local Linear Forests (LLF) as a variant of Generalized Random Forests with greater predictive accuracy for smooth targets. A causal version of LLF is also available in the *grf* package. I thus create a *Super Learner* from the predictions of these 4 algorithms

$$\hat{SL} = \beta_0 + \beta_1 \text{LA}\hat{\text{SSO}} + \beta_2 \hat{\text{RF}} + \beta_3 \hat{\text{LLF}} + \beta_4 \hat{\text{GBM}}, \quad (11)$$

where the weights  $\beta_i$  are determined by a relaxed Elastic Net tuned using 10-fold cross-validation. Each algorithm is also cross-validated/fitted: LASSO is tuned with 10-fold cross-validation using the built-in functionality of the *glmnet* R package (J. Friedman et al., 2010). Random forests and LLF are fit using the *grf* package with honest trees and produce OOB predictions. For GBM, I use XGBoost and employ a 3-fold cross-fitting approach, where an XGBoost model is trained on two folds, using the excluded fold as a validation set for early stopping and then predicting the outcome in the excluded fold. The ensemble estimator of the CAPE [ $\tau()$ ] similarly combines estimates from Causal Forests (Athey et al., 2019) and Local-Linear Causal Forests (Friedberg et al., 2020) obtained via *grf*, with cross-validated/fitted LASSO and XGBoost estimates from minimizing the R-Loss (Eq. 9). A cross-validated relaxed Elastic Net is again used as the final stage estimator and corresponds to the BLP of the CAPE from the 4 constituent estimates. All infrastructure, population, and travel time estimates enter the model in natural logs. This provides a natural interpretation of the CAPE as (semi-)elasticity and improves the fits of the LASSO and LLF models.

The main threat to this DML strategy is the existence of unobserved factors, such as non-physical political, historical, or ethnic contingencies that may affect the local placement of certain infrastructures without affecting the spatial distribution of infrastructure as a whole. Recent

---

<sup>12</sup>The Neyman Orthogonality Condition states that  $\partial_{m,f}(m_0, f_0)E[\epsilon\tilde{W}] = \partial_{m,f}(m_0, f_0)E[(\tilde{Y} - \tau(\mathbf{X}_H)\cdot\tilde{W})\tilde{W}] = 0$ , where  $\partial_{m,f}(m_0, f_0)$  is the (Gateaux) derivative of the moment condition w.r.t. to the nuisance parameters  $m, f$  evaluated at their true values denoted by  $m_0, f_0$ . This is zero, such that the moment conditions are not sensitive to small perturbations (biases) in the nuisance parameter estimates  $\hat{m}, \hat{f}$ . Such moment conditions and the final-stage (GMM) estimators they produce are called Neyman Orthogonal.

<sup>13</sup>The Horvitz-Thompson transformation is given by  $h(\mathbf{X}, W) = (W - e(W))/(e(W)(1 - e(W)))$  where  $e(W)$  denotes the propensity score.

contributions such as Dreher et al. (2019) and Graff (2024) provide evidence of favouritism and colonial legacy having an impact on activity and infrastructure in Africa, but they do not investigate whether this occurs on a local selective basis. To provide an example in the spirit of Dreher et al. (2019): if an African leader supports his/her birth region by only building additional schools, then this may be problematic because the ML prediction of schools from all other observable characteristics would not be able to capture the favouritism in education. If, however, the support is more broad-based and the region has not only more schools but also more hospitals, roads, and access to power, then these other features increase the ML prediction of schools (and of wealth) in that region and thus favouritism is absorbed by the ML control functions. Even if the first case holds, my nuisance models include a spatial lag (spillover version) of the IWI to eliminate spatial autocorrelation, thus favouritism would have to increase wealth only in the current cell.

To provide empirical evidence, I take data on political and ethnic favouritism and colonial railroad construction from Graff (2024). I aggregate the infrastructure data, debiased at high resolution using the ensemble ML estimator, into his  $0.5^\circ$  grid. I then run regressions similar to Tables 1 and 2 in Graff (2024) with both raw and debiased infrastructure quantities as outcome variables. Appendix Table A8/A9 reports results for simple/quantile counts. The top half considers raw infrastructure data and provide strong evidence for colonial history and favouritism extended by leaders towards their birth regions in shaping the spatial distribution of infrastructure. With the debiased data in the bottom half, most coefficients become zero and insignificant. A multiple testing adjustment following Clarke et al. (2020) renders all debiased coefficients insignificant. Thus, I argue that favouritism is an aggregate phenomenon and not a threat to high-resolution causal ML estimates controlling for other infrastructure and spatial autocorrelation.

Another caveat is that infrastructure also enters high-resolution outcome measures such as the IWI (through complex ML models learning from OSM and satellite imagery) or nightlights. However, the *Super Learner* can likely recover these 'unobserved nuisance functions', such that the effect of the treatment infrastructure can still be measured, even though the signal-to-noise ratio in the debiased data may be higher than with direct ground-truth measurements.

To test this, I also estimate (C)APEs using IWI estimates from DHS surveys conducted in SSA since 2010. I obtain 17,396 cells with more than 5 households and 10 persons/ $km^2$ , for which I compute the average household IWI, vs. 89,044 cells with IWI predictions by Lee & Braithwaite (2022). Both estimates are correlated but not perfectly ( $r = 0.814$ ), reflecting slight methodological differences and Lee & Braithwaite (2022)'s substantially higher resolution and use of DHS surveys only from 2017.<sup>14</sup> Appendix Table A10 shows the Median CAPE prediction (for all 103,922 populated cells) obtained from both IWI measures, and Appendix Table A11 shows APE estimates computed using cells where the respective measure is available. Both tables signify quite similar estimates, especially for important predictors such as roads, communications, travel time, accommodation, health and automotive. Curiously, education has no effect with the DHS data. The DHS-based IWI's median absolute coefficient size is slightly larger than the predicted IWI's. This might be due to noise in the predicted IWI, the reduced cell sample for the DHS-based IWI, or differences in methodology or timing, as elucidated above. Notwithstanding, the estimates are highly correlated across feature categories ( $r \geq 0.86$  for the APE). Tables A10 and A11 thus suggest that using a predicted IWI doesn't significantly alter the results.

A final obstacle is reverse causality between infrastructure and wealth/activity. This is impossible to rule out in a pure cross-section. However, controlling for the spatial lags of the IWI and of control infrastructure is likely to mitigate the effects, as similar wealth levels may be present in adjacent cells and wealthy inhabitants of cell  $i$  may also build infrastructure in neighbouring cells.

In summary, I argue that the causal ML strategy accounts for most significant spatial planning decisions, and the first stage residuals from the *Super Learners* are likely to represent some noisy local idiosyncrasies that can be used to estimate at least a partial equilibrium APE for different types of infrastructure. Notwithstanding, as I cannot formally establish causality or identification, the estimates should be interpreted with caution. I now report the estimates.

---

<sup>14</sup>I use DHS surveys from 2010 for this exercise to have more data.

## 4.1 Average Partial Effects

Table 6 reports the APE of the natural log of infrastructure on the IWI. The LHS shows results when features are simply counted in each category and cell, whereas the RHS shows results with quantile counts applied during aggregation, as detailed in Section 2.4. The estimates can be interpreted as the change in IWI points [0, 100] induced by a 100% increase in the corresponding feature intensity. Following Athey & Wager (2019), I also calculate APEs for cells above and below the median CAPE estimate and test for the difference between them as evidence for heterogeneity.<sup>15</sup>

Table 6: Average Partial Effects on IWI

Feature	ATE	Simple Counts			ATE	Quantile Counts		
		High	Low	Diff.		High	Low	Diff.
roads_paved	<b>0.218***</b>	0.283***	0.153***	0.131***	<b>0.216***</b>	0.28***	0.151***	0.128***
power	<b>0.0683***</b>	0.0849***	0.0517***	0.0332	<b>0.0675***</b>	0.0758***	0.0592***	0.0166
education	<b>0.563***</b>	0.958***	0.168***	0.79***	<b>0.436***</b>	0.774***	0.0973***	0.677***
health	<b>0.803***</b>	1.25***	0.359***	0.887***	<b>0.772***</b>	1.15***	0.389***	0.765***
communications	<b>0.689***</b>	0.951***	0.427***	0.525***	<b>0.689***</b>	1.01***	0.371***	0.635***
public_service_utility	<b>0.327***</b>	0.452***	0.203	0.248	<b>0.331***</b>	0.482***	0.179***	0.303*
automotive	<b>0.873***</b>	0.721***	1.02*	-0.299	<b>0.444**</b>	0.668**	0.22	0.447
transport_other	<b>0.332***</b>	0.395***	0.27***	0.125	<b>0.305***</b>	0.402***	0.207***	0.194**
financial	<b>0.626***</b>	0.884***	0.369**	0.515**	<b>0.543***</b>	0.791***	0.295	0.490**
services	<b>1.38</b>	—	1.38	—	<b>0.748*</b>	1.27*	0.222	1.05
ttime_city_1m	<b>-0.513***</b>	-0.335***	-0.691***	0.350***	<b>-0.502***</b>	-0.192***	-0.812***	0.62***
ttime_port_any	<b>-0.293***</b>	—	-0.293***	—	<b>-0.295***</b>	-0.141**	-0.448***	0.307***
residential	<b>-0.0679***</b>	0.0271	-0.163***	0.19***	<b>-0.0519***</b>	0.0391**	-0.143***	0.182***
accommodation	<b>0.807***</b>	0.778**	0.835***	-0.0568	<b>0.771***</b>	0.725***	0.817***	-0.0925
tourism_recreation	<b>0.184**</b>	0.472***	0.147*	0.325*	<b>0.309***</b>	0.509***	0.108	0.401**
mining_industrial	<b>0.541***</b>	0.878***	0.204***	0.674***	<b>0.46***</b>	0.698***	0.221***	0.477***

*Signif. Codes:* \*\*\*: 0.01, \*\*: 0.05, \*: 0.1. A ‘-’ indicates missing estimates (signifying underidentification).

*Notes:* Table shows doubly-robust APE estimates of the log feature intensity (simple counts or quantile counts in each cell, see Section 2.4) on the International Wealth Index [0, 100] by Lee & Braithwaite (2022) (covering 42 SSA countries). The “High” and “Low” estimates report the APE above and below the median CAPE estimate. The “Diff.” column indicates their difference to test for heterogeneity. All terms are tested using a two-sided t-test with standard errors derived from the doubly robust scores following Athey & Wager (2019).

Table 6 suggests that, controlling for all other features, including automotive and public transport facilities and travel time to major cities and ports, paved roads have a moderate partial effect on wealth amounting to a 0.22-point IWI increase to a doubling of the paved road quantity. The effect of power is even smaller at only 0.07. On the other hand education, health, communications, automotive facilities, financial services and accommodation (hotels) have rather large effects around 0.56-0.87. Public services and public transport have smaller effects around 0.33. Travel time to major cities and ports, proxying for exogeneous/extra-cell changes in market access, have the expected negative effect of around -0.5 for major cities and -0.3 for ports. Residential areas also have a slight negative effect around -0.07. This must be understood *ceteris paribus*: in the first-stage models I control for GPW4 population, but this is only available at the smallest administrative units. Thus, residential buildings may proxy for the remaining uncontrolled variation in population. The negative coefficient would then suggest that when all other infrastructure is held fixed, adding people does not increase individual wealth (‘urbanization without growth’ a la Fay & Opal (2000)).

For reference, I have also estimated APEs using the log of nightlights as outcome measures. The results, reported in Table A14, are broadly consistent with Table 6, but emphasize light-generating activities such as automotive facilities (presumably correlated with traffic at night), industrial facilities, and financial services (CBDs), and yield smaller effects of roads, communications, and health/education. Surprisingly, power infrastructure does not have very large effects either.

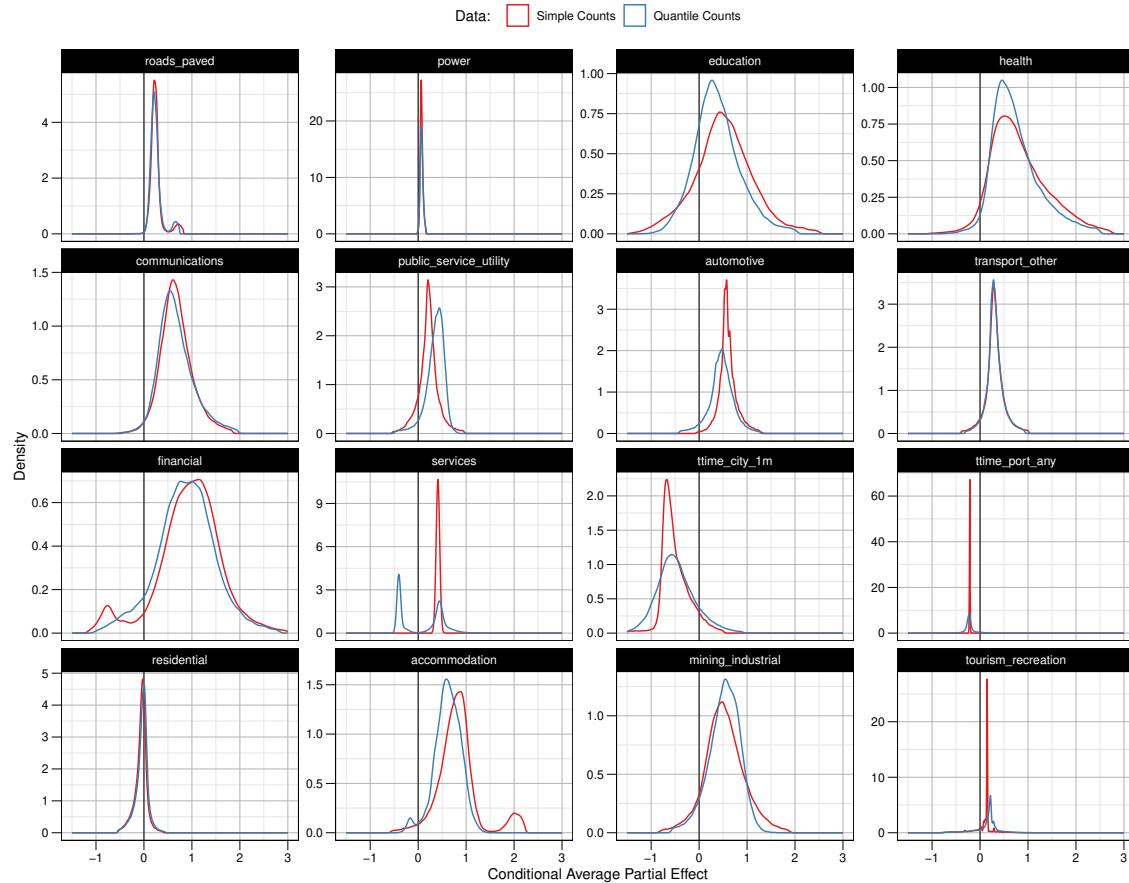
## 4.2 Effect Heterogeneity

The cell-level CAPE estimates are arguably more interesting than an all-Africa infrastructure APE. This subsection investigates their distribution and correlates before the next analyzes their spatial patterns. Figure 12 shows kernel density estimates for the IWI CAPEs. The densities indicate that for most infrastructure categories, there is considerable spatial heterogeneity in the partial effect. Only for paved roads and power, there seem to be no negative CAPEs, and the

<sup>15</sup>Since the reported CATE is already a BLP across multiple CATE estimators, the BLP cannot be used as an additional test for effect heterogeneity.

distributions are very narrow. Most other features except for residential buildings and travel time have largely positive but very heterogeneous effects. For education and financial services there are some negative CAPEs. In theory these shouldn't exist, but ML models are unrestricted. Negative values thus indicate that the effects can be zero or close to zero in some locations.

Figure 12: DML CAPE Kernel Density Estimates for IWI



*Notes:* Figure shows Gaussian kernel density estimates of the Conditional Average Partial Effect (CAPE) of log features' on the International Wealth Index (IWI) [0, 100] by Lee & Braithwaite (2022) (covering 42 SSA countries). Data are aggregated in each cell using either simple counts or quantile counts (see Section 2.4) before taking the natural log.

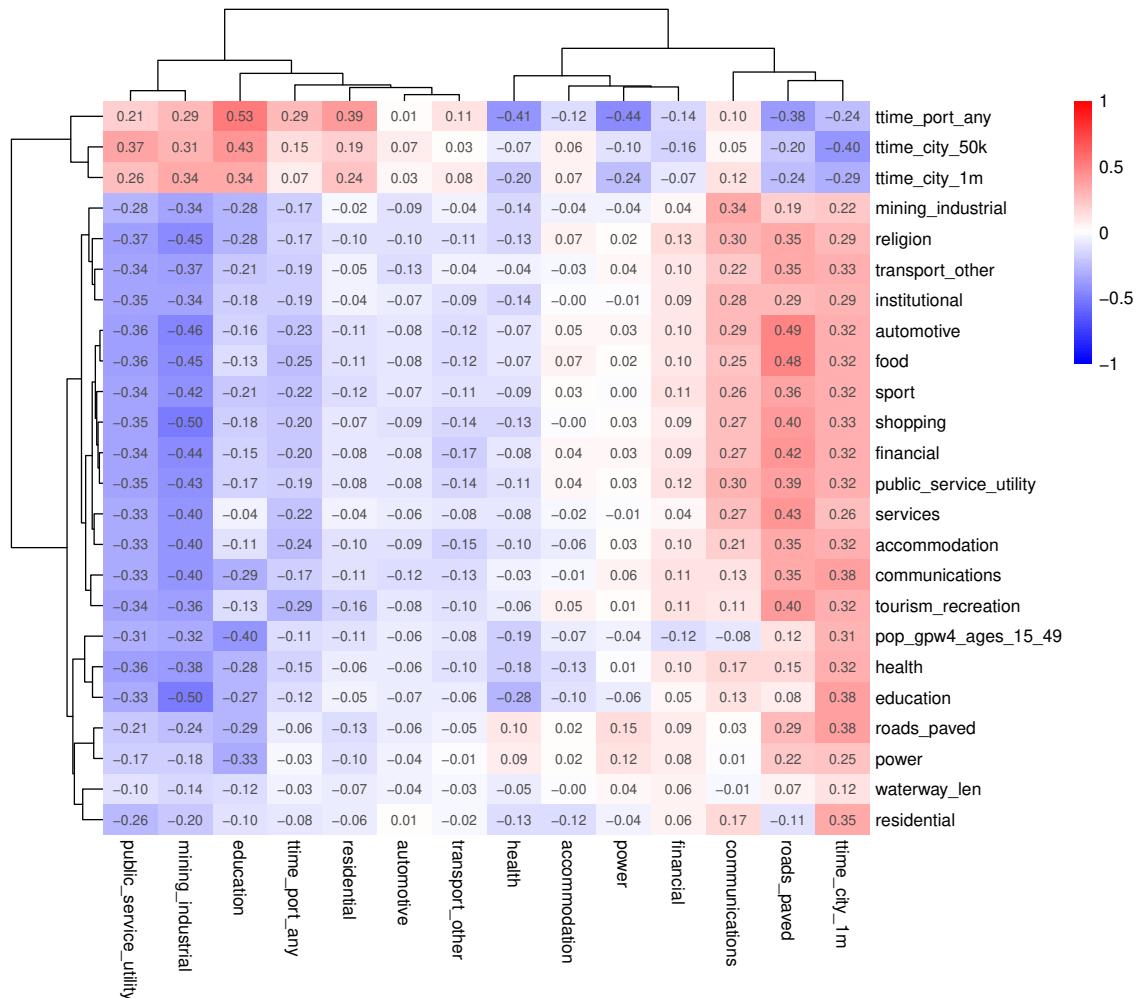
To compactly summarize important determinants of effect heterogeneity, I compute Pearson's correlations between the CAPEs and all variables in logs. I only report the top heterogeneity variables across infrastructure categories based on their average squared correlation with the CAPEs. Figure 13 reports the results, where the columns are the CAPEs, and the rows are covariates in decreasing order of average squared correlations. The number reported is the average correlation across the simple and quantile counts datasets.

At first sight, Figure 13 exhibits rather simple heterogeneity patterns: most CAPEs are either increasing or decreasing in the level of infrastructure/agglomeration. In particular, the CAPEs of paved roads, travel time to cities (market access), communications, and to a lesser extent, power and financial services, are increasing in the level of agglomeration. For other types of infrastructure the opposite appears to be the case. In particular education, public services and utilities, and mining/industrial facilities have stronger marginal effects on household welfare if found in structurally weak/rural areas. Interestingly, this is also the case for travel time to ports. This suggests some form of infrastructure substitution effects governed by agglomeration forces, e.g., in rural areas, travel time to ports becomes a key determinant of market access, and the presence of a school or a mine/industrial plant as employer has larger wealth effects than in cities with high market access and diverse employment opportunities, even for the uneducated.

As a further robustness exercise, Appendix Figures A12 and A13 show analogous CAPE

estimates from direct DHS wealth estimates, and Tables A12-A13 show correlations among CAPEs from both IWI measures. On average, the CAPEs are similar, particularly for important features with a robust effect, such as paved roads, communications and mining/industrial facilities. The median CAPE correlation is only around 0.13, and 0.25-0.3 if spillover variables are removed.

Figure 13: Correlates of IWI CAPE Estimates ( $\mathbf{X}_H$ ): Average Across Datasets



*Notes:* Figure shows average Pearson's correlations across the simple counts and quantile counts datasets of CAPEs (columns) with the features in logs (rows). A positive correlation implies that the CAPE is increasing in feature intensity.

Appendix Figures A14 and A15 show the densities and main correlates of nightlights CAPEs. While the densities are broadly similar, the correlation are almost all negative. This may be due to the unavailability of data in low-lit areas (zero observations are not considered for training).

There is also a strong spatial dimension to CAPEs not conveyed by correlations with infrastructure variables. Therefore, I now proceed with a visual examination of the IWI CAPE estimates.

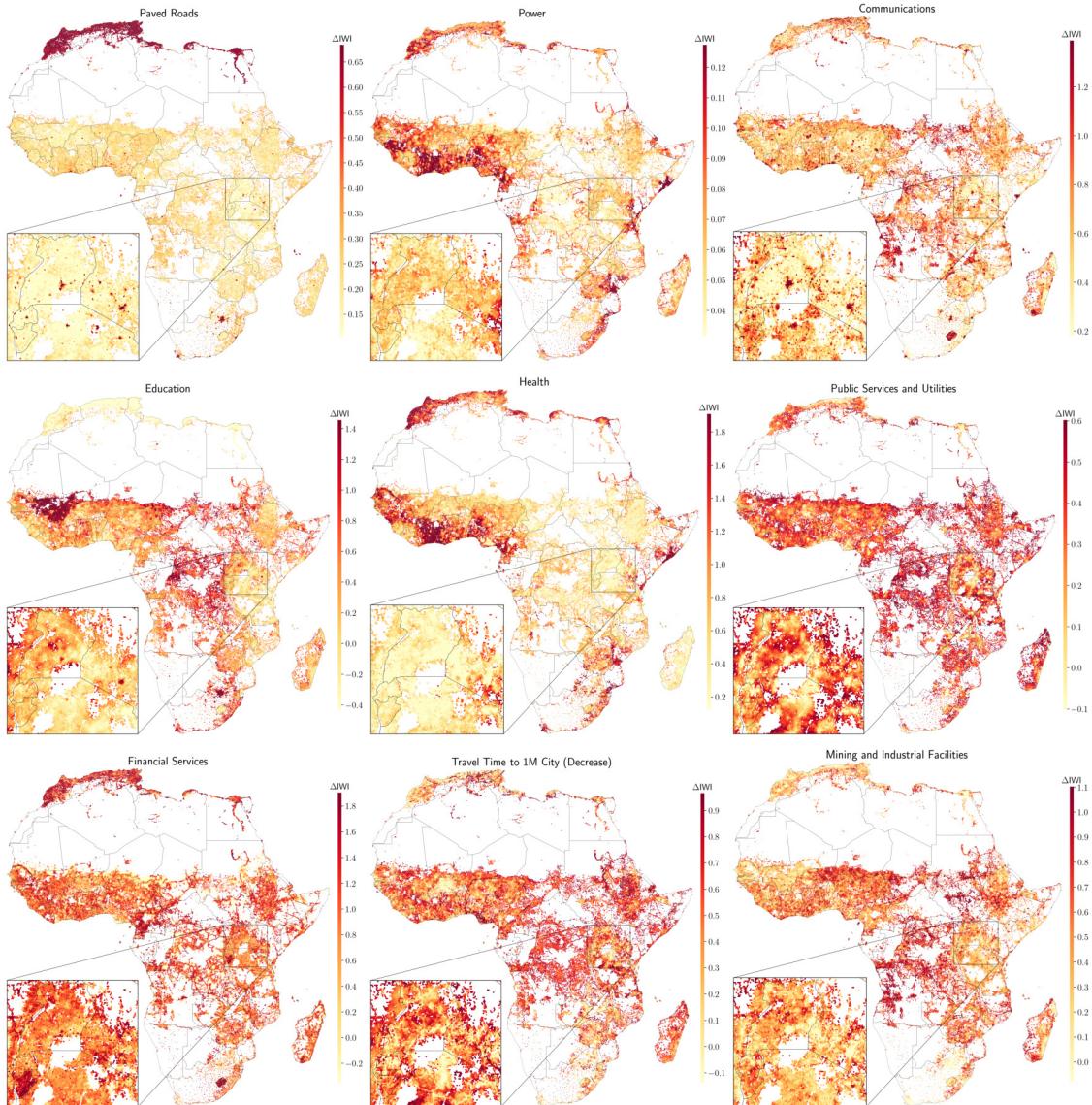
### 4.3 Spatial Examination of Effect Heterogeneity

Figure 14 shows visualizes the geometric mean CAPE across simple and quantile counts-based estimates (to smooth outliers, both estimates are highly correlated at  $r > 0.8$ ) for important features. Overall, there is considerable and complex spatial heterogeneity in the CAPEs. The top panel shows the CAPEs for roads, power, and communications. Exempting power,<sup>16</sup> the effects

<sup>16</sup>In previous versions of this paper, the CAPEs of power was concentrated in cities as well, see Appendix Figure A11. That version only used power infrastructure as reported in OSM. The present version incorporates some higher-resolution grid data from the EU's Joint Research Centre, and yields quite different results. Thus, coverage/definition of infrastructures can significantly affect the outcomes.

are higher in urban areas, particularly for paved roads. Since the IWI is only available for SSA, CAPEs for North Africa are pure predictions from the SSA model. This is problematic for paved roads given their much greater density in North Africa.

Figure 14: Spatial CAPE Estimates: Geometric Mean



*Notes:* Figure shows the Conditional Average Partial Effect (CAPE) of log features' on the International Wealth Index (IWI) [0, 100] by [Lee & Braithwaite \(2022\)](#) (covering 42 SSA countries). A geometric mean across estimates derived from simple and quantile counts data (Section 2.5) is reported to limit outliers. Both are correlated ( $r > 0.8$  for all features).

The middle panel of Figure 14 shows the CAPEs for education, health, and public services and utilities (excl. power). Exempting health,<sup>17</sup> these broadly follow the opposite spatial pattern, being highest in underdeveloped regions such as the Sahel region, Congo, and rural areas in Ethiopia, (South-)Sudan and Somalia. The bottom panel shows CAPEs for financial services, travel time to large cities, and mining and industrial facilities. The financial services CAPEs suggest that some countries like Burundi or Lesotho are financially underdeveloped. Travel time to cities CAPEs are probably the better estimates for what regional connectivity improvements through new roads could yield in terms of welfare. They are large in many areas outside of urban centers. The mining and industrial facilities CAPEs are also high in many structurally weak areas.

<sup>17</sup>In a previous version of this paper (Version 2), the CAPE of health also very much resembled the one of public services and utilities, see Appendix Figure A11, but this version includes significant data on public health facilities in SSA by [Maina et al. \(2019\)](#), yielding quite different results. Comprehensive and uniform data is needed for robust results.

In summary, CAPE estimates indicate varied associations between distinct infrastructure classes and the spatial distribution of economic activity. The estimates for roads and communications generally suggest stronger marginal effects inside urban spaces, whereas education and public service show stronger effects in semi-urban and rural areas. These findings appear consistent with parts of the literature on urbanization summarized by [J. V. Henderson & Turner \(2020\)](#), suggesting that poor service provision in rural areas is a key driver of rural-to-urban migration.

As pointed out before, there are many limitations to these CAPE estimates, starting from an imperfect classification of features, and non-comprehensive coverage of many infrastructures, especially in less populated regions, over a possibly low signal-to-noise ratio given that IWI estimates partly also depend on infrastructure, to possible reverse causality. They also don't consider general equilibrium effects, such as the relocation of populations and economic activities following infrastructure investments. Their highly partial nature, e.g., the partial effect of roads holding fixed market access and automotive facilities, also complicates policy interpretations.

Notwithstanding all of these caveats, the methodology appears to work well with spatial data and yields interesting spatial heterogeneity that is also broadly sensible and consistent with some of the econometric evidence. This allows for a hopeful outlook that rich and comprehensive spatial datasets (already available in the commercial realm e.g. by Google or Dataplor), and a refined spatial ML methodology, could be developed to generate policy-relevant spatial estimates.

From a policy point of view, another limitation is that marginal effects, and semi-elasticities in particular, are difficult to interpret. Policymakers are more interested in counterfactual predictions, such as the simulated effects of different levels of infrastructure in the same locations. It is possible to generate counterfactual predictions with causal ML, and I attempt to do so in the following section, but the properties of such estimates still need to be established.

## 5 Counterfactual Predictions

In this section, I explore counterfactual predictions obtained by increasing infrastructure quantities in the data and comparing ML model predictions from this altered dataset to the baseline prediction. This approach is similar to the single-model or 'S-learner' in the context of CATE estimation with binary treatments - applied by [Hill \(2011\)](#) with Bayesian Additive Regression Trees (BART) and discussed in [Jacob \(2021\)](#). I additionally follow [Facure & Germano \(2021\)](#) in using debiased data to avoid confounding influences in the model. Concretely, I estimate

$$\tilde{Y} = \theta(\tilde{W}, \mathbf{X}_H), \quad (12)$$

where  $\theta()$  is again an ensemble ML estimator (Eq. 11) and  $\tilde{Y}$ ,  $\tilde{W}$  are debiased. In contrast to CAPE estimation, this formulation allows for the effect to be non-linear, i.e., to change with the level of  $\tilde{W}$ . After estimating  $\theta()$  in a cross-fitting manner, I evaluate it at different levels of  $\tilde{W}$ . Since removing infrastructure is not an interesting policy option, I compute the 10%, 25%, 50%, 75%, and 90% quantiles of the positive (non-missing) distribution of  $W$  across cells and alternatingly add them to  $\tilde{W}$  and obtain a prediction. Table 7 shows these quantile increases. The 10% increase amounts to one additional facility in most categories, 59 additional power-related items (transformers, generators), or a 2111m increase in the length of paved roads in each cell. Higher quantiles imply much more substantive increases per cell. The increase is only applied to the respective treatment infrastructure  $\tilde{W}$ ; all other infrastructures  $\mathbf{X}_H$  are unaltered.

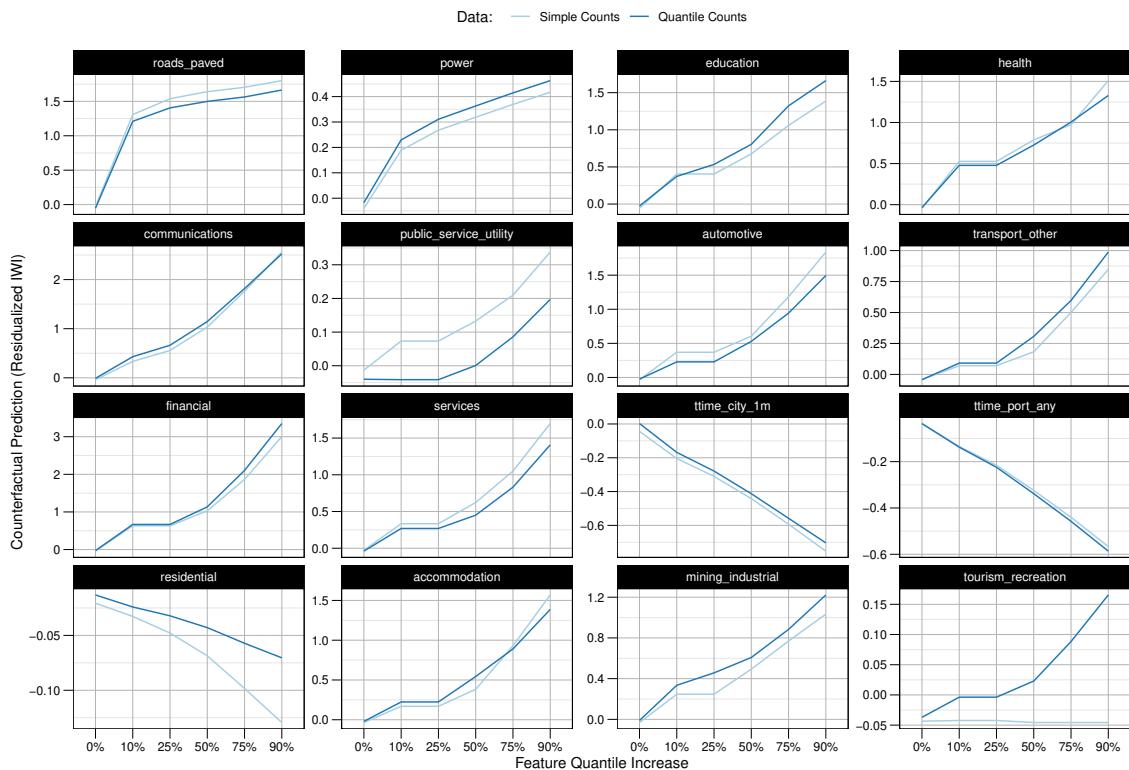
Due to debiasing, the OOB  $R^2$  for the IWI residuals is only 2-5%. The model-based simulation results nevertheless appear meaningful. Figure 15 shows the average predictions with the IWI as outcome measure and Figure A16 with the log of nightlights. Except for residential buildings which have a positive effect with nightlights, the results from both outcomes are very similar. The relative magnitudes are also broadly consistent with the APEs. Appendix Tables A15 and A16-A17 show counterfactual predictions derived from a DHS-based IWI, which are also similar, instilling some confidence in the methodology. Since for many sparse features, both the 10% and 25% treatment levels imply 1 additional feature with count data (see Table 7), the average counterfactual prediction curves in Figure 15 may have some flat segments.

Table 7: Counterfactual Increase in Feature Quantity/Cell for Selected Features

Feature	10%	25%	50%	75%	90%
roads_paved (m)	2111	6295	10371	14493	24119
power	59	168	314	570	980
education	1	1	3	9	24
health	1	1	2	3	8
communications	1	2	6	23	85
public_service_utility	1	1	2	4	10
automotive	1	1	2	7	23
transport_other	1	1	2	7	20
financial	1	1	2	6	21
services	1	1	2	6	24
ttime_city_1m (min)	104	203	363	602	930
ttime_port_any (min)	130	289	564	951	1520
residential	1	3	8	24	84
accommodation	1	1	2	7	24
mining_industrial	1	1	3	8	19
tourism_recreation	1	1	3	8	30

*Notes:* Table shows sample quantiles on simple-counts data, computed across cells with positive infrastructure for each feature category, respectively. The quantiles are rounded to the full number. They are used for counterfactual predictions, i.e., increasing the infrastructure quantity in all cells by this amount and making a prediction using the causal single-learner (Eq. 12).

Figure 15: Average Counterfactual Predictions for IWI



*Notes:* Figure shows average counterfactual predictions (across all cells) derived from simple and quantile counts data.

Overall, the counterfactual predictions in Figure 15 are very consistent with the CAPE estimates. There is, in general, decreasing returns to scale, i.e., building more of the same infrastructure in each cell does not raise wealth proportionately. For some feature categories such as paved roads and power, the 10% increase has a large initial effect, and further increases add much lower wealth effects. The main reason for this is that, as shown in Section 3.1, most cells are scarcely populated, without paved roads or power supply, and do not have many types of features at all. Thus, the

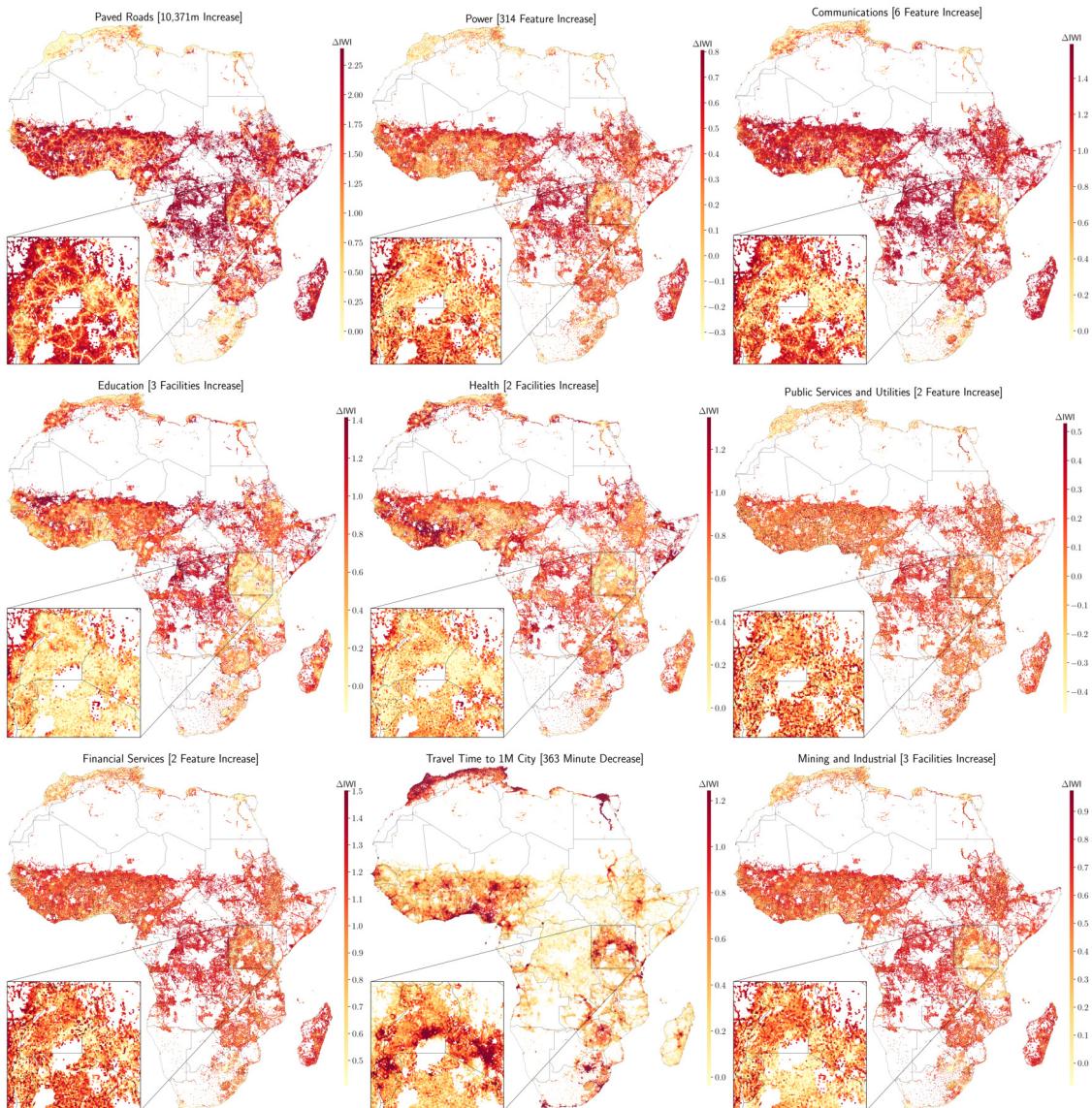
existence of any paved roads or power suggests to the model that more diverse activity than farming takes place in the cell and implies an increase in average wealth levels. This may be unrealistic, that is, even with debiased data, counterfactual predictions may generate locations that are dissimilar to the ones the model has learned, so further increases in roads/power do not increase wealth much.

Because cells have heterogeneous populations, Appendix Figure A17 also provides the average total wealth effect per cell obtained by multiplying the cell-level predictions with the WorldPop 2020 population measure and computing the average across cells. The relative magnitudes are broadly similar to Figure 15.

## 5.1 Spatial Examination of 50% Counterfactual Predictions

It remains to examine the spatial distribution of these counterfactual predictions. Since predictions from simple and quantile count data are very similar, I again report a geometric mean across the two which slightly downweights outliers. Since an appraisal of all different treatment levels in the spatial dimension would be overwhelming, I only plot the 50% level in Figure 16.

Figure 16: Spatial 50% Counterfactual Predictions: Geometric Mean



*Notes:* Figure shows 50% counterfactual predictions of the International Wealth Index (IWI) [0, 100] by Lee & Braithwaite (2022), i.e., the predicted wealth increase (Eq. 12) from an increase in each cell amounting to the median of the non-negative feature density, summarized in Table 7. A geometric mean across estimates derived from simple and quantile counts data (Section 2.5) is reported to limit outliers. Both are correlated ( $r > 0.8$  for all features).

Evidently, counterfactual predictions (CFPRs) are more similar across different infrastructure types than CAPEs, indicating that generally the household wealth returns to building a given amount of infrastructure is higher in rural areas than in cities. A notable exception is travel time to major cities, which the CFPRs suggest should be reduced in the vicinity of such cities. Building infrastructure in rural areas as suggested by most of these maps is, however, not aggregate welfare maximizing since most of these areas are scarcely populated. It likely is more sensible to build infrastructure in populated places even though per-household welfare gains are considerably lower.

To provide at least a heuristic, partial equilibrium appraisal of how a social planner seeking to maximize aggregate welfare might allocate investments, I multiply the 50% CFPRs of Figure 16 by the WorldPop 2020 population measure in each cell. Appendix Figure A19 shows the outcome. As expected, investments in populated areas generally yield higher aggregate welfare returns, but there remains considerable heterogeneity across different infrastructures, for example power and education investments are primarily directed to populated rural areas and less to city centers.

Overall, these counterfactual predictions emphasize the benefits of investments in rural areas towards poverty alleviation. The reader should keep in mind that policymakers may consider not only population but also poverty, inequality, and other social and political objectives in determining optimal infrastructure allocations. With more careful efforts at monetizing them (e.g., considering not only the beneficiaries and policy objectives but also the cost of building infrastructure in different locations), counterfactual predictions appear able to provide useful guidance for spatial planning. They are easier to interpret than a semi-elasticity CAPE, allow simulating different 'treatment levels,' and do not require the restrictive additive separability of treatment assumption in CAPE estimation. On the other hand, the single-learner approach needs to be better studied and has significant shortcomings. In particular, the relevance of the 'treatment variable' is not guaranteed in an unrestricted ML model. Estimating a CAPE using multiple estimators and combining them through a BLP model - as done in this paper - has the clear advantage of at least providing well-calibrated estimates. An equivalent estimation and validation strategy for counterfactual predictions remains to be developed.

## 6 Summary and Conclusion

This paper studies Africa's spatial economy and the interplay of infrastructure with household wealth and economic activity using rich geospatial data compiled from Open Street Map, Overture Maps, and multiple recent research and data contributions. It also maps potential economic benefits from different types of incremental infrastructure investments at high spatial resolutions using causal machine learning methods. The investigation is fruitful in several respects.

It shows that most infrastructure in Africa is highly concentrated in cities - more than population, and, exempting educational facilities, the sum of all features, and network infrastructures such as roads and power, also than GDP. Populated places themselves are very heterogeneous. A case study of 5 African capital cities reveals that some cities, such as Nairobi and Johannesburg, outperform others, such as Lagos, Cairo, and Accra, regarding infrastructure, public services, and amenities per person, even after controlling for wealth, population, and market access. This confirms that other factors, such as city governance and the overall business environment, also shape the distribution of economic activity and infrastructure. Furthermore, in many countries, infrastructure seems inefficiently allocated, with population, roads and, power infrastructure, and economic activities often found in disparate locations. Countries with higher levels of spatial efficiency, characterized by the average proximity of population, infrastructure, and activities, have higher average levels of development, a stronger business environment, and greater logistic performance.

Training tree-based ensemble ML models to predict wealth and economic activity from infrastructure yields that roads, communications, education, power infrastructure, and residential areas are the most important predictors of wealth. Other features such as accommodation (hotels), automotive facilities, sports facilities, commercial buildings, public services and utilities, and health facilities are also important. Models including population and travel time to major cities and ports rank them as top predictors, indicating that market access is similarly essential to physical infrastructure. The relative importance of these variables varies across locations. In cities, the quantity

of paved roads and power infrastructure is very important for wealth prediction. In contrast, education, health, and public service facilities are of greater relevance in structurally weaker areas.

Similar patterns are evident in marginal effects ((semi-)elasticities) of household wealth to infrastructure, computed using causal ML methods. Investments in education, paved roads, power, communications, industrial facilities, health, automotive facilities, transport, public services, and hotels yield high average but spatially very heterogeneous returns. Doubly robust estimates of the Average Partial Effect (APE) of infrastructure on wealth range between 0 and 1 point increases in the International Wealth Index (IWI) in response to the doubling of infrastructure within a specific category. The APE of roads is robustly around 0.22 IWI points, market access held fixed. The Conditional Average Partial Effect (CAPE) for roads is also positive everywhere but highest in urban spaces, implying that more roads are always a good idea, especially in urban agglomerations. Power has a smaller APE of 0.07. Education, on the other hand, has a large APE between 0.44 and 0.56, with the largest CAPEs realized in rural areas lacking schools. Communications also has a large APE of 0.696; the CAPEs are generally higher in urban areas. Public services have an APE of around 0.33, with higher CAPEs in less developed areas. Industrial facilities have an APE around 0.5, with, surprisingly, higher CAPEs in rural areas, presumably because they are not major income generators in African cities. Market access itself also has a sizeable impact, with a doubling of travel time to major cities ( $> 1$  million inhabitants) decreasing the IWI by -0.5 points, and a doubling of travel time to ports having a similar effect of -0.3 IWI points. Residential buildings also have a small negative APE between -0.05 and -0.07, indicating that, all other infrastructure held fixed, individual wealth is an inverse function of the number of people using these infrastructures. These marginal effects are partial equilibrium effects that do not consider the spatial relocation of economic activity and populations following such investments. They may also suffer from reverse causality.

Counterfactual predictions, obtained by applying a spatially uniform increase in infrastructure to the data (e.g., one additional school in all cells) and predicting the outcome (e.g., wealth) using a debiased ML model, yield greater effects in rural areas compared to the CAPEs (which are semi-elasticities). However, the number of beneficiaries in such areas is smaller. Factoring in the different populations of locations generally emphasizes investments in more populated locations, especially those lacking the respective infrastructure. The correct use of counterfactual predictions - even if perfectly causal - thus critically depends on the objective of policymakers: welfare maximization, poverty alleviation, and inequality aversion objectives imply different spatial policies derived from the same set of predictions. Apart from these policy challenges, the less robust estimation methodology and lack of theoretical literature on counterfactual predictions are clear disadvantages despite their more natural interpretation and less restrictive theoretical assumptions.

In conclusion, the findings of this paper need to be digested with caution. They are exploratory and not sufficiently robust or causal to be used for policy advice. However, given the elusiveness of comprehensive causal empirical evidence, they offer a significant advance toward data-driven evidence on the local and global returns of infrastructure and public/private services towards activity and wealth generation/poverty alleviation. Many methodological improvements are conceivable, e.g., using more sophisticated spatial weighting matrices or ML methods explicitly capable of learning spatial dependencies (such as Convolutional Neural Networks). A refined methodology and comprehensive geospatial data may generate infrastructure potential maps that could provide helpful guidance for spatial planning and gauging the returns to small public infrastructure investments. As geospatial data is becoming increasingly rich and the measurement of wealth and activity across space continues to improve, the pioneering results presented in this paper allow for a hopeful outlook on causal machine learning's potential to enhance spatial economic and development planning.

## References

- ADB. (2018). Africa's infrastructure: great potential but little impact on inclusive growth. *African Economic Outlook*.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining*.
- Akpandjar, G., & Kitchens, C. (2017). From darkness to light: The effect of electrification in ghana, 2000–2010. *Economic Development and Cultural Change*, 66(1), 31–54.
- Allcott, H., Collard-Wexler, A., & O'Connell, S. D. (2016, March). How do electricity shortages affect industry? evidence from india. *American Economic Review*, 106(3), 587-624. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/aer.20140389> doi: 10.1257/aer.20140389
- Apley, D. W., & Zhu, J. (2016). Visualizing the effects of predictor variables in black box supervised learning models. *arXiv preprint arXiv:1612.08468*.
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360.
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148.
- Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational studies*, 5(2), 37–51.
- Athey, S., & Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1), 133–161.
- Barnes, R. (2020). dggridr: Discrete global grids [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=dggridR> (R package version 2.0.4)
- Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115, 103124.
- Bayer, P., Kennedy, R., Yang, J., & Urpelainen, J. (2020). The need for impact evaluation in electricity access research. *Energy Policy*, 137, 111099.
- Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2021). Deep neural networks and tabular data: A survey. *arXiv preprint arXiv:2110.01889*.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Byers, L., Friedrich, J., Hennig, R., Kressig, A., Li, X., McCormick, C., & Valeri, L. M. (2018). A global database of power plants. *World Resources Institute*, 18.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5), 261–265.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1), C1-C68. Retrieved from <https://doi.org/10.1111/ectj.12097> doi: 10.1111/ectj.12097
- Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india* (Tech. Rep.). National Bureau of Economic Research.

- Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., & Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4), 1501–1535.
- Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3), e2113658119.
- CIESIN. (2016). *Gridded population of the world, version 4 (gpwv4): Population count* (Tech. Rep.). NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from <http://dx.doi.org/10.7927/H4X63JVC>
- Clarke, D., Romano, J. P., & Wolf, M. (2020). The romano–wolf multiple-hypothesis correction in stata. *The Stata Journal*, 20(4), 812–843.
- Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899–934.
- Donaldson, D., & Hornbeck, R. (2016). Railroads and american economic growth: A “market access” approach. *The Quarterly Journal of Economics*, 131(2), 799–858.
- Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171–198.
- Dreher, A., Fuchs, A., Hodler, R., Parks, B. C., Raschky, P. A., & Tierney, M. J. (2019). African leaders and the geography of china’s foreign assistance. *Journal of Development Economics*, 140, 44–71.
- Faber, B. (2014, 03). Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System. *The Review of Economic Studies*, 81(3), 1046-1070. Retrieved from <https://doi.org/10.1093/restud/rdu010> doi: 10.1093/restud/rdu010
- Facure, M., & Germano, M. (2021). *Python Causality Handbook: First Edition*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.4445778> doi: 10.5281/zenodo.4445778
- Fay, M., & Opal, C. (2000). *Urbanization without growth: A not so uncommon phenomenon* (Vol. 2412). World Bank Publications.
- Foster, V., & Briceño-Garmendia, C. (2010). *Africa’s infrastructure: a time for transformation*. World Bank.
- Friedberg, R., Tibshirani, J., Athey, S., & Wager, S. (2020). Local linear forests. *Journal of Computational and Graphical Statistics*, 30(2), 503–517.
- Friedman, J., Tibshirani, R., & Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. doi: 10.18637/jss.v033.i01
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
- Gennaioli, N., La Porta, R., Lopez-de Silanes, F., & Shleifer, A. (2013). Human capital and regional development. *The Quarterly journal of economics*, 128(1), 105–164.
- Gibson, J., Olivia, S., & Boe-Gibson, G. (2020). Night lights in economics: Sources and uses 1. *Journal of Economic Surveys*, 34(5), 955–980.
- Goldbeck, M., & Lindlacher, V. (2021). *Digital infrastructure and local economic growth: Early internet in sub-saharan africa*. ifo Institute for Economic Research, University of Munich.
- Graff, T. (2024). Spatial inefficiencies in africa’s trade network. *Journal of Development Economics*, 103319.
- Henderson, J. V., & Turner, M. A. (2020). Urbanization in the developing world: too early or too slow? *Journal of Economic Perspectives*, 34(3), 150–173.

- Henderson, V., Storeygard, A., & Weil, D. N. (2011). A bright idea for measuring economic growth. *American Economic Review*, 101(3), 194–199.
- Henderson, V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, 102(2), 994–1028.
- Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217–240.
- Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164, 73–84.
- Jacob, D. (2021). Cate meets ml: Conditional average treatment effect and machine learning. *Digital Finance*, 3(2), 99–148.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.
- Jedwab, R., & Storeygard, A. (2022). The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960–2010. *Journal of the European Economic Association*, 20(1), 1–38.
- Kakoulaki, G., & Moner-Girona, M. (2020). *Electricity grid africa* [Dataset]. <http://data.europa.eu/89h/624c6e71-3b9c-4f48-8c67-645911798d41>. European Commission, Joint Research Centre (JRC).
- Kmoch, A., Matsibora, O., Vasilyev, I., & Uuemaa, E. (2022). Applied open-source discrete global grid systems. *AGILE: GIScience Series*, 3, 41.
- Krantz, S. (2023). osmclass: Classify open street map features [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=osmclass> (R package version 0.1.3)
- Kummu, M., Taka, M., & Guillaume, J. H. (2018). Gridded global datasets for gross domestic product and human development index over 1990–2015. *Scientific data*, 5(1), 1–15.
- Lee, K., & Braithwaite, J. (2022). High-resolution poverty maps in sub-saharan africa. *World Development*, 159, 106028.
- Lee, K., Miguel, E., & Wolfram, C. (2020, February). Does household electrification supercharge economic development? *Journal of Economic Perspectives*, 34(1), 122–44. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/jep.34.1.122> doi: 10.1257/jep.34.1.122
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1), 56–67.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Maina, J., Ouma, P. O., Macharia, P. M., Alegana, V. A., Mitto, B., Fall, I. S., ... Okiro, E. A. (2019). A spatial database of health facilities managed by the public health sector in sub saharan africa. *Scientific data*, 6(1), 134.
- Moneke, N. (2020). *Infrastructure and structural transformation: evidence from ethiopia* (Unpublished doctoral dissertation). London School of Economics and Political Science.
- MSI. (2019). *World port index 2015* (Tech. Rep.). Retrieved from <https://msi.nga.mil/Publications/WPI>
- Nelson, A. (2022, 10). *Travel time to cities and ports in the year 2015* (Tech. Rep.). Retrieved from [https://figshare.com/articles/dataset/Travel\\_time\\_to\\_cities\\_and\\_ports\\_in\\_the\\_year\\_2015/7638134](https://figshare.com/articles/dataset/Travel_time_to_cities_and_ports_in_the_year_2015/7638134) doi: 10.6084/m9.figshare.7638134.v4
- Nelson, A., Weiss, D. J., van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019). A suite of global accessibility indicators. *Scientific data*, 6(1), 1–9.

- Nie, X., & Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2), 299–319.
- Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N. M., Mohammed, M., ... Weiss, J. (2006). The g-econ database on gridded output: methods and data. *Yale University, New Haven*, 6, 11.
- Peng, C., & Chen, W. (2021). Roads to development? Examining the Zambian context using AI-Sat.
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427), 846–866.
- Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 931–954.
- Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., ... others (2018). Nasa's black marble nighttime lights product suite. *Remote Sensing of Environment*, 210, 113–143.
- Sahr, K. (2022). User documentation for discrete global grid generation software. *Southern Oregon Univ., Ashland, OR, USA, Tech. Rep. Dggrid version, 7.5*.
- Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, 30(2), 121–134.
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest: non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.
- Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-saharan africa. *The Review of economic studies*, 83(3), 1263–1295.
- Tibshirani, J., Athey, S., Sverdrup, E., & Wager, S. (2023). grf: Generalized random forests [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=grf> (R package version 2.3.0)
- Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, 6(1).
- Weiss, D. J., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., ... others (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688), 333–336.

## Appendix

### Constructing the Africa Infrastructure Database

The basis for the database is the Open Street Map (OSM) of Africa from April 2024, downloaded from Geofabrik.de. OSM has three basic data structures (nodes, ways, and relations) labeled through tags. A tag consists of a key and a value. Under the free tagging system, an object can have unlimited tags. However, the community agrees on certain key-value combinations for the most commonly used tags, which act as informal standards. In particular, the [OSM Feature Documentation](#) lists 29 primary tags, such as amenity, building, highway, water, landuse, shop, craft, etc. These primary tags are used with specified values to classify certain features, for example, amenity = school, shop = bakery, or building = hotel, and often accompanied by supplementary tags providing more precise information about a feature, e.g., name, description, denomination, etc.

There are three main obstacles to reconciling this tagging system with the economic significance of map features: (1) a feature can be classified according to multiple primary tags, e.g., amenity = school, building = education, landuse = education or amenity = hospital, healthcare = laboratory, emergency = yes; (2) sometimes classifications based on primary tags can conflict, e.g., for a religious school amenity = school and religion = christian, or for a hotel with restaurant building = hotel and amenity = restaurant; (3) the proximity of the object to the economic activity concerned may not be very clear, e.g., amenity = school and landuse = education both signify the primary use of an object for educational purposes, but the 'school' tag is a lot more specific. The tagging system of OSM is also subject to changes, and due to the crowdsourced nature of the map, some features are not classified according to current standards.

Constructing a functional classification of OSM is thus more an art than a science. One needs to devise a system of economic categories and lay out all tags according to which an economic category is to be assigned and the order in which these tags and categories are to be matched. This needs to be informed by both the OSM tagging standards and empirical accounts of current mapping practice. Once a classification scheme has been specified and applied to the map, features assigned to multiple categories need to be investigated, and the classification needs to be iteratively refined to minimize overlap and misclassification.

Following this process, I have developed a classification scheme to classify point and polygon (closed ways) map features into 33 economic categories based on 33 (mostly primary) tags and 341 values to be matched, including matching on any value for specific tags like 'sport' or 'power.' Within each category, tags providing precise information about the nature of the feature are matched first, and more general tags like 'building' or 'landuse' are matched last. The classification excludes natural features like mountains or lakes, natural or administrative boundaries, and minor features with little economic significance, such as traffic signs or flag poles. Minor infrastructure related to power, telecommunications, and military purposes is, however, included.

The [osmclass](#) R package developed for this purpose helps apply such classifications - defined as nested lists of categories, tags, and values - to OSM PBF files imported as spatial data frames. It includes the classifications used in this paper. Table A1 summarizes the classification and features extracted from the Africa OSM of April 2024, sorted by the number of features on the map.<sup>18</sup>

---

<sup>18</sup>This is not the order in which categories were matched, which was chosen to minimize misclassification in Africa. A detailed view of the classification is provided in the R package at <https://github.com/SebKrantz/osmclass/blob/main/R/classifications.R>.

Table A1: Classification of Africa OSM Point and Polygon Features: April 2024

Category	NTags	NVals	Tags and Number of Matched Values	N
residential	3	11	building (8), building:use (2), landuse (1)	6,382,886
power	4	4	power (all), utility (1), building (1), tower:type (1)	1,672,204
farming	4	19	place (1), man_made (1), building (9), landuse (8)	1,252,658
construction	2	2	building (1), landuse (1)	566,912
transport	12	49	amenity (21), highway (9), railway (all), aerialway (all), waterway (6), aeroway (all), public_transport (all), bridge (all), junction (all), office (2), man_made (2), building (3)	453,510
shopping	4	9	amenity (2), shop (2), building (4), landuse (1)	358,407
education	3	7	amenity (3), building (3), landuse (1)	328,016
sports	3	13	leisure (7), sport (all), building (5)	177,134
facilities	2	19	amenity (18), building (1)	161,083
religion	6	21	amenity (6), building (11), office (1), landuse (1), religion (all), denomination (all)	118,861
food	1	7	amenity (7)	115,087
health	3	11	amenity (9), healthcare (all), building (1)	106,353
utilities_other	5	18	man_made (12), water (1), office (1), building (3), landuse (1)	95,801
commerical	2	2	building (1), landuse (1)	90,549
industrial	4	7	industrial (all), man_made (2), building (1), landuse (3)	84,928
recreation	4	20	amenity (5), leisure (13), landuse (1), building (1)	79,447
historic	1	1	historic (1)	75,332
accommodation	2	8	tourism (7), building (1)	67,054
craft	1	1	craft (all)	59,246
tourism	3	4	tourism (1), shop (1), office (2)	47,617
financial	2	11	amenity (6), office (5)	46,040
institutional	3	7	office (5), building (1), landuse (1)	42,775
public_service	2	12	amenity (10), building (2)	36,658
mining	2	5	man_made (4), landuse (1)	29,711
communications	7	17	amenity (2), telecom (all), communication (all), utility (1), man_made (6), office (1), tower:type (5)	28,740
storage	3	6	man_made (2), building (3), landuse (1)	28,308
office_other	2	2	office (all), building (1)	26,158
waste	4	8	amenity (5), water (1), man_made (1), landuse (1)	21,313
education_alt	3	9	amenity (6), office (2), building (1)	13,852
military	3	5	military (all), building (3), landuse (1)	12,696
entertainment	2	17	amenity (13), leisure (4)	6,028
emergency	1	1	emergency (1)	4,622
creativity	3	8	amenity (2), leisure (1), office (5)	2,142
SUM	33	341		12,592,128

Table A2 shows a similar classification of line-based features. Residential roads are excluded as they are irrelevant to trade and strongly overlap with residential buildings recorded on the map. Smaller natural water features such as streams, wadis, and ponds are also excluded.

Table A2: Classification of Africa OSM Line Features: April 2024

Category	NTags	NVals	Tags and Number of Matched Values	N	Length (Km)
road	1	10	highway (10)	763,912	1,621,144
waterway	3	12	waterway (8), water (1), man_made (3)	359,756	1,507,112
power	1	1	power (all)	120,052	422,504
railway	1	1	railway (all)	84,707	128,408
aeroway	1	1	aeroway (all)	27,360	11,019
pipeline	1	1	man_made (1)	9,453	55,394
storage	1	2	man_made (2)	8,551	389
ferry	1	1	route (1)	2,412	48,259
aerialway	1	1	aerialway (all)	171	175
telecom	2	2	telecom (all), communication (all)	87	28,682
SUM	11	32		1,376,805	3,834,579

### Harmonized Classification

I then combine the point/polygon data from OSM with the 824 thousand POIs from the Overture Places layer to create a uniform classification. This involves creating 47 detailed categories and matching features to them based on OSM tags or primary POI categories in Overture maps, of which there are more than 1000. To simplify the dataset again for most analytical use cases and ensure that each category has sufficient features, I combined some of the 47 categories, yielding a simplified classification of 26 categories. The other POI data, e.g., from All The Places, is more limited and easier to classify (e.g., mostly shops, restaurants, and hotels clearly tagged) and can seamlessly be assigned to the 47 categories.

### Deduplication

Having classified POIs from 11 different sources (Table 1) into 47 categories, it remains to sort out duplicates across sources. The order of precedence is to favour curated data (such as Global Integrated Power Tracker or health facilities by [Maina et al. \(2019\)](#)) above OSM, which in turn takes precedence over Overture places (which various [online appraisals](#) found less accurate). POIs are then deduplicated within each category and 10m square. This resolution is motivated by considering a dense shopping mall where stores may be only 10m apart. POIs are resolved to grid cells by dividing their coordinates by the degree-equivalent of 10m at the equator  $[10/(40075017/360)] \approx 9e-5$ , subtracting the modulus of this division from the coordinates and using them to group and deduplicate features. To ensure equal distance representation across Africa, longitudes are multiplied beforehand by  $\cos(lat \times \pi/180)$ , where  $lat$  is the latitude. After a deduplication round, the coordinates are incremented by 1m degree-equivalent ( $\approx 9e-6$ ), and the process is repeated. A sequence of such 1m nudges to the  $lon$  and  $lat$  coordinates is used to 'shift the grid' across space in a structured way until no more duplicates can be found. Since which POIs are first compared depends on the initial position of the 'grid,' there is some path dependence in this process. However, the clear hierarchy across sources ensures that curated datasets are generally fully retained.

Table A3 summarizes the finally classified and deduplicated POI data, including the corresponding number of POIs and the share of OSM and OSM polygons (tagged buildings).

Linestring (network) data is mainly taken from OSM, which is a reliable data source for roads, waterways, and railways. However, its coverage of power lines is limited to major lines in most regions. To increase granularity, I add Africa electricity grid maps from the EU's Joint Research Center ([Kakoulaki & Moner-Girona, 2020](#)) and the [World Bank](#), which more than doubles the total length of power lines observed from 423 thousand km in OSM to 967 thousand km (Table 2) following harmonization. To combine/harmonize the linestrings, I compute a geometric union between all three data sources. This obscures definitions of individual power lines from different datasets and breaks up the linestrings into smaller segments, but for the purposes of this research, only the total length of lines per cell matters. Table 2 summarizes the final lines dataset.

Table A3: Africa Infrastructure Database: Places Dataset by Category

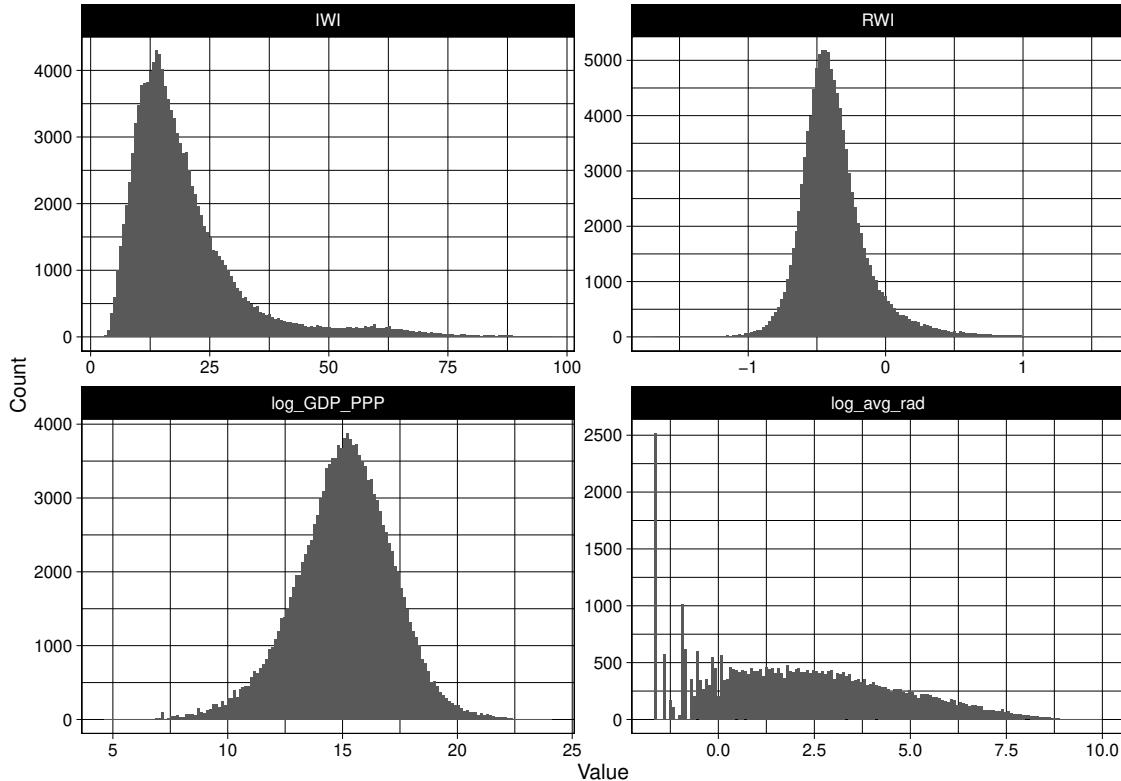
detailed (47)	simplified (26)	count	perc	polygons	poly_perc	osm	osm_perc	src_cat
residential	residential	6,276,302	41.42	6,256,643	99.69	6,267,969	99.87	12
communications_network	communications	1,914,448	12.64	502	0.03	20,092	1.05	31
communications_other	communications	14,245	0.09	826	5.80	8,239	57.84	31
power	power	1,654,373	10.92	1,711	0.10	1,651,175	99.81	92
farming	farming	1,228,969	8.11	1,219,018	99.19	1,223,781	99.58	28
construction	construction	585,168	3.86	565,752	96.68	574,167	98.12	90
transport_other	transport	303,186	2.00	21,166	6.98	293,233	96.72	188
automotive	transport	164,845	1.09	61,571	37.35	118,311	71.77	84
education_essential	education	392,361	2.59	216,007	55.05	326,848	83.30	27
education_other	education	27,943	0.18	2,520	9.02	10,811	38.69	36
shopping_essential	shopping	227,820	1.50	43,393	19.05	173,233	76.04	42
shopping_other	shopping	176,718	1.17	5,459	3.09	106,535	60.29	1,239
wholesale	shopping	9,994	0.07	5,094	50.97	6,185	61.89	18
facilities	public_service_utility	173,669	1.15	50,344	28.99	170,252	98.03	37
utilities_other	public_service_utility	109,465	0.72	18,261	16.68	108,419	99.04	48
public_service	public_service_utility	58,369	0.39	18,571	31.82	40,192	68.86	32
health_essential	health	247,386	1.63	33,318	13.47	95,917	38.77	233
health_specialized	health	8,295	0.05	38	0.46	538	6.49	51
health_other	health	5,722	0.04	125	2.18	1,037	18.12	68
sport	sport	181,758	1.20	123,156	67.76	158,453	87.18	242
industrial	mining_industrial	138,407	0.91	62,032	44.82	121,246	87.60	156
mining	mining_industrial	30,209	0.20	22,378	74.08	29,824	98.73	10
SEZ	mining_industrial	387	0.00	0	0.00	0	0.00	6
religion	religion	161,716	1.07	59,682	36.91	116,015	71.74	84
food	food	160,645	1.06	7,602	4.73	74,966	46.67	141
financial	financial	151,493	1.00	3,289	2.17	42,018	27.74	37
accommodation	accommodation	123,248	0.81	19,853	16.11	61,228	49.68	34
parks_and_nature	tourism_recreation	66,925	0.44	43,918	65.62	57,798	86.36	32
tours_and_sightseeing	tourism_recreation	40,796	0.27	1,698	4.16	26,200	64.22	42
museums	tourism_recreation	5,096	0.03	812	15.93	1,805	35.42	28
beaches_and_resorts	tourism_recreation	4,856	0.03	171	3.52	312	6.43	8
outdoor_activities	tourism_recreation	3,110	0.02	157	5.05	1,033	33.22	53
commercial	commercial	109,580	0.72	97,251	88.75	108,351	98.88	11
historic	historic	100,738	0.66	27,458	27.26	72,683	72.15	94
beauty	beauty	65,711	0.43	769	1.17	27,305	41.55	29
professional_services	services	30,948	0.20	413	1.33	2,427	7.84	65
home_services	services	27,575	0.18	391	1.42	5,466	19.82	78
business_services	services	6,168	0.04	56	0.91	382	6.19	31
institutional	institutional	62,313	0.41	9,974	16.01	44,450	71.33	65
drinking	entertainment	35,314	0.23	2,256	6.39	23,858	67.56	18
performing_arts	entertainment	11,584	0.08	1,415	12.22	2,785	24.04	19
nightlife	entertainment	7,865	0.05	302	3.84	2,011	25.57	14
gaming	entertainment	1,425	0.01	65	4.56	250	17.54	13
storage	storage	28,077	0.19	22,761	81.07	27,757	98.86	17
military	military_emergency	13,104	0.09	9,953	75.95	12,441	94.94	41
emergency	military_emergency	3,348	0.02	75	2.24	3,200	95.58	39
port	port	244	0.00	0	0.00	0	0.00	9
total	total	15,151,918	100.00	9,038,206	59.65	12,221,198	80.66	35

*Notes:* Table summarizes places of interest (POIs) data collected from different sources (Table 1), classified into 47 economic categories and deduplicated by category and location (as described above). Column 'count' records the number of POIs, 'perc' the percentage in total POIs, 'polygons' the number of tagged OSM polygons (buildings) interpreted as POIs, 'poly\_perc' the percentage of POIs that are polygons, 'osm' the number of POIs taken from OSM, 'osm\_perc' the fraction of POIs from OSM, and 'src\_cat' the number of primary categories (across sources) mapping to a specific detailed category (e.g., there are 1,239 different kinds of small/specialized shops mapping to the 'shopping\_other' category).

## Descriptive Statistics

Table A4: Dataset of Spatial Predictors (Simplified Categories) over 96km<sup>2</sup> Grid, N = 160,499

#	Variable	Ndist	Mean	SD	Min	25%	50%	75%	Max
<i>POIs: simple counts combined across detailed categories</i>									
1	accommodation	224	0.8	9.8	0.0	0.0	0.0	0.0	1399.0
2	automotive	307	1.0	17.7	0.0	0.0	0.0	0.0	2114.0
3	beauty	207	0.4	10.3	0.0	0.0	0.0	0.0	1382.0
4	commercial	205	0.7	34.7	0.0	0.0	0.0	0.0	8264.0
5	construction	536	3.6	86.6	0.0	0.0	0.0	0.0	13209.0
6	farming	770	7.7	49.8	0.0	0.0	0.0	1.0	2581.0
7	financial	276	0.9	18.1	0.0	0.0	0.0	0.0	2208.0
8	food	316	1.0	19.0	0.0	0.0	0.0	0.0	1696.0
9	historic	183	0.6	15.6	0.0	0.0	0.0	0.0	5580.0
10	institutional	169	0.4	6.6	0.0	0.0	0.0	0.0	794.0
11	port	4	0.0	0.0	0.0	0.0	0.0	0.0	3.0
12	power	2435	115.9	319.9	0.0	0.0	0.0	20.0	14536.0
13	religion	248	1.0	11.0	0.0	0.0	0.0	0.0	1079.0
14	residential	1739	39.1	488.8	0.0	0.0	2.0	10.0	65378.0
15	sport	280	1.1	31.9	0.0	0.0	0.0	0.0	5112.0
16	storage	133	0.2	5.9	0.0	0.0	0.0	0.0	765.0
17	transport_other	383	1.9	32.3	0.0	0.0	0.0	0.0	2999.0
18	communications	1194	14.9	198.7	0.0	0.0	0.0	0.0	22867.0
19	education	363	2.6	20.9	0.0	0.0	0.0	0.0	2414.0
20	health	332	1.9	22.9	0.0	0.0	0.0	1.0	3521.0
21	entertainment	420	1.7	32.4	0.0	0.0	0.0	0.0	3944.0
22	services	197	0.4	9.4	0.0	0.0	0.0	0.0	1573.0
23	shopping	698	5.5	126.8	0.0	0.0	0.0	0.0	21543.0
24	public_service_utility	140	0.4	5.7	0.0	0.0	0.0	0.0	952.0
25	mining_industrial	272	1.2	11.5	0.0	0.0	0.0	0.0	1027.0
26	military_emergency	75	0.1	1.7	0.0	0.0	0.0	0.0	249.0
27	tourism_recreation	288	1.0	15.8	0.0	0.0	0.0	0.0	2441.0
<i>Lines: total length in km in each cell</i>									
28	dam_len	1868	29.5	431.6	0.0	0.0	0.0	0.0	79889.0
29	aeroway_len	2607	66.6	805.7	0.0	0.0	0.0	0.0	68671.0
30	ferry_len	1362	57.4	958.9	0.0	0.0	0.0	0.0	95065.0
31	pipeline_len	4031	281.0	2675.5	0.0	0.0	0.0	0.0	331527.0
32	railway_len	7707	776.7	4975.3	0.0	0.0	0.0	0.0	706405.0
33	roads_paved	18393	2825.1	9028.5	0.0	0.0	0.0	0.0	465351.0
34	roads_unpaved	27757	6361.4	10741.4	0.0	0.0	0.0	10768.0	625255.0
35	waterway_len	36356	23069.8	150937.0	0.0	0.0	0.0	13141.5	11753756.0
<i>Raster Layers: sum of population and mean of travel time and internet speed (bytes/s) in each cell</i>									
36	internet_speed	53324	20800.5	22664.3	151.0	6209.0	12031.0	27115.0	278876.0
37	pop_gpw4_ages_0_14	14147	2323.1	9063.3	0.0	192.0	716.0	2013.0	729958.0
38	pop_gpw4_ages_15_49	15038	2786.9	14553.7	0.0	207.0	726.0	2071.0	1503562.0
39	ttime_city_50k	3380	281.8	457.2	0.0	76.0	156.0	312.0	9303.0
40	ttime_city_1m	4333	610.5	631.5	0.0	245.0	451.0	779.0	9868.0
41	ttime_port_any	4723	811.5	758.8	0.0	316.0	602.0	1054.0	10065.0
42	ttime_port_ml	5229	1093.5	909.3	1.0	453.0	849.0	1472.0	10216.0

Figure A1: Histogram of Wealth/Activity Measures in  $96\text{km}^2$  Grid, Avg. Obs.: 101,417Table A5: Pearson's Correlations of Wealth/Activity Measures in  $96\text{km}^2$  Grid

		log_GDP_PPP	log_avg_rad	IWI	RWI
Pooled Across Countries	log_GDP_PPP	1 (126123)			
	log_avg_rad	.453* (33293)	1 (36656)		
	IWI	.397* (98948)	.576* (25608)	1 (105924)	
	RWI	.449* (95295)	.577* (31858)	.550* (88101)	1 (101456)
Scaled and Centered by Country	log_GDP_PPP	1 (126123)			
	log_avg_rad	.436* (33293)	1 (36656)		
	IWI	.363* (98948)	.578* (25608)	1 (105924)	
	RWI	.349* (95295)	.560* (31858)	.533* (88101)	1 (101456)

Notes: N. obs in parentheses, a '\*' denotes significance at the 5% level.

Table A6: Detailed Category Combination Weights ( $\beta_{jp}$ ) for 3 Categories

<b>communications</b>	communications_network	communications_other	telecom_len
IWI	1.00	5.70	0.24
RWI	1.00	13.65	0.10
GDP_PPP	1.00	8.86	0.02
avg_rad	1.00	18.65	0.03
<b>Average</b>	<b>1.00</b>	<b>11.72</b>	<b>0.098</b>
<b>mining_industrial</b>	industrial	mining	SEZ
IWI	1.00	3.43	89.13
RWI	1.00	0.93	66.93
GDP_PPP	1.00	0.00	52.49
avg_rad	1.00	0.86	23.97
<b>Average</b>	<b>1.00</b>	<b>1.31</b>	<b>58.13</b>
<b>shopping</b>	shopping_essential	shopping_other	wholesale
IWI	1.00	1.46	15.41
RWI	1.00	1.46	14.73
GDP_PPP	1.00	3.82	52.37
avg_rad	1.00	2.57	18.71
<b>Average</b>	<b>1.00</b>	<b>2.33</b>	<b>25.31</b>

*Notes:* The weights are normalized, then averaged across outcomes.

Table A7: Correlations of Total Infrastructure with Wealth/Activity and Population

	IWI	RWI	GDP_PPP	avg_rad	pop_gpw4	pop_wpop
<i>Cell-level, N = 88,961 populated cells (&gt;10 persons/km<sup>2</sup>)</i>						
Total PP Feature Count	0.302	0.401	0.481	0.484	0.546	0.562
Total PP Features ERFPC	0.249	0.351	0.586	0.590	0.603	0.617
Total Line Length	0.519	0.578	0.477	0.623	0.491	0.515
Total Line Length EW	0.558	0.588	0.482	0.640	0.490	0.515
<i>Country-level, N = 55</i>						
Total PP Feature Count	-0.009	-0.263	0.538	-0.055	0.800	0.799
Total PP Features ERFPC	-0.010	-0.265	0.535	-0.057	0.799	0.799
Total Line Length	0.412	-0.040	0.730	0.265	0.719	0.721
Total Line Length EW	0.530	0.049	0.769	0.332	0.671	0.667

ERFPC = excluding residential buildings, farmland, power and construction

EW = excluding waterways

Figure A2: Clustered Correlations of Variables

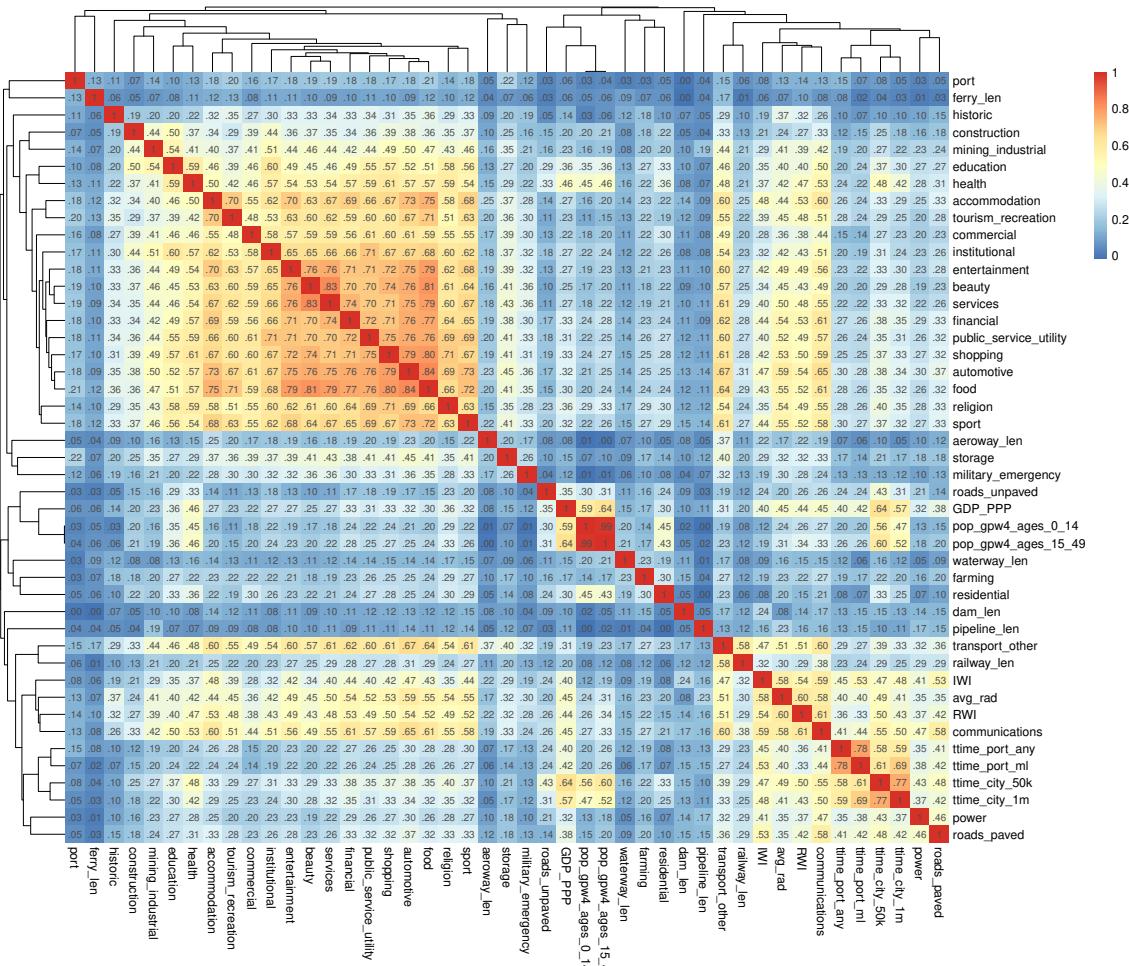


Figure A3: Index of Spatial Efficiency

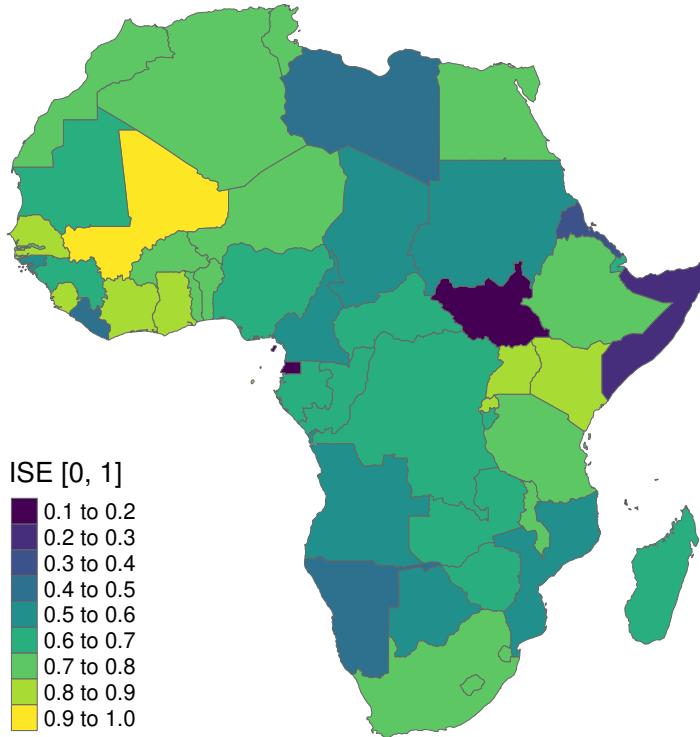
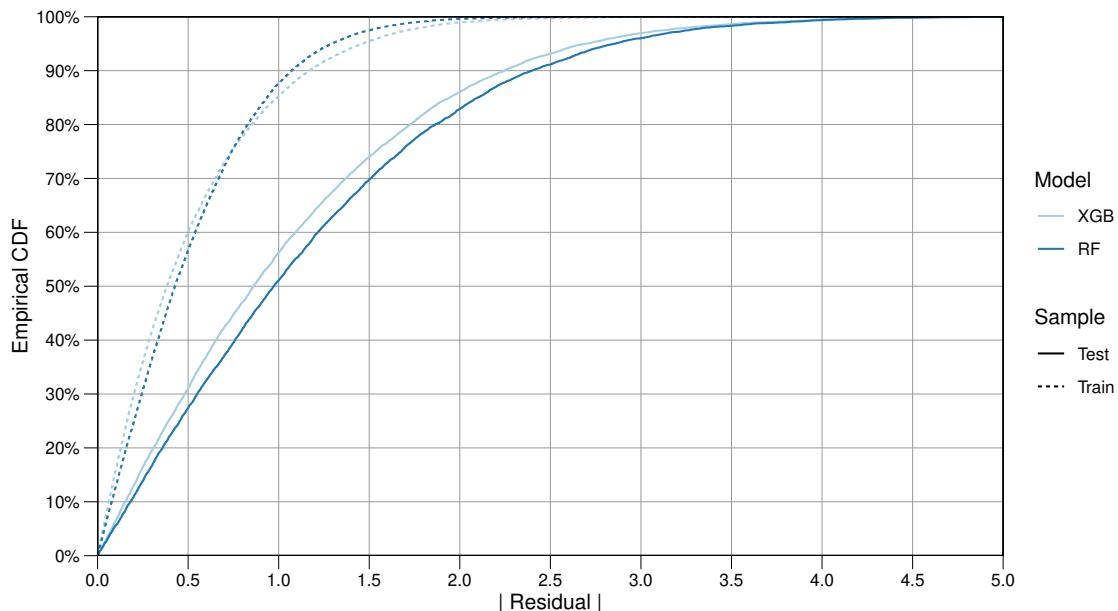


Figure A4: Empirical CDF's of Residuals from ML Models Predicting the Log of Nightlights 2022



*Notes:* Dataset includes 42 predictors, excluding spillover variables. The training set has 28,976 observations, the test set 9,655. The training set  $R^2$  is 91.8% for XGB and 92.7% for RF. The test set  $R^2$  is 68.4% for XGB and 63.6% for RF.

## SHAP Values for Other Outcomes

Figure A5: SHAP Variable Importance: Log Nightlights 2022

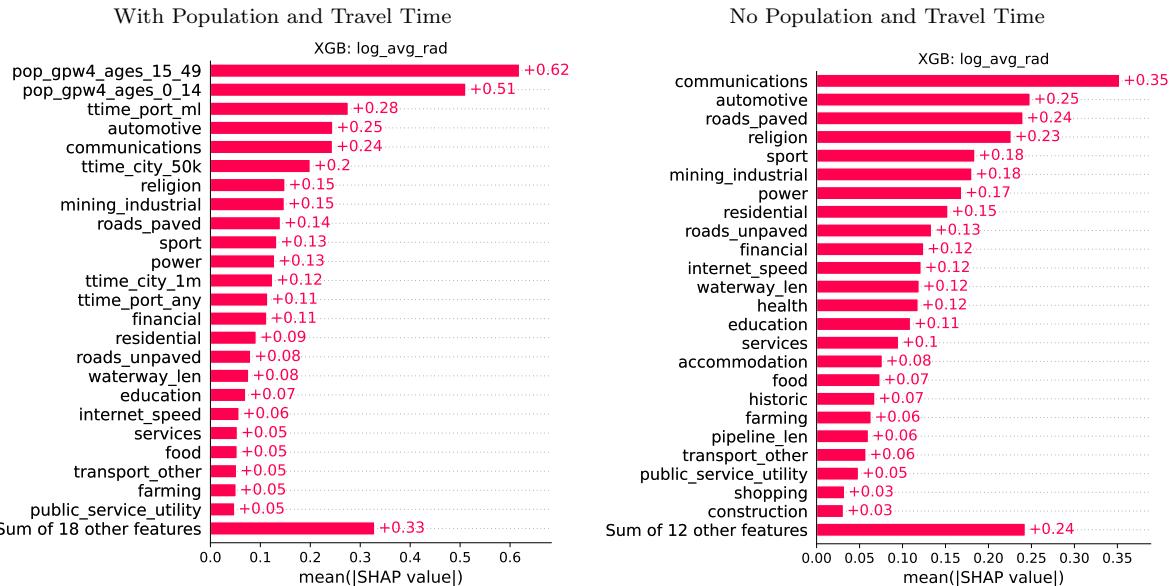
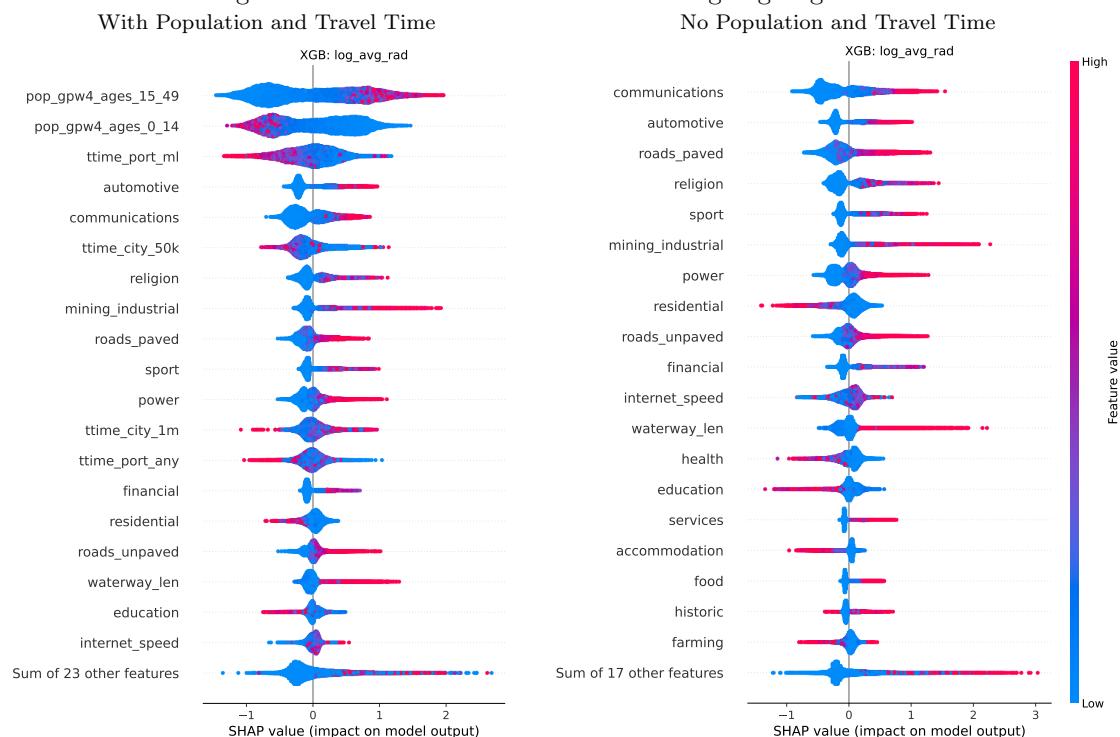


Figure A6: SHAP Effect Distribution: Log Nightlights 2022



### SHAP Values from Feature Counts

Figure A7: SHAP Variable Importance: Simple Feature Counts  
IWI, no POPTT      Log Nightlights, no POPTT

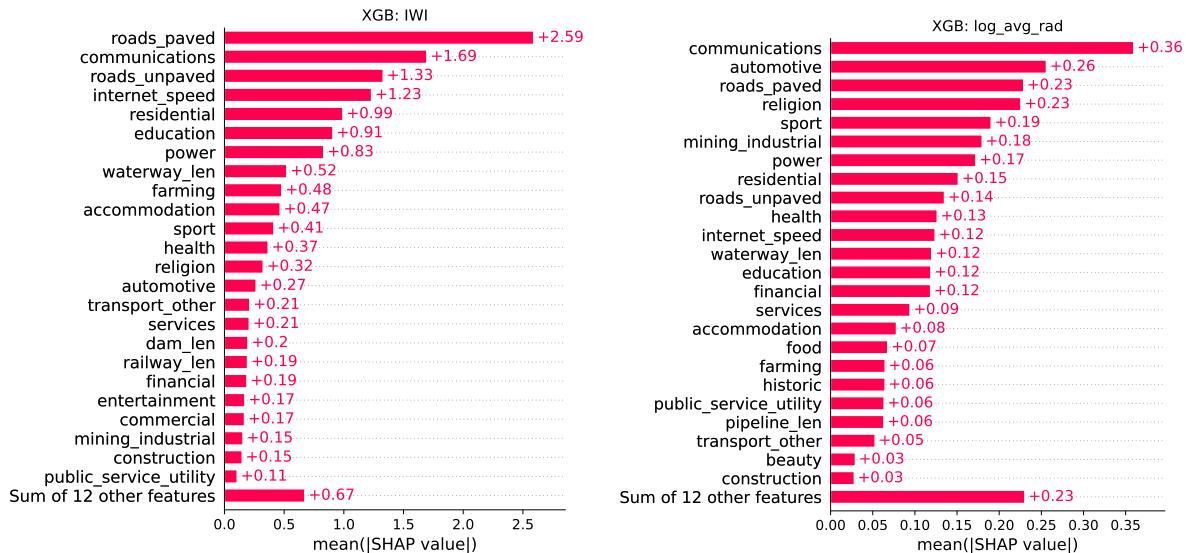
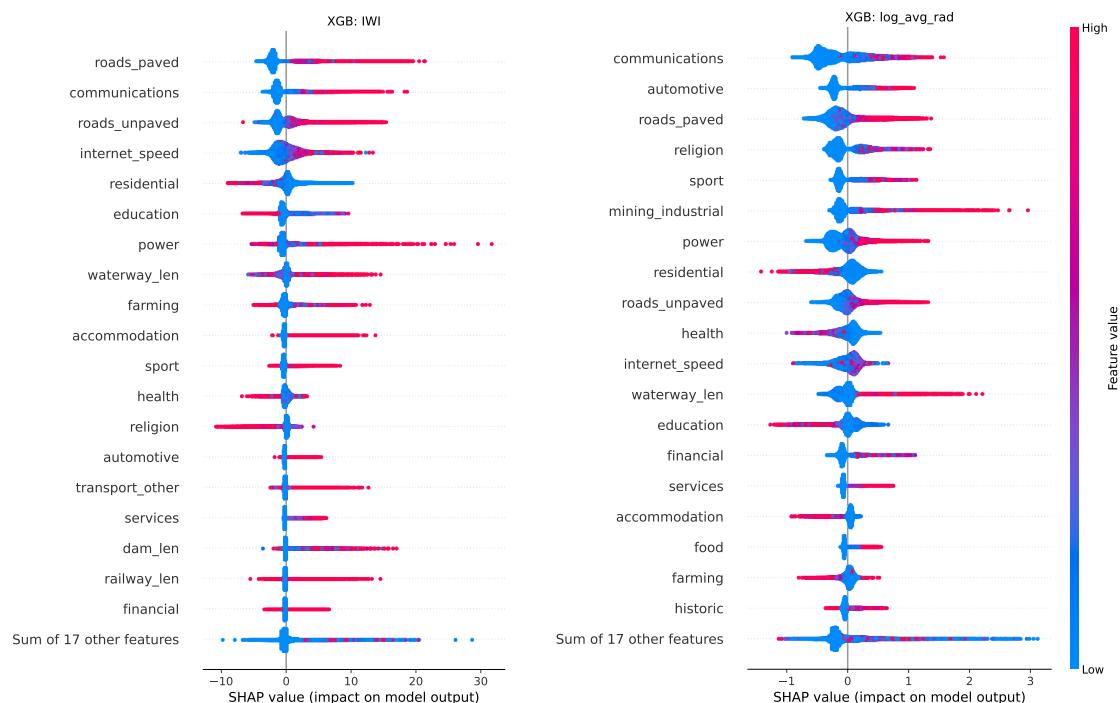


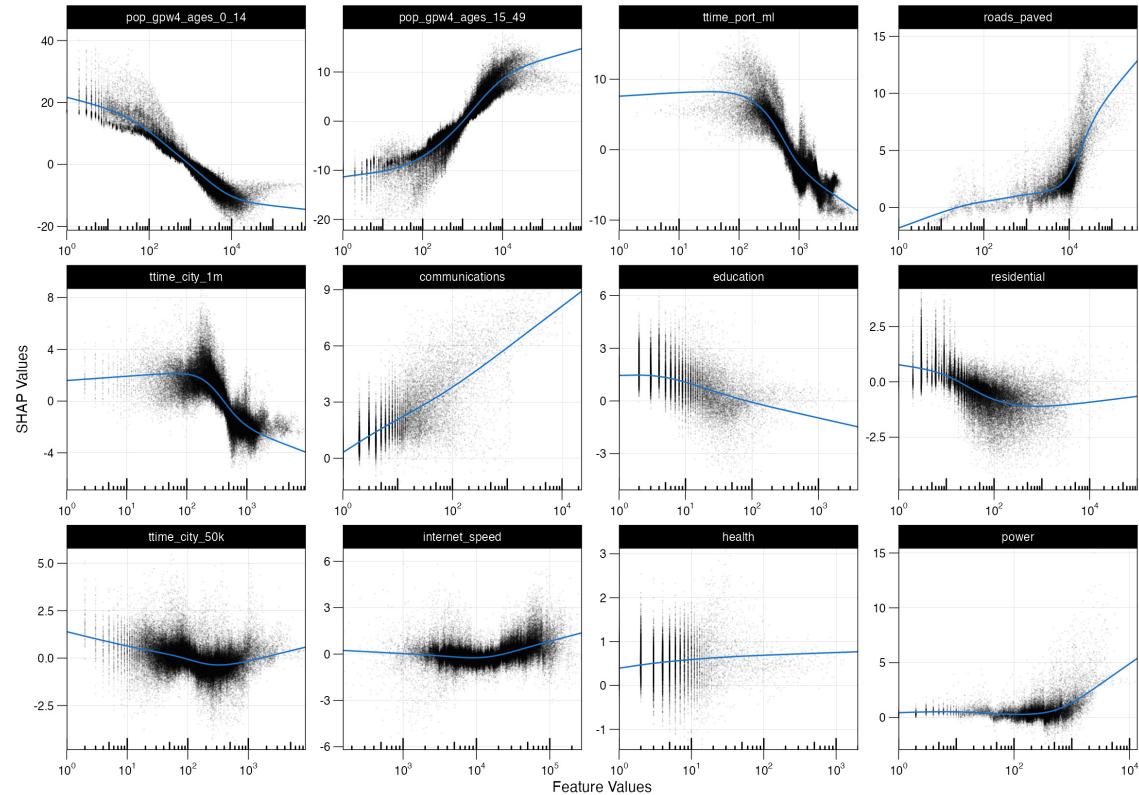
Figure A8: SHAP Effect Distribution: Simple Feature Counts  
IWI, no POPTT      Log Nightlights, no POPTT



### SHAP Values Scatter (ALE) Plots

Figure A9: IWI: SHAP Values and Feature Levels - Top 12 Predictors

With Population and Travel Time



No Population and Travel Time

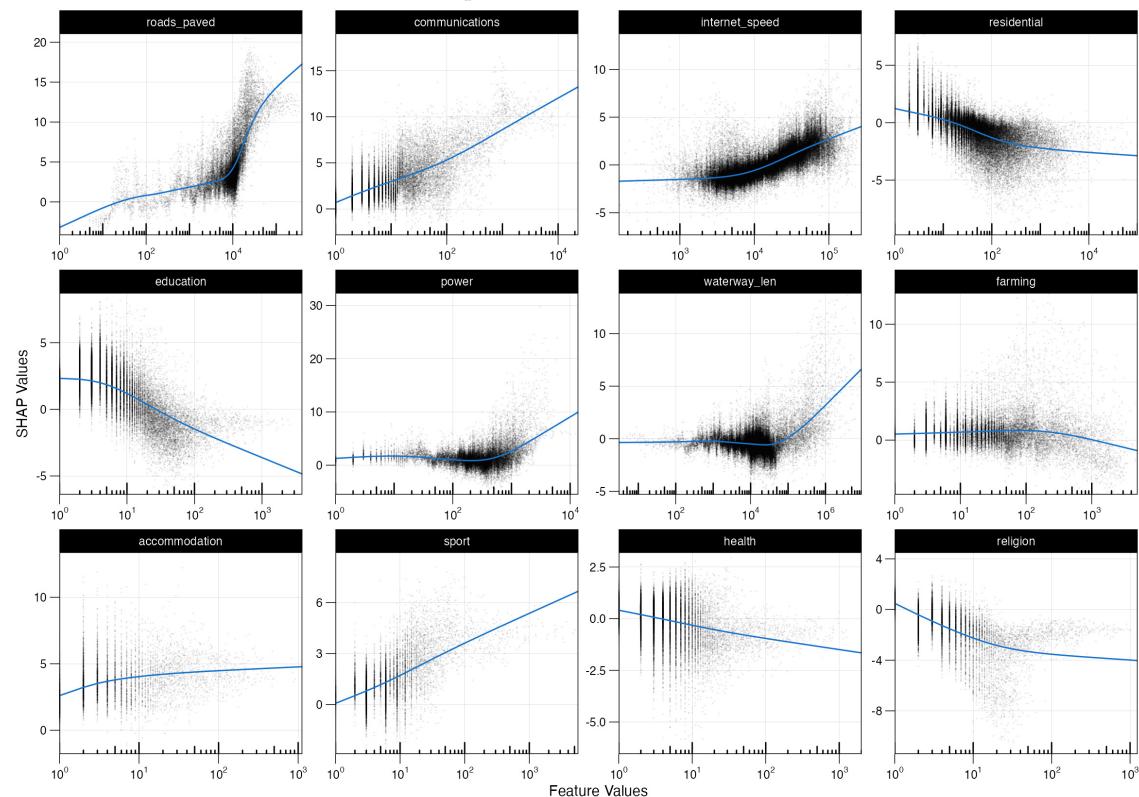
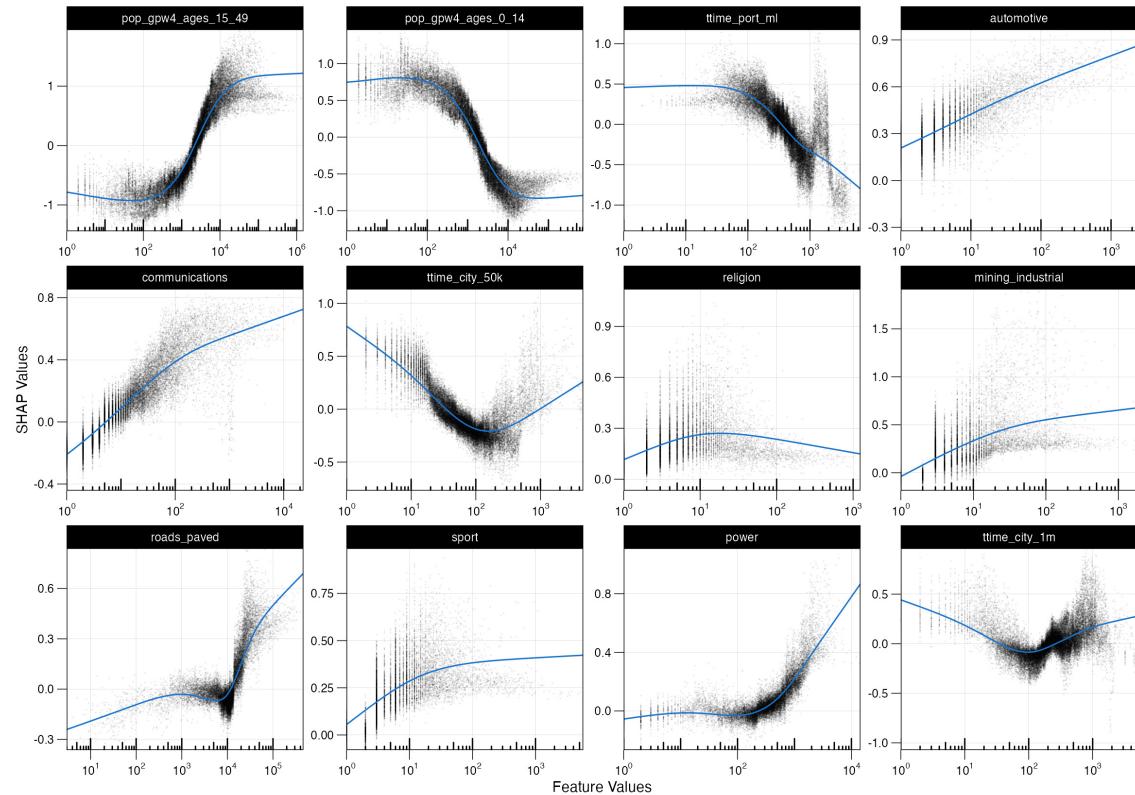
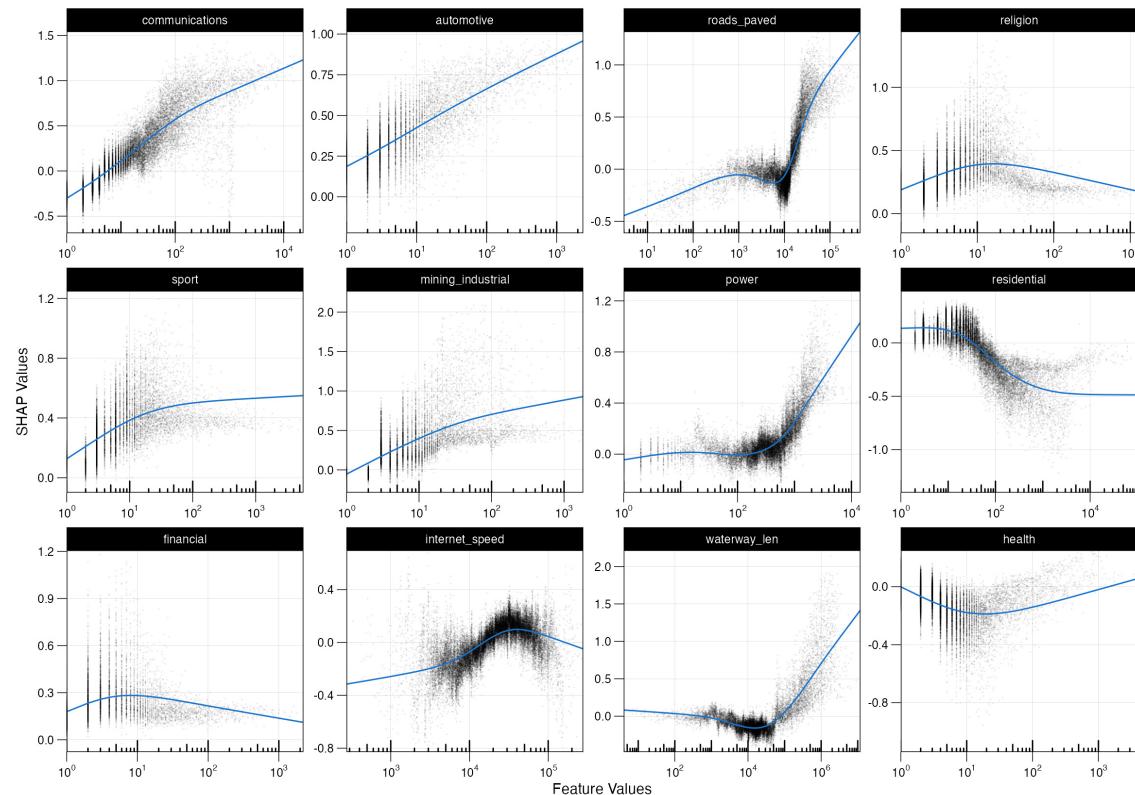


Figure A10: Log Nightlights 2022: SHAP Values and Feature Levels - Top 12 Predictors

With Population and Travel Time



No Population and Travel Time



## Robustness Checks and Additional Results

Figure A11: CAPE Estimates for Power and Health using Only OSM Data (February 2024 Version)

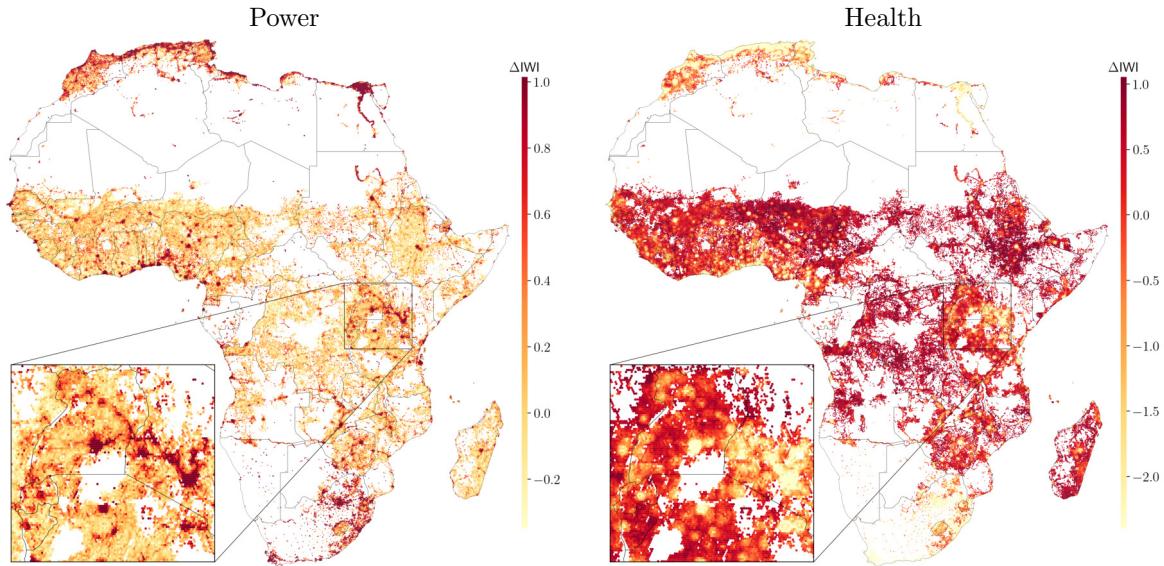


Table A8: ML Debiased Infrastructure and Favouritism: Simple Feature Counts

feature	Colonial Railroads		Political Favouritism				Ethnic Favouritism			
	Ln Kilometers Raw	Ctrl	Ln Years in Power Raw	Ctrl	Ever in Power Raw	Ctrl	Discriminated Raw	Ctrl	Excl. from Gov. Raw	Ctrl
<i>Raw Infrastructure Data, in Natural Logs</i>										
roads.paved	.517***	.29***	.521***	.02	1.506***	.355	-.164	.225	-.674***	-.061
power	.374***	.287***	.417***	.066	1.024***	.224	.571**	.304	-.186	.338*
education	.097***	.041***	.149***	-.02	.369***	-.005	.052	.038	-.29***	.001
health	.076***	.029***	.122***	-.048**	.324***	-.076	.048	-.05	-.128***	.037
communications	.212***	.115***	.236***	-.013	.622***	.078	.106	.118	-.229***	.133
public_service.utility	.037***	.02***	.063***	-.016	.161***	-.018	.01	.006	-.018	.037*
automotive	.062***	.036***	.08***	-.023	.203***	-.031	.045	.054	-.021	.065**
transport_other	.142***	.102***	.114***	-.021	.292***	-.013	.06	.071	-.031	.027
financial	.057***	.025***	.067***	-.049***	.169***	-.091**	.066**	.031	-.002	.009
services	.039***	.015***	.045***	-.034***	.104***	-.074***	.017	.03*	-.003	.028*
ttime.city.1m	-.162***	-.088***	-.189***	.051*	-.477***	.076	.098	.108	.316***	.066
ttime_port.any	-.121***	-.069***	-.21***	-.017	-.585***	-.104*	-.003	-.004	.366***	-.128***
residential	.032**	.052***	.167***	-.075	.444***	-.129	-.085	-.007	-.257**	.019
accommodation	.074***	.036***	.081***	-.013	.24***	.019	.07*	.07*	-.025	.008
<i>Debiased Data via Ensemble ML Models</i>										
roads.paved	.015	.012	-.031	-.067	.025	-.063	-.034	.191	.006	-.186
power	0	.003	.023	.028	.021	.024	.205**	.186	.048	.051
education	0	-.003	-.014	-.011	-.048	-.031	-.011	.002	0	-.007
health	-.005*	-.007**	-.005	-.007	-.005	-.014	-.014	-.046*	-.005	-.021
communications	-.007*	-.007	-.001	.016	-.005	.043	.018	-.011	-.032	-.039
public_service.utility	-.003*	-.001	.005	.006	.012	.015	-.003	-.009	.006	.022*
automotive	-.002	0	-.002	0	-.021	-.013	.018*	.019	-.005	.002
transport_other	.003	.003	-.004	-.001	-.02	-.012	.042***	.017	.021*	-.005
financial	-.002	-.001	-.012*	-.016**	-.024	-.034*	.018*	.02	.002	-.011
services	.001	-.002*	0	0	.001	.003	.002	.008	.001	.008
ttime.city.1m	-.005	-.002	.002	.002	-.025	-.024	.011	.015	.012	.046
ttime_port.any	-.014***	-.006*	-.033**	-.015	-.125***	-.081**	.008	.046	.034	-.003
residential	-.002	-.002	-.011	-.044	-.053	-.108	-.039	-.035	-.027	-.008
accommodation	.005***	.004*	0	0	.008	.007	.007	.027	-.003	-.01
Observations	5346	5346	5346	5346	5346	5346	385	385	385	385
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Geographic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Simulation Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1; based on cluster-robust standard-errors

Notes: See Graff (2024) Tables 1 and 2 for further details about the variables and controls.

Table A9: ML Debiased Infrastructure and Favouritism: Quantile Feature Counts

feature	Colonial Railroads		Political Favouritism				Ethnic Favouritism			
	Raw	Ctrl	Ln Kilometers	Ln Years in Power	Ever in Power	Discriminated	Excl. from Gov.	Raw	Ctrl	
<i>Raw Infrastructure Data, in Natural Logs</i>										
roads_paved	.517***	.29***	.521***	.02	1.506***	.355	-.164	.185	-.674***	-.067
power	.373***	.286***	.416***	.065	1.022***	.222	.571**	.273	-.186	.31*
education	.111***	.046***	.186***	-.012	.465***	.025	.047	.014	-.341***	.034
health	.079***	.031***	.137***	-.047*	.375***	-.063	.04	-.083	-.147***	.023
communications	.212***	.115***	.237***	-.012	.624***	.078	.106	.072	-.229***	.086
public_service_utility	.045***	.026***	.079***	-.011	.213***	.01	.037	.005	-.026	.029
automotive	.069***	.04***	.093***	-.022	.238***	-.022	.062*	.069*	-.02	.07**
transport_other	.146***	.105***	.119***	-.02	.306***	-.009	.075	.074	-.027	.048
financial	.058***	.026***	.067***	-.049***	.171***	-.091**	.064**	.028	-.005	.007
services	.039***	.015***	.045***	-.035***	.103***	-.075***	.017	.027	-.004	.024
ttime_city_1m	-.162***	-.088***	-.189***	.051*	-.477***	.076	.098	.116	.316***	.089
ttime_port_any	-.121***	-.069***	-.21***	-.017	-.585***	-.104*	-.003	-.002	.366***	-.124***
residential	.01	.049***	.153**	-.102	.413**	-.195	-.094	-.058	-.176	.021
accommodation	.076***	.037***	.088***	-.01	.261***	.031	.07*	.06	-.03	.012
mining_industrial	.073***	.045***	.096***	-.022	.219***	-.054	-.077	.009	-.151***	.17***
tourism_recreation	.044***	.027***	.086***	-.012	.241***	.009	0	.022	-.031	.014
construction	.056***	.027***	.114***	-.012	.286***	-.004	-.056	.085	-.359***	.107
<i>Debiased Data via Ensemble ML Models</i>										
roads_paved	.018	.015	-.025	-.06	.027	-.059	-.057	.161	.002	-.159
power	-.005	-.002	.041	.043	.076	.07	.246***	.202	.073	.087
education	.001	-.003	-.019	-.018	-.068	-.057	-.026	-.029	-.009	-.024
health	-.006*	-.006*	-.012	-.015	-.02	-.033	-.019	-.068**	.001	-.019
communications	-.01***	-.009**	-.004	.013	-.016	.03	.035	.012	-.033	-.053*
public_service_utility	-.003	-.001	.011	.012	.035*	.04*	.007	-.001	-.004	.002
automotive	-.003	-.001	.001	.004	-.016	-.008	.026**	.028*	-.001	.003
transport_other	.003	.003	-.004	0	-.023	-.015	.052***	.021	.029**	.006
financial	-.002	-.001	-.009	-.014**	-.02	-.03*	.02**	.02	.003	-.009
services	0	-.002*	-.001	0	-.002	.001	-.002	.004	0	.007
ttime_city_1m	-.006**	-.004	-.002	-.003	-.033	-.034	.007	.002	.009	.053*
ttime_port_any	-.015***	-.006*	-.033**	-.017	-.127***	-.082**	-.002	.048	.032	.001
residential	-.003	-.001	-.021	-.058	-.068	-.134	-.034	-.091	-.016	-.062
accommodation	.005**	.004	.002	.004	.014	.017	.004	.022	-.004	-.004
mining_industrial	-.003	-.002	.003	.004	-.023	-.023	-.045*	-.04	.01	.031
tourism_recreation	-.004*	-.001	0	-.002	.005	-.005	-.018	-.009	-.011	-.022
construction	-.006	-.003	-.005	.006	-.023	.001	.044	.117*	-.007	.126**
Observations	5346	5346	5346	5346	5346	5346	385	385	385	385
Country FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Geographic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Simulation Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1; based on cluster-robust standard-errors

Notes: See Graff (2024) Tables 1 and 2 for further details about the variables and controls.

Table A10: Median CAPE: DHS IWI vs. IWI Prediction by [Lee & Braithwaite \(2022\)](#)

Feature	Simple Counts		Quantile Counts	
	DHS	L&B	DHS	L&B
roads_paved	0.248	0.231	0.179	0.225
power	0.197	0.068	0.221	0.071
education	0.009	0.509	-0.032	0.347
health	0.932	0.724	1.033	0.661
communications	0.798	0.655	0.571	0.624
public_service_utility	0.770	0.214	0.149	0.381
automotive	0.615	0.580	0.437	0.468
transport_other	0.593	0.300	0.686	0.296
financial	3.054	0.992	2.302	0.891
services	1.766	0.412	0.651	0.262
ttime_city_1m	-1.285	-0.566	-1.323	-0.521
ttime_port_any	-0.965	-0.215	-1.003	-0.229
residential	-0.068	-0.047	0.007	-0.025
accommodation	1.507	0.802	1.356	0.654
mining_industrial	0.841	0.509	0.497	0.544
tourism_recreation	-0.061	0.149	-0.185	0.198
construction	0.246	0.153	0.093	0.180
Median Abs. Coef.	0.770	0.412	0.497	0.347
Corr. Spearman		0.865		0.809
Corr. Pearson		0.857		0.902

*Abbreviations:* DHS = IWI estimate from Demographic and Health Survey; L&B = Predicted IWI estimate by [Lee & Braithwaite \(2022\)](#).

*Notes:* the DHS-based IWI estimate is computed from all DHS surveys conducted in SSA since 2010. It is averaged across  $96\text{km}^2$  hexagonal grid cells just like the IWI estimate of [Lee & Braithwaite \(2022\)](#). Grid cells with less than 5 households or less than 10 people per  $\text{km}^2$  population density are excluded. This yields 17,396 cells with DHS-based IWI estimates used for training, versus 89,048 cells available with the predicted IWI by [Lee & Braithwaite \(2022\)](#). After the ensemble CAPE model is trained, a CAPE prediction is made for 103,922 cells which have a population density above 10 persons/ $\text{km}^2$  and any POI feature. The median of these CAPE estimates is reported in this table.

Table A11: Average Partial Effect: DHS IWI vs. IWI Prediction by [Lee & Braithwaite \(2022\)](#)

Feature	Simple Counts		Quantile Counts	
	DHS	L&B	DHS	L&B
roads_paved	0.202	0.218	0.173	0.216
power	0.087	0.068	0.114	0.068
education	0.031	0.563	-0.073	0.436
health	0.585	0.803	0.645	0.772
communications	1.037	0.689	1.031	0.689
public_service_utility	0.842	0.327	0.470	0.331
automotive	1.482	0.873	0.952	0.444
transport_other	0.382	0.332	0.324	0.305
financial	1.857	0.626	1.641	0.543
services	1.911	1.380	4.506	0.748
ttime_city_1m	-1.158	-0.513	-1.174	-0.502
ttime_port_any	-1.014	-0.293	-0.937	-0.295
residential	-0.097	-0.068	-0.101	-0.052
accommodation	1.518	0.807	1.073	0.771
mining_industrial	0.740	0.541	0.508	0.460
tourism_recreation	0.137	0.184	0.094	0.309
construction	0.211	0.161	0.154	0.186
Median Abs. Coef.	0.740	0.513	0.508	0.436
Corr. Spearman		0.863		0.870
Corr. Pearson		0.895		0.874

*Abbreviations:* DHS = IWI estimate from Demographic and Health Survey; L&B = Predicted IWI estimate by [Lee & Braithwaite \(2022\)](#).

*Notes:* the DHS-based IWI estimate is computed from all DHS surveys conducted in SSA since 2010. It is averaged across  $96\text{km}^2$  hexagonal grid cells just like the IWI estimate of [Lee & Braithwaite \(2022\)](#). Grid cells with less than 5 households or less than 10 people per  $\text{km}^2$  population density are excluded. This yields 17,396 cells with DHS-based IWI estimates used for training, versus 89,048 cells available with the predicted IWI by [Lee & Braithwaite \(2022\)](#). After the ensemble CAPE model is trained, a CAPE prediction is made for 103,922 cells which have a population density above 10 persons/ $\text{km}^2$  and any POI feature. The APE is obtained through augmented inverse probability weighting following Eq. 10, and thus only considers cells where outcome data is available. Differences in the estimates are thus expected because with L&B much more cells are available for APE calculation.

Table A12: CAPE Correlations: DHS IWI and IWI Prediction by [Lee & Braithwaite \(2022\)](#)

Feature	Simple Counts	Quantile Counts
roads_paved	0.593	0.119
power	0.157	0.040
education	0.317	0.024
health	0.013	-0.038
communications	0.271	-0.171
public_service_utility	0.100	0.386
automotive	0.108	0.114
transport_other	0.126	0.001
financial	0.287	0.362
services	0.000	-0.006
ttime_city_1m	0.038	-0.022
ttime_port_any	-0.118	-0.025
residential	-0.000	0.383
accommodation	0.246	0.013
mining_industrial	0.422	0.271
tourism_recreation	0.076	0.000
construction	0.488	-0.213
Median Corr.	0.126	0.0128
Median Abs. Corr.	0.126	0.040

*Notes:* the DHS-based IWI estimate is computed from all DHS surveys conducted in SSA since 2010. It is averaged across  $96\text{km}^2$  hexagonal grid cells just like the IWI estimate of [Lee & Braithwaite \(2022\)](#). Grid cells with less than 5 households or less than 10 people per  $\text{km}^2$  population density are excluded. This yields 17,396 cells with DHS-based IWI estimates used for training, versus 89,048 cells available with the predicted IWI by [Lee & Braithwaite \(2022\)](#). After the ensemble CAPE model is trained, a CAPE prediction is made for 103,922 cells which have a population density above 10 persons/ $\text{km}^2$  and any POI feature. Pearson's correlation of these CAPE estimates is reported in this table.

Table A13: CAPE Correlations: DHS IWI and IWI Prediction by [Lee & Braithwaite \(2022\)](#): Previous Estimates without Separate Spillover Variables (One Variable Per Infrastructure)

	Count	Count + SS	Tag Weights	Weights + SS
roads_paved	0.213	0.305	0.245	0.246
power	-0.337	0.249	-0.001	0.078
education	0.550	0.379	0.632	-0.338
health	0.319	0.280	0.251	0.708
communications	0.559	0.271	0.399	0.305
public_service_utility	0.292	0.542	0.802	0.267
automotive	0.301	-0.070	0.575	0.161
transport_other	-0.083	0.329	-0.031	0.181
financial	0.253	0.128	0.402	0.303
services	0.231	0.003	0.025	-0.313
ttime_city_1m	0.325	0.353	0.288	0.419
ttime_port_any	0.500	0.257	0.508	0.375
residential	0.556	0.135	0.454	-0.009
accommodation	0.305	0.481	0.325	0.260
industrial	0.132	0.196	0.727	0.510
tourism_recreation	0.158	0.069	0.374	0.373
construction	0.296	-0.153	0.037	-0.000
Median Corr.	0.294	0.244	0.284	0.195
Median Abs. Corr.	0.298	0.244	0.299	0.232

*Abbreviations:* SS = Spatial Spillovers.

Figure A12: DML CAPE Kernel Density Estimates for DHS-Based IWI

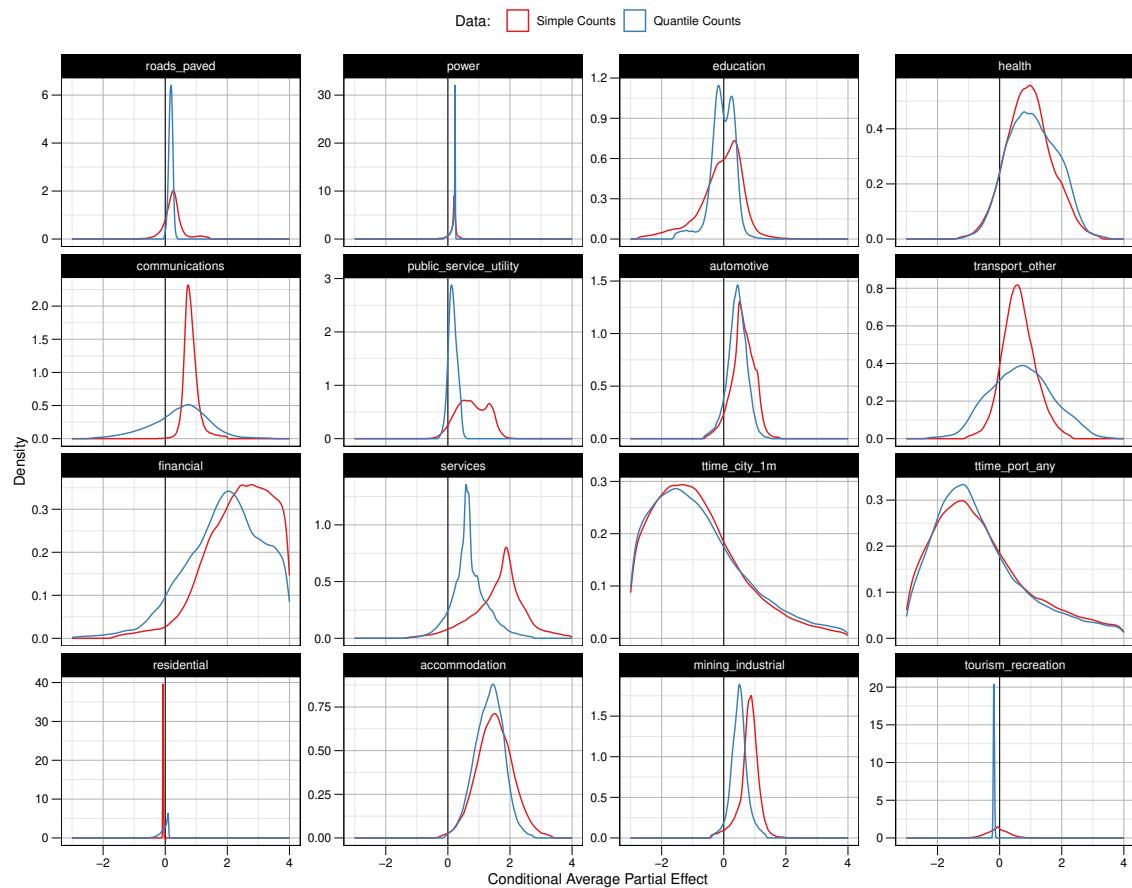


Figure A13: Top 25 Correlates of DHS-Based IWI CAPE Estimates: Average Across Datasets

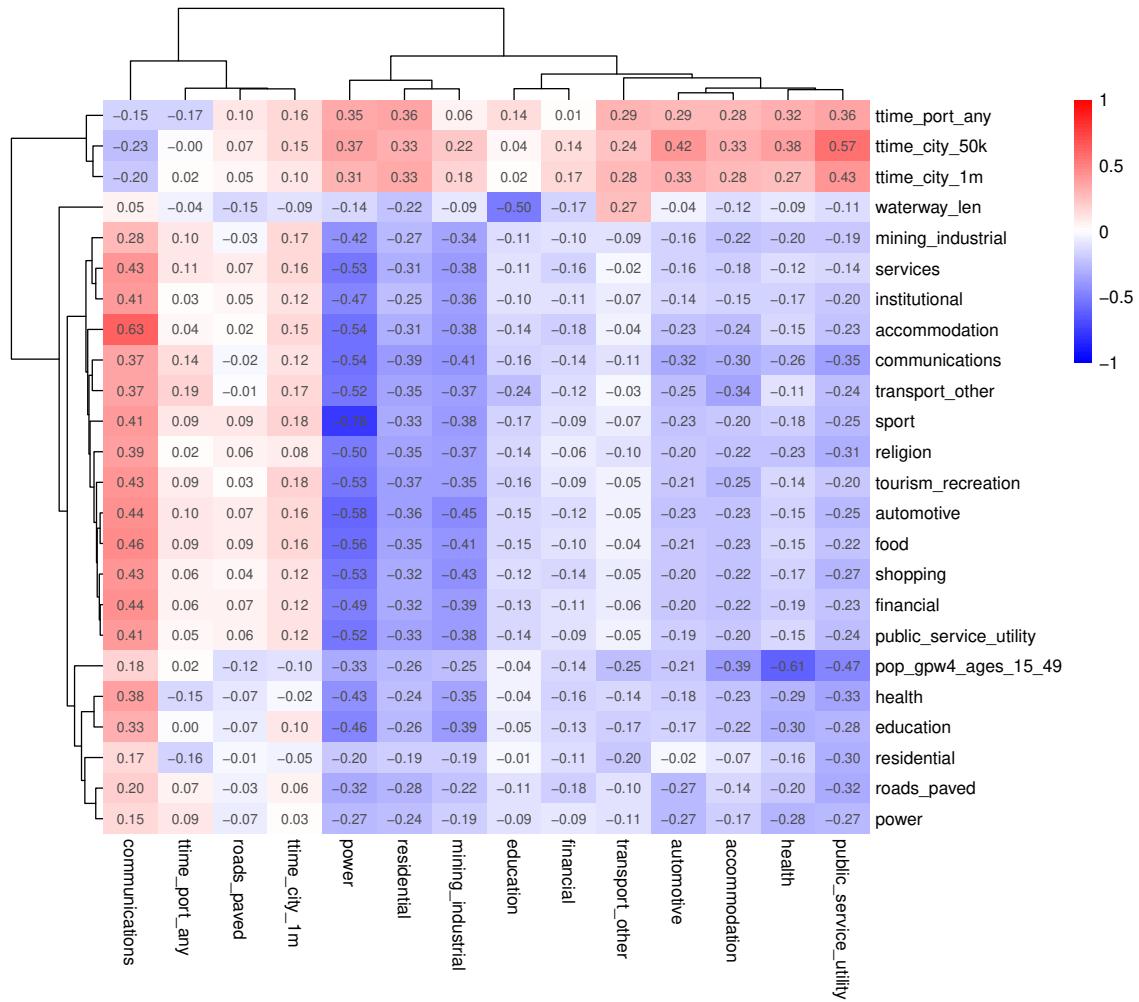


Table A14: Average Partial Effects on Log Nightlights 2022

Feature	ATE	Simple Counts			ATE	Quantile Counts			Diff.
		High	Low	Diff.		High	Low	Diff.	
roads_paved	<b>0.0198***</b>	0.0328***	0.00682**	0.026***	<b>0.0193***</b>	0.0351***	0.00358	0.0315***	
power	<b>0.0235***</b>	0.0436***	0.00338	0.0402***	<b>0.0247***</b>	0.0445***	0.00497	0.0395***	
education	<b>0.0825***</b>	0.144***	0.0209	0.123***	<b>0.0532***</b>	0.104***	0.00198	0.102***	
health	<b>0.107***</b>	0.252***	-0.0382**	0.29***	<b>0.101***</b>	0.243***	-0.0405**	0.284***	
communications	<b>0.17***</b>	0.212***	0.128***	0.0847***	<b>0.168***</b>	0.226***	0.111***	0.115***	
public_service_utility	<b>0.112***</b>	0.214***	0.00991	0.204***	<b>0.106***</b>	0.21***	0.00209	0.208***	
automotive	<b>0.247***</b>	0.313***	0.181***	0.132**	<b>0.201***</b>	0.272***	0.131***	0.141***	
transport_other	<b>0.107***</b>	0.144***	0.0696***	0.0744**	<b>0.111***</b>	0.157***	0.0646***	0.0928***	
financial	<b>0.246***</b>	0.414***	0.0787***	0.336***	<b>0.231***</b>	0.451***	0.0109	0.441***	
services	<b>0.184**</b>	0.277**	0.0908	0.186	<b>0.133**</b>	0.214**	0.0527	0.161	
ttime_city_1m	<b>-0.0285</b>	0.0928***	-0.15***	0.243***	<b>-0.0119</b>	0.129***	-0.152***	0.281***	
ttime_port_any	<b>-0.0795***</b>	0.0781***	-0.237***	0.315***	<b>-0.0764***</b>	0.0507**	-0.204***	0.254***	
residential	<b>0.0105*</b>	0.0346***	-0.0137	0.0483***	<b>0.016***</b>	0.0444***	-0.0124	0.0568***	
accommodation	<b>0.1***</b>	0.197***	0.0035	0.194***	<b>0.0877***</b>	0.206***	-0.0302	0.236***	
mining_industrial	<b>0.235***</b>	0.372***	0.097***	0.275***	<b>0.185***</b>	0.282***	0.0878***	0.195***	
tourism_recreation	<b>0.0139</b>	0.0769**	-0.0491*	0.126***	<b>-0.000934</b>	0.0523	-0.0542*	0.107**	
construction	<b>0.0505***</b>	0.0921***	0.00898	0.0831***	<b>0.0515***</b>	0.093***	0.0101	0.0829***	

Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Notes: Table shows doubly-robust APE estimates of the log feature intensity (simple counts or quantile counts in each cell, see Section 2.4) on the log of NASA nightlights 2022 (Román et al., 2018). The "High" and "Low" estimates report the APE above and below the median CAPE estimate. The "Diff." column indicates their difference to test for heterogeneity. All terms are tested using a two-sided t-test with standard errors derived from the doubly robust scores following Athey & Wager (2019).

Figure A14: DML CAPE Kernel Density Estimates for Log Nightlights 2022

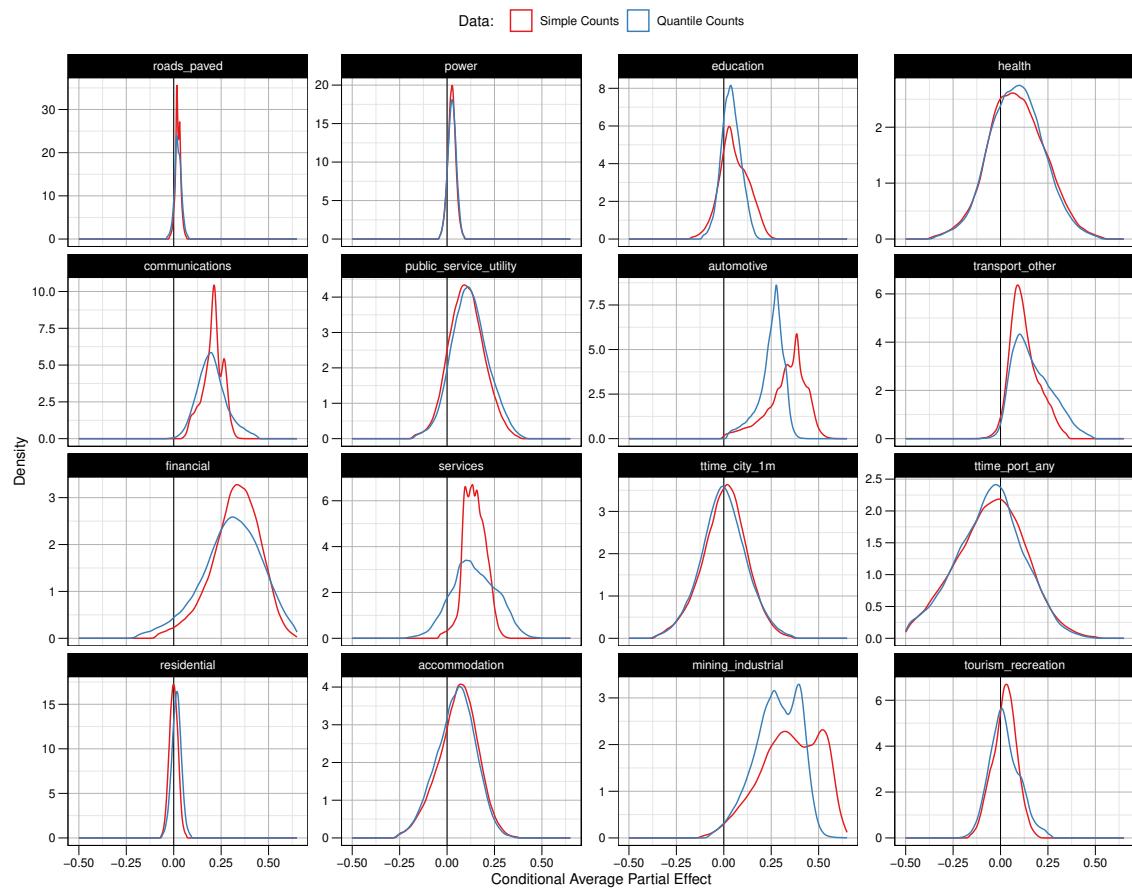


Figure A15: Top Correlates of Nightlights CAPE Estimates: Average Across Datasets

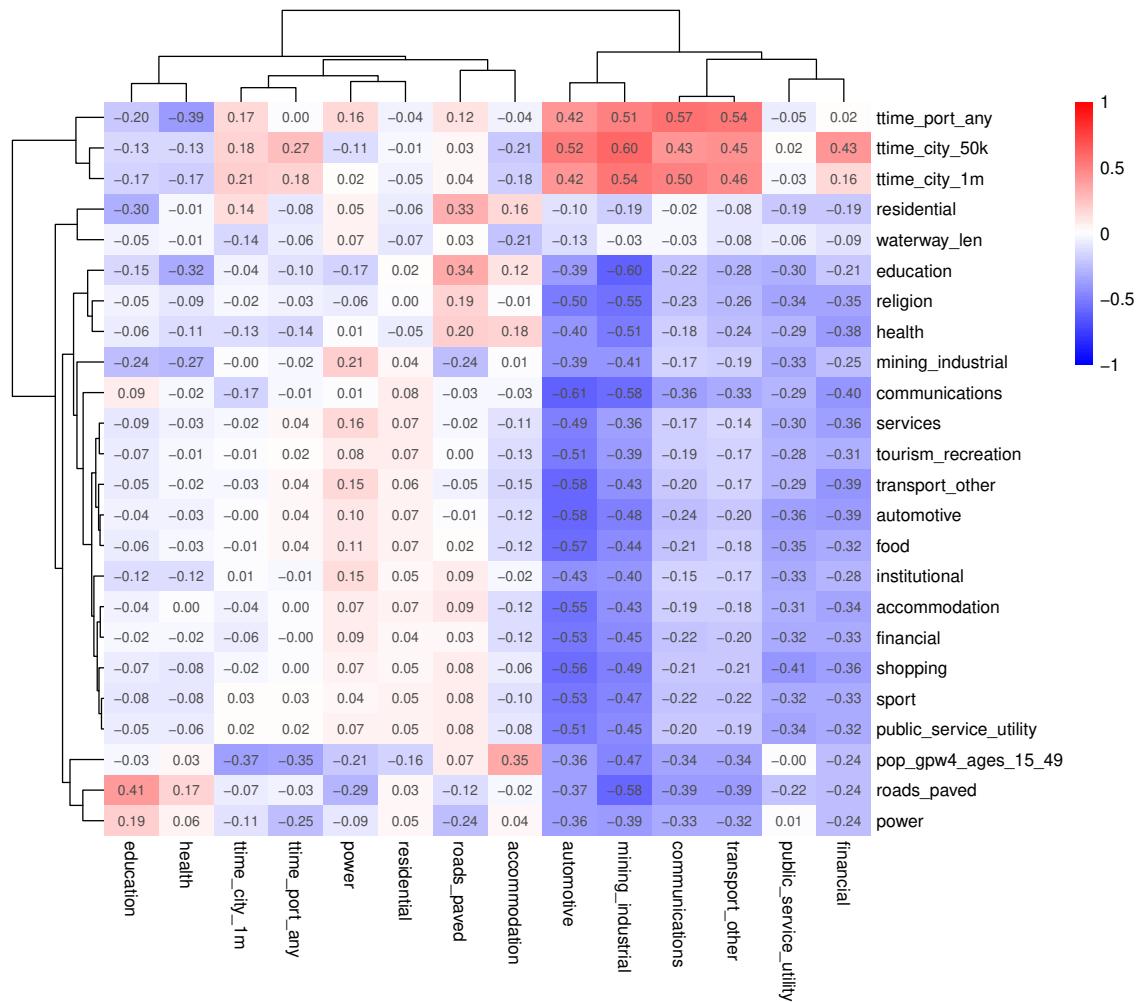


Figure A16: Counterfactual Predictions for Log of Nightlights

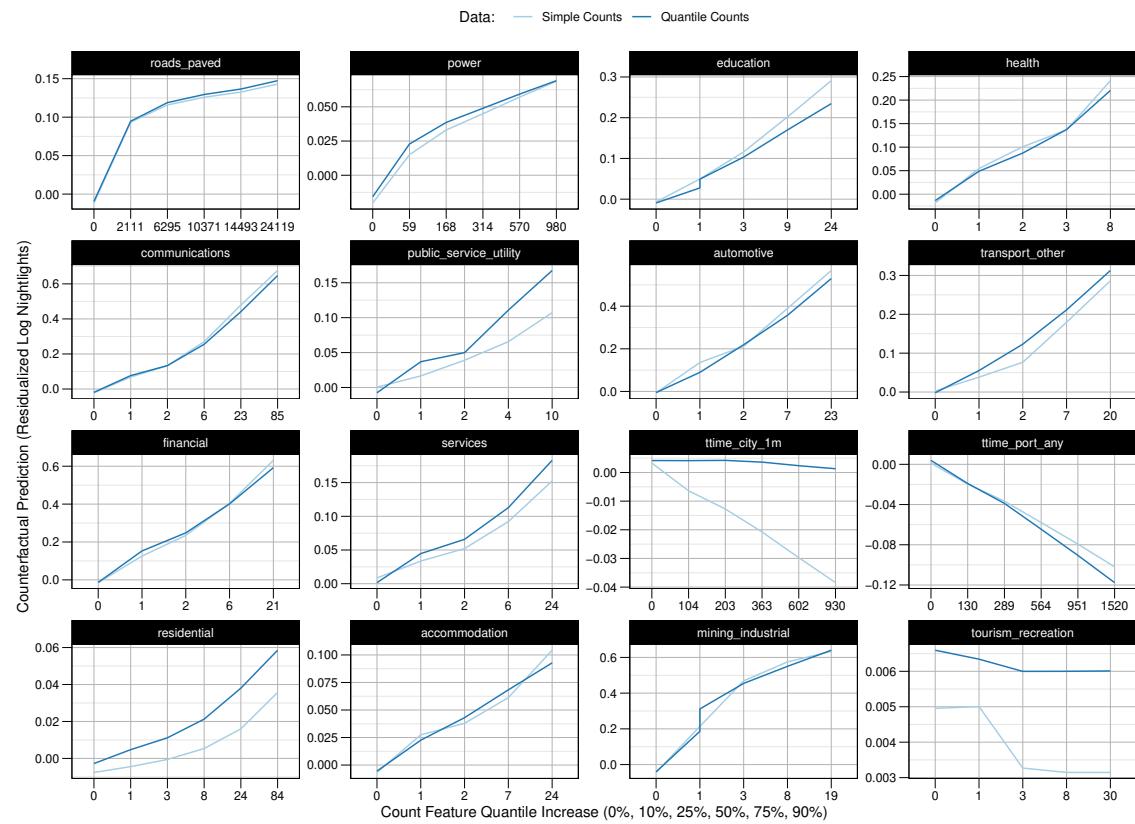


Figure A17: Counterfactual Predictions for IWI: Average Total Wealth Effects per Cell

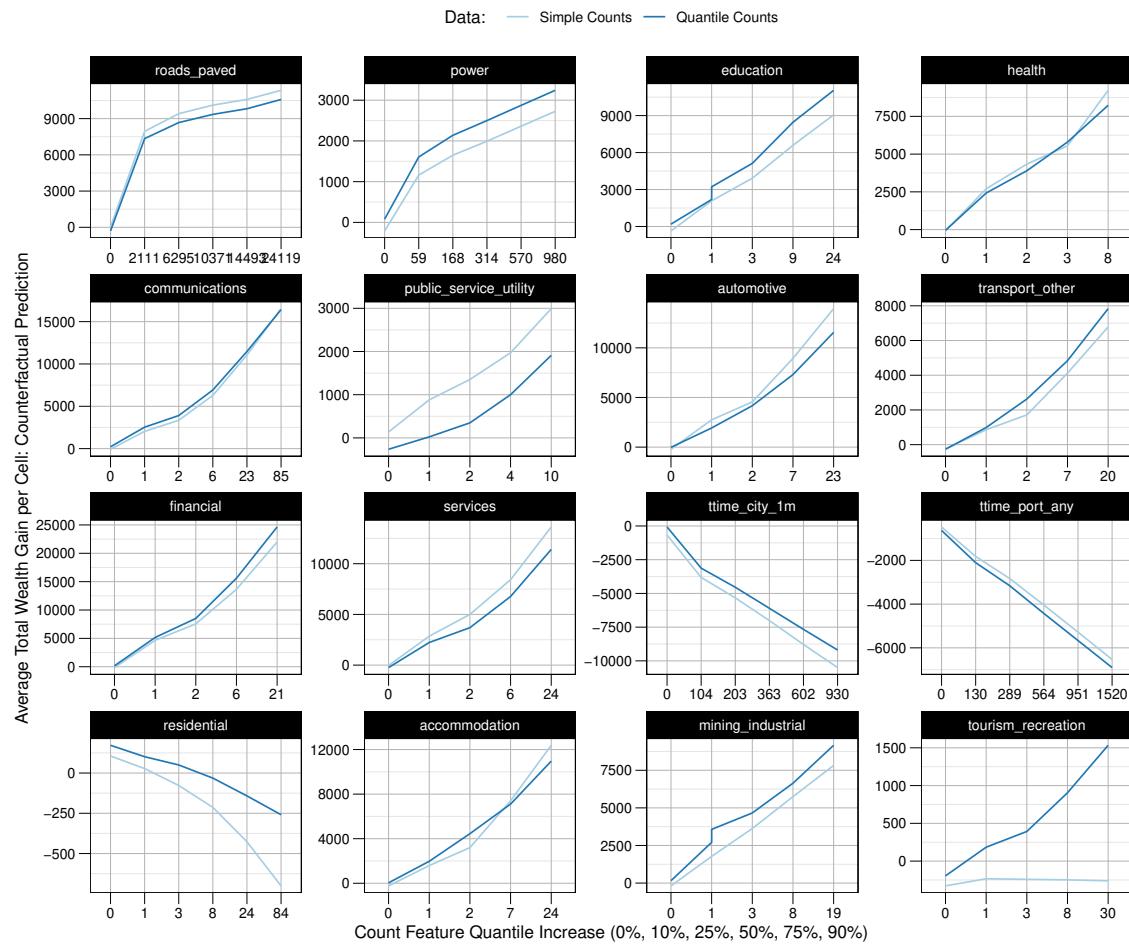


Figure A18: Counterfactual Predictions for DHS-Based IWI

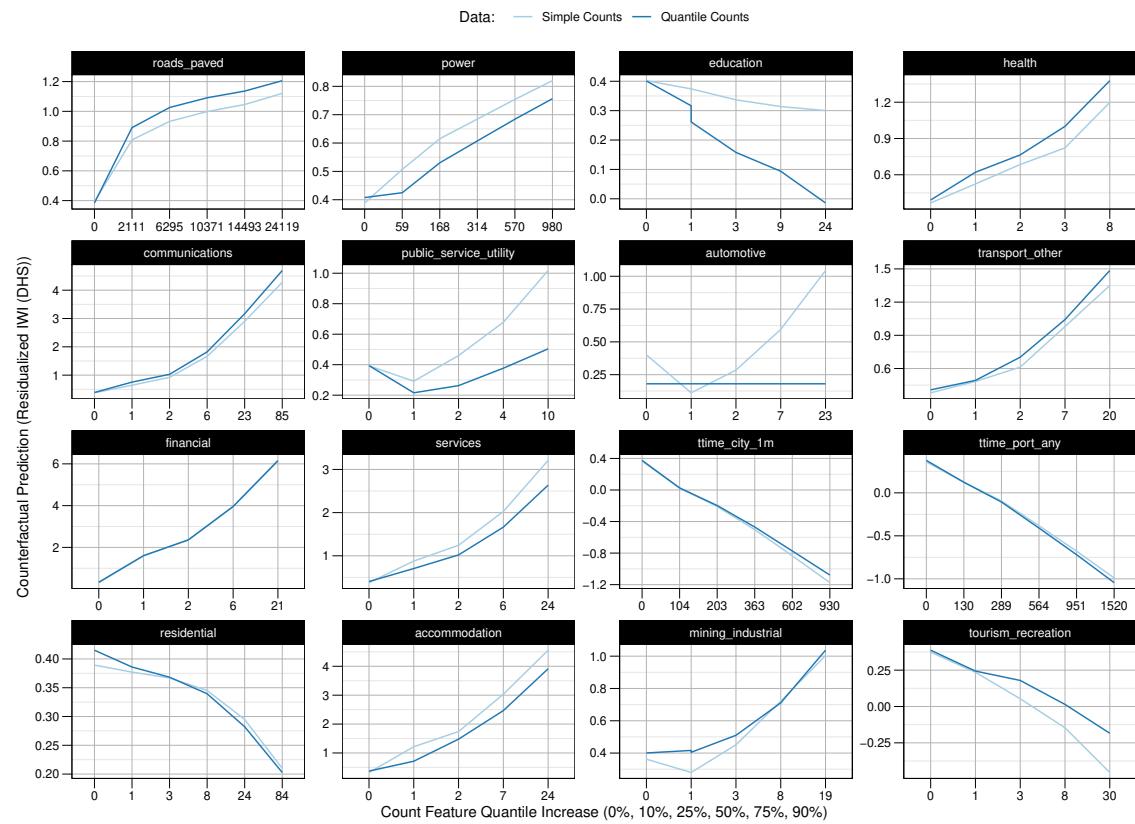


Table A15: Median 50% Counterfactual Prediction: DHS IWI vs. IWI by [Lee & Braithwaite \(2022\)](#)

Feature	Simple Counts		Quantile Counts	
	DHS	L&B	DHS	L&B
roads_paved	1.087	1.622	1.165	1.470
power	0.762	0.302	0.694	0.354
education	0.405	0.662	0.242	0.802
health	0.739	0.776	0.821	0.713
communications	1.710	0.996	1.871	1.112
public_service_utility	0.534	0.131	0.330	-0.007
automotive	0.363	0.583	0.176	0.512
transport_other	0.674	0.171	0.773	0.291
financial	2.368	1.000	2.377	1.112
services	1.285	0.609	1.068	0.435
ttime_city_1m	-0.505	-0.481	-0.469	-0.442
ttime_port_any	-0.365	-0.351	-0.386	-0.371
residential	0.398	-0.068	0.406	-0.044
accommodation	1.762	0.371	1.517	0.533
mining_industrial	0.506	0.477	0.575	0.595
tourism_recreation	0.108	-0.058	0.232	0.012
Median Abs. Coef.	0.604	0.479	0.635	0.477
Corr. Spearman		0.715		0.724
Corr. Pearson		0.700		0.715

*Abbreviations:* DHS = IWI estimate from Demographic and Health Survey; L&B = Predicted IWI estimate by [Lee & Braithwaite \(2022\)](#).

*Notes:* the DHS-based IWI estimate is computed from all DHS surveys conducted in SSA since 2010. It is averaged across  $96\text{km}^2$  hexagonal grid cells just like the IWI estimate of [Lee & Braithwaite \(2022\)](#). Grid cells with less than 5 households or less than 10 people per  $\text{km}^2$  population density are excluded. This yields 17,396 cells with DHS-based IWI estimates used for training, versus 89,048 cells available with the predicted IWI by [Lee & Braithwaite \(2022\)](#). Counterfactual predictions are made for 103,922 cells which have a population density above 10 persons/ $\text{km}^2$  and any POI feature. The median of the 50% increase counterfactual prediction is reported in this table.

Table A16: Correlations of Counterfactual Predictions: DHS IWI and IWI by [Lee & Braithwaite \(2022\)](#)

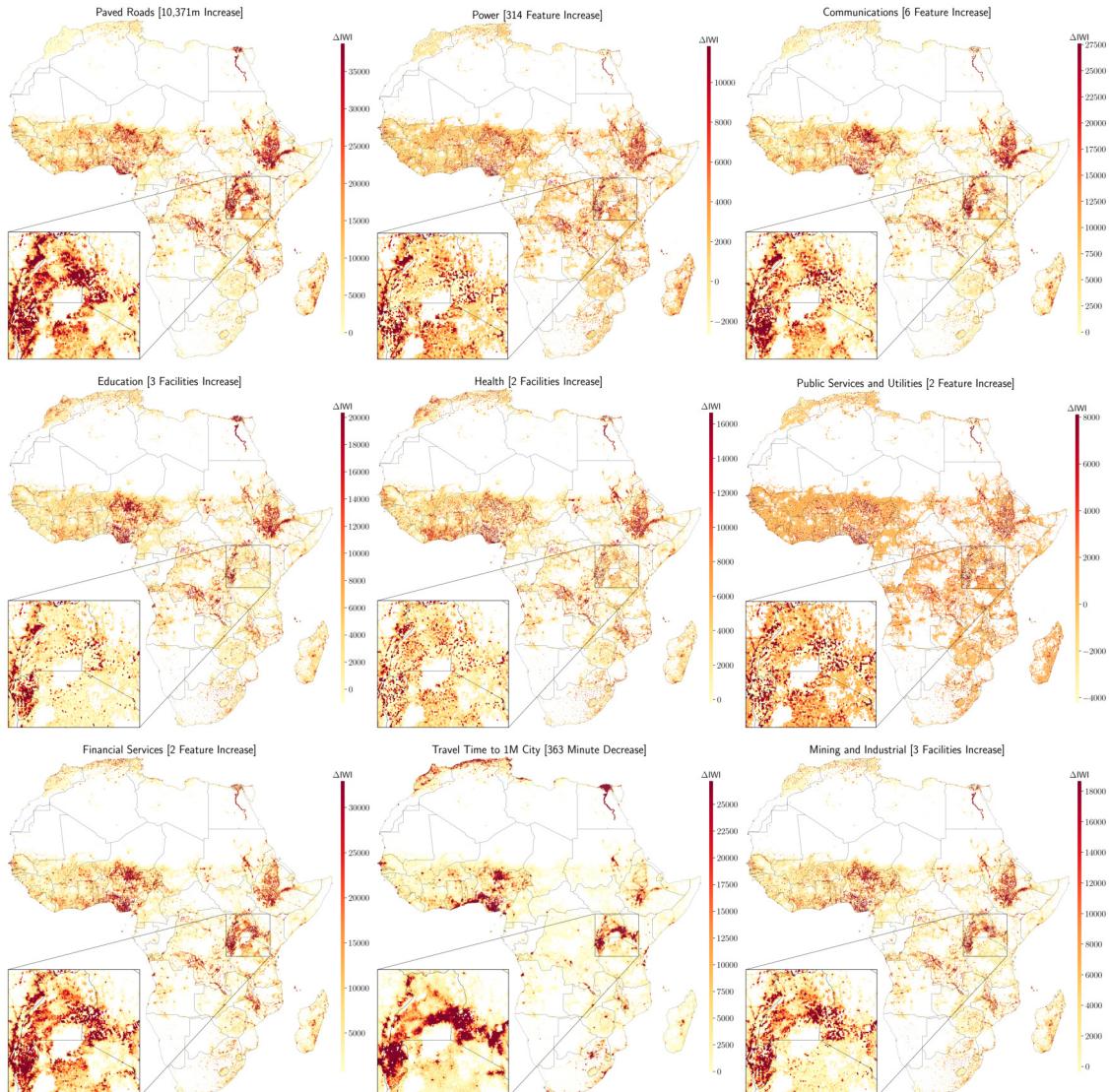
Feature	Simple Counts	Quantile Counts
roads_paved	-0.045	-0.115
power	-0.019	0.136
education	0.069	0.349
health	0.272	0.219
communications	0.281	0.410
public_service_utility	0.299	0.029
automotive	-0.016	0.323
transport_other	0.109	0.006
financial	0.385	0.505
services	0.177	0.058
ttime_city_1m	0.345	0.540
ttime_port_any	0.127	0.118
residential	0.304	0.271
accommodation	0.298	0.364
mining_industrial	0.145	0.208
tourism_recreation	0.062	0.091
Median Abs. Corr.	0.161	0.214
Median Abs. Corr.	0.161	0.214

*Notes:* the DHS-based IWI estimate is computed from all DHS surveys conducted in SSA since 2010. It is averaged across  $96\text{km}^2$  hexagonal grid cells just like the IWI estimate of [Lee & Braithwaite \(2022\)](#). Grid cells with less than 5 households or less than 10 people per  $\text{km}^2$  population density are excluded. This yields 17,396 cells with DHS-based IWI estimates used for training, versus 89,048 cells available with the predicted IWI by [Lee & Braithwaite \(2022\)](#). Counterfactual predictions are made for 103,922 cells which have a population density above 10 persons/ $\text{km}^2$  and any POI feature. The average Pearson's correlation of these counterfactual predictions across the two IWI estimates is reported in this table.

Table A17: Correlations of Counterfactual Predictions: DHS IWI and IWI by [Lee & Braithwaite \(2022\)](#): Previous Estimates without Separate Spillover Variables (One Variable Per Infrastructure)

	Count	Count + SS	Tag Weights	Weights + SS
roads_paved	0.396	0.743	0.471	0.610
power	0.313	0.141	0.166	0.195
education	0.054	0.297	0.238	0.378
health	0.147	0.091	0.152	0.336
communications	0.184	0.282	0.264	0.345
public_service_utility	0.237	0.166	0.193	0.321
automotive	0.129	0.502	0.215	0.497
transport_other	0.267	0.165	0.218	0.240
financial	0.269	0.057	0.376	0.244
services	0.206	0.302	0.325	0.368
ttime_city_1m	0.216	0.131	0.257	0.125
ttime_port_any	0.226	0.036	0.240	0.108
residential	-0.060	0.057	-0.032	0.081
accommodation	0.233	0.574	0.363	0.651
industrial	0.147	0.218	0.146	0.170
tourism_recreation	0.239	0.161	0.308	0.302
Median Corr.	0.216	0.163	0.248	0.298
Median Abs. Corr.	0.216	0.163	0.248	0.298

*Abbreviations:* SS = Spatial Spillovers.

Figure A19: Spatial 50% CFPRs (Geometric Mean)  $\times$  WorldPop 2020 Population

*Notes:* Figure shows 50% counterfactual predictions of the International Wealth Index (IWI) [0, 100] by Lee & Braithwaite (2022), i.e., the predicted wealth increase (Eq. 12) from an increase in each cell amounting to the median of the non-negative feature density, summarized in Table 7, multiplied with the WorldPop 2020 population measure. It is thus an estimate of the partial effect of the investment in a cell on total welfare.

# Optimal Investments in Africa's Road Network

Sebastian Krantz\*

July 23, 2024

## Abstract

This paper characterizes economically optimal investments in Africa's road network in partial and general equilibrium - based on a detailed topography of the network, road construction costs, frictions in cross-border trading, and economic geography. Drawing from data on 144 million trans-continental routes, it first assesses local and global network efficiency and market access. It then derives a large network connecting 447 cities and 52 ports along the fastest routes, devises an algorithm to propose new links, analyzes the quality of existing links, and estimates link-level construction/upgrading costs. Subsequently, it computes market-access-maximizing investments in partial equilibrium and conducts cost-benefit analysis for individual links and several investment packages. Using a spatial economic model and global optimization over the space of networks, it finally elicits welfare-maximizing investments in spatial equilibrium. Findings imply that cross-border frictions and trade elasticities significantly shape optimal road investments. Reducing frictions yields the greatest benefits, followed by road upgrades and new construction. Sequencing matters, as reduced frictions generally increase investment returns. Returns to upgrading key links are large, even under frictions.

**Keywords:** African roads, spatially optimal investments, big data, PE and GE analysis

**JEL Classification:** O18; R42; R10; O10

## 1 Introduction

For several decades, development economists and policymakers have highlighted insufficient or inefficient transport infrastructure, coupled with border frictions and high trading costs, as a major obstacle to African economic development and regional integration. A World Bank report from 2010 states that Sub-Saharan Africa (SSA) has 31 paved roads *km* per  $100km^2$  of land, compared to  $134km^2$  in other low-income countries, estimates the annual infrastructure gap at \$93 billion, and urges SSA countries to spend one percent of GDP on roads (Foster & Briceño-Garmendia, 2010). This is echoed by the African Development Report 2010, attesting a financing gap at 5 percent of GDP to overhaul the infrastructure sector (ADB, 2010). More extensive and detailed work by Teravanithorn & Raballand (2009) and Lamarque & Nugent (2022) unveils the nature of transportation and trucking in Africa, the determinants of transport costs and frictions, and discusses the potential and policy options for development along specific transport corridors. However, little remains known about the economic optimality of such investments. Most academic and policy literature analyzes a single country or corridor in isolation and relies on heuristic calculations or strongly reduced road network graphs. Equilibrium concepts and implications for optimal routing and traffic flow, including along non-targeted roads, are often neglected.

This paper instead presents a comprehensive, data-driven, and quantitatively rigorous assessment of Africa's road network and optimal investments into it at the trans-African scale. It (1) conducts a detailed empirical stocktake of Africa's present road network; (2) characterizes global economic optimality in network investments from a market access and welfare perspective; (3) investigates the effects of cross-border frictions on optimal spatial allocations; and (4) presents cost-benefit analyses and evidence on the macroeconomic feasibility of large scale (optimal) investments. This agenda mirrors questions of international policymakers who, in theory, can invest in roads anywhere on the continent and may ask: (1) is it still important to invest in African roads?; (2) where exactly should we invest in African roads?; (3) are investments sensible with/without regulatory reforms?; and (4) are they macroeconomically feasible? The paper engages this ambitious

---

\*Kiel Institute for the World Economy  
 Address: Haus Welt-Club, Duesternbrooker Weg 148, D-24105 Kiel  
 E-mail: sebastian.krantz@ifw-kiel.de  
 Latest version available at <https://sebastiankrantz.com/research>

agenda by building realistic economic topographies informed by routing engines, geospatial big data, and surveys, and analyzing them with tools from optimal transport and spatial economic modeling. It satisfactorily answers these questions in light of present empirical possibilities.<sup>1</sup>

Several significant contributions have already been made in economic literature. [Atkin & Donaldson \(2015\)](#) show that the effect of log distance on trade costs within Ethiopia or Nigeria is 4-5 times larger than in the US and  $\leq 30\%$  of this effect is explained by the availability and quality of roads. Using ten years of disaggregated monthly price data, [Porteous \(2019\)](#) estimates a dynamic model of agricultural storage and trade comprising 6 grains traded across 230 regional markets in all 42 SSA countries, which are connected in a network with 413 overland links to 30 international ports. He finds median trade costs five times higher than elsewhere in the world. Reducing trade costs to the world average would yield a 46% reduction in grain prices and a welfare gain equivalent to 2.17% of GDP. He further shows that 88% of the welfare gain can be achieved by lowering trade costs through ports and links representing just 18% of the trade network, supporting transport-corridor-based approaches to network improvement.

[Graff \(2024\)](#) studies spatial inefficiency in African national road networks using a regular  $0.5^\circ$  grid-network parameterized with Open Street Map (OSM) routes between adjacent cells. Employing data on population, nightlights (productivity) and ruggedness ([Nunn & Puga, 2012](#)), and the quantitative framework of [Fajgelbaum & Schaal \(2020\)](#), he generates ideal national networks, and estimates that Africa would gain 1.3% of total welfare from better organizing its national road systems (with national welfare gains ranging between 6.6% for Somalia and 0.5% for South Africa), and 0.8% from a 10% optimal expansion. He also documents that colonial infrastructure projects skewed trade networks towards a sub-optimal equilibrium and that regional favoritism and inefficient aid provision hinder optimal network investments ([Graff, 2024](#)). The latter is consistent with findings by [Jedwab & Moradi \(2016\)](#); [Jedwab et al. \(2017\)](#), [Burgess et al. \(2015\)](#), and [Bonfatti et al. \(2019\)](#). Similarly, [Gorton & Ianchovichina \(2021\)](#) analyze trade network inefficiencies and optimal expansions for Latin America closely following [Fajgelbaum & Schaal \(2020\)](#). They find average welfare gains of 1.6% from optimal reallocation and 1.6%/1% from a 50%/10% optimal expansion of national road networks. They also consider transnational networks in the MERCOSUR and Andean Community trade blocs. Disregarding borders, they find that optimal expansions better connect major cities across countries at welfare gains of 1.5-1.9%. Throughout, they find that optimal network investments reduce spatial wealth disparities.

Many further works such as [Jedwab & Storeygard \(2022\)](#), [Peng & Chen \(2021\)](#), [Abbasi et al. \(2022\)](#), [Fiorini et al. \(2021\)](#), and [Moneke \(2020\)](#) study the economic effects of road interventions in Africa. The latter two find that road network expansion in Ethiopia enhanced the impact of trade liberalization/electrification on local firm productivity/industrialization, respectively. A vast further literature studies the effects of road interventions in other world regions, particularly in China ([Faber, 2014](#); [Baum-Snow et al., 2020](#)) and India ([Asher & Novosad, 2020](#); [Alder, 2016](#)).

Apart from the partial equilibrium simulations by [Porteous \(2019\)](#) on a heuristic agricultural trade network, and the work by [Graff \(2024\)](#) examining hypothetically optimal reshuffling and 10% expansion of national road networks on a regular  $0.5^\circ$  grid parameterized in a way that has limited resemblance of optimal routing through real road networks (see Section 4), the literature has not yet (to my knowledge) seen a detailed quantitative attempt to estimate what should be done to improve or extend Africa's continental road network. This paper attempts to fill this void.

Starting from a precise continental grid of 12,016 cells with centroids exactly 50.28km apart, I obtain 144 million OSM routes connecting all points and compute distance-weighted local measures of network quality and continental market access (MA). Using data on trading across borders, which I translate into road distance/travel time, I characterize how cross-border trade frictions impede cell-MA. A trade across the median African border is equivalent to  $\sim 1000$ km of national roads or 3.2 road days. My results imply that particularly cells across the border from major markets such as Nigeria suffer  $\leq 80\%$  MA losses due to frictions. Deriving an optimal graph representation of the network where links connecting adjacent cells are weighted to minimize the

---

<sup>1</sup>Nearly all African countries lack subnational trade and (disaggregated) productivity data. General equilibrium simulations also impose a tight computational tradeoff between network size and number of traded products.

routing error through the entire network, I simulate global MA returns from improvements in local connectivity around individual cells. Better connecting cells surrounding urban centers or between major agglomerations along existing transport routes yields sizeable marginal MA gains.

In the second part of the paper, I identify 447 large cities with populations above 100,000 and 52 international ports with a quarterly outflow above 100,000 TEU and obtain routes between them. I process these routes into a realistic representation of the African transport network comprising 1379 nodes and 2344 edges representing 315,000km of road segments. The network connects 444 million people, or 86% of the African urban population. I parameterize the network with fresh routing information along all links and develop an algorithm to find 481 new links (104,000km), yielding route efficiency gains of  $\geq 50\%$  (verified with OSM). Using this network graph, I simulate partial equilibrium MA returns from extending and/or upgrading the entire network or individual links. I find that continental MA, denominated in travel time, can be improved by  $\leq 78\%$  through eliminating cross-border frictions. Upgrading the entire network to allow travel speeds of 100km/h yields a 42% frictionless MA gain and a 27% MA gain under current frictions. MA gains from new roads are lower and also decrease under frictions. Spatially, including frictions favors investments in national links, significantly reducing the value of transnational and remote links. I also find sizeable differences in the marginal values and cost-benefit ratios of links and propose three investment packages containing high-return links with and without border frictions. These packages cost between 17 and 61 billion 2015 USD, and generate MA gains between 20% and 38%. They are macroeconomically feasible if they can raise the aggregate economic growth rate by  $\leq 0.54\%$ .

The final part of the paper moves this to a general equilibrium setting following [Fajgelbaum & Schaal \(2020\)](#) and [Graff \(2024\)](#). I first let cities of different sizes trade with each other and with port cities, resembling optimal regional investments at a continental scale. To simulate optimal investments in trans-African roads, I then reduce the network to links connecting the 47 largest (port-)cities and let 17 megacities produce differentiated products traded across the continent. On both networks, I let welfare maximizing social planners spend budgets covering  $<1/3$  of all work - 10 and 50 billion USD'15, respectively. I simulate cases with/without frictions, imports, increasing returns to infrastructure, and inequality aversion in the planner's objective. The regional \$50B planner optimally connects large cities with ports and surrounding cities. Investments are redistributional as in [Gorton & Ianchovichina \(2021\)](#). Under increasing returns to infrastructure, this planner upgrades many more roads and expands into remote regions, yielding higher and more equitable welfare gains. Similarly, the trans-African \$10B planner primarily upgrades road segments within different populated regions, enhancing connectivity between major cities and with ports. An inequality-averse planner also upgrades an entire corridor crossing southern central Africa from Luanda/Kinshasa to Mombasa/Dar es Salaam. Lowering the elasticity of substitution between different goods to a value of 2 creates incentives to also build corridors spanning northern central Africa, better connecting Rwanda/Bukavu and Khartoum to western Africa. Under border frictions, investments in connections through Congo and into Nigeria are reduced. Prioritizing the central African infrastructure gap thus requires inequality aversion, optimism about trans-African trade elasticities, and/or efforts to reduce border frictions. The results overall demonstrate that optimal \$10B investments in trans-African roads are non-trivial and differ significantly from AU Trans-African Highways, Chinese BRI investments, or EU Global Gateway corridors.

Methodologically, in this paper, I advance a rigorous quantitative approach to characterizing optimal road network investments. I demonstrate the use of geospatial big data and routing engines to algorithmically create large and accurate network graphs, including potential new roads, and simulate investments on them via established quantitative frameworks. I am also the first to account for cross-border frictions and trade through ports and study their implications on optimal spatial allocations. Detailed estimation of road construction/upgrading costs facilitates interpretation and enables link-level cost-benefit analysis. A corollary of the work is an open-source Julia translation of [Fajgelbaum & Schaal \(2020\)](#)'s MATLAB toolbox presently available on [GitHub](#).<sup>2</sup>

The paper begins with a high-resolution examination of Africa's road network (Section 2) and cross-border frictions (Section 3). It then proceeds via optimal network representations (Section 4) towards partial equilibrium analysis (Section 5) (including cost-benefit analyses) and culminates with general equilibrium analysis (Section 6) (including trans-African network trade simulations).

---

<sup>2</sup><https://github.com/SebKrantz/OptimalTransportNetworks.jl>

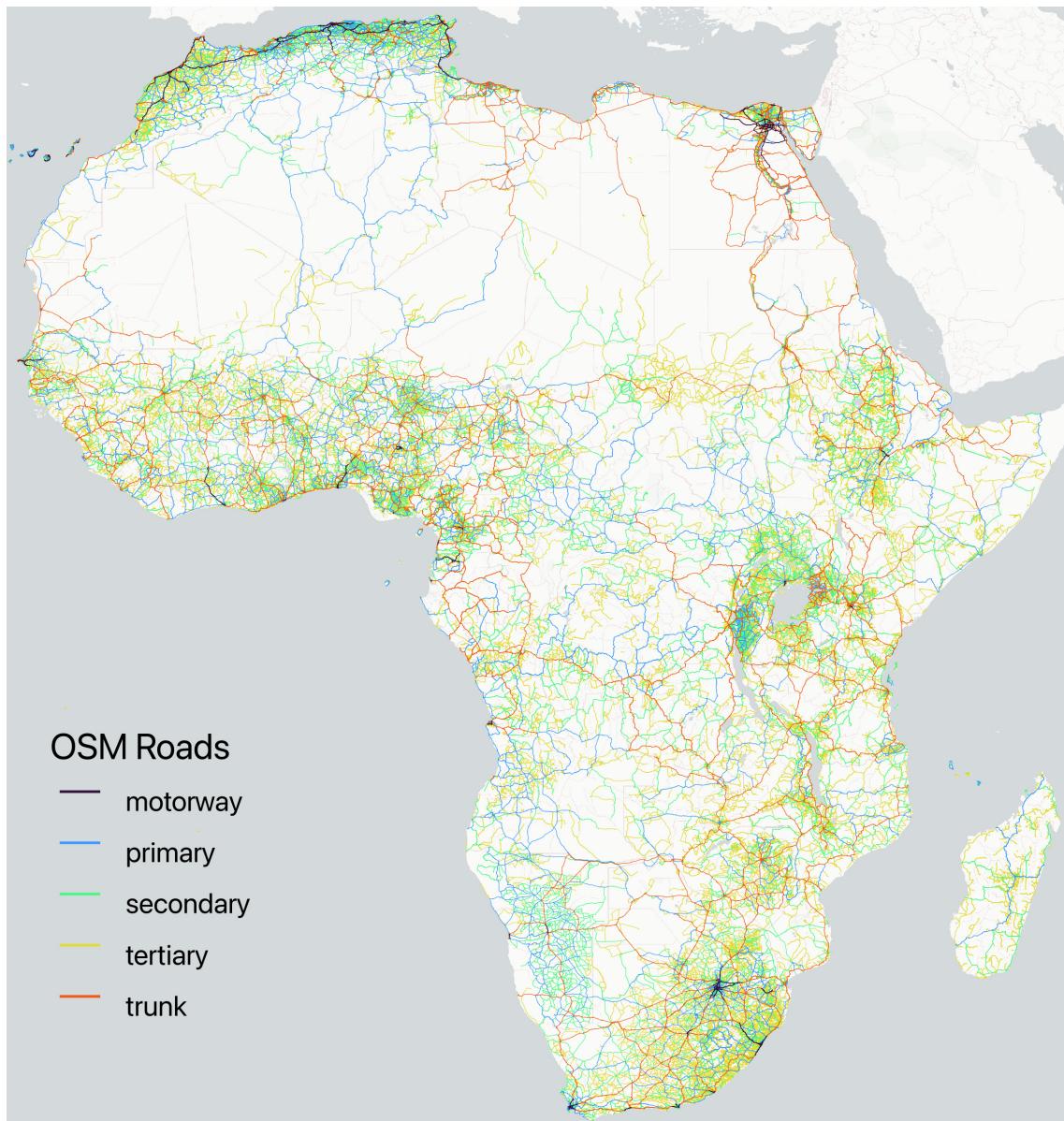
## 2 Africa's Present Road Network

Africa's road network, like any road network, is multi-layered, consisting of major transport routes between cities across countries, smaller roads connecting villages and towns, inner-city roads, and residential roads. Open Street Map (OSM) has an established hierarchy of tags to classify roads comprising motorway > trunk > primary > secondary > tertiary > residential roads. Table 1 shows a summary of all non-residential roads (excluding short links connecting these roads), and Figure 1 a corresponding map. Combined there are 673,045 roads with a total length of 1,611,912km.

Table 1: African Roads (OSM April 2024)

Highway Tag	N	Length in km	Paved Length in km
motorway	14,986	22,400	17,744 (79.2%)
trunk	75,265	181,459	128,978 (71.1%)
primary	124,547	252,047	136,452 (54.1%)
secondary	158,674	365,939	97,904 (26.8%)
tertiary	299,573	790,067	94,373 (12.0%)
Total	673,045	1,611,912	475,450 (29.5%)

Figure 1: African Roads (OSM April 2024)



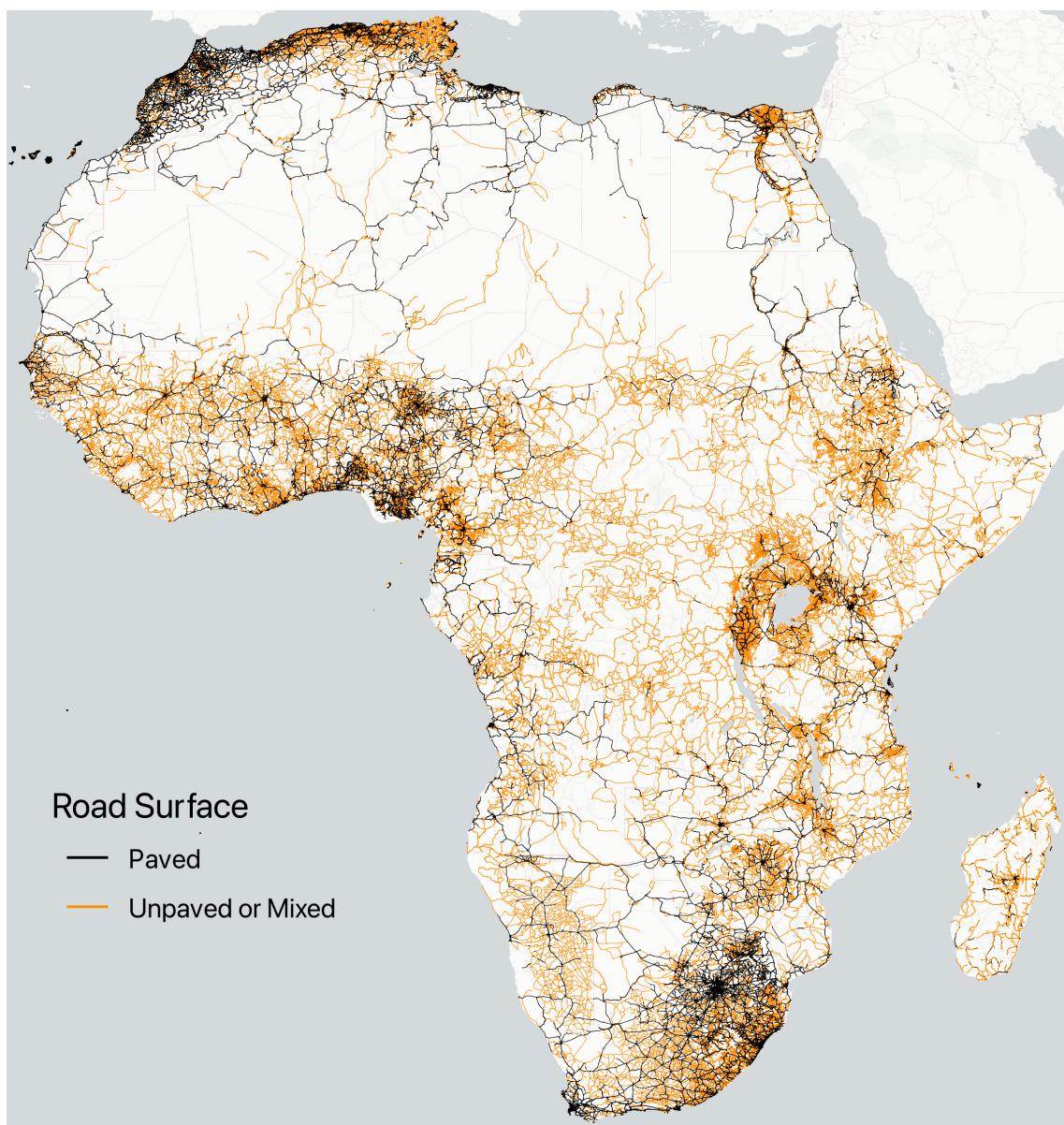
*Notes:* Figure shows all non-residential roads in the Africa OSM (April 2024) colored by type (OSM 'highway' tag).

Is this a comprehensive account of the network? There is reason to believe so: in 2022, Microsoft Research released a [global dataset of segmented roads](#) indicating that globally, 2.3% of roads are missing from OSM, and in Africa around 2.6%. Presumably, most of these are tertiary roads.

Apart from the sheer extent of the network, the surface material and quality of roads determine their efficacy. OSM has a detailed tagging system for roads, including tags such as 'surface,' 'smoothness,' 'tracktype,' 'lanes,' 'lit,' 'maxspeed,' etc. To keep things simple, I develop a binary classification based on whether the 'surface' tag unambiguously indicates that a road is paved, which I summarize on the right hand side (RHS) of Table 1, and plot in Figure 2.

In early 2024, 29.6% of African transport roads are paved, amounting to 475,450km. The actual fraction is presumably a few percentage points higher due to missing 'surface' tags for 38% of roads, all of which are classified as unpaved. But it may also be that a few roads labeled as 'paved' are actually mixed. It is probably relatively safe to assume that at least 1/3 of the network is paved. Among motorways  $\geq 79\%$ , trunk roads  $\geq 71\%$ , and primary roads  $\geq 54\%$  are paved. Significantly lower shares of secondary and tertiary roads are paved;  $\leq 88\%$  of tertiary roads are dirt roads. Paved roads are also concentrated in northern, southern, and western Africa, and are largely missing in central Africa, the Horn of Africa, and Madagascar.

Figure 2: Surface of African Roads (OSM April 2024) | Paved: 475,450km (29.6%)



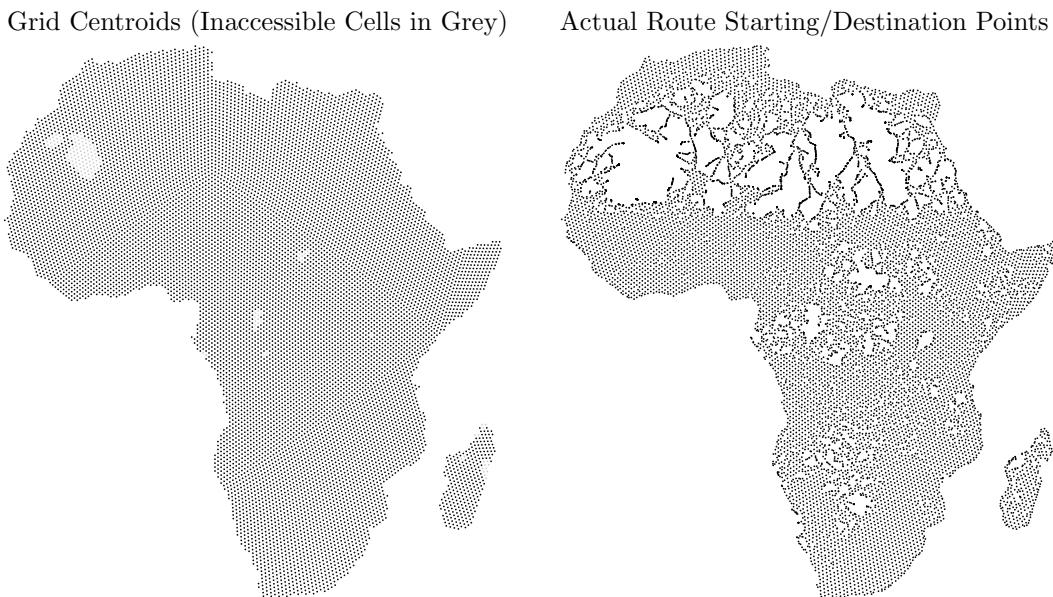
*Notes:* Figure shows all non-residential roads in the Africa OSM (April 2024) colored by the 'surface' tag.

## 2.1 Routing Data

Plotting the road network, while informative about its extent, does not permit a rigorous assessment of its efficiency. Routing engines<sup>3</sup> provide more precise information on how cars can use the network to get from A to B. In my case, A and B could be anywhere on the continent. Two approaches invite themselves to choose points: (1) using a uniform spatial grid and taking the cell centroids, or (2) selecting the centers of important cities/places. In this paper, I employ both approaches, but for spatial analysis, the grid approach is more suitable. Hence, I begin with it.

Accurate spatial analysis at the continental scale warrants using an equally spaced grid. Discrete Global Grid Systems (DGGS) provide these via hierarchical tessellation of cells partitioning the globe. Sahr et al. (2003) propose the Icosahedral Snyder Equal Area Aperture 3 Hexagon (ISEA3H) as a good general-purpose Geodesic DGGS.<sup>4</sup> The grid is implemented as part of the DGGRID C++ library (Sahr, 2022), accessed via the 'dggridR' R package (Barnes, 2020). To limit complexity, I choose a resolution 9 ISEA3H grid, with a cell area of 2591.4km<sup>2</sup> and a centroid spacing of 50.28km. Partitioning the African mainland and Madagascar with this grid yields 12,016 cells.

Figure 3: Grid Cell Centroids and OSRM Route Starting and End Points



*Notes:* Figure shows centroids of 50.28km ISEA3H grid (LHS) and the nearest road segment to each centroid (RHS).

To compute a large number of routes, I configure a server with the Open Source Routing Machine (OSRM)<sup>5</sup> and a copy of the Africa OSM from the 5th of June 2023, and compute fastest car routes between all cell centroids. This amounts to  $12,016^2 - 12,016 = 144,372,240$  routes, for which I obtain both the road distance in m and the estimated travel time in minutes. OSRM 'snaps' cell centroids to the nearest road, yielding divergent route start/endpoints. Figure 3 shows the outcome of this process. For some cells in the Sahara, the Congo Basin, and Madagascar, colored in light grey on the left-hand side (LHS), OSRM cannot find the nearest road and returns 'NA'. For many further cells, the nearest road snapped to is outside of the cell, as shown on the RHS.

I clean the data by first mapping inaccessible cells to the nearest route start/endpoints. I then follow Graff (2024) in adding the spherical centroid-start/end distance to all routes. This increases the length of routes by 0.53% on average. I also adjust travel time estimates by assuming a travel speed of 10km/h on these straight passages, increasing travel time by 3.4% on average.<sup>6</sup>

<sup>3</sup>Large computer programs used for navigating (routing) through digitized transport networks.

<sup>4</sup>Sahr et al. (2003) vindicate ISEA3H as: "... due to its lower distortion characteristics we choose the icosahedron for our base platonic solid. We orient it with the north and south poles lying on edge midpoints, such that the resulting DGGS will be symmetrical about the equator. Next, we select a suitable partition. The hexagon has numerous advantages, and we choose aperture 3, the smallest possible aligned hexagon aperture. Because equal-area cells are advantageous, we choose the inverse ISEA projection to transform the hexagon grid to the sphere..."

<sup>5</sup>A high-performance open routing engine in C++ optimized for OSM. Website: <https://project-osrm.org/>.

<sup>6</sup>In this regard I differ from Graff (2024), which assumes walking speed of 4km/h. My main reason is that the

## 2.2 Efficiency of Africa's Road Network

With realistic road distance and travel time estimates for 144 million trans-African routes, I measure the African network's local and global efficiency. Following Wolfram (2021), I first compute two metrics to assess the efficiency of individual routes: (1) route efficiency (RE), which divides the geodesic distance  $s_{ij}$  between two locations  $i$  and  $j$  by the road distance  $r_{ij}$ , resulting in an index between 0 and 1 with 1 implying a direct straight road; and (2) time efficiency (TE), which divides  $s_{ij}$  by the travel time along the fastest route  $t_{ij}$ , and thus measures the average travel speed in the direction of the destination. Formally, for any route between two cells  $i$  and  $j$ ,

$$\delta_{ij}^r = s_{ij}/r_{ij} \quad (\text{Route Efficiency}) \quad (1)$$

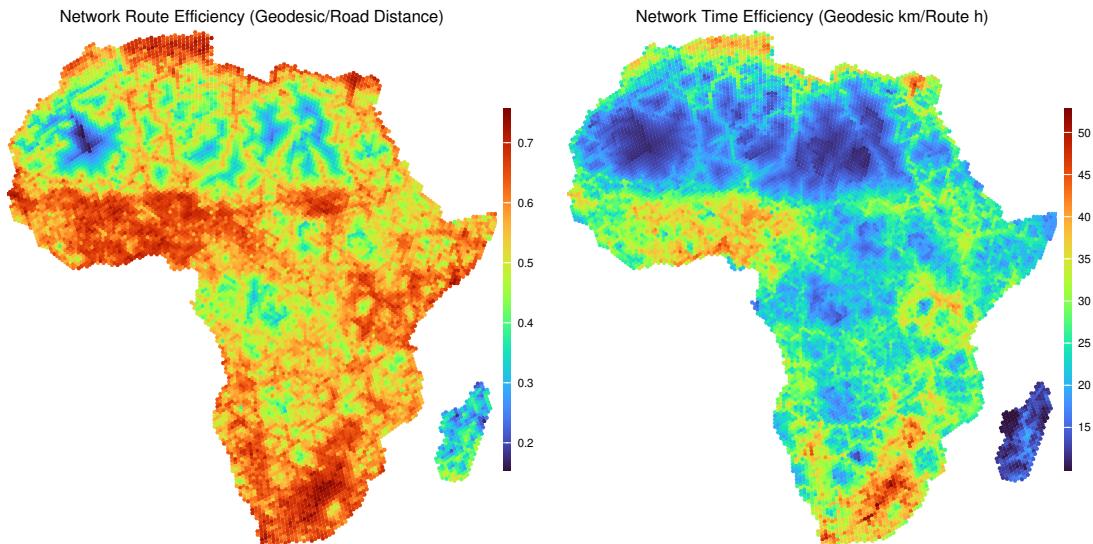
$$\delta_{ij}^t = s_{ij}/t_{ij} \quad (\text{Time Efficiency}). \quad (2)$$

To visualize these metrics, I compute a weighted average across all routes starting in cell  $i$ , using the inverse area of an expanding circle around the cell as weight.<sup>7</sup> For each cell  $i$ , I obtain

$$\kappa_i^m = \sum_{j \neq i} \frac{\delta_{ij}^m}{s_{ij}^2} / \sum_{j \neq i} \frac{1}{s_{ij}^2} \quad m \in r, t. \quad (3)$$

Figure 4 reports the cell-level route and time efficiency estimates. Appendix Figure A1 additionally shows the average network speed (NS), computed as  $r_{ij}/t_{ij}$ .

Figure 4: Local Route and Time Efficiency of Africa's Road Network



*Notes:* Figure shows local route efficiency (Eq. 1) and time efficiency (Eq. 2) for each cell  $i$ , computed as an inverse-squared spherical distance weighted average across all routes starting in  $i$  (Eq. 3).

All three metrics are higher in cities or close to major transport routes. RE is less concentrated than TE and NS, indicating larger heterogeneity in road quality. Particularly outside of major cities, the difference between route and time efficiency is significant. For example, the Southern part of Sudan has, according to Figure 2, a quite dense network of mostly dirt roads, permitting, as Figure 4 suggests, route-efficient navigation at slow driving speeds. Since OSRM does not take into account traffic volumes, TE estimates inside cities and populated areas are too high.

A simple average of these cell-level estimates ( $\kappa_i$ ) yields a NS of 40.3km/h (maximum 69.7km/h), RE of 0.534 (maximum 0.756), and TE of 24.9km/h (maximum 52.8km/h). As  $\kappa_i$  is low in

---

centroid of such cells is unlikely of economic significance, and assuming a travel speed of 4km/h increases travel time by 8.5% on average, whereas assuming 10km/h only increases travel time by 3.4% on average. Furthermore, since most cells where the distance between the cell centroid and the route starting/endpoints is large are in the Sahara, I believe that motorized transport with specialized vehicles (4x4) is possible, and thus a travel speed of 10km/h might be more realistic.

<sup>7</sup>The reason for considering the area ( $\pi s_{ij}^2$ ,  $\pi$  cancels out) as a weight on individual routes is that with increasing distance from  $i$  there are more cells  $j$ , and each of these cells is a less important destination for cell  $i$  dwellers.

uninhabited areas, I also compute a population-weighted average using 2020 total population counts from [WorldPop](#), shown in Appendix Figure A1. This yields a NS of 53.2km/h, RE of 0.625, and TE of 34.4km/h. [Wolfram \(2021\)](#) computes global efficiency directly by probabilistically sampling route start/end points using population weights.<sup>8</sup> To approximate this strategy, I compute a global weighted average across routes using Newton's Law of Gravity to weight each route by the product of the start/end cell populations divided by their squared spherical distance. This yields NS of 56.7, RE of 0.668, and TE of 38.8. Since research on inter-city migration gravity models such as [Zipf \(1946\)](#) or [Poot et al. \(2016\)](#) suggests that migration varies according to the distance rather than the squared distance, I also consider weights denominated by the simple distance. This yields even higher NS of 64.2, RE of 0.68, and TE of 43.9, indicating that efficiency is higher on longer routes - as [Wolfram \(2021\)](#) finds in the US and Europe. But even the most favorable RE estimate of 0.68 is far below [Wolfram \(2021\)](#)'s estimates of 0.843/0.767 for US/European road networks.

## 2.3 Market Access

Whereas route and time efficiency are important from a transportation standpoint, the spatial economics literature such as [Donaldson & Hornbeck \(2016\)](#), [Jedwab & Storeygard \(2022\)](#) or [Peng & Chen \(2021\)](#) has used market access (MA), typically defined as an inverse-distance weighted sum of population, or a proxy therefore such as built-up in the case of [Peng & Chen \(2021\)](#), to evaluate economic gains from transport investments. The MA measure for cell  $i$  is

$$\text{MA}_i = \sum_{j \neq i} L_j \tau_{ij}^{-\theta}, \quad (4)$$

where  $L_j$  is a measure of population and  $\tau$  a measure of distance, travel time, or transport cost, weighted by an elasticity  $\theta$  measuring how trade volumes fall as distance/time/costs increase. The value of  $\theta$  is contentious, for example [Jedwab & Storeygard \(2022\)](#) use  $\theta = 3.8$  whereas [Peng & Chen \(2021\)](#) use  $\theta = 1$ . In both cases,  $\tau_{ij}$  measures travel time.

To appropriately measure continental MA implied by the road network, I incorporate spatial differences in purchasing power into  $L_j$ . [Kummu et al. \(2018\)](#) provide a spatial estimate of GDP in 2015 USD PPP terms by distributing national GDP estimates from the World Development Indicators at 5 arc-min (0.0833 degrees or 9.3km at the equator) resolution using subnational value-added estimates from [Gennaioli et al. \(2013\)](#) and population from the HYDE 3.2 database. For robustness, I also consider a predicted International Wealth Index (IWI) by [Lee & Braithwaite \(2022\)](#), which employ day- and nighttime satellite imagery and OSM features to predict the IWI - a comparable asset-based wealth index calculated from DHS Surveys for 25 countries in SSA conducted since 2017. Appendix Figure A2 shows the raw estimates from [Kummu et al. \(2018\)](#) and [Lee & Braithwaite \(2022\)](#).<sup>9</sup> To map these estimates into the grid, I sum GDP and average the IWI in each cell. I multiply the IWI with the WorldPop cell-population to estimate total wealth.

Using Eq. 4, I compute MA maps using both measures of economic mass ( $L_j$ ), and trade elasticities of  $\theta = 1$  ([Peng & Chen, 2021](#)) or  $\theta = 3.8$  ([Jedwab & Storeygard, 2022](#)). As Eq. 4 excludes cell  $i$  itself, the centers of several large cities have a lower MA measure than adjacent cells. To remedy this, I add  $i$ 's GDP/wealth to the MA measure using a distance  $\delta_{ij} = 25\text{km}/\kappa_i$ , where  $\kappa_i$  is cell  $i$ 's RE/TE measure from Figure 4. Figure 5 shows the resulting maps with GDP and Appendix Figure A4 with total IWI-based wealth.

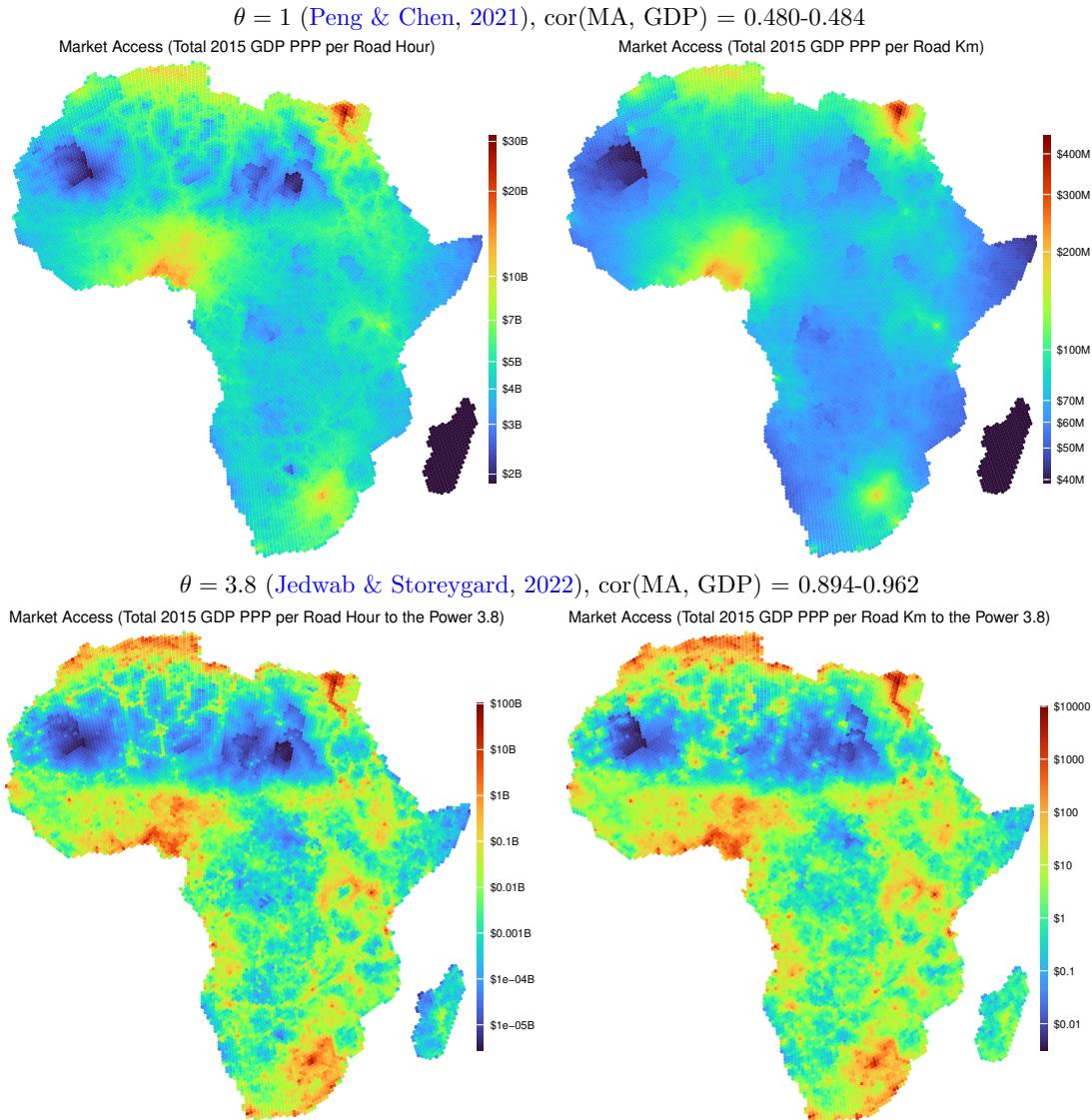
Evidently,  $\theta = 3.8$  yields a much stronger concentration of MA around major cities and a high correlation of the MA measure with cell GDP. Both  $\theta$  values elicit northern Egypt, Nigeria, and, to a lesser extent, greater Johannesburg as high MA areas. Cairo has the greatest MA. The  $\theta = 1$  estimates imply that  $>\$30\text{B}$  of GDP can be reached within one road hour from Cairo. Wealth-based estimates in Appendix Figure A4 are very similar.

---

<sup>8</sup>Concretely, [Wolfram \(2021\)](#) computes the route efficiency of the US road network using OSRM-generated routes between 2000 population-weighted random points in the US and finds a median route efficiency of 0.843 with a standard deviation of 0.045 across 4 million routes.

<sup>9</sup>Since the IWI is not available for North Africa, I predict it at 96km<sup>2</sup> resolution using a rich gridded infrastructure database developed in [Krantz \(2023\)](#) with OSM as principal data source. Imputation using the *MissForest* algorithm ([Stekhoven & Bühlmann, 2012](#)) via the *missRanger* R package ([Mayer, 2023](#)) reaches an  $R^2$  of 97%, as shown in Appendix Figure A3.

Figure 5: Market Access Maps using GDP in 2015 USD PPP



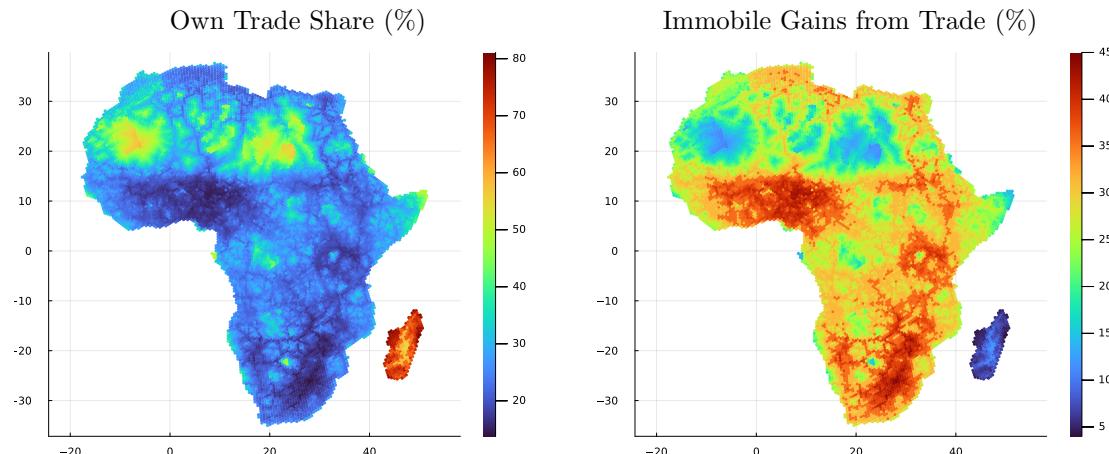
Notes: Figure shows local market access using Eq. 4 where  $L_j$  is cell-GDP in 2015 USD PPP from Kummu et al. (2018).

## 2.4 Gravity Implied Trade and Welfare Gains

A yet more elaborate way to synthesize the data is through the lens of a quantitative spatial model. I take the canonical model of Redding (2016) with fixed labor and calibrate it with iceberg trade costs  $d_{ij} = 1 + 0.1 \times \sqrt{\text{travel time in minutes}_{ij}}$ , population from WorldPop, the imputed IWI as productivity measure, and residential land area from the Overture Maps Buildings Dataset. Using the parameterization of Redding (2016), I solve the model and compute trade flows between all locations, consumer and land prices, output, consumption, and welfare. Of particular interest are the trade share of a location with itself and the welfare gains from trade vis-a-vis autarky. Figure 6 visualizes these two, indicating that populated, productive, and well-connected locations have higher trade shares with other locations and experience greater gains from trade. In contrast to MA maps, the model emphasizes the large connected economic areas in western, southern, and, to a lesser extent, eastern Africa as exhibiting the greatest potential gains from inner-African trade.

All of these maps, indicative of Africa's potential to trade and benefit from a single market, assume that transport is the only obstacle to accessing distant markets and that there is no congestion. To add some realism, I proceed to examine barriers to cross-border trading in Africa. Section 6 additionally introduces congestion forces through a general equilibrium framework.

Figure 6: Own Trade Share and Welfare Gains from Trade Following Redding (2016)



*Notes:* Figure shows cell-own trade share and immobile gains from trade in equilibrium of a quantitative spatial model introduced by Redding (2016) (CRS version) using OSRM travel time estimates to inform iceberg trade cost parameters.

### 3 Barriers to Cross-Border Trade

Africa has 'thick borders,' and any evaluation of continental roads should at least be mindful of them. Trade barriers can broadly be distinguished into three categories: (1) traditional trade tariffs, (2) traditional non-tariff measures such as quantity or quality restrictions (e.g., sanitary and phytosanitary measures), and (3) regulatory obstacles and frictions at the border. This paper will concentrate on type (3) barriers as types (1) and (2) tend to be product- and quantity-specific. But, for the sake of completeness, I begin with a brief overview type (1) and (2) barriers.

#### 3.1 Trade Tariffs

The World Development Indicators provide aggregate estimates of the extensive and intensive margins of trade tariffs applied by each country.<sup>10</sup> Table 2 summarizes these indicators for different country groups. While SSA countries only apply specific (non-ad-valorem) rates on 0.1% of tariff lines, 33.9% of African tariffs exceed 15%. When weighted by the product import shares corresponding to each partner country, the average SSA country applies an 8.2% import tariff and a 7.5% tariff on manufactured products. This is comparatively high and only exceeded by South Asia and the low income country (LIC) average. Averaging across all African countries using 2020 GDP in PPP terms as weights yields very similar outcomes, at an average tariff rate of 8.7%.

The African Continental Free Trade Agreement (AfCFTA), concluded in 2019 and ratified by 47 African states by the end of 2023, aims to facilitate trade by eliminating 97% of tariff lines between members alongside most NTBs. Yet, trading under the agreement has started sluggishly; at present, only 8 countries trade 96 products under the Guided Trade Initiative<sup>11</sup>. With LDCs given 10+ years to remove tariff lines and the sluggish start of trade under the agreement, it is expected that the full benefits of the agreement will only accrue by the middle of the century.

#### 3.2 Non-Tariff Measures

Non-tariff measures (NTMs) are still difficult to quantify comprehensively, particularly in terms of their depth and effects. A simple database that just counts the incidence of NTMs is the WIIW NTM Database by Ghodsi et al. (2017). According to it, the average African country in 2019 had 101 general NTMs (with all WTO members), whereas the median African country only had 18. Countries with the most NTMs were Uganda (1268), Kenya (492), and South Africa (418). In comparison, the average non-African country had 316 NTMs, and the median non-African country had 86. The country with the most NTMs is the USA (5471), followed by Brazil

<sup>10</sup>Based on the World Integrated Trade Solution system with data from United Nations Conference on Trade and Development's Trade Analysis and Information System (TRAITS) database and the World Trade Organization's (WTO) Integrated Data Base (IDB) and Consolidated Tariff Schedules (CTS) database.

<sup>11</sup><https://au-afcfta.org/>, <https://www.macmap.org/en/learn/afcfta>

Table 2: World Development Indicators: Tariffs Applied in 2020

Region/Income Status	All Products				Manufactured Products			
	SSR	SIP	TR	WTR	SSR	SIP	TR	WTR
OECD	5.7	2.7	2.0	1.6	0.3	0.6	1.5	1.5
Non-OECD	0.5	19.1	8.0	6.3	0.2	17.8	7.5	6.0
High income	4.3	4.7	3.0	2.6	0.2	3.0	2.6	2.4
Low income	0.0	34.3	11.1	9.3	0.0	32.4	10.6	8.4
Europe & Central Asia	5.0	2.4	2.5	1.9	0.3	0.6	2.2	1.9
Latin America & Caribbean	0.0	19.2	8.0	6.7	0.0	17.7	7.5	6.6
South Asia	0.6	28.6	10.6	8.8	0.6	27.8	10.2	9.9
Middle East & North Africa	0.8	9.3	6.4	5.1	0.1	8.8	5.6	4.8
Sub-Saharan Africa	0.1	33.9	10.8	8.2	0.1	32.5	10.5	7.6
All Africa wtd. GDP PPP	0.0	26.9	11.2	8.7	0.0	26.1	9.0	7.6

SSR = Share of Tariff Lines with Specific Rates (%)

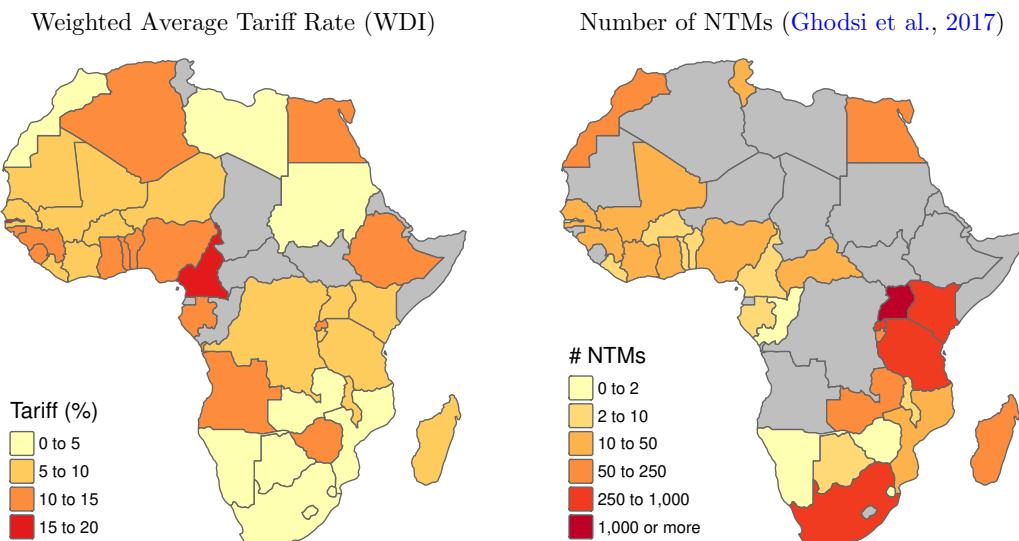
SIP = Share of Tariff Lines with International Peaks (%)

TR = Tariff Rate, Applied, Simple Mean (%)

WTR = Tariff Rate, Applied, Weighted Mean (%)

(2598) and Canada (2100). In addition, the median African country had 1 bilateral NTM with another African country (1.1 on average). Overall, this suggests that most African countries are not imposing a lot of classical NTMs; they are rather underutilizing them. Figure 7 shows the latest available estimates for tariffs and NTMs by country.

Figure 7: Trade Tariffs and Non-Tariff Measures



*Notes:* Figure shows weighted average tariff rate from the World Development Indicators and number of non-tariff measures from the WIIW NTM Database (Ghodsi et al., 2017). Latest estimates from 2019 are shown (if available).

### 3.3 Regulatory Obstacles and Border Frictions

The World Bank's Doing Business (DB) ranking recorded the time and costs (excluding tariffs) required for trading across borders<sup>12</sup> in all African countries, with the latest surveys done in 2019. The surveys record separately the time and cost in USD of documentary compliance (DC) and border compliance (BC) for both exporting and importing. Table 3 summarizes these results. Overall, SSA is at the bottom of the ranking, and the MENA region is not much advanced either. Weighting the estimates by countries' 2019 GDP in PPP terms also yields no improvements. The weighted estimates imply an average cost of 808 USD for exporting a container in Africa, most of which is payable at the border, and an average time investment of 170 hours, 81 of which are spent

<sup>12</sup><https://archive.doingbusiness.org/en/data/explore/topics/trading-across-borders>

at the border. The costs of importing are even worse, at an average of 1177 USD per container and an average time commitment of 292 hours, 124 of which are spent at the border.

Table 3: World Bank Doing Business: Trading Across Borders in 2019

Region/Income Status	Exporting				Importing			
	USD BC	Hours DC	USD BC	Hours DC	USD BC	Hours DC	USD BC	Hours DC
OECD	149.7	34.8	12.8	2.5	106.5	26.5	9.4	3.8
Non-OECD	453.8	145.0	64.0	55.8	531.5	191.4	82.5	66.5
High income	228.4	68.0	24.8	12.6	252.1	74.3	22.1	16.0
Low income	528.8	197.1	84.1	78.9	670.7	348.0	125.9	124.3
Europe & Central Asia	117.8	47.1	11.9	12.2	91.8	42.1	10.4	11.2
South Asia	310.6	157.9	53.4	73.7	472.9	261.7	85.7	93.7
Latin America & Caribbean	513.2	99.5	55.7	36.4	625.3	106.5	55.8	44.3
Middle East & North Africa	447.0	239.7	54.1	63.1	526.0	261.8	97.3	72.4
Sub-Saharan Africa	603.1	172.5	97.1	71.9	690.6	287.2	126.2	96.1
All Africa wtd. GDP PPP	628.9	178.8	89.3	80.5	717.6	459.6	167.9	123.9

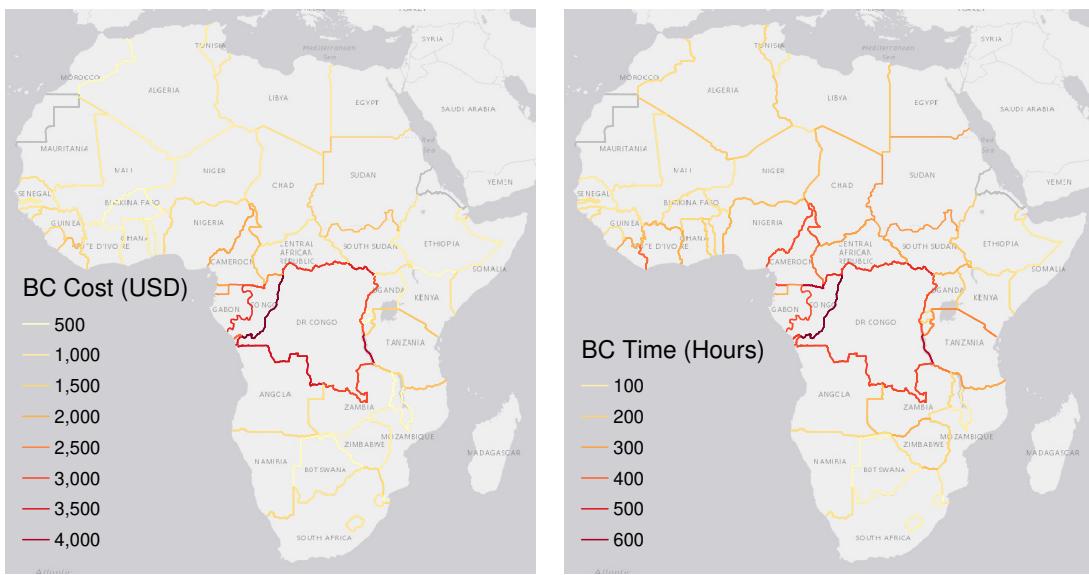
DC = Documentary Compliance, BC = Border Compliance

Figure 8 shows symmetrized exporting + importing border compliance times and costs for each border. In central African countries such as Congo, the costs are exorbitant, whereas particularly in southern Africa, they are quite efficient, coming as low as 509 USD for trades across the Rwanda-Burundi border and 40 hours for trades between Mozambique and Eswatini.

Figure 8: World Bank Doing Business: Border Compliance Frictions in 2019

Border Compliance Cost, Median: 1222\$

Border Compliance Time, Median: 224h



Notes: Figure shows average (bi-directional) 2019 border compliance cost/time of trading a standardized product.

### 3.4 Border Frictions Adjusted Routes

It is difficult to take tariffs and NTMs into account when evaluating transport networks and market access because their effects depend on the value, type, and quantity of products being traded. In addition, AfCFTA has set clear aims to eliminate most inner-African tariffs in the foreseeable future. Regulatory and border frictions on the other hand are more uniformly applicable.

A critical question then is how the estimates from Table 3 can be sensibly added to road distance and travel time estimates. Surely time spent preparing documents or waiting at the border is

different from driving time. To better understand the relative cost of these times, I collect data on domestic transport costs from the 2019 DB survey for a sample of 23 (mostly landlocked) African economies where exporting/importing involved domestic transportation of 100km or more. This information is not standardized across countries and thus not published in the DB indicators. For each country surveyed, the distance in km to the border, the time in hours to the border, and the cost in USD of exporting/importing a representative product are recorded alongside the published border and documentary compliance times and costs.

To estimate the average cost of an hour spent on the road, I regress the cost of the trip on the travel time in hours while controlling for the log of the velocity and a dummy indicating whether the transport was import related.<sup>13</sup> I do the same for the distance traveled. Since I am more interested in converting times into each other rather than converting cost to time, I also regress the cost of border and documentary compliance from Table 3 on the respective time required using data for all 53 available African economies. Table 4 reports the results.

Table 4: Relative Cost of Exporting/Importing Times

Cost in USD: Model:	Transport (1)	Border (3)	Documentary (4)
Constant	-711.6** (285.9)	276.7 (317.4)	217.8*** (65.28)
Travel Time (Hours)	15.79*** (2.074)		
Travel Distance (Km)		1.560*** (0.2987)	
Border Time (Hours)			3.988*** (0.6781)
Documentary Time (Hours)			1.655*** (0.3954)
Log Velocity (Km/Hour)	346.3*** (85.84)	-43.95 (102.9)	
Import Dummy	106.6 (133.3)	107.7 (135.9)	-44.92 (66.38)
Observations	46	46	106
R <sup>2</sup>	0.448	0.415	0.471
Adjusted R <sup>2</sup>	0.408	0.373	0.461

*Heteroskedasticity-robust standard-errors in parentheses.*

*Signif. Codes:* \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

Evidently, time spent on the road, valued at 15.8 USD per hour in the average African country, is more costly than time spent at the border (4\$/h) or filing documents (1.7\$/h). The estimates imply that in the average African country, road time is  $15.79/3.988 = 3.959 \approx 4$  times more costly than border time, and  $15.79/1.655 = 9.540785 \approx 10$  times more costly than documentary time. To create the NTB-adjusted travel time matrix, I divide border compliance time by 4 and documentary compliance time by 10 before and add them to the OSRM estimates. Using the coefficient on distance indicating an average cost of 1.6\$/km, I also obtain an adjusted road distance matrix by dividing documentary and border compliance costs by 1.6 and adding them to the OSRM distance matrix. These estimates are roughly consistent with Djankov et al. (2010)'s result that a day of border delay is equivalent to a country distancing itself from its trade partners by  $\sim 70$ km since  $24h \times 4\$/h \div 1.6\$/km = 60$ km.

This accounting likely underestimates the true extent of frictions as I only convert time/cost to travel time/distance rather than converting both measures. Table 4 would allow converting both, but I suspect cost estimates to include labor time. Transportation economics also typically considers transport costs a function of distance rather than time whereas for personal transport

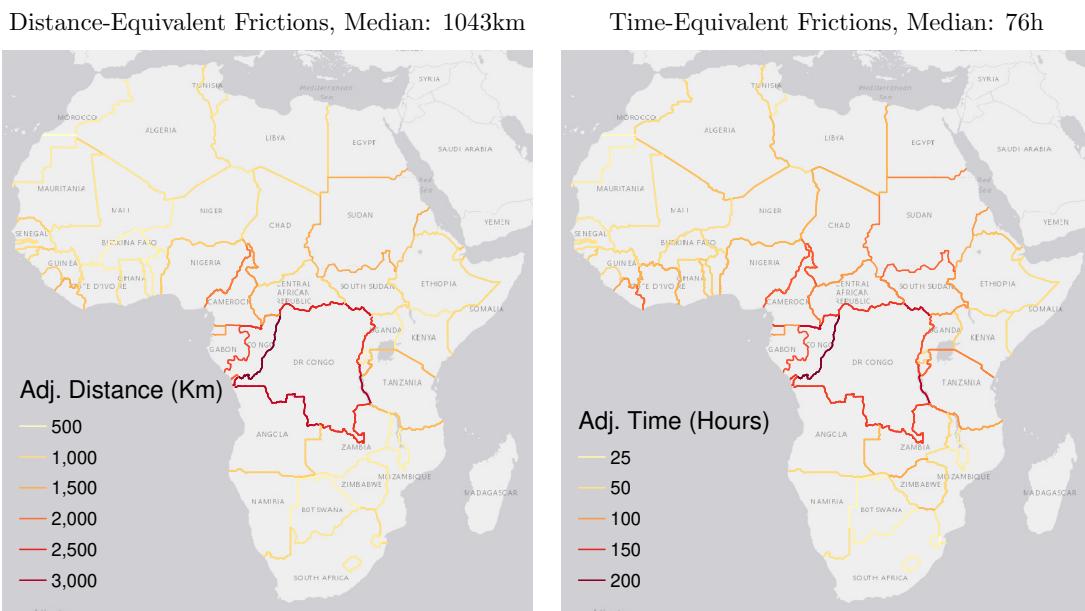
<sup>13</sup>I also test the interaction of travel time and the import dummy, but the term is insignificant; thus export and import time do not differ in terms of cost.

time may be more important than costs. Having two different adjusted MA measures can thus capture different aspects of frictions. Another consideration is that OSM travel speeds are too fast since they ignore congestion; thus, very large time-denominated frictions would overstate their effects. African borders may also have improved since 2019. In any case, I find that the frictions accounted in this way are sufficient to dramatically impede MA.

It remains to specify how to adjust routes between cells that are not in neighboring countries. The DB surveys do not give information about the time and cost of transiting through a country. Efficient calculation of multiple routes with OSRM also does not return the geometries of routes but only the total distance and travel time. I thus opt for a simple approach and stipulate a lump-sum transit time of 2 days on all routes transiting through 3rd countries. Using the estimates from Table 4, this yields a  $48\text{h} / 4\$/\text{h} = 12\text{h}$  increase in travel time and or a  $48\text{h} \times 4\$/\text{h} \div 1.6\$/\text{km} = 120\text{km}$  increase in road distance. The reader scrutinizing this generic adjustment should remember that long routes do not impact market access calculations much.

Figure 9 shows the frictions denominated in travel distance and time implied by each border crossing - again after imposing symmetry regarding the direction of travel. The median African border imposes an additional cost equivalent to 1043km of road travel or  $76\text{h} = 3.2$  road days.

Figure 9: Road Distance and Time Equivalent 2019 Border Frictions



*Notes:* Figure shows average (bi-directional) 2019 border and documentary compliance cost/time converted to road distance/travel time following Table 4, i.e., dividing costs by 1.6 and documentary/border times by 10/4, respectively.

### 3.5 Border Frictions Adjusted Market Access Maps

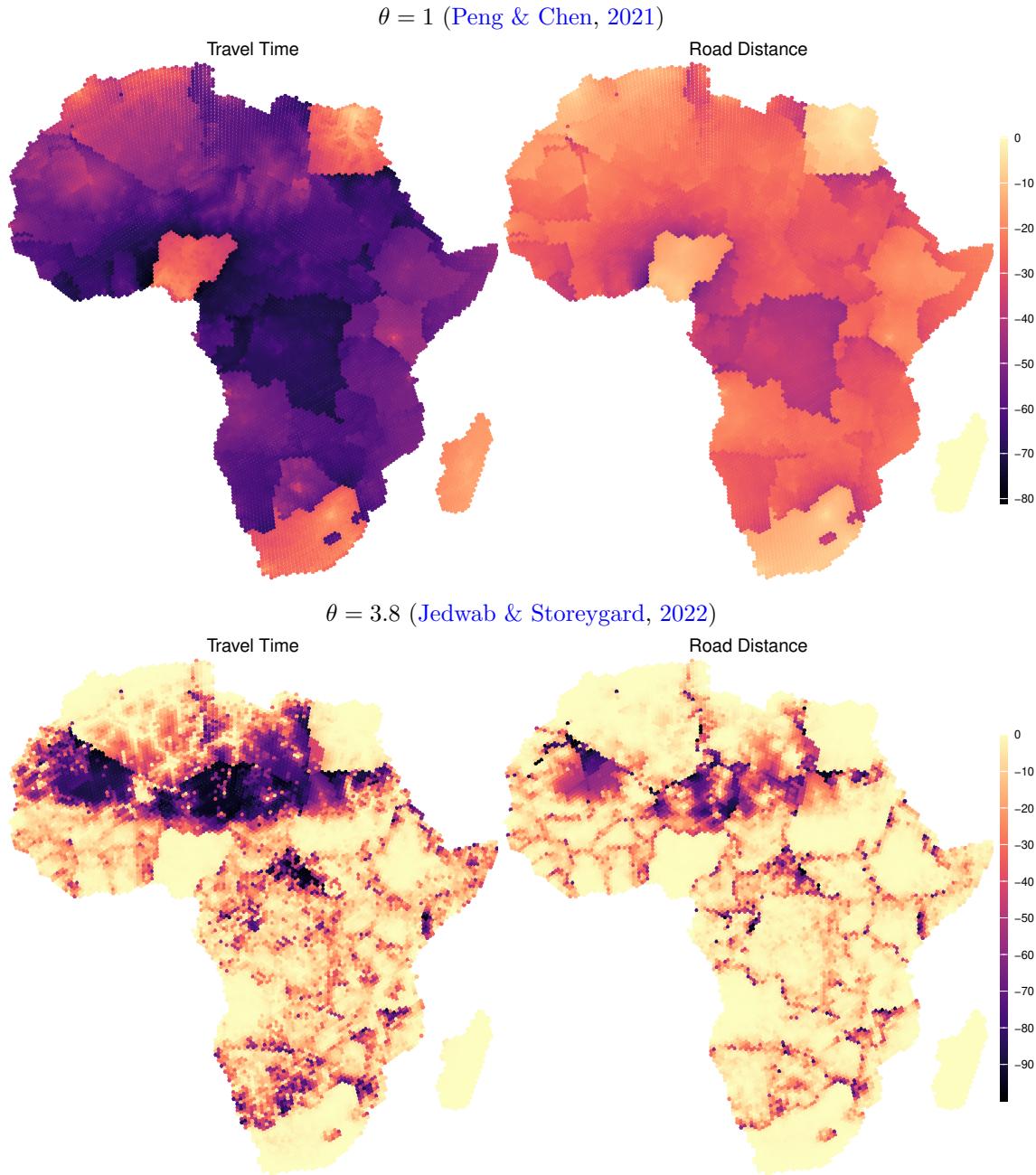
Appendix Figure A5 shows the frictions adjusted market access maps using GDP, and Appendix Figure A6 with IWI-based wealth. They indicate that, especially when  $\theta$  is low, border frictions significantly obstruct access to markets in neighboring countries. The adverse effects are especially strong for cells surrounding major African markets such as Nigeria, Egypt, and South Africa.

Summing together the MA of all locations, I estimate the total cost of these frictions. Under  $\theta = 1$ , the estimated travel times imply a total MA of 980 billion USD/min in 2015 PPP terms, which is reduced by 52.3% to 467 billion USD/min under the 2019 trade frictions. Considering total IWI-based wealth (Figure A6) yields a similar MA reduction by 53.8%. When considering only distance, with  $\theta = 1$  MA is 922 billion USD/km, and reduced by 25.5% to 687 billion USD/km under 2019 frictions. Using wealth yields a similar reduction of 27.2%. Thus, under a low trade elasticity, the African continental market could be up to twice its effective size, even with the current transport network, if regulatory and border frictions impeding cross-border trading could be eliminated. Under  $\theta = 3.8$ , the reduction in MA shrinks to 1.13%-1.19% in all cases, implying

that with a high trade elasticity, cross-border frictions become largely irrelevant. Evidence on inter-city migration and by Peng & Chen (2021) suggests that a lower elasticity is more plausible.

Cell-level MA losses due to border frictions range between -6.6% and -81.1% in GDP/min terms with  $\theta = 1$ , and between -0.1% and -99.82% with  $\theta = 3.8$ . Figure 10 shows their spatial distribution. Under  $\theta = 1$ , large parts of central Africa incur 50-80% time- and 20-50% distance-denominated losses. Cells across the border from major markets also incur very large losses. Under  $\theta = 3.8$ , losses are large in some cells close to borders but negligible in most populated cells.

Figure 10: Market Access Loss (%) from Border Frictions using GDP in 2015 USD PPP



*Notes:* Figure shows local MA %-loss from frictions (using Eq. 4 where  $L_j$  is cell-GDP from Kummu et al. (2018)).

In summary, assuming  $\theta \ll 3.8$ , many regions could gain significant MA from lower border frictions. In the following sections, I show that these frictions also reduce the value of (trans-African) road investments. Thus, their reduction should be a top policy priority. To simulate such investments, however, I first need to create suitable graph representations of Africa's road network.

## 4 Optimal Network Representations and Extensions

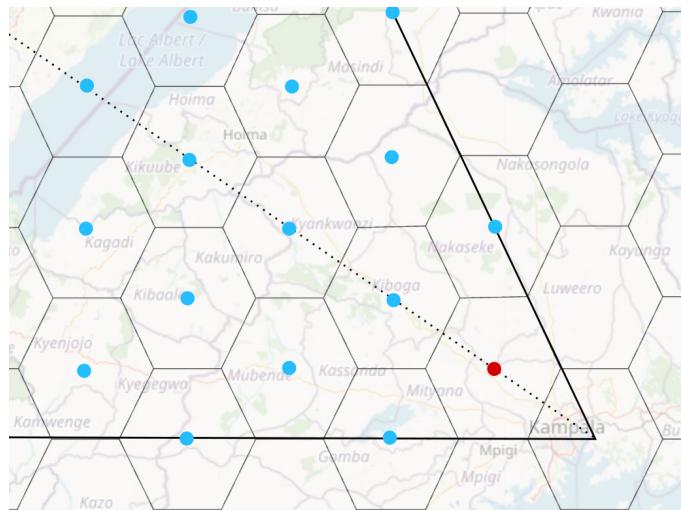
The central aim of this paper is to provide quantitative evidence on which roads should be enhanced or built. This requires accurate representations of the road network via graphs to simulate the effects of network investments. Most research creates these graphs in an ad-hoc fashion by either connecting adjacent cells in a gridded representation and parameterizing them by the connectivity along these links, or by choosing the largest roads and their intersections as nodes/edges of the graph. This section instead uses a routing engine (OSRM) to create graphs that optimally represent the road network as interpreted by the engine, implying that the optimal network graph produces road distance/travel time estimates that closely approximate those of the routing engine.

Section 4.1 demonstrates that it is possible to reduce the matrix of 144 million routes to a graph network with  $\sim 12K$  nodes and  $\sim 70K$  edges that reasonably approximates it. Computational constraints and tractability warrant a further reduction of the network to specific roads connecting important cities and ports. In Section 4.2, I thus also create a reduced (sparse) graph precisely representing the transport network. I then extend this graph with high-value new links connecting previously unconnected nodes (Section 4.3) and map out major transport routes (Section 4.4).

### 4.1 An Optimal Graph of Africa's Entire Road Network

Previous research such as [Graff \(2024\)](#) has created a graph representation of Africa's road network by dividing the continent into cells and computing routes connecting adjacent cells. However, cell centroids may not be of economic interest or well-connected at all. Using the full matrix of trans-African routes, I can empirically improve on this design by computing optimal edge weights that take routes to cells in the same direction as the neighboring cell into account.

Figure 11: Optimal Graph Representation of Africa's Road Network



*Notes:* Figure shows 50.28km ISEA3H grid covering Uganda. In an optimal directed graph, the weight of the edge connecting Kampala to the red dot takes into account routes to all cells in this direction (red and blue dots).

Figure 11 illustrates this idea using Uganda as an example. Suppose I want to optimally weight the link from the center of Kampala to the cell with the red dot. I could simply take the road distance from the center of Kampala to the red dot, and then from the red dot to the next blue dot, etc., but then adding those distances may not well reflect routes from Kampala to blue dots. Thus, I let the direct OSRM routes to the blue dots also inform the weight attached to the link from Kampala to the red dot. Specifically, I calculate a directional route/time efficiency

$$\kappa_{ik}^m = \frac{\sum_j \delta_{ij}^m s_{ij}^{-\gamma}}{\sum_j s_{ij}^{-\gamma}} \quad \forall i, j, k \quad \text{where } \alpha_{ik,ij} < \alpha \quad \text{and } i \neq j, \quad m \in r, t \quad (5)$$

for all neighbouring cells  $i$  and  $k$ .  $\gamma$  is a parameter that governs spatial decay and  $\alpha$  the (two-sided) angle up to which further points  $j$  (red dot  $k$  itself or blue dots in the direction of  $k$ ) are considered.

Using spherical distances  $a = i \rightarrow k$  (Kampala  $\rightarrow$  red dot),  $b = i \rightarrow j$  (Kampala  $\rightarrow$  blue dot  $j$ ), and  $c = k \rightarrow j$  (red dot  $\rightarrow$  blue dot  $j$ ) the angle  $\alpha_{ik,ij}$  is calculated as

$$\alpha_{ik,ij} = \alpha_{a,b} = \arccos\left(\frac{a^2 + b^2 - c^2}{2ab}\right) \frac{180}{\pi} \quad (\text{Triangle Equation}). \quad (6)$$

The edge  $ik$  then receives a weight  $\omega_{ik}^m = \kappa_{ik}^m s_{ik}$ . I numerically determine  $\alpha$  and  $\gamma$  using a quasi-newton algorithm by computing all shortest paths on the graph (Floyd-Warshall algorithm) and comparing it to the full 144 million road distances/travel times matrix generated by OSRM. To limit the influence of 'remote regions' on the optimal parameter choice, I exclude Madagascar and only use routes between cells where the OSRM start/end points lie within the cell.<sup>14</sup>

For travel time, this yields optimal values  $\alpha^* = 29.88^\circ$  and  $\gamma^* = 1$ . The average shortest path takes  $1.38\times$  longer than the actual route versus  $2.26\times$  longer using a simple graph a la [Graff \(2024\)](#). Computing an inverse-squared distance weighted average of these ratios yields 0.93 for the optimized graph and 1.59 for the simple graph, implying that the former is slightly more efficient than the actual network at short distances to generate more realistic estimates on longer routes. When using simple inverse distance weights, the simple/optimized graphs are 2.08/1.25 times less efficient than the real network. [Wolfram \(2021\)](#) documents that road networks are more efficient on longer routes. Thus, to optimally represent a real network by a dense graph, the graph edges need to be slightly more efficient than reality.<sup>15</sup> The optimal graph, which excludes Madagascar, has 11,582 nodes and 68,338 edges. Appendix Figure A8 plots it with duration edge weights.

## 4.2 An Optimal Graph of Africa's Transport Network

A sparse network graph is also needed to reduce computational complexity and increase interpretability. I construct it by sampling important cities and ports and letting OSRM compute routes between them, which I then process into an accurate graph. To sample cities, I use the updated (March 2024) world cities database from [simplemaps.com](#) sourced from the US National Geospatial Intelligence Agency.<sup>16</sup> I derive, in multiple steps,<sup>17</sup> the largest cities with a population of more than 100,000 within a radius of 70km in continental Africa.<sup>18</sup> This yields 447 cities. I add populations of smaller cities within 30km of these cities to create agglomeration data. In addition, I use data on international ports compiled by the World Bank<sup>19</sup> to select ports with a deployed capacity of more than 100,000 TEU in Q1 of 2020, which are 57 in Africa. From these, I take the largest within 100km, yielding 52 large ports at least 100km apart. I then match these ports to my 447 large cities within a 30km radius and am able to match 46. To the remaining 6 (among which Walvis Bay in Namibia), I add the population of cities/towns within a 40km radius (including Swakobsmund for Walvis Bay) to create more loosely defined 'port cities'. In total, I obtain  $447 + 6 = 453$  important (port-)cities in continental Africa. Figures 12 and 14 plot them.

I then let OSRM compute precise routes between all 453 (port-)cities that are spherically less than 2000km apart. This yields 25,742 routes, shown on the LHS of Figure 12. To generate an undirected graph representation, I apply several algorithms<sup>20</sup> to intersect and segmentize the

<sup>14</sup>In particular, on this subset of 'well-defined routes,' I minimize a simple average of the ratios of the reconstructed route lengths/durations to the real ones, restricting  $\gamma \in [1, 5]$  and  $\alpha \in [1^\circ, 40^\circ]$ .

<sup>15</sup>As a robustness exercise, I also minimize a weighted average across route-ratios with inverse squared spherical distance weights. This does not significantly alter the results,  $\gamma^*$  remains 1 and  $\alpha^*$  increases slightly to  $30.07^\circ$ . Thus, this representation seems optimal for both medium and long routes.

<sup>16</sup>The Africapolis project (<https://africapolis.org/>) is more prevalent among researchers but does not provide precise city centroids, and the 2015 edition available at the time of writing is a bit dated.

<sup>17</sup>Concretely, I first find the largest cities within a 30km radius through an iterative algorithm that finds the largest city within 30km of each city, drops all non-selected cities, and repeats. I weight the population of cities with a higher administrative function by 5, such that national and administrative capitals are selected unless they are more than 5 times smaller than close cities. I repeat this process with an increased radius of 50km. Then, I add the populations of all cities within 30km of the selected (largest in 50km) cities to create urban agglomerations. I then remove agglomerations with a population below 100,000 and apply the algorithm one last time to find the largest agglomerations within 70km. In this last run, no additional weight is given to cities with higher administrative functions, i.e., only total population is used to select the largest urban agglomerations within 70km.

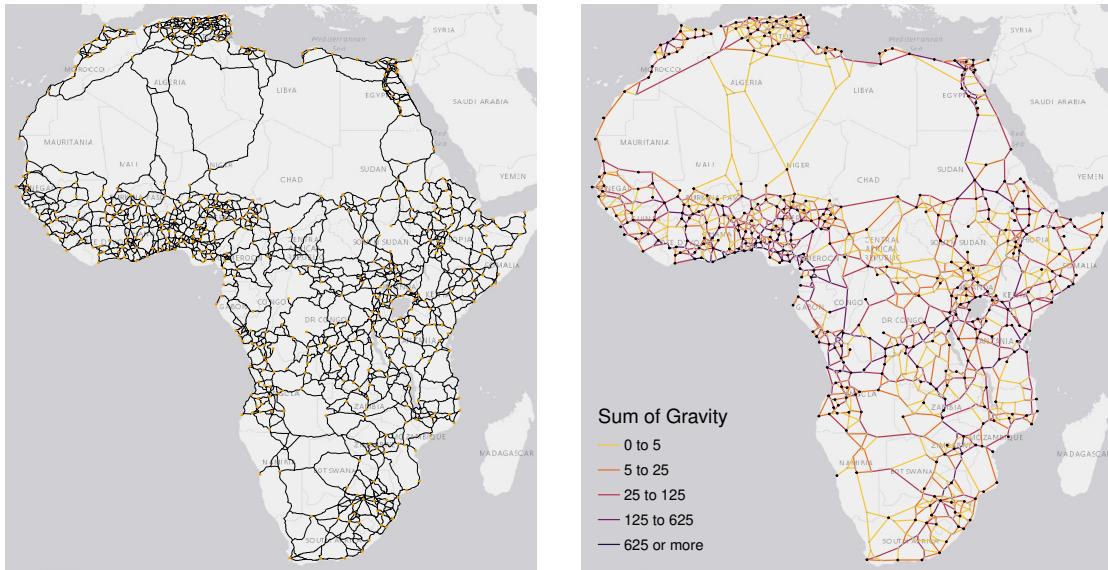
<sup>18</sup>I.e., excluding small island states, Madagascar, and also excluding cities not connected to the mainland via roads, such as Zanzibar City.

<sup>19</sup><https://datacatalog.worldbank.org/search/dataset/0038118/Global—International-Ports>

<sup>20</sup>Available in R packages *rmapshaper*, *stplanr*, *sfnetworks*, *tidygraph/igraph*, and *dbscan*.

routes, smooth them, and finally contract intersection points within 30km of each city/node.<sup>21</sup> This yields a discretized network representation comprising 1,379 nodes and 2,344 edges - shown on the RHS of Figure 12. Appendix Figure A9 provides an overlay version of Figure 12, illustrating the accuracy of the network representation thus created.

Figure 12: Computing an Optimal Graph of Africa's Transport Network  
25,742 Routes Connecting 453 (Port-)Cities



*Notes:* The LHS shows 25,742 fastest car routes connecting all 453 (port-)cities less than 2000km spherical distance apart. The RHS shows an undirected graph comprising 1,379 nodes and 2,344 edges generated from these intersecting routes.

The overall length of the transport network is 315,000km. The edges/links on the RHS of Figure 12 are colored by the appropriately scaled 'sum of gravity' across all 25,742 routes. Each route connects two cities  $i$  and  $j$  with populations  $P_i$  and  $P_j$  and road distance  $D_{ij}$  (in km) between them. Following Poot et al. (2016), migration flow can be modeled as

$$M_{ij} = G \frac{P_i^\alpha P_j^\beta}{D_{ij}^\gamma}. \quad (7)$$

In their New Zealand case study, the parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  are all around 0.8-0.9. A famous early study by Zipf (1946) on US inter-city migration showed that  $\alpha = \beta = \gamma = 1$  gives a good fit to the data. For simplicity, I adopt this specification and set  $G = 1e-9$  to reduce the size of the estimates. The RHS of Figure 12 shows the sum of  $M_{ij}$  across all routes using the link. Darker edges are thus expected to receive higher traffic volumes. The highest volumes are expected in the Nile Delta and along the West African coastline. Due to the border closure between Algeria and Morocco since 1994, OSRM routes all car traffic between both countries through Western Sahara, yielding an unrealistic medium-sized gravity-implied flow along this link.

To parameterize the network, I add city populations within 30km of the  $1379 - 453 = 926$  navigational nodes, yielding 530 or 57% of navigational nodes with populations. Figure 15 shows these as small dark nodes. In total,  $453 + 530 = 983$  nodes, comprising 71.3% of the 1,379 nodes, are associated with cities. Compared to the initial network of 453 (port-)cities (large dark nodes), the population represented by this network increases by 14.7% to 444.2 million, which, according to the simplemaps world cities dataset, represents 86.3% of the African city population of 515 million.<sup>22</sup> I also use OSRM to obtain road distance and travel time along each link, yielding edge weights representing the real road network. Figure 15 visualizes the corresponding segments.

<sup>21</sup>In the contraction step I first contract all nodes/intersections within 30km of the 453 cities and ports to the respective city/port centroid. I then cluster the remaining nodes with a 30km radius using density-based spatial clustering and contract these clusters to the most important node, which is the node used by the routes with the greatest 'gravity implied flows,' as further elaborated below.

<sup>22</sup>The Africapolis 2015 estimate sums to 574 million across 9186 urban agglomerations, versus 4339 cities in the simplemaps dataset summing to 515 million. I use simplemaps because it is more up-to-date and precise.

### 4.3 Optimal Network Extensions

Africa's transport network is underdeveloped, and, with an average route efficiency of 0.68 vs. 0.84 for the US, far from ideal. To characterize optimal network investments on a sparse graph, including the possibility of new roads, I first need to find sensible new links and add them to the graph as potential (presently impassable) edges. I devise a network finding algorithm tailored to this task and establish its properties by letting it draw an ideal hypothetical transport network. I then apply it to the transport network graph to select high-value new links.

The algorithm begins with a complete graph where all nodes are connected and eliminates edges intercepted by other nodes which should be connected first. I illustrate it by letting it generate an ideal network between the 453 (port-)cities with a geodesic US route efficiency of  $\geq 0.843$  and a maximum  $20^\circ$  deviation from the direction of travel. Figure 13 helps explain the algorithm. It first computes the geodesic distance between some point A and another point B. Then, it computes the distance from A to all other points (cities), such as C, D, E, and F, and, using the triangle equation (Eq. 6), the angle  $\alpha$  between AB and AC, AD, AE, and AF. It then drops points where the distance is greater than AB or  $\alpha > 20^\circ$ . For the remaining points (C and F), it computes the distance ACB and AFB, and compares ratios AB/ACB and AB/AFB to the US route efficiency measure of 0.843. If either C or F are found route-efficient, the algorithm concludes that there should be no direct AB link because these points should be connected via C or F. The same process is repeated taking point B as a vantage point, such that if the algorithm finds either AB or BA intercepted by other route-efficient points, this link is dropped from the undirected graph.

Figure 13: Network Finding Algorithm Illustration

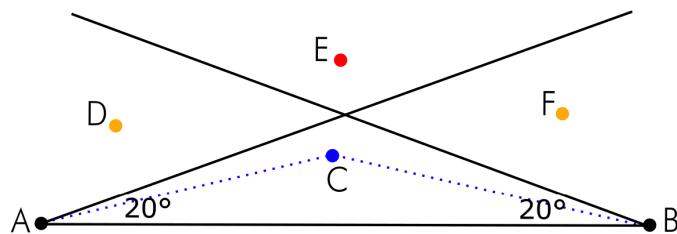


Figure 14 shows the thus generated US-grade hypothetical network connecting the 453 (port-)cities. It has 1459 edges, some of which intersect and should be broken up further.

Figure 14: Demonstration of Network Finding Algorithm

543 (Port-)City Sizes (Population)



US-Grade Hypothetical Network

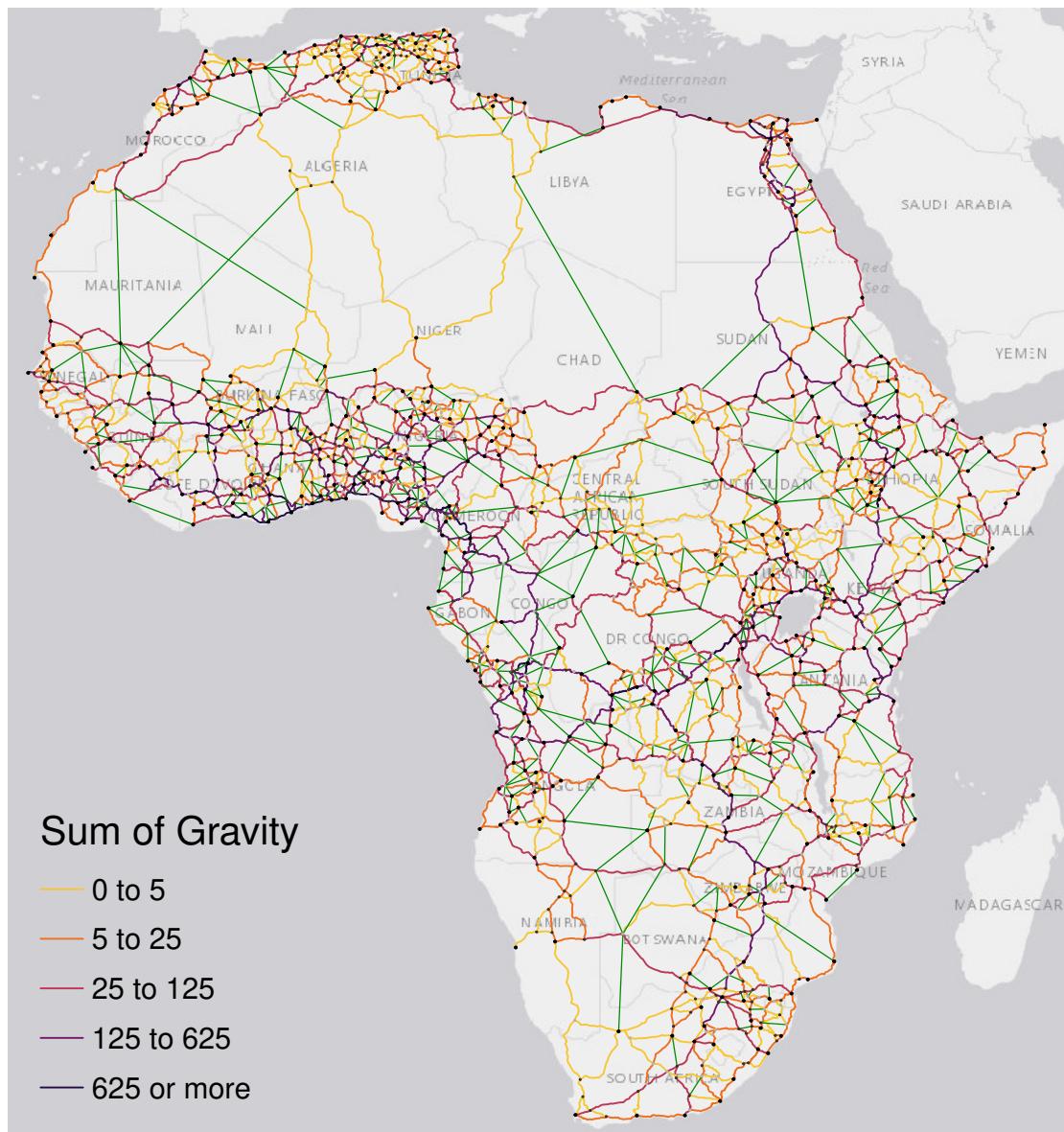


*Notes:* The LHS shows the 453 (port-)cities; circles reflect population. The RHS shows the result of applying the network finding algorithm described above, requiring US route efficiency of 0.834 (Wolfram, 2021) and a maximum  $20^\circ$  deviation.

To propose optimal network extensions, I apply the algorithm to a complete transport network graph where all 1379 nodes are fully connected. Using the US network efficiency and  $20^\circ$  angle criterion as illustrated above yields a staggering 2,031 new edges, shown in Appendix Figure A10. Since this is close to the number of existing edges, and many of the new links overlap, I adjust the parameters of the algorithm to generate fewer potential links. Concretely, I increase the maximum angle for intercepting nodes to  $\alpha = 45^\circ$ , and use the European route efficiency estimate of 0.767 by Wolfram (2021), which is lower than the US value of 0.843 partly due to Europe's more fragmented (water-bound) geography. In addition, I let OSRM compute a full road distance matrix between all 1,379 nodes and remove connections that already have a route efficiency of  $2/3$  or higher. The 516 links retained by this modified algorithm, if straight, thus reduce travel distance by  $\geq 50\%$ .

I then verify whether these links can be constructed and remove 46 links significantly intercepted by waterbodies or protected areas. Some proposed links across the closed Morocco-Algeria border already exist. I keep them and reduce their construction costs in the network optimization process. Finally, I manually add 11 links that I consider important for improved transport flow. In total, I retain 481 new links. Figure 15 draws them as thin green lines. Exempting one crossing in northern Mali, where adding a node would add no value for routing purposes, and a similar case in central Kenya, these links don't intersect and thus imply no additional nodes to the graph.

Figure 15: Discretized and Calibrated African Transport Network incl. High-Value New Links



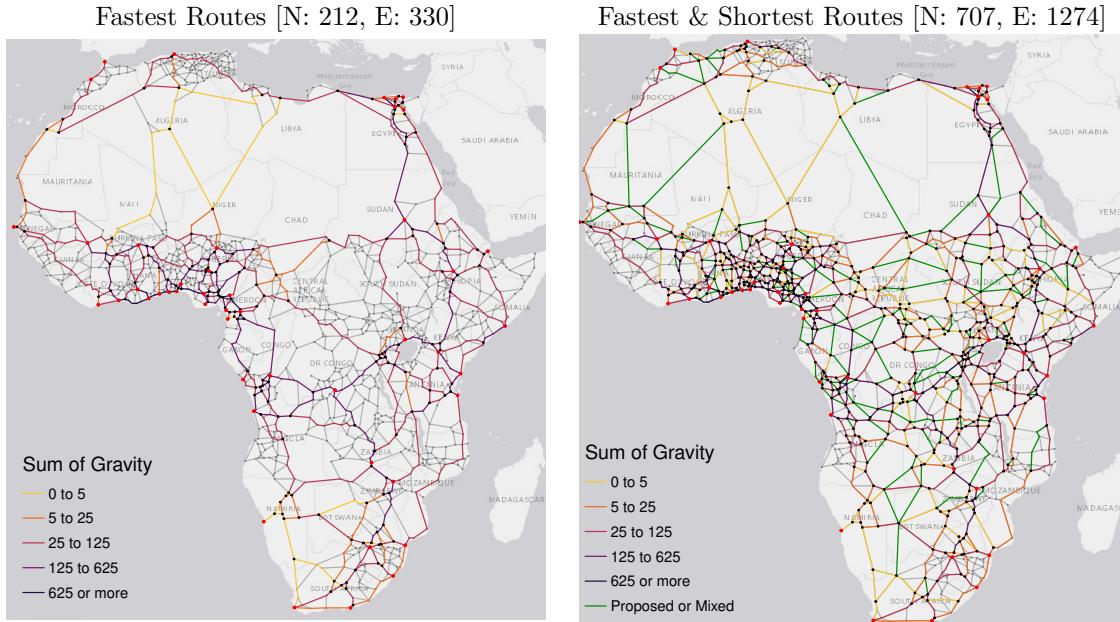
*Notes:* Figure shows optimal graph with 1,379 nodes and 2,344 edges (= links/road segments), including 481 feasible new links proposed by the network finding algorithm. The sum of gravity (Eq. 7) is calculated across all routes using a link.

In reality, roads constructed along these links will not be perfectly straight. I assume that new links have the same average route efficiency as existing links, which is 0.83.<sup>23</sup> Thus, I divide the spherical distance spanned by new links by 0.83 to estimate their road distance (edge weight), yielding 104,000km in total. In the remainder of this paper, I show simplified graphs with straight edges in all smaller plots to make them more legible. All computations, however, use edge weights corresponding to actual road distances and travel times, with existing network as in Figure 15.

#### 4.4 Optimally Connecting Major Cities and Ports

Of particular interest to trans-African policymakers and traders are important transport routes connecting major (port-)cities across the continent. Affording special consideration to these, particularly through general equilibrium global optimization in Section 6, requires a further network reduction. To elicit these routes, I consider the subset of major agglomerations with  $>2$  million inhabitants and ports with  $>1$  million TEU in 2020Q1, yielding 47 major (port-)cities. I then use the parameterized network developed above to compute three sets of shortest paths between these (port-)cities: (1) fastest routes along the existing network, (2) shortest routes along the existing network, and (3) shortest routes including proposed links at route efficiency of 0.83. Figure 16 shows the outcome after (1) and the union of (1)-(3). In both cases, I contract the network, removing unnecessary intermediate nodes. Since the original graph optimally connects 453 (port-)cities, the reduced graph also optimally connects the 47 largest (port-)cities.

Figure 16: Optimally Connecting 47 Major (Port-)Cities



*Notes:* The LHS shows the fastest routes connecting 47 major (port-)cities with  $>2$  million people or  $>1$  million TEU in 2020Q1, yielding a reduced graph. The RHS shows an equivalent graph accommodating both fastest and shortest routes.

The reason for adding (2) and (3) to the network is to provide the GE planner of Section 6 with additional choices for roads to build or improve to enhance connectivity between major (port-)cities. As evident from the RHS of Figure 16, the planner has quite a few options to do that, the retained network has half the size of the full transport network.

Having thus created detailed graph representations of Africa's road network, I now examine counterfactual investments in the network and compare them in search of optimality. Section 5 characterizes MA-maximizing investments in partial equilibrium, including cost-benefit analysis, whereas Section 6 computes welfare-maximizing investments in spatial general equilibrium.

<sup>23</sup>The route efficiency across individual links of 0.83 is larger than the overall network route efficiency of 0.68 because the latter considers full trans-African routes comprising multiple links.

## 5 Optimal Network Investments in Partial Equilibrium

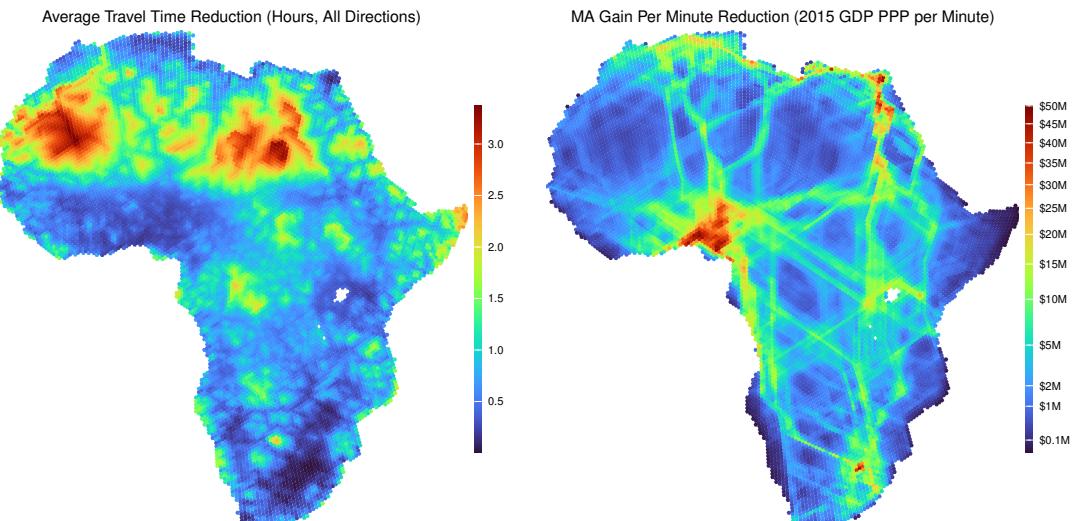
This section characterizes optimal road network investments in partial equilibrium (PE). PE estimates capture the effects of changes to the network on overall network efficiency and continental MA, keeping fixed the distribution of economic activity. Via shortest-path algorithms, PE estimates account for traders' reactions to such changes but ignore any resulting long-term shifts in activity. PE tools can also be used to compare investments in individual links or predefined investment packages. They cannot determine how a given infrastructure budget should be optimally spent on improving the network - a case considered in Section 6 characterizing optimal network investments in spatial general equilibrium (GE). This section, however, reaches important conclusions using much fewer assumptions than required for GE simulation and offers some insightful cost-benefit analyses. The PE objective of MA maximization is also fundamentally distinct from welfare maximization considered in GE. Both objectives and approaches thus produce complementary insights, and policymakers are advised to consider both rather than relying on a single approach, particularly since many difficult-to-estimate parameters are needed to produce GE results.

### 5.1 Local Road Network Investments

On the entire network graph (Section 4.1 and Figure A8), I simulate MA gains from small (local) road investments by enhancing the network around each node, one node at a time, and computing continental MA gains resulting from that. In the duration weighted graph, I adjust the travel time on all links such that  $\min(\kappa_{ik}^t) = 50\text{km/h}$ , implying  $t'_{ik} = \min(t_{ik}, s_{ik}/50\text{km/h})$  for all  $k$  bordering  $i$ . 50km/h is a sensible upper bound for time efficiency as 99% of links are slower. [Wolfram \(2021\)](#) also finds very few US routes with time efficiency greater than 50km/h.

The LHS of Figure 17 shows average travel time reductions from these local investments. It ranges between 0 and 3.5 hours, which is sensible as links are 50.3km long, and a minimum travel speed of 10km/h was assumed. The LHS is thus a local time efficiency improvement potential map and summarizes the underlying graph. The RHS shows the MA gain from each local improvement divided by the investment, i.e., by the LHS. Each data point combines MA gains measured across all 11582<sup>2</sup> routes; thus, it required 11582<sup>3</sup> shortest paths, about 1.6 trillion, to produce the plot. It indicates that MA gains per minute reduction are highest in large city centers, ranging up to 50 million USD'15. Large MA gains from local travel time reductions are evident all across Nigeria, the Nile Delta down to Khartoum, and greater Johannesburg. More interestingly, marginal MA gains are also high when connectivity between the major agglomerations like Cairo, Nigeria, the Great Lakes region, and Johannesburg is improved. The bright lines map out areas where the existing road network is already somewhat dense. Thus, the map suggests transport corridors in a purely data-driven way (although marginal PE gains cannot support large-scale investments).

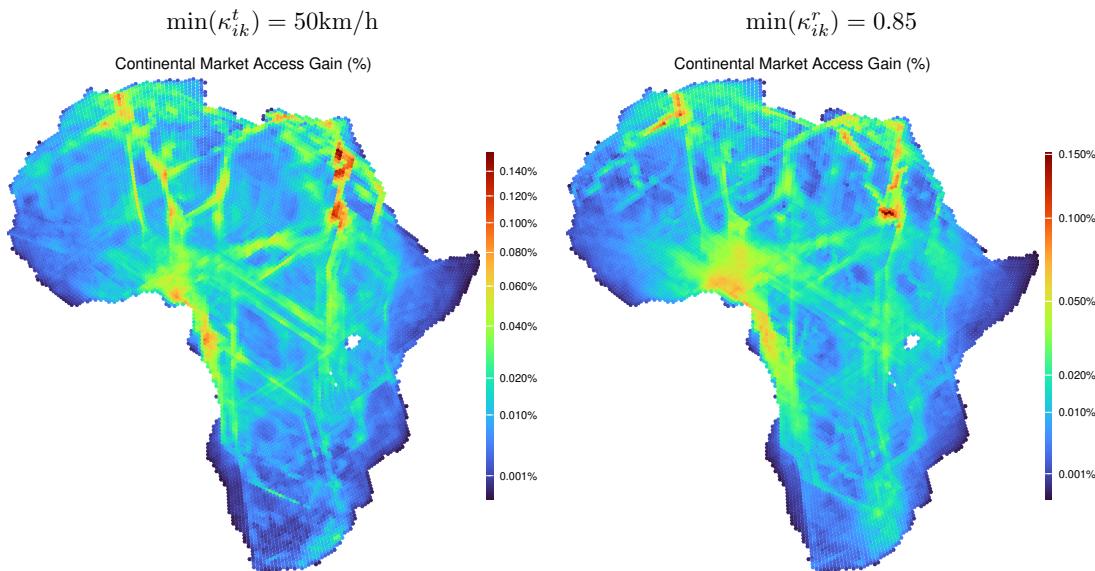
Figure 17: Local Travel Time Reduction:  $\min(\kappa_{ik}^t) = 50\text{km/h}$



*Notes:* The LHS shows the average travel time reduction across links connecting to a node from increasing their time efficiency (Eq. 2) to  $\geq 50\text{km/h}$ . The RHS shows the resulting marginal MA gain measured (summed) across all trips.

A caveat is that the map does not consider MA through ports, which will only be considered in GE with several goods. Appendix Figure A11 shows the graph computed for a local road distance reduction, imposing  $\min(\kappa_{ik}^r) = 0.85$ . It is similar but slightly more diffuse, reflecting greater local variation in route efficiency. Further insights can be gained refraining from division by the investment and just plotting the total MA gains from heterogeneous investments in Figure 18. This map shows, for both route and time efficiency, that the largest local MA gains can be reaped by improving connectivity along the corridor between Kinshasa and Lagos, around Khartoum and upwards in the Nile Delta, and by reopening the border between Morocco and Algeria, which is closed since 1994. Simulations were also done with border frictions but did not yield interesting results at this fine local scale. Frictions do however have significant effects on marginal investments in the sparse transport network graph which I now turn to.

Figure 18: Total MA Gain from Heterogeneous Local Investments



*Notes:* Figure shows total MA gains measured (summed) across all trips from increasing local time efficiency to  $\geq 50\text{km/h}$  (LHS) and local route efficiency to  $\geq 0.85$  (RHS). All links/edges connecting to a node are improved.

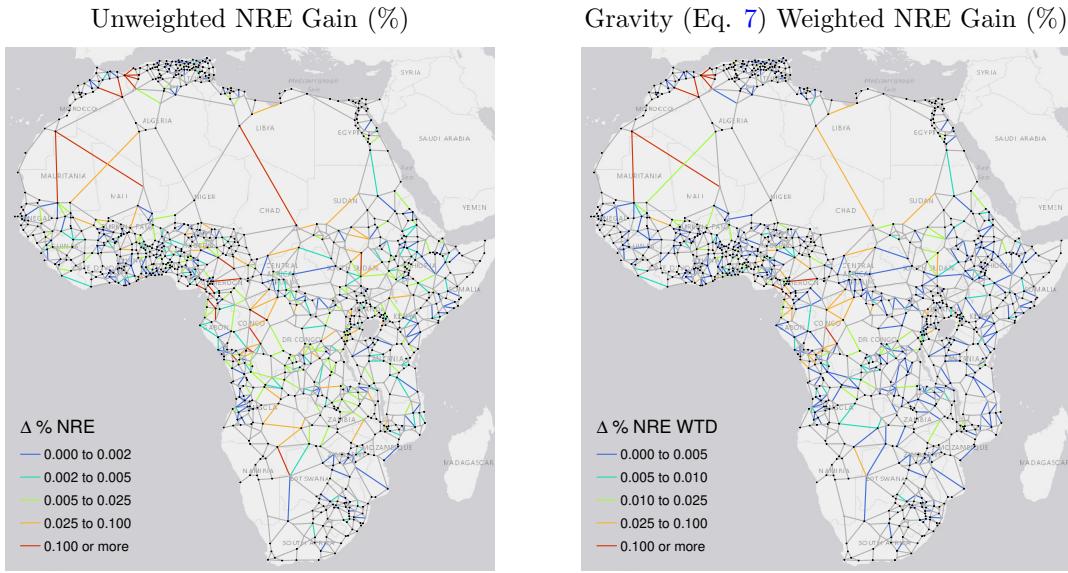
## 5.2 Investments in New Links

Considering the full transport network, I begin with a set of transport-oriented simulations to gauge the value of adding the 481 proposed links. Computing the route efficiency across all routes with and without the new links yields a 7.1% increase in average direct RE, from 0.68 to 0.72. When weighting routes by gravity (Eq. 7 but with the spherical distance), RE increases by 5% from 0.73 to 0.76, suggesting that travel between populated areas is more efficient (ignoring congestion).

By virtue of the smaller network size, I can also run these computations on a link-by-link basis by adding a link to the network, recomputing all shortest paths, and then computing the percentage change to the existing RE. Figure 19 shows the outcome of this process both with and without gravity weights. The results are similar regardless of weighting. The most valuable links with an impact of more than 0.1% on overall RE (up to a maximum of 0.8%) reconnect Morocco to Algeria, followed by improved trans-Saharan connections through Mauritania, Mali, and Chad. Also vital are additional links connecting southern Africa (Johannesburg) to western Africa and new links in (South-)Sudan facilitating north-south and east-west connections.

I also examine MA gains from the proposed links by computing the GDP per capita of each city as the inverse-distance weighted average of observations within 30km using data from Kummu et al. (2018) and multiplying it by city (agglomeration) population. I then calculate MA for all cities using the network and Eq. 4, and sum them to measure total MA. Setting  $\theta = 1$ , the total MA implied by this transport network is 1.42 billion USD/m in 2015 PPP terms, short \$1.42B/m, or around \$1M/m in the average node. With high trade elasticity ( $\theta = 3.8$ ), total MA is below \$1/m<sup>3.8</sup>. The MA gain from adding all proposed links is 5.5% if  $\theta = 1$  and 1.9% if  $\theta = 3.8$ .

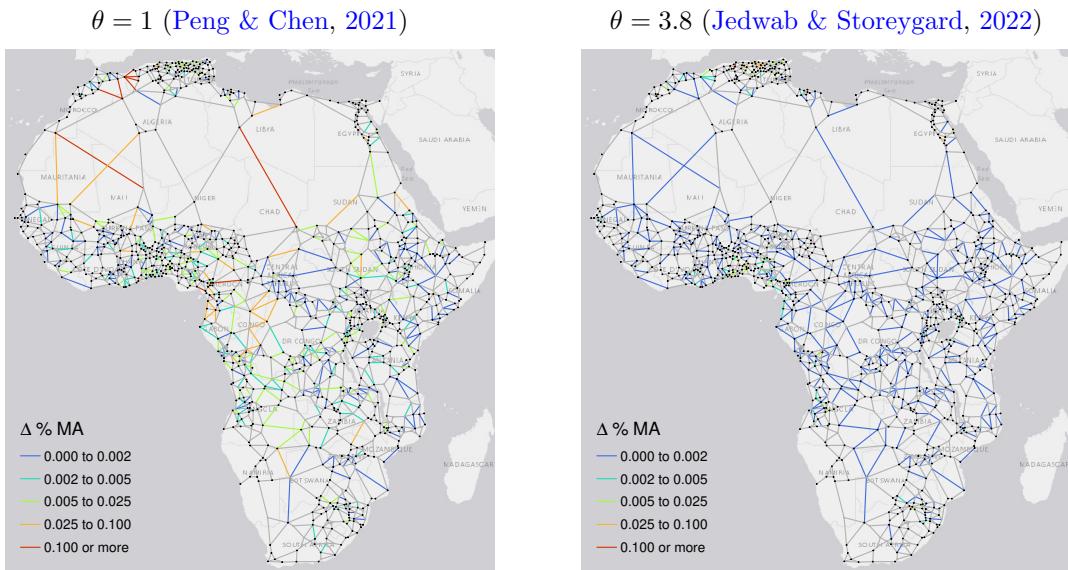
Figure 19: Percentage Gain in Network RE from Constructing Each Proposed Link



Notes: Figure shows (weighted) average route efficiency (RE) gains across all trips from building a specific link/edge.

The MA gains from adding individual links, reported in Figure 20, are similar to the route efficiency gains if  $\theta = 1$ . With high trade elasticity ( $\theta = 3.8$ ), only links near agglomerations yield high returns. For the remainder of this section, I set  $\theta = 1$  to limit the complexity of exposition and facilitate the interpretation of results. There are also substantive reasons for adopting this value.<sup>24</sup>

Figure 20: Percentage Gain in Market Access (GDP) from Constructing Each Proposed Link



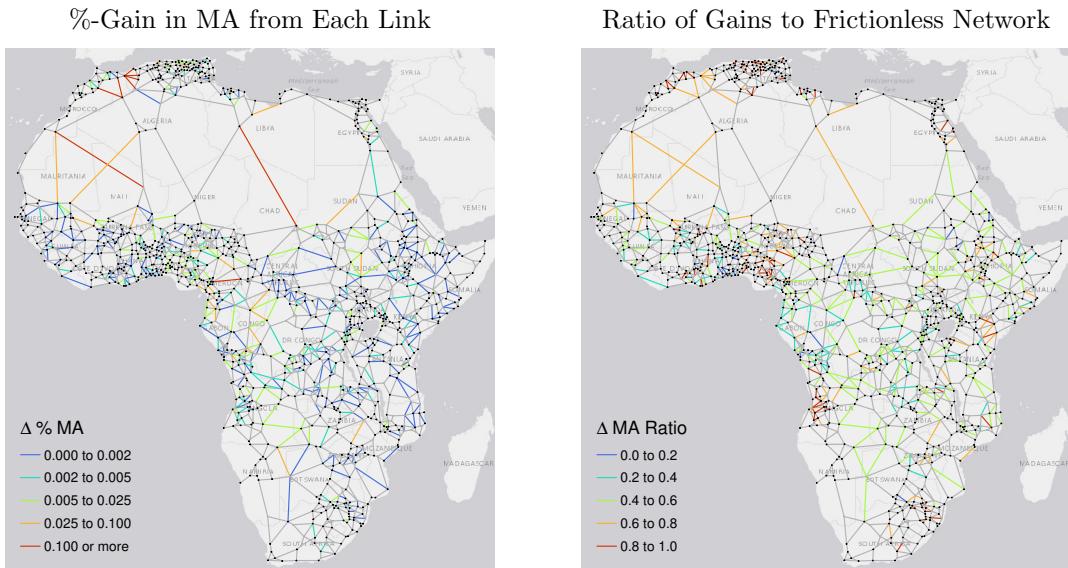
Notes: Figure shows total road-distance denominated MA gains measured (summed) across all trips from building a link.

To investigate how the presently high trade costs between borders may impact returns to link building, I add the respective export/import documentary and border compliance costs of the start and end country to each route - divided by 1.6 to turn them into kilometers following Table 4 - and also add a transit cost of  $(48h \times 4\$)/1.6 = 120\text{km}$  for routes spanning any third country.

<sup>24</sup>Using the Bayesian Information Criterion (BIC), Peng & Chen (2021) show that  $\theta < 3$  yields decent models predicting various outcomes (incl. built-up and nightlights) by the log of MA, choosing  $\theta = 1$  as their preferred value. This value is also more consistent with the evidence on inter-city migration (Poot et al., 2016; Zipf, 1946). Once considering road upgrades in Section 5.5, I find that MA gains under  $\theta = 3.8$  are significantly higher than under  $\theta = 1$  because upgrades also increase travel speeds at short distances in populated areas. In contrast, new links primarily improve connectivity through remote regions. Since total MA gains from road network improvements are much larger vis-a-vis new roads, setting  $\theta = 1$  yields a lower bound on total MA gains from road network improvements.

Consequently, the total MA implied by the network decreases from  $\$1.42B/m$  to  $\$1.08B/m$ , and the total gain from building all proposed links reduces to 3.9%, versus 5.5% without frictions. Thus, border frictions reduce the aggregate returns to building new roads. They also imply relatively greater gains from building domestic links. The LHS of Figure 21 shows MA gains under frictions, and the RHS the ratio of the gains with frictions to the frictionless gains (LHS of Figure 20).

Figure 21: Percentage Gains in Market Access (GDP) under Border Frictions



*Notes:* Figure shows absolute and relative distance-denominated MA gains from building a link under trip-level frictions.

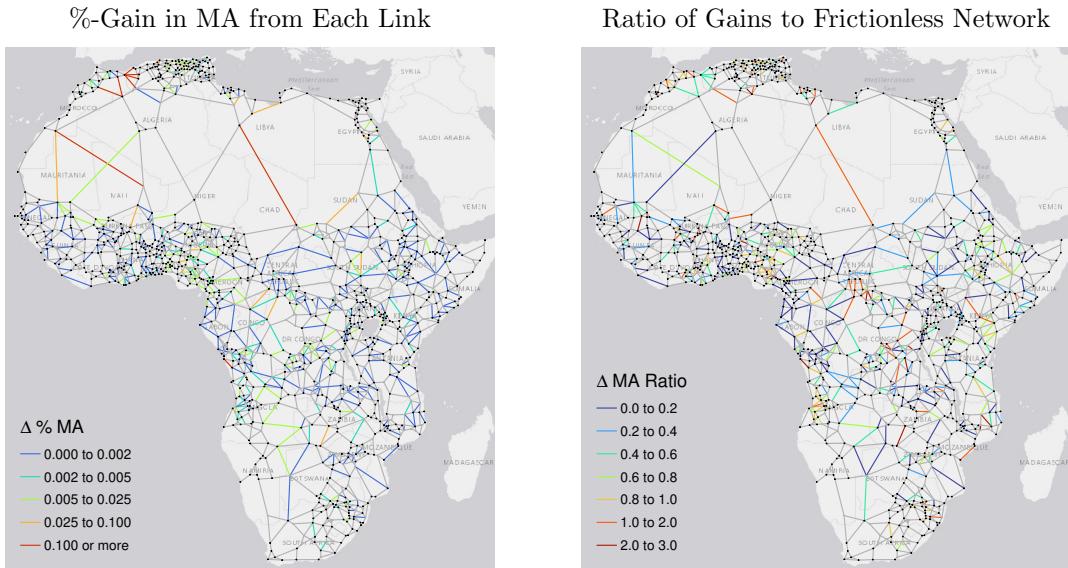
Ostensibly, frictions always reduce the gains from building links but mainly affect transnational links and links in remote areas (which are also more likely to cater to transnational traffic). New national links in populated areas (in red) yield almost the same MA gains compared to a frictionless environment. The median ratio is 0.52, implying a 48% lower return on the median link. Nevertheless, the LHS of Figure 21 suggests that certain trans-national links, especially additional trans-Saharan connections, bring the greatest MA gains even under high frictions. Thus border frictions considerably reduce the value of trans-national links and links in remote areas but do not appear capable of dramatically altering the overall distribution of returns.

However, this conclusion is tentative as it only measures border frictions at the origin and destination country border (+ lump-sum 24h transit frictions), giving traders no incentive to deviate from the frictionless route. In reality, traveling through many countries may be significantly more costly, and traders are reasonably assumed to optimize their route. To approximate this case, I ignore transit costs and add each border's average export and import cost (in km) to the respective border-crossing link. This parameterization thus conflates exporting, importing, and transit but introduces cumulative frictions prompting traders to avoid expensive borders.

Computing all shortest paths along this network has a substantial impact on trade costs. Whereas under the previous assumption routes are on average 32% more expensive than in a frictionless scenario, building these costs into the network makes them 117% more expensive, despite optimizing behavior. Total MA is reduced to  $\$0.91B/m$ , versus  $\$1.08B/m$  under the simple transit scenario. Yet, building all proposed links in this network still increases total MA by 4%.

Figure 22 visualizes the corresponding MA gains from building each link and the ratio of MA gains to frictionless gains. Similar to Figure 21, under frictions, most gains are lower, especially from border-crossing links. The median ratio is 0.43, implying a 57% reduction in value. The RHS of Figure 22 shows that, with optimizing agents, the ratio of MA gains to the frictionless network is not bounded above by 1. Some links, colored in dark orange/red, yield greater MA gains with border frictions than without. Thus, if traders are optimizing, border frictions not only reduce average gains but also complicate the evaluation of road network investments.

Figure 22: Gains in Market Access (GDP) with Optimizing Agents



*Notes:* Figure shows absolute and relative distance-denominated MA gains from building a link under link-level frictions.

### 5.3 Road Building Costs

A central consideration for any infrastructure project is its cost-benefit ratio. In this subsection, I thus factor in heterogeneous link-building costs. The most well-known empirical work on road construction costs in developing countries is Collier et al. (2016), who analyze unit costs from 3,322 activities across 99 countries obtained from the World Bank's Road Costs Knowledge System (ROCKS) database. They find that the cost of an asphalt overlay of 40-59mm can vary by a factor of 3-4 between countries. Regressing this cost on various geographic and non-geographic factors, they find that terrain ruggedness and population density increase unit costs, whereas long roads ( $>50\text{km}$ ) decrease it. Furthermore, conflict, corruption, the business environment, and the origin of construction firms affects unit costs. ROCKS version 2.3 used by Collier et al. (2016) runs up to 2008. In 2018, the World Bank's transport unit released an update including only projects with complete information (Bosio et al., 2018). Unfortunately, most projects in this database are road upgrades, but a few new road projects are included as well and summarized in Table 5.

Table 5: Road Building Costs: Road Costs Knowledge System (ROCKS) - 2018 Update

Work Type	N	Length (km)	Cost (M\$'15)	Cost/km
<i>Continental Africa</i>				
New 1L Road	1	99.0	21.15	0.214
New 2L Highway	4	122.0	59.53	0.611
<i>All Low and Lower-Middle Income Countries</i>				
New 1L Road	6	59.5	7.65	0.135
New 2L Highway	6	84.5	44.09	0.611
New 4L Expressway	1	84.8	328.08	3.869
<i>All Countries</i>				
New 2L Highway	7	50.0	43.03	0.682
New 4L Expressway	24	74.9	213.01	5.152

*Notes:* Statistics are aggregated across projects using the median. Costs are in millions of 2015 USD. The raw costs in current USD were deflated at their project end date using the US GDP deflator with base year 2015.

According to the final column of Table 5, which provides median costs per km in millions of constant 2015 USD, the median new African road in the 2-lane highway category (implying one lane in each direction) costs \$611K/km. This estimate, though derived from only four projects, is consistent with other estimates: a 2014 report by the African Development Bank (African

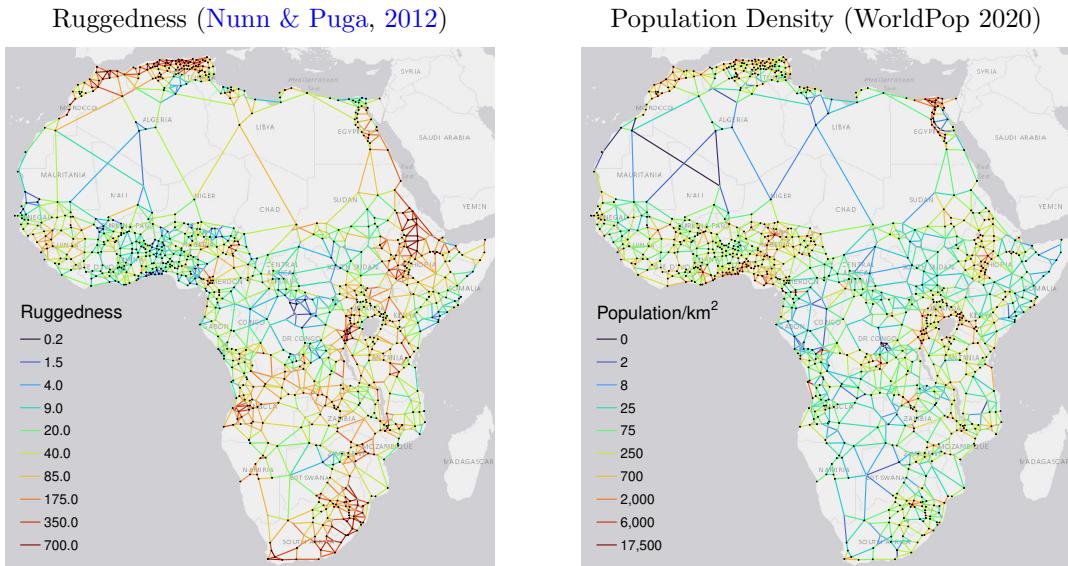
Development Bank, 2014) analyzes 172 Projects contracted between 2000 and 2008 and reports a median cost of 227,800 2006 USD per lane-km. This figure includes road upgrades. Multiplying by 2 for a 2-lane highway and inflating by 16% to 2015 USD yields 528,800 USD per lane-km.

Following Collier et al. (2016), I allow for heterogeneity in road construction costs. Since it is difficult to measure (ever-changing) political conditions across the continent, I follow Fajgelbaum & Schaal (2020) and Graff (2024) and only consider average returns to ruggedness and distance, as well as population. My preferred specification is

$$\ln\left(\frac{\text{cost}}{\text{km}}\right) = \ln(120,000) - 0.11 \times (\text{dist} > 50\text{km}) + 0.12 \times \ln(\text{rugg}) + 0.085 \times \ln\left(\frac{\text{pop}}{\text{km}^2} + 1\right). \quad (8)$$

The first part is taken from Eq. (21) of Fajgelbaum & Schaal (2020). The 120,000 USD/km constant makes the mean cost/km across all existing links approximately equal to \$611K/km. The population coefficient is averaged across Tables 4 and 5 of Collier et al. (2016), following Fajgelbaum & Schaal (2020), which obtain the other coefficient in this way. The tables suggest that Collier et al. (2016) regressed the log of cost on the population density (in 100 people/km<sup>2</sup>), but applying the coefficient to the levels density yields peak construction costs above \$10M/km near large cities, which I regard as unrealistic. Thus, I use the natural log of the population density and add 1 to avoid negative values in remote regions. This effectively reduces the relevance of this term and remains close to Fajgelbaum & Schaal (2020). I extract ruggedness from the original 30 arc-second layer by Nunn & Puga (2012) within a 3km buffer around existing and proposed links, and do the same for population from the WorldPop 2020 population layer at 1km resolution.<sup>25</sup> Figure 23 shows ruggedness and population density thus estimated for each link.

Figure 23: Ruggedness and Population Density within 3km Buffer around Links



*Notes:* Figure shows average ruggedness and total population within a 3km (two-sided) buffer around (straight) links.

Appendix Figure A12 shows the corresponding cost estimates from applying Eq. 8.<sup>26</sup> I divide the cost of building Algeria-Morocco links by 3 as these are, to a large extent, already constructed. The distribution of costs is roughly consistent with African Development Bank (2014), whose 1st and 3rd quartile estimates, appropriately converted, are 386,000 and 987,500 USD'15/km. Reconstructing the total existing network comprising 315K km of roads is estimated to cost \$186B USD'15, and the proposed extension of 104K km costs \$56B. Figure A12 suggests that roads in unpopulated and relatively flat areas such as Congo and the Sahara are cheapest. This may be erroneous since both exhibit rather extreme weather conditions and human security risks. Collier et al. (2016) unfortunately don't consider temperature, humidity, and landcover variables. More research is thus needed to refine estimates of road construction costs in developing countries.

<sup>25</sup>It is possible to track the terrain around existing roads more closely, but being more realistic with existing links complicates comparison with potential new links. Thus, I opt for a simpler approach of considering terrain within a broad 3km buffer in the direction of travel for both existing and potential new links.

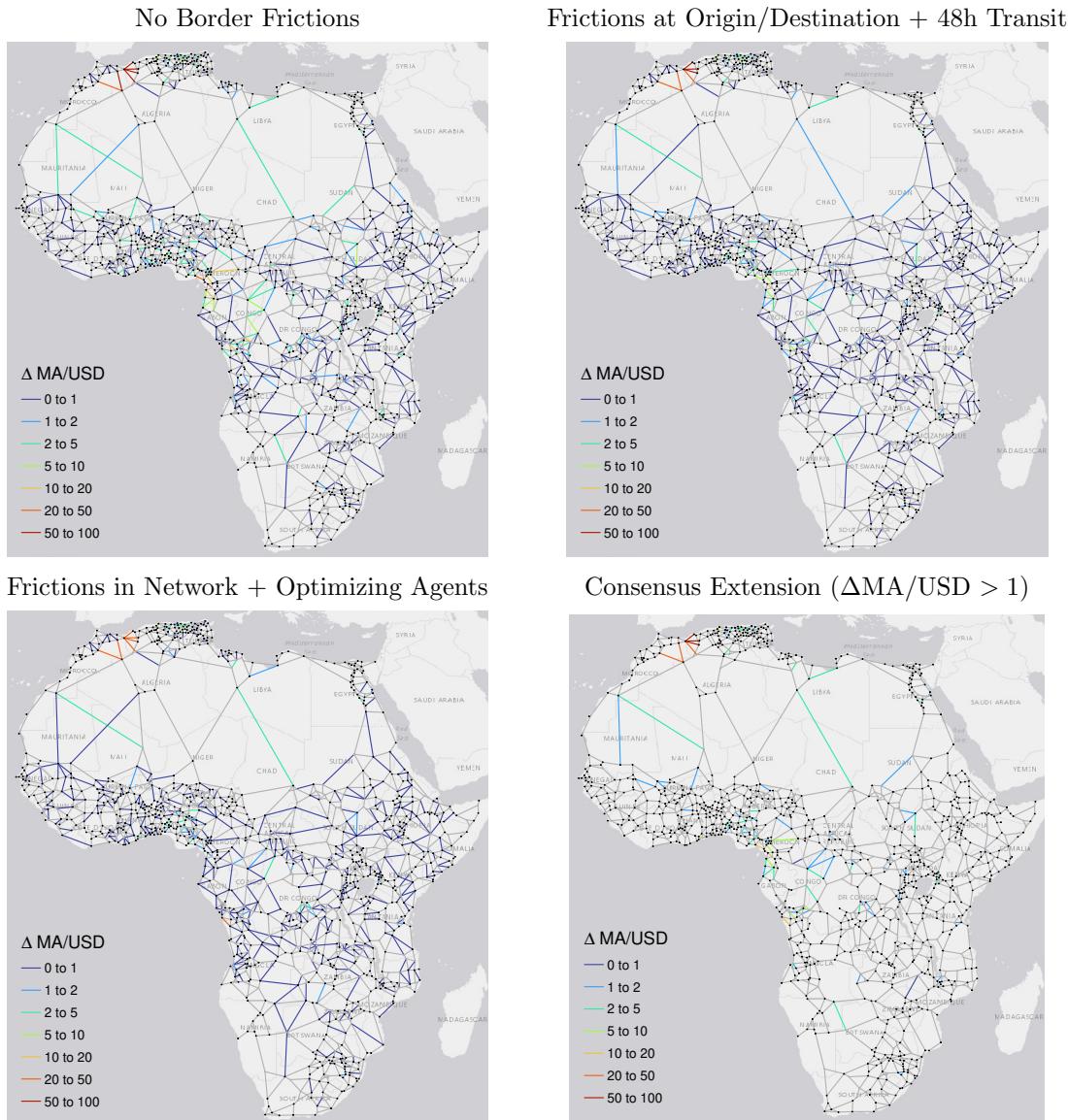
<sup>26</sup>Figure 31 provides the same information, including upgrading costs. I thus relegate Figure A12 to the Appendix.

## 5.4 Cost-Benefit Analysis: Building New Links

Comparing the proposed extensions at cost of \$56B to the total MA gain of 5.5% of \$1.42B/m, amounting to \$77.6M/m or \$77.6B/km, yields a MA return of \$1.4/km per USD invested, in short, \$1.4/km/\$. If border frictions are added, the MA gain from the network extensions is only 3.9% or \$41B/km, implying a return of \$0.76/km/\$. If frictions are built into the network and agents optimize, the MA gain is 4% or \$36.4B/km, implying a return of \$0.65/km/\$. Either way, current frictions reduce the gains from the proposed investment package by around 50%.

Since the total MA gains are below \$1/km/\$ under frictions, a limited investment package focusing on critical links is more sensible. Figure 24 shows the gain in \$/km/\$ of different links under the three assumptions about border frictions, and a consensus extension including only links above \$1/km/\$ in the frictionless case and at least one of the two scenarios with frictions.

Figure 24: Cost-Benefit Analysis (USD/km MA Gain per USD Invested)



*Notes:* Figure shows MA returns in USD/km per USD invested in link building under different types of border frictions.

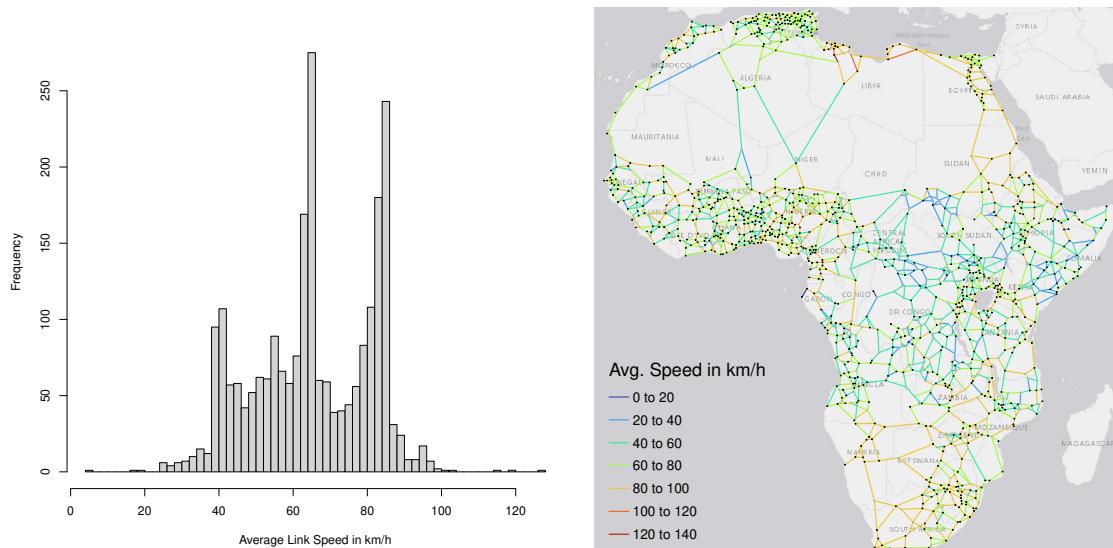
Figure 24 indicates that, even without frictions, 66.5% of the proposed links yield a MA return of <\$1/km/\$, and only 19.5% of links yield MA returns >\$2/km/\$. Some links, especially in West Africa (Cameroon) and Congo (Kinshasa), yield up to \$20/km/\$ returns. Reopening the Morocco-Algeria border would yield even higher returns on any border-crossing link. The consensus extension combining at least break-even links in MA terms includes 105 links (21.8% of the proposed

481). This package's estimated cost is \$12.1B. In the frictionless scenario, it yields a 4% MA gain amounting to \$57.5B/km, implying a return of \$4.75/km/.<sup>27</sup> Excluding the 6 links bridging the Morocco-Algeria border still yields a 2.9% MA gain of \$41.8B/km or \$3.45/km/. Under the most restrictive cumulative frictions scenario, the MA gain (excl. Morocco-Algeria links) drops to \$22B/km (2.4%) or \$1.82/km/. Thus, this package, which includes 4 additional trans-Saharan links, is only sensible if MA gains generate high economic gains (more in Section 5.9).

## 5.5 Upgrading of Existing Links

Since constructing new roads is expensive and the quality of African roads is low in many regions, upgrading the existing network is another vital policy option. While routing engines cannot perfectly consider road quality, OSM has an elaborate tagging system to indicate the surface and smoothness of roads, which is taken into account in OSRM routing and travel time estimates. Figure 25 shows the average travel speed in km/h along each link, indicating sizeable heterogeneity. The distribution of travel speeds is trimodal, with slow links around 40km/h, medium links around 65 km/h, and fast links around 85km/h. The three fastest links at travel speeds around 120km/h are in Libya, which satellite imagery reveals are newly asphalted straight roads through flat desert. The slowest link at a travel speed of 5km/h is in Gabon and connects the coastal cities of Libreville and Port-Gentil via a ferry (in absence of a viable road). Apart from these extremes, most links in populated areas have travel speeds  $\geq 60$  km/h. Slower links with travel speeds  $< 60$ km/h, or even  $< 40$  km/h, are prevalent in central Africa (Congo, Central African Republic, Chad, (South-)Sudan) and around the Horn of Africa (Northern Kenya, Somalia, South-Sudan, peripheral Ethiopia). The three trans-Saharan links connecting Algeria and Libya to Mali and Niger also have average travel speeds around 50km/h. Thus, Figure 25 (and Appendix Figure A1) suggests that mainly roads connecting the different populated macro-regions in Africa need upgrading.

Figure 25: Average Link Speeds in Africa's Transport Network



*Notes:* Figure shows average link speeds calculated at the link-level using the OSRM R API (accessed June 2024).

The MA measure implied by the current network speed is \$1.75T/min. Letting all links have a speed of  $\geq 100$ km/h yields a MA of \$2.48T/m, a 42.1% increase.<sup>28</sup> This is dramatically higher than the 5.5% gain from adding all proposed links (disregarding speed). To determine whether this discrepancy is an artefact of the way of denominating MA, I also compute MA gains from the proposed links, assuming average speeds of either 100km/h or 65km/h, the median link speed in the existing network. Adding all new links at 65km/h yields a 5.7% MA gain - comparable to the 5.5% measured earlier - whereas assuming a speed of 100km/h yields a 14.1% increase. This highlights the importance of road quality considerations. Since new roads are typically high quality, the 100km/h assumption and associated 14.1% MA gain without border frictions may be

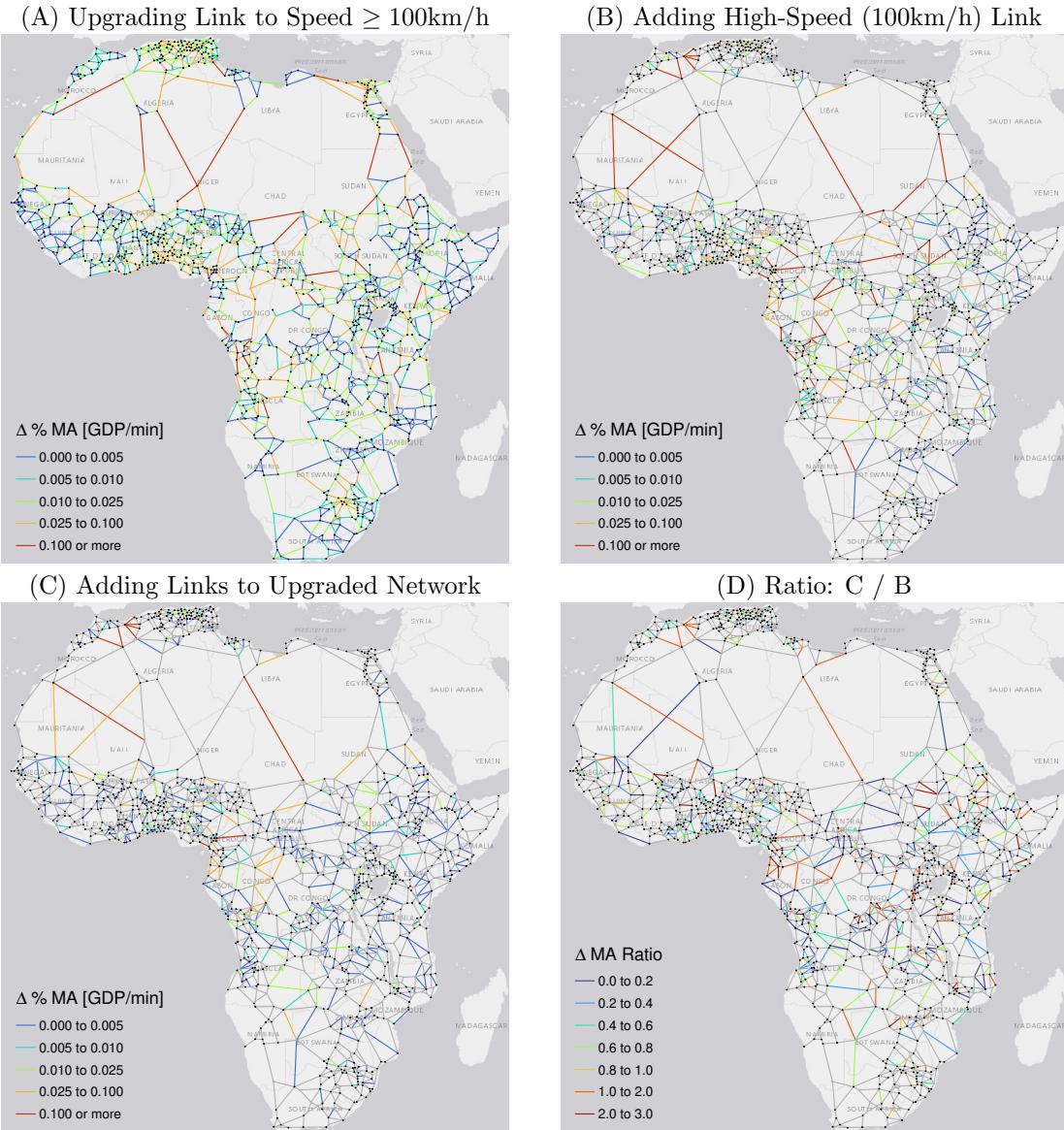
<sup>27</sup>These returns are based on recalculating shortest paths and MA after adding all 105 links.

<sup>28</sup>The Histogram in Figure 25 suggests 100km/h as a natural efficiency target for transport roads. This target is also sensible as different maximum speed regulations may exist beyond it.

more realistic. In total, eliminating all frictions, upgrading all existing links to  $\geq 100\text{km/h}$ , and adding the proposed links at  $100\text{km/h}$ , would increase MA to \$2.62T/min, a 50% improvement. This affirms that the largest gains (42%) can be reaped from improving existing roads. Gains from new roads are higher without improvements to existing roads due to optimizing behavior.

All of these scenarios can be simulated on a link-by-link basis. Figure 26 reports three scenarios that are policy relevant: upgrading existing links to  $\geq 100\text{km/h}$ , building new links at  $100\text{km/h}$  with the existing network unchanged, and building new links at  $100\text{km/h}$  on an upgraded network. The final panel shows a ratio indicating the relative MA gains under the latter two scenarios.

Figure 26: MA Gains (GDP/min) from Upgrading Links and Adding High-Speed Links

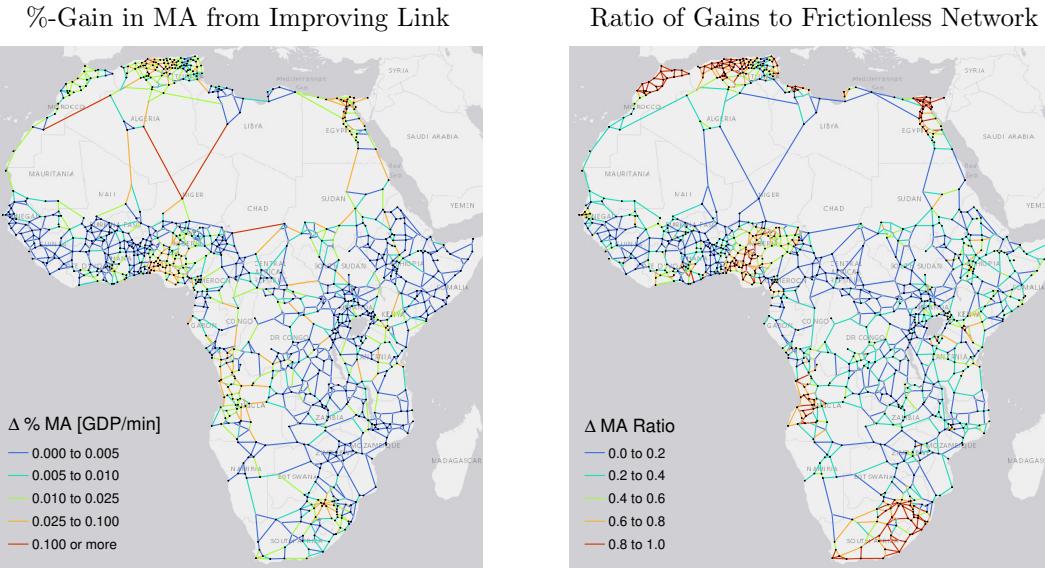


*Notes:* Figure shows total MA gains from upgrading a link and building a new link before/after upgrading all links.

Panel A signifies that improving major north-south/east-west roads and roads in metropolitan areas yields significant MA gains. Panels B and C are congruent to previous results that new high-speed north-south/east-west roads are most impactful, with lower average gains under the upgraded network. Panel D signifies that, whereas most links yield smaller returns under an upgraded network ( $\text{ratio} < 1$ ), some links, including re-opening the Morocco-Algeria border and the two proposed north-west south-east oriented trans-Saharan connections, yield higher returns if the existing network is fully upgraded. Notably, these are also particularly high-value links from a pure route efficiency standpoint and included in the consensus extension of Figure 24.

Another major area of necessary improvement is border frictions. Applying current frictions, i.e., dividing border time by 4 and documentary time by 10 (Table 4) and adding a 48h transit through 3rd countries, yields a MA of \$982B/min, which is only 56.1% of the \$1.75T/min without frictions. When improving all links to a speed of  $\geq 100\text{km/h}$ , total MA increases to \$1.24T/min, which is 50% of the \$2.48T/min without frictions and only 27% more than the \$982B/min under the current network, compared to a 42% increase without frictions. Figure 27 reports link-level gains and a comparison to frictionless gains (Panel A of Figure 26). The results are unequivocal: border frictions particularly diminish MA gains from upgrading remote and transnational links.

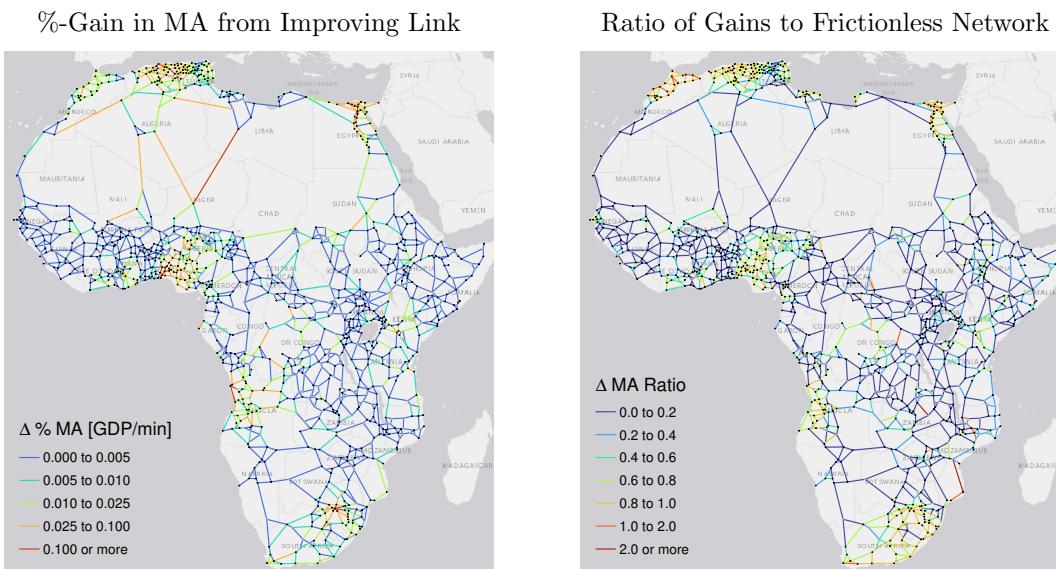
Figure 27: Percentage Gains in Market Access (GDP/min) under Border Frictions



*Notes:* Figure shows absolute and relative MA gains from upgrading a link under trip-level frictions.

The exercise is again repeated by incorporating average export/import frictions into border-crossing links. This reduces MA to \$820B/min, or 47% of the frictionless value. With all links upgraded to 100km/h, the MA is \$1.04T/min, only 42% of the MA gain without frictions, but still 27.3% more MA than without road improvements. These numbers reflect an average 6x travel time increase under frictions, which rises to 8.6x with an upgraded 100km/h network. Figure 28 shows gains under the cumulative frictions case. The effects are similar and stronger than in Figure 27. Because of the optimizing behavior away from expensive links, the RHS of Figure 28 again includes a few links whose upgrading under frictions yields greater MA gains than without frictions.

Figure 28: Gains in MA (GDP/min) under Border Frictions with Optimizing Agents



*Notes:* Figure shows absolute and relative MA gains from upgrading a link under link-level frictions.

## 5.6 Road Upgrading Costs

For cost-benefit analysis, I also estimate the cost of upgrading roads. The ROCKS database distinguishes several types of road improvements. Table 6 summarizes the four most common ones: 'Upgrading', 'Reconstruction', 'Asphalt Mix Resurfacing', and 'Strengthening'. To obtain a summary statistics for the cost of upgrading, I compute a weighted average of the median costs/km across types, using the number of works (N) times the median length of the work in km as weights. This yields an average upgrade cost estimate of \$326M/km. Thus, upgrades are, on average, only about half as costly as new highways priced at \$611M/km (see Table 5). However, road upgrades in Africa are  $\sim 60\%$  more costly than the low- and lower-middle income country average of \$205M.

Table 6: Road Upgrading Costs: Road Costs Knowledge System (ROCKS) - 2018 Update

Work Type	N	Length (km)	Weight	Cost (M\$'15)	Cost/km
<i>Continental Africa</i>					
Upgrading	28	103.5	2898	49.73	0.513
Reconstruction	27	33.0	891	20.20	0.116
Asphalt Mix Resurfacing	14	133.8	1874	16.93	0.143
Strengthening	11	110.0	1210	42.97	0.316
Weighted Average		104.0		35.77	0.326
<i>All Low and Lower-Middle Income Countries</i>					
Upgrading	68	61.8	4199	20.63	0.320
Reconstruction	53	73.0	3869	20.20	0.162
Asphalt Mix Resurfacing	35	143.9	5036	15.50	0.090
Strengthening	26	92.0	2392	20.89	0.316
Weighted Average		95.9		18.90	0.205

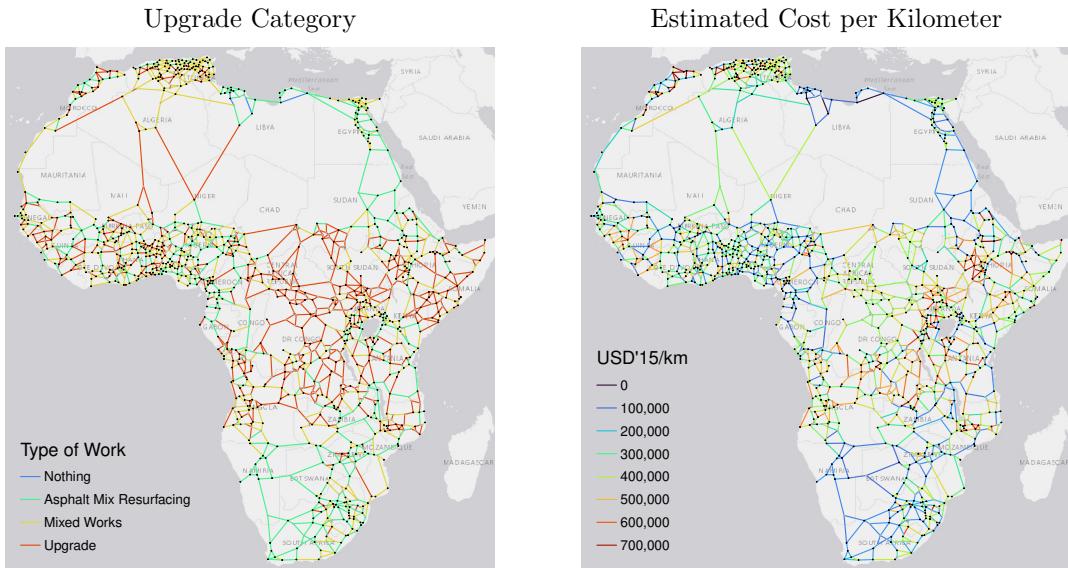
*Notes:* Statistics are aggregated across projects using the median. The 'Weight' column is the number of projects (N) times the median length (Length). This is used to compute a weighted average across different work types. Costs are in millions of 2015 USD. The raw costs in current USD were deflated at their project end date using the US GDP deflator with base year 2015.

These estimates are consistent with [African Development Bank \(2014\)](#), who report an average cost of 180,300 USD'06/lane-km for a category called 'Rehabilitation,' yielding 418,500 USD'15/km for 2-lane highways. Their 1st and 3rd quartile estimates translate to 254,900 and 673,100 USD'15/km, respectively. As mentioned earlier, [African Development Bank \(2014\)](#) lump together 'Upgrading' with new construction at a median price of 528,800 USD'15/km. Their estimate is thus between the 'Upgrading' estimate of \$513K in Table 6 and new construction at \$611K from Table 5.

Upgrading a link to allow speeds of 100km/h should require different types of road work depending on the existing condition of the road. It is unrealistic to assume that this is a continuous function of the distance of travel speed from 100km/h. For example, two roads with travel speeds of 85 and 90km/h likely both require a full asphalt resurfacing to bring them to 100km/h. Since the distribution of speeds in Figure 25 is trimodal, I propose the following three categories: (1) roads with a travel speed below 60km/h require a full 'Upgrade' at an average cost of \$513K/km; (2) roads with travel speeds of 60-79km/h may require mixed works at an average cost of \$326K/km; (3) roads with travel speeds of 80km/h and higher require an 'Asphalt Mix Resurfacing' at \$143K/km on average. To introduce heterogeneity in these cost estimates, I use Eq. 8 and modify the intercept such that the median (across all routes, regardless of their upgrade type) is close to the respective category medians: For category (1), the intercept is ln(101,600), for category (2) it is ln(64,600), and for category (3), it is ln(28,400). In addition, I clip estimates at the 2.5th and 97.5th percentile to guard against outliers, such that, e.g., 'Asphalt Mix Resurfacing' costs remain between \$94K/km and \$206K/km. Figure 29 reports the classification and estimated unit costs.

The entire central-African network and trans-Saharan connections via Mali and Niger, as well as most of the network around the horn of Africa, is classified as requiring a full upgrade, at costs around \$400K/km due to mostly flat terrain and small population density. The total cost of upgrading the entire 315K km transport network is thus estimated at \$106B (in 2015 USD), comprising \$57.8B full road upgrades, \$35.6B mixed works, and \$12.3B mixed asphalt resurfacing.

Figure 29: Estimated Network Upgrading Cost per Kilometer



*Notes:* Figure shows stipulated work required to upgrade a link to speed  $\geq 100\text{km/h}$  (LHS) and estimated cost/km (RHS).

## 5.7 Cost-Benefit Analysis: Road Upgrading

Comparing this cost to the total MA gain of 42% or \$701B/min implies a return of \$6.6/min/\$. With border frictions, the MA gain is reduced to 27% or between 214 and 253 billion USD/min in the two scenarios, implying a 2/2.4 \$/min/\$ return in the cumulative/simple frictions scenarios, respectively. Thus, road upgrades yield significantly higher average returns than fresh construction.

It remains to gauge the returns to upgrading individual links. Figure 30 shows estimates following the same schema as Figure 24, signifying that upgrading links in populated areas and critical trans-continental connections yields significant returns. The consensus package again includes links where the MA return in \$/min/\$ is greater than 1 in the frictionless case and at least one of the scenarios with frictions. It suggests prioritizing upgrading roads near urban centers and important trans-continental links. The package comprises 1255 links, 53.5% of all links, and costs \$45B. It yields a frictionless 33.1% MA gain of \$552B/min or 12.3\$/min/\$.

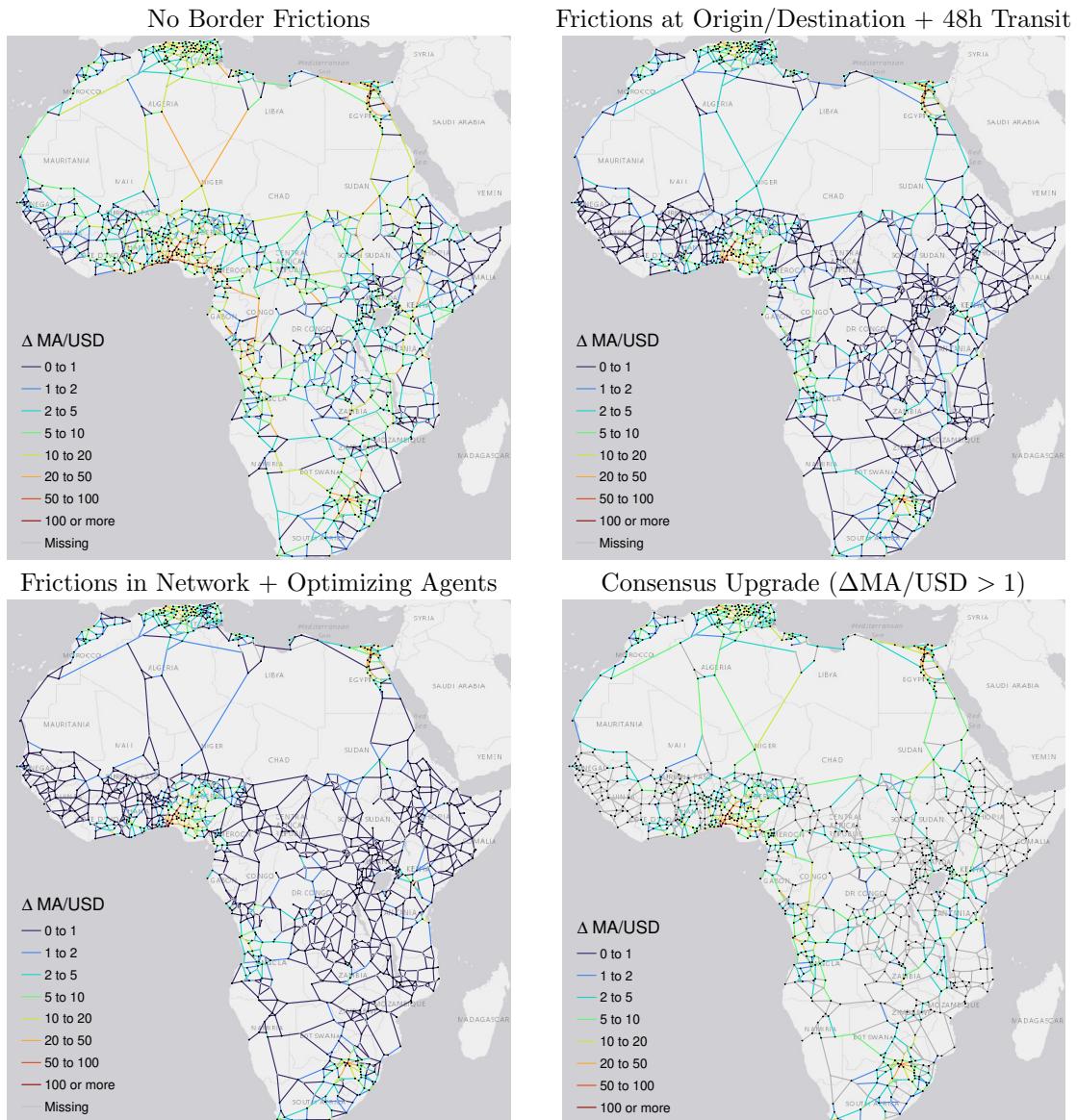
## 5.8 Cost-Benefit Analysis: Joint Scenarios

Combining road construction and upgrading costs (see Figure 31), I also perform cost-benefit analysis for joint scenarios where building new links and upgrading existing ones are part of the choice set. Due to the different intercepts in Eq. 8, the cost of building a new link is always greater than upgrading an existing one *ceteris paribus*.

All proposed road works combined cost 161 billion USD'15 (around 200 billion in current USD). They yield a frictionless MA gain of 50%, amounting to \$831B/min or \$5.2/min/\$. The more restrictive cumulative frictions case yields a 31% MA gain, amounting to \$241B/min or \$1.5/min/\$. Figure 32 provides the full link-level analysis, including three consensus packages with MA gains above 1, 2 and 4 \$/min/\$, at costs of 60.9, 36.6, and 17 billion USD'15, respectively:

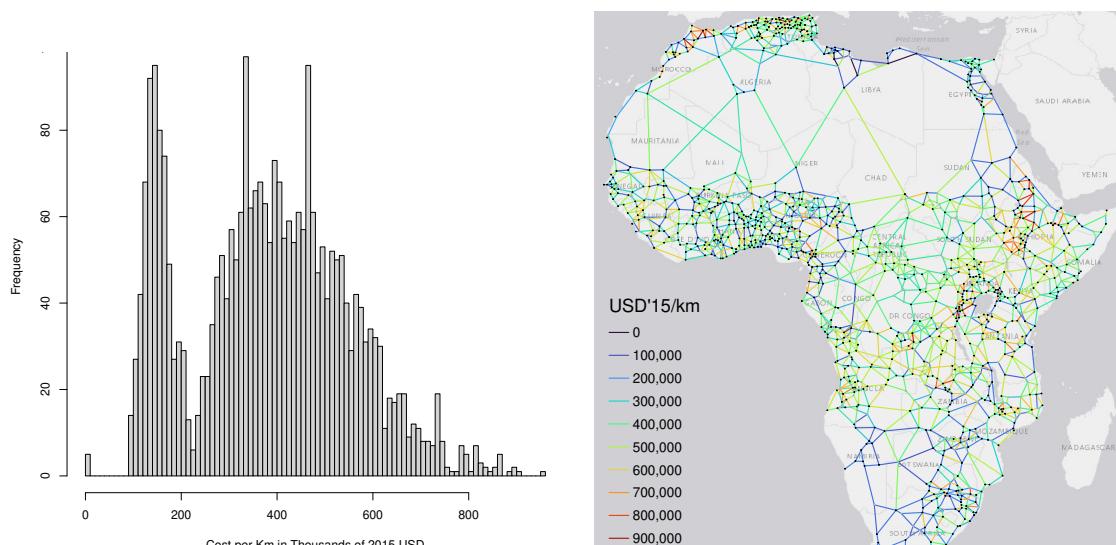
- The > \$1 package comprises upgrading 54% of links and building 30% of proposed links. It yields a 38% MA gain of \$636B/min or \$10.4/min/\$ without frictions and 26.8% | \$251B/min | \$4.1/min/\$ or 28% | \$221B/min | \$3.6/min/\$ under simple or cumulative frictions.
- The > \$2 package comprises upgrading 37% of links and building 18% of proposed links. It yields a 30.6% MA gain of \$511B/min or \$14/min/\$ without frictions and 24% | \$225B/min | \$6.1/min/\$ or 25.7% | \$201.5B/min | \$5.5/min\$ under simple or cumulative frictions.
- The > \$4 package comprises upgrading 22.4% of links and building 8.9% of proposed links. It yields a 20% MA gain of \$340B/min or \$20/min/\$ without frictions and 19.5% | \$182B/min | \$10.7/min/\$ or 21.3% | \$167B/min | \$9.8/min/\$ under simple or cumulative frictions.

Figure 30: Cost-Benefit Analysis (USD/min MA Gain per USD Invested)



*Notes:* Figure shows MA returns in USD/min per USD invested in link upgrading under different types of border frictions.

Figure 31: Estimated Network Building/Upgrading Cost per Kilometer



*Notes:* Figure shows combined cost/km of building new (fast) links and upgrading existing links to speed  $\geq 100\text{km/h}$ .

The first two packages are prone to generate a much greater developmental impact than the high-yield package, which primarily addresses low connectivity in the already large markets.

## 5.9 Macroeconomic Cost-Benefit Analysis

A critical question neglected so far is whether these investments are macroeconomically feasible. How do MA gains translate into medium-term economic gains, and are these gains large enough to justify such expenses, potentially even on a debt-financing basis?

The medium-term economic returns to increased MA are very difficult to estimate, especially on such a large scale. The only large-scale analysis I am aware of is [Donaldson & Hornbeck \(2016\)](#). The authors regress changes on the log of land values across 2,327 US counties between 1870 and 1890 on changes in the log of MA, and find an elasticity of 0.5, which is quite robust to different choices of  $\theta$ . This suggests that in the medium run MA translates into economic activity at a ratio of 2:1. However, present-day Africa is very different from the historical US.

Thus, instead of estimating economic returns to MA gains in Africa, I present an attempt to determine what the returns would need to be in order to justify the investment. This will take the form of a reverse macroeconomic present discounted value (PDV) calculation. African GDP in 2022 was 2.81 trillion 2015 USD. The median growth rate in 2000-2022 was 4.1%. If we are pessimistic about African growth, let it be 3% going forward. I discount the future by 10%. This is a reasonable rate given that stock market returns are generally below this - the MSCI World has a long-term return rate of 8-9%. Emerging market bonds yield even lower returns, around 6%, according to the EMBI index by JP Morgan. So at the macroeconomic scale, it is very difficult to invest with more than 10% returns. I also assume that the horizon of policymakers is 30 years and disregard benefits accruing after 30 years. With these assumptions, I calculate the change in the macroeconomic growth rate of Africa that would be necessary to break even in PDV terms for each of the three scenarios proposed above as well as the full/consensus construction and upgrade scenarios considered in Sections 5.4 and 5.7. Table 7 reports these calculations.

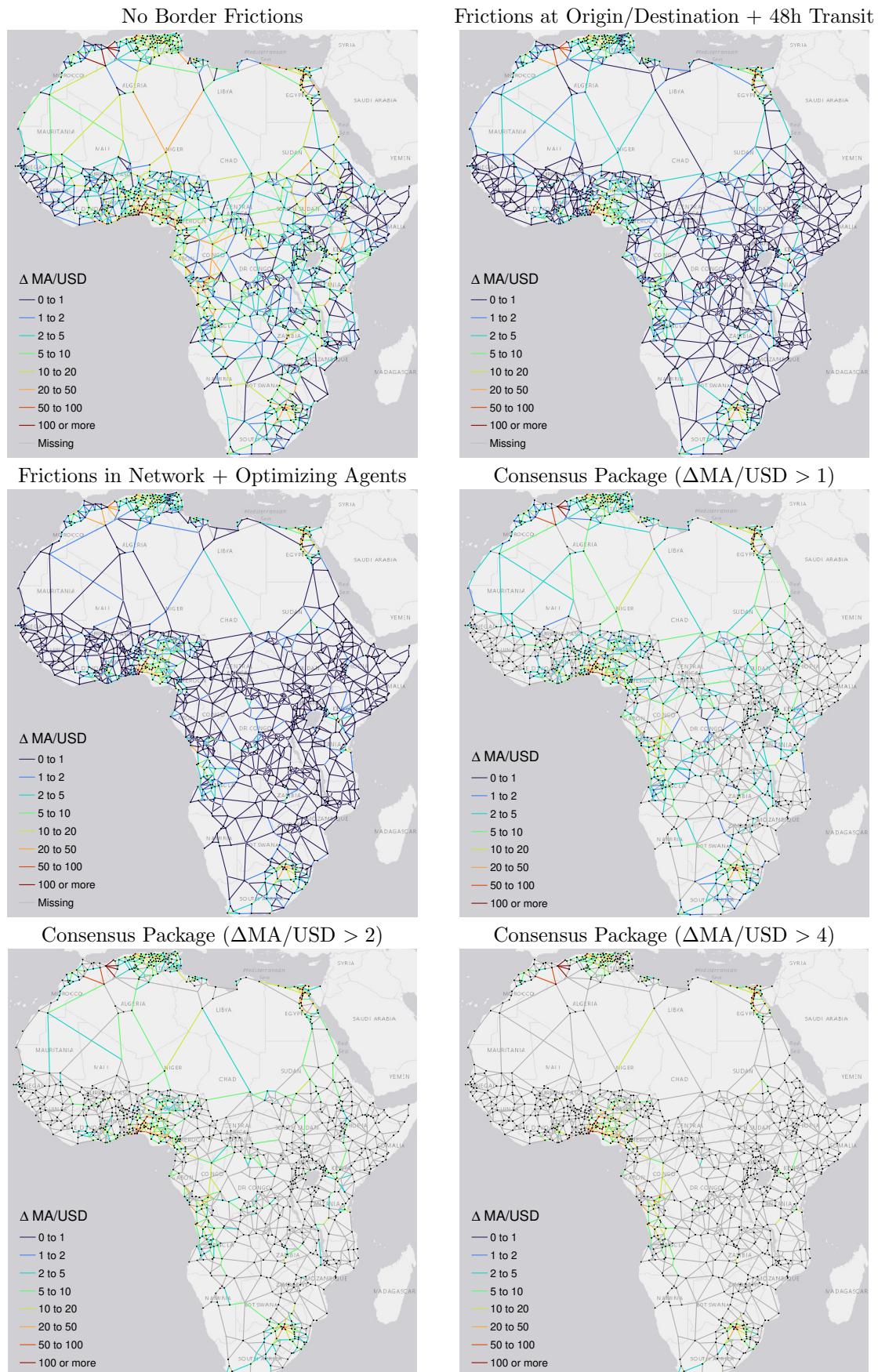
Table 7: Break Even Changes in the Macroeconomic Growth Rate for Different Investments

Work Package	Cost	%ΔMA	BEG (4.1%)	BEG (3%)
Full Extension (↑RE)	\$56.1B	3.9 – 5.5	0.31% → 4.113	0.50% → 3.015
Consensus Extension (↑RE)	\$12.1B	2.4 – 4.0	0.07% → 4.103	0.11% → 3.003
Full Upgrade (↑TE)	\$105.7B	27.0 – 42.1	0.58% → 4.124	0.94% → 3.028
Consensus Upgrade (↑TE)	\$45.0B	21.2 – 33.1	0.25% → 4.110	0.40% → 3.012
All Links > \$1/min/\$	\$60.9B	26.8 – 38.1	0.33% → 4.114	0.54% → 3.016
All Links > \$2/min/\$	\$36.6B	24.0 – 30.6	0.20% → 4.108	0.32% → 3.010
All Links > \$4/min/\$	\$17.0B	19.5 – 20.4	0.09% → 4.104	0.15% → 3.005

*Notes:* Table shows the cost in 2015 USD PPP of different investment packages, their MA %-gain under the worst frictions and frictionless scenarios, and the change in Africa's macroeconomic growth rate needed to break even after 30 years under realistic (4.1%) and pessimistic (3%) baseline growth.

I illustrate their interpretation using the largest consensus package containing all links with  $>\$1/\text{min}/\$$  marginal returns. It costs \$60.9B and yields a MA gain between 26.8% (worst frictions scenario) and 38.1% (no frictions). Under the 4.1% baseline growth scenario, the growth rate needs to change by 0.33% to 4.114% to yield a return of \$60.9B in PDV terms after 30 years. Comparing the 26.8% lower bound to 0.33% yields a ratio of 81:1. Thus, if changes in MA translate to changes in the growth rate at 81:1 or higher, the package is macroeconomically feasible. This appears very plausible given the results of [Donaldson & Hornbeck \(2016\)](#). However, many things can go wrong. First, growth could be slower at 3%, requiring a 0.54% increase in the growth rate or a return ratio above 50:1. Then, the package could end up being more expensive for various reasons related to terrain, weather, corruption or delays. Assuming it turns out twice as expensive, this would require a ratio above 25:1. Finally, if the investment is financed through debt, it could adversely affect both the total costs and the macroeconomic growth rate by limiting fiscal space and worsening credit ratings. Such forces could lower the required return ratio to as low as 10:1. Nevertheless, this still appears reasonably achievable; thus, overall, these back-of-the-envelope calculations suggest that large-scale *optimal* investments in Africa's road network are macroeconomically feasible.

Figure 32: Cost-Benefit Analysis: Joint Scenarios (USD/min MA Gain per USD Invested)



*Notes:* Figure shows MA returns in USD/min per USD invested in link building and/or upgrading under different types of border frictions. It also shows 3 different 'consensus packages' containing high-returns links with and without frictions.

## 6 Optimal Network Investments in General Equilibrium

Having simulated many scenarios in partial equilibrium by modifying one link at a time or jointly adding/improving a selection of (high marginal value) links and computing MA returns, this section goes a bold step further and considers welfare-maximizing network investments in general equilibrium (GE). In particular, using a quantitative framework established by [Fajgelbaum & Schaal \(2020\)](#), I optimize over the space of networks after setting an infrastructure budget. The framework accommodates trade in multiple goods and requires exogenous productivities for the production of each good in each location. Cities must have different productive endowments to create incentives for trade. For my application to the African network, this is a strong limitation, firstly because such city-level productivity data is unavailable, and secondly because global optimization with many goods traded on a large network is computationally infeasible.

To remedy the lack of detailed productivity data, I follow [Armington \(1969\)](#), [Fajgelbaum & Schaal \(2020\)](#), and [Graff \(2024\)](#) and assume that products are differentiated by city of origin and each city produces just one good. To overcome the dimensionality problem, I perform simulations on two different network graphs: (1) on the full transport network from Section 4.3 with 1379 nodes and  $2344 + 481 = 2825$  edges, I let small, medium, large, and port-cities produce different goods. This creates incentives for regional trade between cities of different sizes and with coastal ports, but limited incentives for trans-African trade. (2) On the reduced network connecting 47 major cities and ports from Section 4.4 with, in the simplest case, only 212 nodes and 330 edges, I let the largest 17 cities produce differentiated products. Other (port-)cities are categorized.

In my model specification and parameterization, I largely follow [Graff \(2024\)](#). I set the existing level of infrastructure along each link from city  $j$  to city  $k$ ,  $I_{jk}$ , to the average link speed in km/h

$$I_{jk} = \text{speed}_{jk} \quad \forall j, k. \quad (9)$$

$I_{jk}$  is bounded from below by the present speed ( $I_{jk}^{\min} = I_{jk}$ ), which is zero for proposed links. It is also bounded from above by the greater of 100km/h and the present speed, denoted by  $I_{jk}^{\max} = \max(I_{jk}^{\max}, 100)$ . I then calculate the (per km/h) infrastructure building cost  $\delta_{jk}^I$  as

$$\delta_{jk}^I = \text{upgrade cost}_{jk} / (I_{jk}^{\max} - I_{jk}), \quad (10)$$

where upgrade cost $_{jk}$  corresponds to the cost per km of building each link or upgrading it to 100km/h from Figure 31, multiplied by the link length in km. Since in the model of [Fajgelbaum & Schaal \(2020\)](#) the lower bound on  $I_{jk}$  is simply a constraint on how a given infrastructure budget may be spent, I set the baseline budget equal to  $\delta_{jk}^I \times I_{jk}$ . I then consider a social planner with a budget of approx. 1/3 of the cost of all proposed work (upgrading and construction) in addition to the baseline budget. The planner can freely choose which links to build/upgrade at cost  $\delta_{jk}^I$ .

In reality, most road upgrade/construction decisions are discrete, but the quantitative framework does not allow a discretization of optimal infrastructure investments. The results may instead be interpreted as 'partial' upgrades/completion. For example, suppose a road at 80km/h has a speed of 90km/h after the optimization. In that case, we may conclude that the planner decided to upgrade  $(90-80)/(100-80) = 50\%$  of the road.<sup>29</sup> The model also features iceberg trade costs of the form

$$\tau_{jk}^n(Q_{jk}^n, I_{jk}) = \delta_{jk}^\tau \frac{(Q_{jk}^n)^\beta}{I_{jk}^\gamma}, \quad (11)$$

such that traders need to send  $1 + \tau_{jk}^n$  units of a good for one unit to arrive. Following [Graff \(2024\)](#), who averages estimates from [Atkin & Donaldson \(2015\)](#) for Nigeria and Ethiopia, I set

$$\delta_{jk}^\tau = 0.1159 \times \ln(\text{distance in miles}_{jk}). \quad (12)$$

---

<sup>29</sup>While this interpretation is plausible, the cost per km of building/upgrading segments decreases for roads longer than 50km (Eq. 8, following [Collier et al. \(2016\)](#)), and it is thus not realistic that the planner considers prices  $\delta_{jk}^I$  inelastic to the quantity (length) constructed. In addition, if the planner partially builds a new links, we must assume that the remainder corresponds to a dirt road that can be constructed free of cost. Thus, the framework may slightly overstate the amount of infrastructure that can be constructed on a given budget.

The trade cost positively depends on the flow of good  $n$  across link  $j \rightarrow k$ ,  $Q_{jk}^n$ , subject to an elasticity  $\beta$ , an effect Fajgelbaum & Schaal (2020) describe as 'congestion forces'. It also depends negatively on the level of infrastructure  $I_{jk}$ , subject to an elasticity  $\gamma$ . Following Graff (2024), I set  $\gamma = 0.946$  and  $\beta = 1.177$ .<sup>30</sup> As discussed at length by Fajgelbaum & Schaal (2020), the lagrangian objective of the model is only convex if  $\beta > \gamma$ , in which case a unique global solution to the network design problem can be computed. If  $\beta < \gamma$ , the problem becomes non-convex and yields a local optimum, which Fajgelbaum & Schaal (2020) further refine using simulated annealing methods. They denote this case as 'increasing returns to infrastructure' (IRS) and show that the optimal network is more tree-like, with larger investments in fewer links. Graff (2024)'s parameterization implies  $\beta > \gamma$ . This may, however, not hold for large transnational networks; thus, following Fajgelbaum & Schaal (2020), I also examine the IRS case assuming  $\gamma' = \beta^2/\gamma = 1.465$ .<sup>31</sup>

Fajgelbaum & Schaal (2020)'s model has a normalized inequality-averse Cobb-Douglas utility function giving the utility of a single worker in location  $j$  in terms of consumption and housing

$$u_j(c_j, h_j) = \left[ \left( \frac{c_j}{\alpha} \right)^\alpha \left( \frac{h_j}{1-\alpha} \right)^{1-\alpha} \right]^{1-\rho} / (1-\rho), \quad (13)$$

where  $\rho$  controls the level of inequality aversion. Following Graff (2024), I assume  $\alpha = 0.7$ , which is based on estimates by Porteous (2022), and  $\rho = 0$  (no inequality aversion), but I also explore  $\rho = 2$  (moderate inequality aversion). Following Graff (2024), I set  $h_j = 1 - \alpha$ , i.e., the local housing availability is proportional to the current population. Total consumption in location  $j$  is  $C_j = L_j c_j$ , which in turn is a CES aggregate over  $N$  traded goods with elasticity of substitution  $\sigma$

$$C_j = \left[ \sum_{n=1}^N (C_j^n)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}. \quad (14)$$

In line with the international trade literature, Fajgelbaum & Schaal (2020) and Graff (2024) use  $\sigma = 5$ , following recommendations by Atkin & Donaldson (2022). Both assume that the largest  $N - k$  cities produce their own good and mostly conduct single-country simulations.<sup>32</sup> Since I have few cities per country, it is also plausible to interpret  $\sigma$  as an Armington (1969) elasticity, and here a meta-study by Bajzik et al. (2020) suggests a median value of 3.8. These values apply to the trans-African network simulation case where I let 17 megacities produce differentiated products. In the case of only 4 goods produced by small, medium, large, and port cities, such elasticities generate near autarky equilibria. I thus assume  $\sigma$  values as low as 1.5 to generate realistic trade flows in this case. Goods are produced according to a linear technology

$$Y_j^n = Z_j^n L_j^n. \quad (15)$$

Due to lack of better measures,<sup>33</sup> I set city productivity  $Z_j^n$  equal to the predicted International Wealth Index of Lee & Braithwaite (2022) (Figure A2). Since the original estimate is only available for SSA, I use the imputed version of Krantz (2023) (Figure A3) and compute the productivity of each city as the inverse-distance-weighted average of cells within 30km of the city centroid. I assume

<sup>30</sup>The value for  $\gamma$  is informed by estimates of the costs of road delays to African truckers by Teravaninthorn & Raballand (2009). The value of  $\beta$  is informed by estimates on the cost of traffic congestion by Wang et al. (2011). Fajgelbaum & Schaal (2020) use  $\beta = 0.13$  and  $\gamma = 0.1$ , but their characterization of infrastructure ( $I_{jk}$ ) in terms of road-lane km yields significantly larger values than the formulation in terms of km/h adopted from Graff (2024).

<sup>31</sup>Fajgelbaum & Schaal (2020) page 1439 simulate an IRS case by inverting the ratio of  $\beta$  and  $\gamma$  keeping  $\beta$  fixed.

<sup>32</sup>Graff (2024) sets  $N = 6, k = 2$  (with one international good produced outside the country and one agricultural good produced by smaller cities) and conducts simulations for every African country. Fajgelbaum & Schaal (2020) set  $N = 10, k = 1$  and conduct simulations for France, Spain, and Western Europe.

<sup>33</sup>Graff (2024) uses nightlights divided by city population to form the productivity measure. While this is a viable approach, the task of measuring differences in city productivity across countries encourages the use of richer data sources. Lee & Braithwaite (2022) combine daylight satellite imagery and OSM data in an iterative ML framework that has proven able to predict wealth within and across SSA countries reliably. As an attempt to verify, taking a city-population weighted average of the IWI by countries yields a correlation of 0.62 with the 2020 World Bank national GDP/capita estimate and a correlation of 0.73 with the log of GDP/capita. Using NASA's Black Marble nightlights, on the other hand, only yields a correlation of 0.15 between city-population weighted nightlights/capita and national GDP/capita. However, using the log of nightlights/capita yields a 0.58 correlation with GDP/capita and a 0.65 correlation with the log of GDP/capita. While this appears more promising, a visual comparison, provided in Appendix Figure A13 for the interested reader, shows that the IWI is higher in larger cities in richer countries - in line with much scientific evidence. The nightlights measure appears quite noisy and imprecise.

that average household wealth, aggregated at the city level, approximates worker productivity. Footnote 33 shows that national average urban wealth strongly correlates with GDP per capita.

I also calibrate the productivity of international ports so that overall production matches national accounts data. Averaging estimates from the UN National Accounts Main Aggregates database between 2020 and 2022 yields that imports were equal to 26.5% of African GDP. Since my network does not include foreign locations,<sup>34</sup> the imported goods will have to be produced by port cities. Assuming that port outflows in 2020Q1 are proportional to the inflows, multiplying outflows from the 52 largest ports by a factor of 37 yields (precisely) an imports measure summing to 26.5% of the total production volume using the IWI as productivity measure. These imports need to be divided by port city population to yield productivities. Since the framework of Fajgelbaum & Schaal (2020) endogenously allocates labor to different tasks, letting port cities produce both a domestic and international variety would obscure this calibration. Thus, I compound both tasks so that port cities use all labour to produce a mixed international variety at productivity

$$Z_j^{\text{intl.}} = \text{IWI}_j + \frac{37 \times \text{Outflow in TEU } 2020 \text{ Q1}_j}{\text{City Population}_j}. \quad (16)$$

The resulting increase in purchasing power implies that a part of the equilibrium consumption of port cities is exported. For simplicity, I assume the export share is equal to the share of international productivity in total city productivity and adjust consumption and welfare accordingly. Another implication of this calibration is that I am constrained to fixed labor scenarios. The mobile labor case, examined by Graff (2024) for national African road networks and by Fajgelbaum & Schaal (2020) for France, Spain, and Western Europe, lets workers reallocate to more productive and better-connected locations - implying potentially large changes in the production of imports.

## 6.1 Optimal Regional Network Investments

Due to computational constraints, I parameterize the full transport network graph from Section 4.3 with 1379 nodes and 2825 edges using only  $N = 4$  traded goods, three of which are produced by cities (agglomerations) of sizes 0-200K, 200K-1M, and >1M, respectively, and the fourth, international good, is produced by port cities. This incentivizes regional trade between larger and smaller cities, hinterland cities and coastal ports. I think of these goods as consumption bundles. Small cities may specialize in agriculture, medium-sized cities in simple manufactured goods, and large cities in diverse manufacturing. Ports provide a bundle of international goods.

Due to their compound nature, the elasticity of substitution between them should be significantly lower. My benchmark specification is  $\sigma = 1.5$ , which is the lowest value where the model delivers robust solutions. In equilibrium, all cities still consume > 85% of their own output, and large/port cities > 96%. This seems much when aggregate data suggests that at least 25% of consumption is imports, but such are the limitations of using the framework on a large network.

Table 8 summarizes city populations and productivities. The majority of people, 283 million, live in large interior cities (agglomerations) with >1 million people, and nearly 120 million live near large ports. Port cities are the most productive, even without international (import-related) productivity. When adding intl. productivity, their total production volume almost doubles from 7 to 12.7 billion IWI points, higher than the 9.3 billion IWI points produced by major interior cities. The median intl. productivity is 27 IWI points, but the mean is much higher at 2748, indicating high import volumes in some port cities with small populations. Top of the list is Egypt's Sokhma Port (towards the Red Sea), with a population of 1000 and intl. productivity of 134,656, followed by Egypt's Port Said, with a population of 894,850 (within a 30km radius) and intl. productivity of 1,264. Other ports with intl. productivity above 500 include Tanger Med and Walvis Bay. Since

---

<sup>34</sup>It would be difficult to add foreign markets to an already large network. In theory, this could be done by adding additional nodes for ports and separating the city from the port in this way, but for any goods to flow to these locations, they need to have populations. Since flows are an equilibrium outcome, they would have to be calibrated to match the data by reverse engineering the model, which is quite complicated solving for populations of multiple ports. The framework may also have difficulties computing optimal flows along zero-cost links.

utility functions are mildly concave ( $\alpha = 0.7$ ) and the planner maximizes utility, the comparatively small populations of these port cities downweights their importance for network optimization.

Table 8: Productivity and Population by City Type

City Type Count	1-200K 625	200K-1M 248	>1M 59	Port 51
<i>Population</i>				
Sum	48.9M	112.2M	283M	117.2M
Mean	77.5K	422K	3.29M	2.3M
Median	66.1K	361K	1.64M	1.04M
<i>Productivity</i>				
Total	1.50B	3.76B	9.28B	12.67B
Total Domestic	1.50B	3.76B	9.28B	6.97B
Domestic Mean	30.37	35.48	43.64	52.87
Domestic Median	26.58	31.49	43.10	53.64
International Mean	0	0	0	2784.44
International Median	0	0	0	27.15

*Notes:* Table shows population and productivity (IWI) summarized for cities of different sizes and port-cities. International productivity (for imports) is the second term in Eq. 16.

Figure 33 provides an overview of the calibrated network structure. Like Fajgelbaum & Schaal (2020), I conduct all simulations assuming cross-good-congestion, as it is realistic to assume that the same roads are utilized to transport different goods. The maximum investment volume in this network comprising all proposed roads and upgrades is 161 billion USD'15. To create an interesting allocation problem, I assume the social planner has \$50B, around 1/3 of the needs.

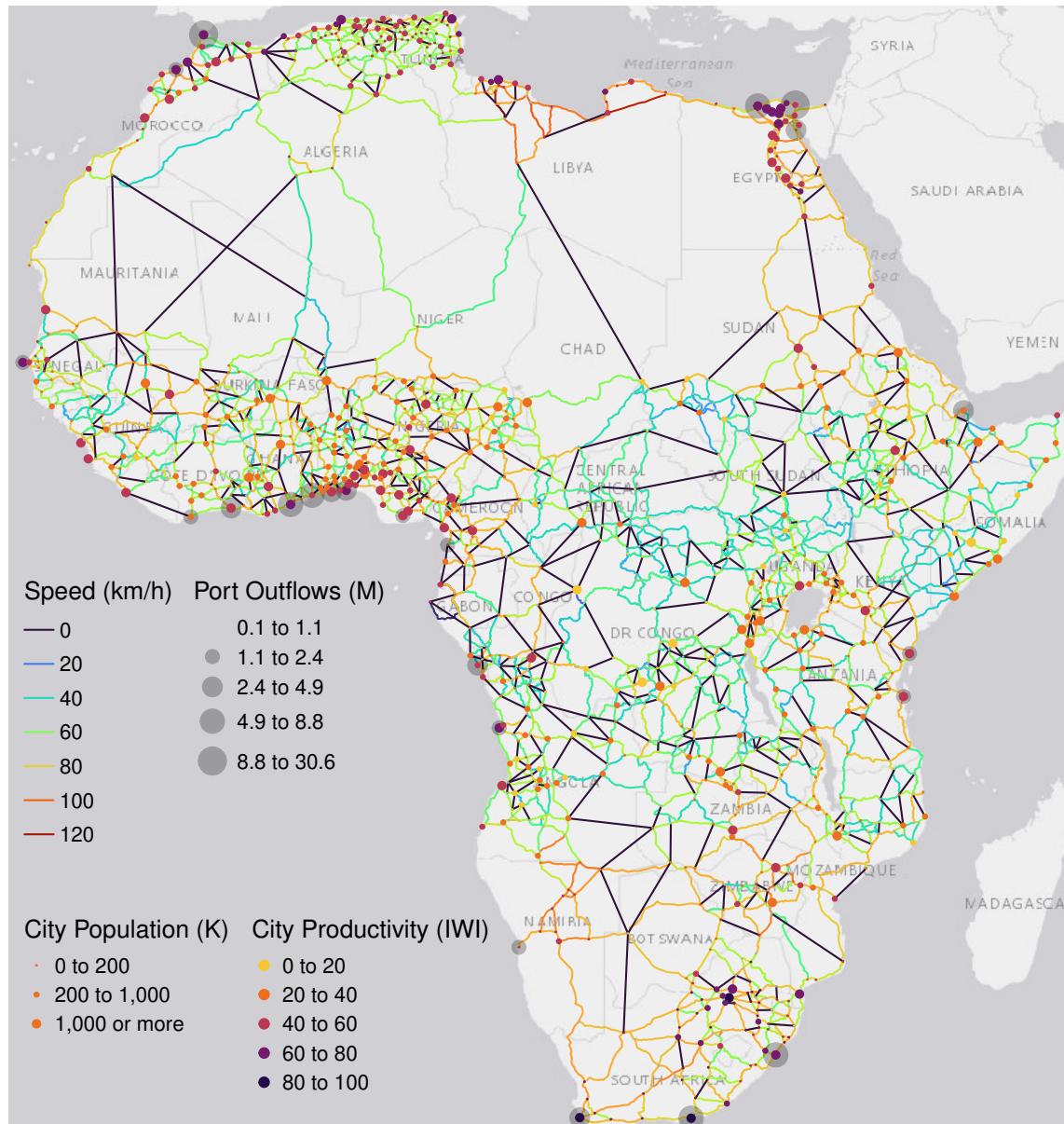
### 6.1.1 Decreasing Returns to Scale in Infrastructure

Under decreasing returns (DRS) ( $\gamma < \beta$ ), the \$50B social planner focuses on connecting high-productivity large cities and port cities to surrounding smaller cities, particularly where road connectivity is inferior. Fajgelbaum & Schaal (2020) find that, in equilibrium, more infrastructure is allocated to regions with lower levels of infrastructure, higher population, and greater worker productivity. Figure 34 visualizes the optimal allocation. The left panel shows network speed post optimization, the middle panel the speed difference to the current network, and the right panel the percentage to which the stipulated work on each link was completed. Thus, the right panel measures the extensive margin of road work, the middle panel the intensive margin.

Overall, this planner commissions road activities equivalent to 130,562km of fully built/upgraded road segments, of which 94,008km are upgrades at a cost of \$29B, and the remaining 36,554km are fresh construction at a cost of \$21B. These activities induce a non-risk-averse model-implied utilitarian welfare gain of 5.8% and a market access gain of 23.2%. The welfare gain is proportional to a consumption gain of 8%. Both are spatially heterogeneous, with utility per worker gains ranging from a 15% loss (1st percentile) to a 158% gain (99th percentile), at a median gain of 12%. The gains are negatively correlated with the IWI ( $r = -0.31$ ) and also with population ( $r = -0.14$ ), indicating that they are mildly redistributive. The MA gains are lower than the consensus packages of Section 5.8, where the second package at a cost of \$36.6B yields a 30.6% MA gain. This difference is due to the welfare maximizing motive and limited incentives for continental trade. Notably, PE packages invest more in major markets and trans-African links connecting them.

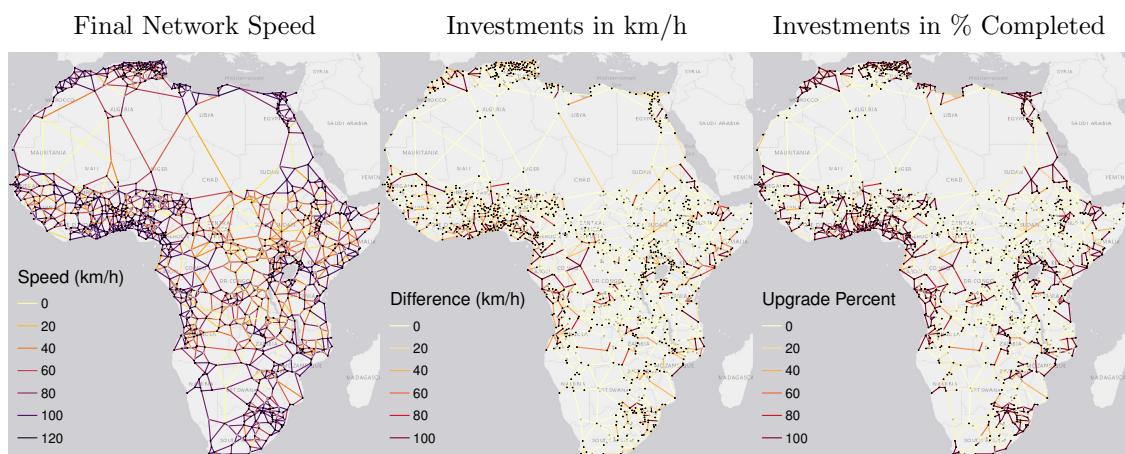
To better understand the planner's optimal allocation, I also examine the flow of goods. Figure 35 visualizes the average flow (across both directions) along each link. The scales are logarithmic, owing to significant differences in the magnitudes of flows. The geographic reach of flows differs by city size. Most small and medium-sized cities' trade is 'local', whereas large cities and ports send their goods to far-away destinations. Only these goods are shipped on trans-Saharan roads, explaining why the planner hardly invests in them.

Figure 33: Calibrated Network for Regional Trade (4 Goods) General Equilibrium Simulations



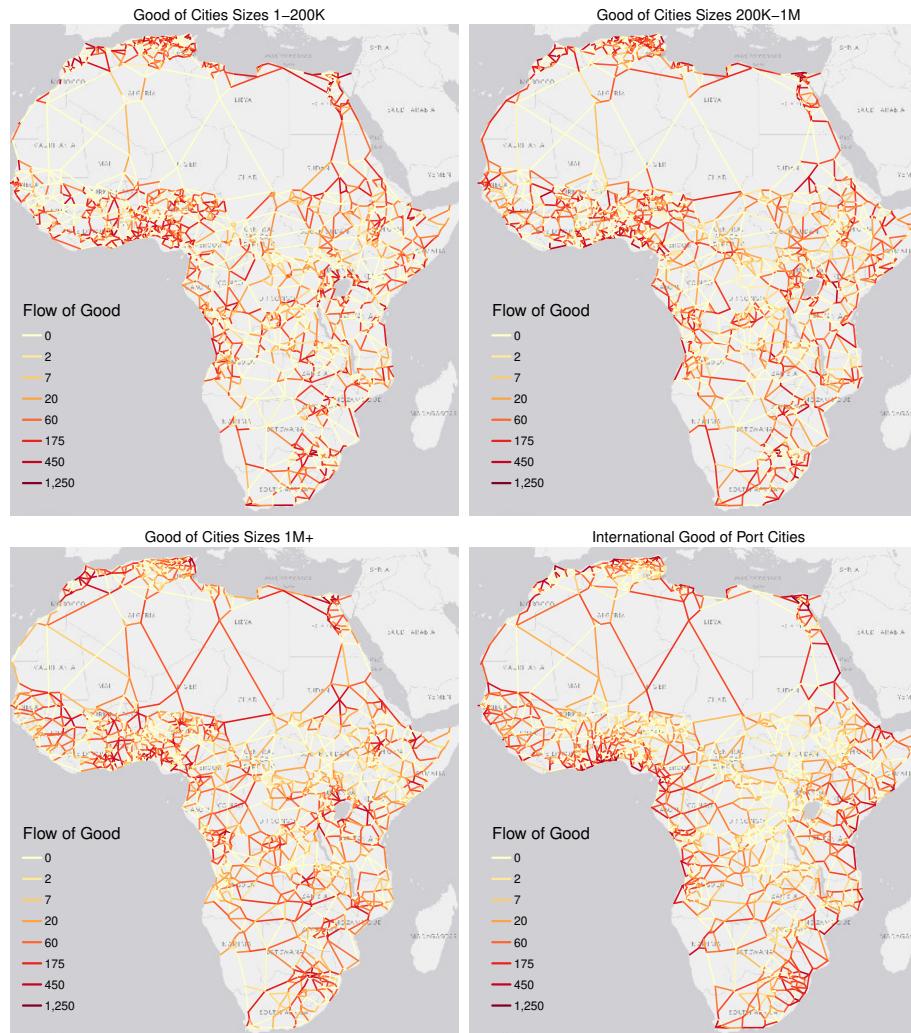
Notes: Figure shows network calibrated with speed, city population and productivity (IWI) and port outflows → Eq. 16.

Figure 34: Optimal \$50B Regional Network Investments under Decreasing Returns



Notes: Figure shows the optimal spatial allocation of a \$50B social planner. Investments in km/h are bounded above by 100km/h. The %-completed is the invested amount (speed) divided by the difference of the initial speed from 100km/h.

Figure 35: Flow of Goods Following \$50B Investments under Decreasing Returns



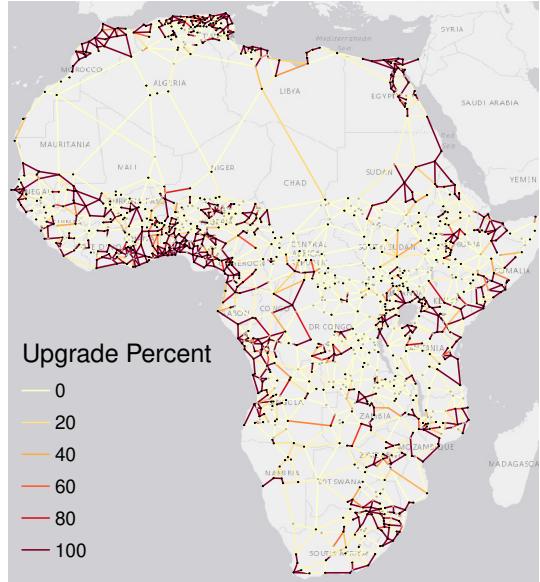
*Notes:* Figure shows the equilibrium flow of goods produced by difference (port-)cities on a logarithmic scale.

Despite the relatively low elasticity of substitution ( $\sigma = 1.5$ ), all cities primarily consume their output. The median small (0-200K) city's own consumption share is 85.5%; for medium-sized cities (200K-1M), it is 93.5%; large cities (>1M) 96.7%, and for port cities, even 98.9% own consumption. Further lowering  $\sigma$  gives smaller own consumption shares but introduces numerical stability issues in the more complex increasing returns case. Yet, optimal infrastructure is relatively insensitive to  $\sigma$ . Appendix Figure A14 shows optimal network investments in percent of work completed, and below it the flow of major-city goods and local welfare gains, for  $\sigma$  values of 1.5, 2, and 4.

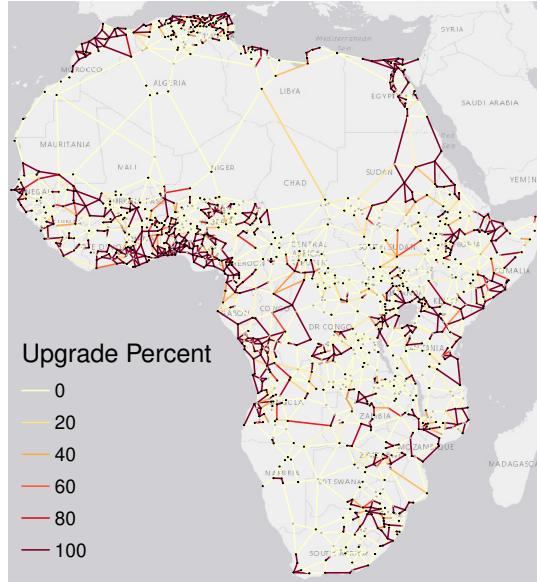
Higher elasticities induce more localized infrastructure investments in the vicinity of large cities, more concentrated trade around these centers, and reduced welfare gains, both overall and particularly in remote locations. As alluded to earlier, assuming  $\sigma = 4$  yields a near-autarky equilibrium where small/large cities consume 97%/99.8% of their output. The overall welfare gain of 0.19% is minimal and also non-redistributional, at a correlation of  $r = 0.25$  with the IWI versus  $r = -0.31$  under the benchmark elasticity. At  $\sigma = 2$ ,  $r = 0.07$ , indicating a relatively even distribution of welfare gains. Yet the optimal infrastructure allocation is largely unaffected, except for specific locations such as Juba in South Sudan. Under  $\sigma = 4$ , the planner fully connects Juba to surrounding cities through four new roads and one upgrade (see Figure 33). Under lower elasticities, these works are only partially completed. Another notable example is the proposed trans-African connection through Chad, connecting Libya with central and eastern Africa. Under  $\sigma = 1.5$ , the planner chooses to construct ~30% of this road, but under  $\sigma = 4$ , no work is done. Thus, trans-African connections are less important under high  $\sigma$ . The optimal allocation is also relatively insensitive to the presence of imported goods produced by ports. Appendix Figure A15 shows that without imports, investments shift slightly towards interior roads.

Figure 36: Optimal \$50B Network Investments under Decreasing Returns: Different Scenarios

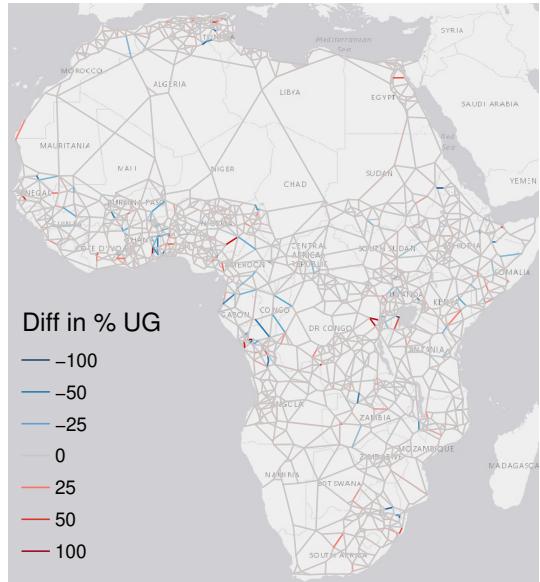
Non-Inequality Averse Planner, No Frictions  
Builds 130,562km - 72% Upgrades at \$29B



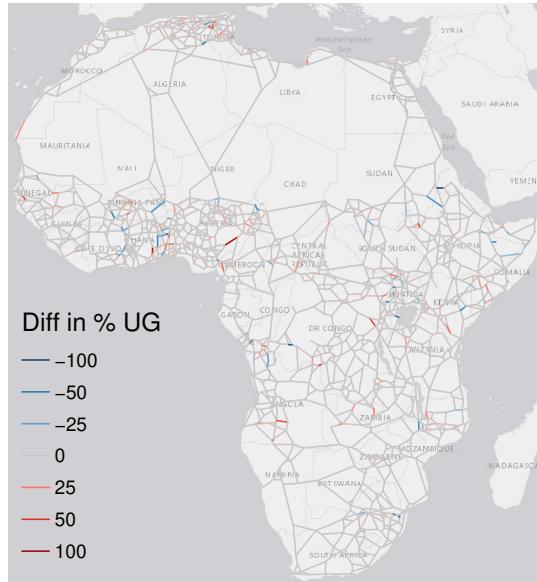
Inequality Averse Planner, No Frictions  
Builds 126,989km - 69.4% Upgrades at \$27.8B



Non-Inequality Averse Planner, With Frictions  
Builds 130,930km - 72.5% Upgrades at \$29.3B



Inequality Averse Planner, With Frictions  
Builds 127,101km - 69.7% Upgrades at \$28B



*Notes:* Figure compares optimal allocations with/without an inequality averse planner and border frictions. Friction results show the difference in upgrade percentage points (% completed) to the frictionless case.

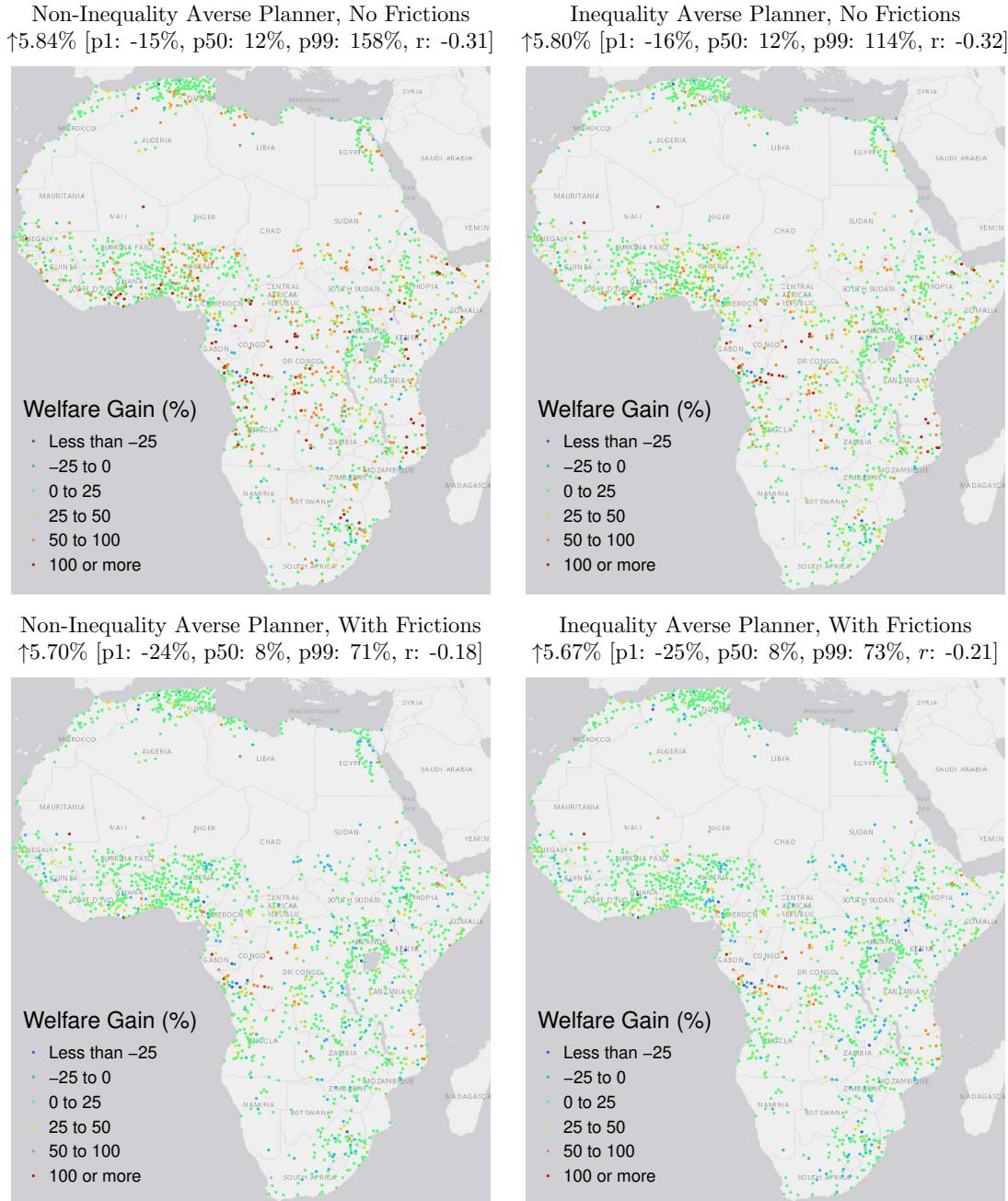
Another important factor influencing optimal investment allocations is the planner's regard for alleviating inequalities ( $\rho$ ). The benchmark assumes  $\rho = 0$  (no inequality aversion). Unfortunately, setting  $\rho = 2$  or even 1.5 is computationally infeasible on the large network.<sup>35</sup> However, since housing ( $h_j$ ) is fixed, the consumption share  $\alpha$  is an alternative way of controlling the planner's regard for inequality - albeit less effective than the  $\rho$  parameter. Figure 36 shows optimal network investments under  $\alpha = 0.1$ , which simulations on small random networks reveal has almost the effect of  $\rho = 2$ . Under inequality aversion, investments in peripheral links increase slightly. For example, the planner better connects Juba and upgrades the link connecting Khartoum to Egypt.

What happens under border frictions? I investigate this by adding the distance-equivalent border frictions (Figure 9) to  $\delta_{jk}^\tau$  (Eq. 12). Surprisingly, optimal regional network investments change very little; the planner slightly reduces investments on certain links, many of which are

<sup>35</sup>Similar to lowering  $\sigma$  too much, inequality aversion makes the planner's objective too complex.

close to borders, and expands others instead. The bottom half of Figure 36 shows the difference to the frictionless case. However, their effects on trade flows and welfare are considerable. Appendix Figure A16 shows the optimal flows ratio. To say the least, the adjustment is complex; the ratio is not bounded above by 1. Rather, flows through expensive border-crossing links are redirected to less expensive border-crossing links. Overall consumption and welfare gains are reduced slightly, and particularly in remote cities which would trade more without frictions.

Figure 37: Local Utilitarian Welfare Gains under Decreasing Returns



*Notes:* Figure shows welfare gains from allocations with/without an inequality averse planner and border frictions.

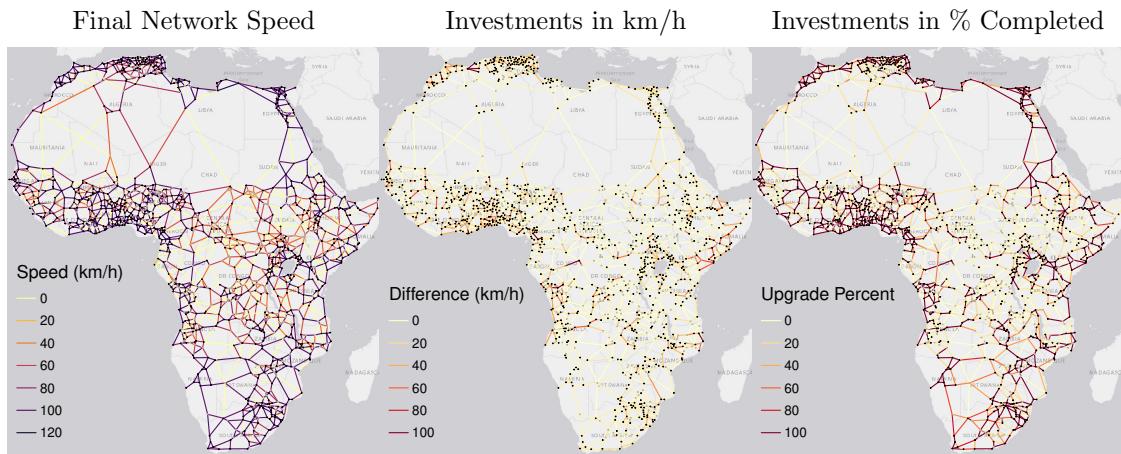
Figure 37 shows the bespoke heterogeneity in welfare gains. For the inequality-averse planner ( $\alpha = 0.1$ ), I recalculated welfare gains by re-solving the model under  $\alpha = 0.7$  using the optimal investments of the  $\alpha = 0.1$  planner. In both scenarios, workers in Congo, Zambia, Mozambique, Nigeria, Cote d'Ivoire, Ghana, Sudan and Somalia gain the most. In the inequality-averse case, the gains are slightly more equitable at  $r = -0.32$  (vs.  $-0.31$ ) correlation with the IWI. The bottom panel shows that median welfare gains are reduced from 12% to 8% under frictions. Remote regions benefit less from the investments, and gains are less redistributive at  $r = -0.18$  /  $-0.21$ .

### 6.1.2 Increasing Returns to Scale in Infrastructure

More than any parameter considered so far, the returns to infrastructure govern its optimal allocation. The benchmark specification is decreasing returns ( $\gamma < \beta$ ). [Fajgelbaum & Schaal \(2020\)](#) make this assumption drawing on a study of the determinants of driving speed in large U.S. cities by [Couture et al. \(2018\)](#). [Graff \(2024\)](#)'s calibration also implies decreasing returns, but his parameters are derived very indirectly from the literature (not from econometric estimates) and thus form a sparse evidential basis. It may well be that trans-African connections, much less prone to congestion than inner-city roads, exhibit increasing returns ( $\gamma > \beta$ ).

Follow [Fajgelbaum & Schaal \(2020\)](#), I invert the ratio of  $\beta$  and  $\gamma$  keeping  $\beta$  fixed, setting  $\gamma' = \beta^2/\gamma = 1.465$ . I then approximate the global optimum of the now non-convex optimization problem with simulated annealing and [Fajgelbaum & Schaal \(2020\)](#)'s random rebranching algorithm, whose desirable properties they establish. Figure 38 shows optimal network investments under IRS.

Figure 38: Optimal Network Investments under Increasing Returns



*Notes:* Figure shows the optimal allocation of a \$50B social planner under increasing returns to infrastructure ( $\gamma > \beta$ ).

The 50 billion IRS planner commissions 166,549km of road work, considerably more than the DRS planner at 130,562km. Of these, 163,701km or 98.3% are upgrades at a cost of \$48.2B, and only 2,848km is new construction at \$1.8B. Compared to the DRS case (Figure 34), the investments are more dispersed, and the planner builds/upgrades fewer roads near large cities. The difference is particularly manifest in southern Africa, where the IRS planner chooses to resurface roads in the entire region. In contrast, the DRS planner only invests locally around Cape Town and Johannesburg. The total welfare gain of 7.51% is higher than the 5.84% gain under DRS, consistent with [Fajgelbaum & Schaal \(2020\)](#) and [Gorton & Ianovichina \(2021\)](#). The IRS investment is also better for MA, yielding 26.9% gain in total MA, versus 23.2% under DRS. It is also more equitable, with welfare gains ranging from -6% (p1) to 103% (p99) at median gain of 12% as in the DRS case, and slightly more redistributional at  $r = -0.32$  IWI correlation. However, not everyone benefits under this scenario. For example, the IRS planner does not build new roads to better connect Juba, which are deemed relatively too expensive compared to upgrades in other regions.

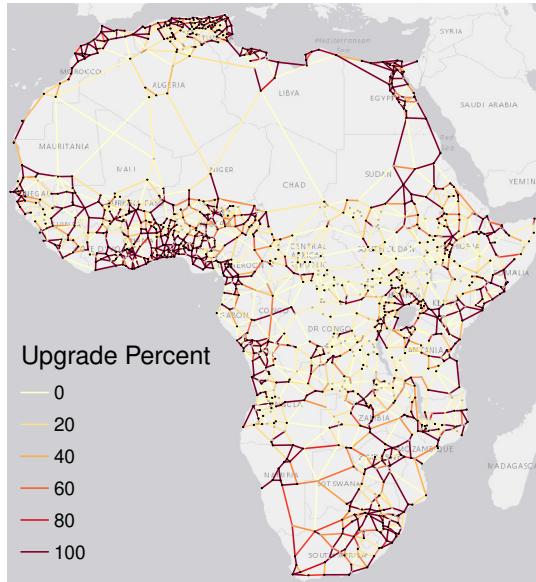
Under IRS, consumption flows in all goods, reported in Appendix Figure A17, are much larger than under DRS. This is also reflected in cities' own consumption shares. Under the benchmark  $\sigma = 1.5$ , small/medium/large/port-cities consume 52%/80%/88%/96% of their produce, considerably less than under DRS. Appendix Figure A18 shows that optimal investments are again relatively insensitive to  $\sigma$ . Raising  $\sigma$  lets them become more concentrated around cities; for example, under  $\sigma = 1.5$ , the planner builds/upgrades a long road into central DRC, unlike under  $\sigma = 4$ . Trade flows again become more local, and welfare gains decrease but remain larger than under DRS and, importantly, redistributional, at an IWI correlation of  $r = -0.2$  even under  $\sigma = 4$ . The optimal allocation mildly reacts to ignoring imported goods, as Appendix Figure A19 shows.

Can a more inequality-averse planner still improve the spatial allocation? Figure 39 shows the

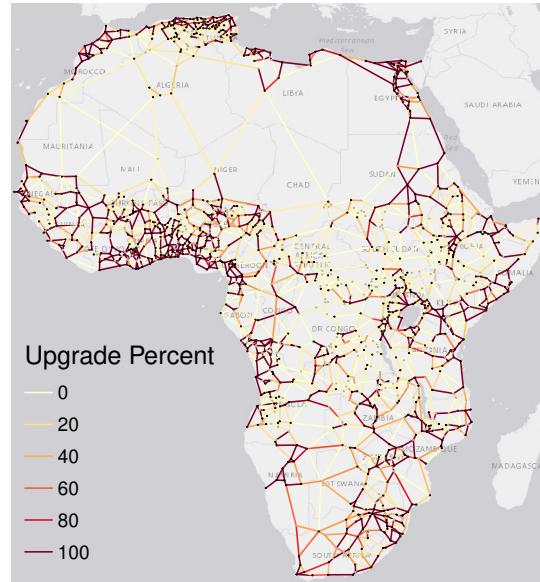
optimal investments.<sup>36</sup> The averse planner invests slightly more in countries like Angola, Congo, Sudan, and Somalia and less in northern, southern, and western Africa. Notably, (s)he upgrades an additional road in Congo but reconnects Morocco to Algeria using one instead of three links and does not upgrade a link connecting Algeria to Libya. Border frictions added at the bottom of Figure 39 (difference view) have a relatively small impact on the allocation but significantly change optimal flows (see Appendix Figure A20), and reduce welfare gains.

Figure 39: Optimal \$50B Network Investments under Increasing Returns: Different Scenarios

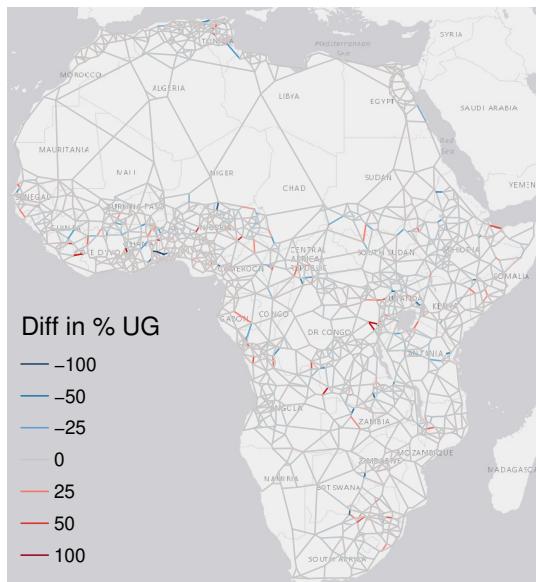
Non-Inequality Averse Planner, No Frictions  
Builds 166,549km - 98.3% Upgrades at \$48.2B



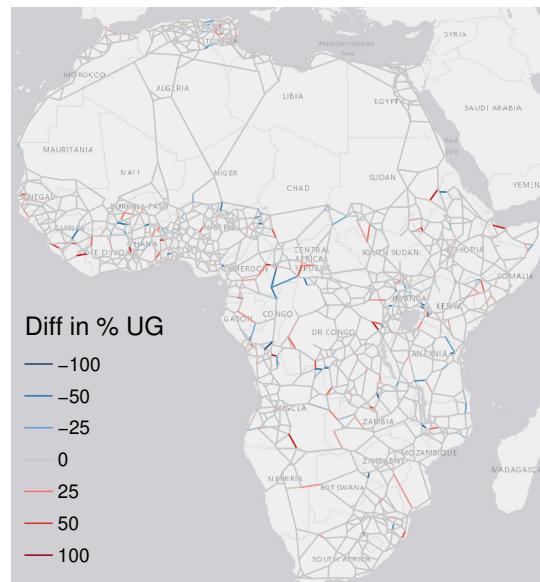
Inequality Averse Planner, No Frictions  
Builds 165,798km - 98.5% Upgrades at \$48.6B



Non-Inequality Averse Planner, With Frictions  
Builds 166,337km - 98.4% Upgrades at \$48.3B



Inequality Averse Planner, With Frictions  
Builds 165,605km - 98.4% Upgrades at \$48.5B



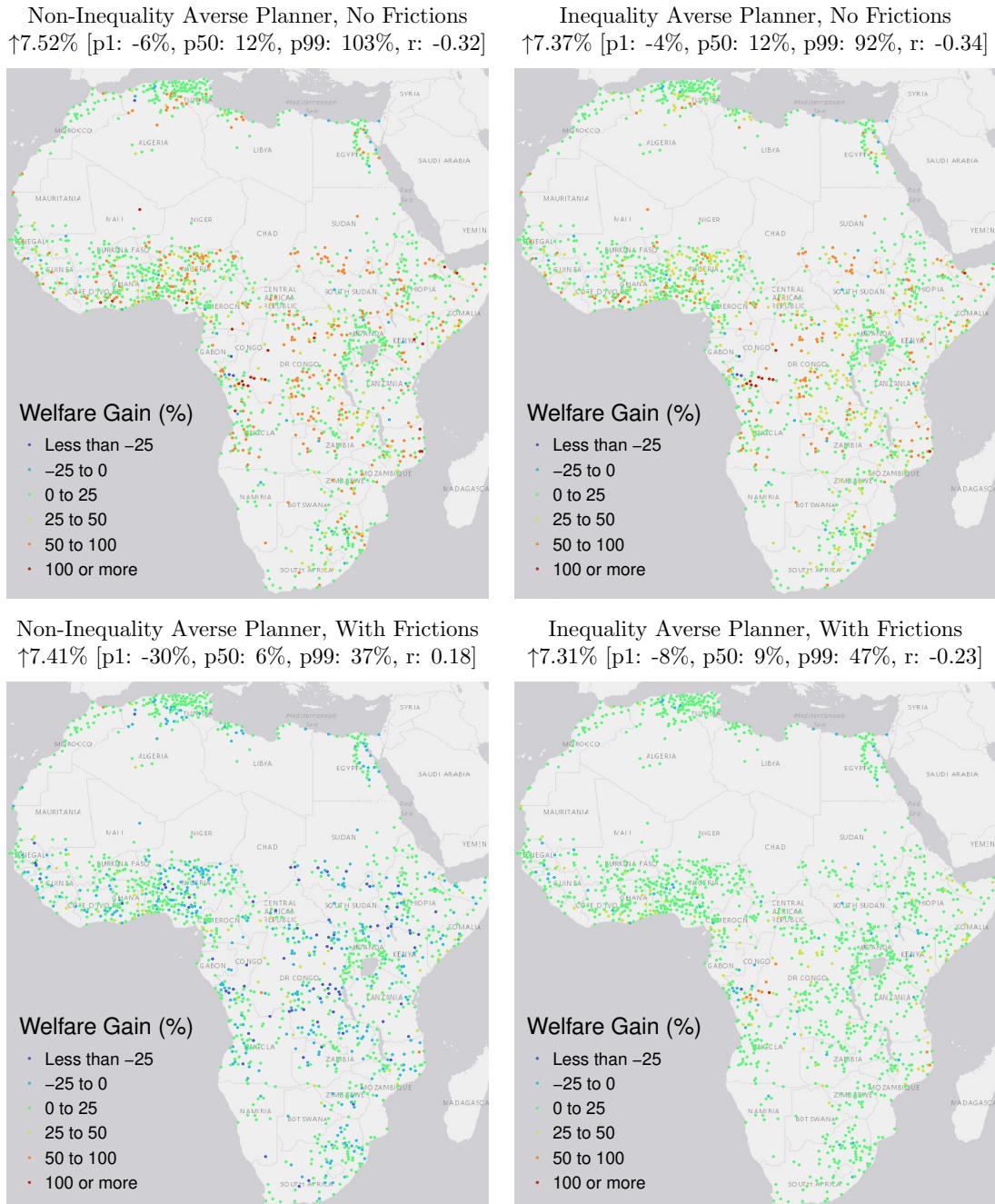
*Notes:* Figure compares optimal allocations under IRS with/without an inequality averse planner and border frictions.

Figure 40 shows welfare gains for IRS scenarios. Gains under the inequality-averse planner are again slightly lower on average but more equal and redistributive at  $r = -0.34$ , up from  $r = -0.32$ . Due to the concave utility of consumption, the welfare gains are highest in more remote cities, which are now better connected. These results are consistent with [Gorton & Ianchovichina](#)

<sup>36</sup>Due to bespoke numerical challenges, I again test this setting  $\alpha = 0.1$  to compute the optimal infrastructure allocation and re-solve the spatial model with  $\alpha = 0.7$  afterwards.

(2021). The frictions case under IRS is particularly interesting. The non-averse planner achieves an aggregate welfare gain of 7.41%, but the median city only gains 6%, with welfare losses up to -30% in remote cities due to trade diversion. The gains become non-redistributional at IWI correlation of  $r = 0.18$ . In stark contrast, the averse planner achieves a slightly lower aggregate welfare gain of 7.31%, but the median city gains 9%, the overall distribution of city-level gains is much more positive, and the gains are redistributional at  $r = -0.23$ .<sup>37</sup> Overall, the IRS allocation with inequality-averse planner yields socially optimal outcomes and significantly enhances regional connectivity through >160K km of road upgrades and 2.5K km of strategic new roads. Currently, there is no empirical evidence supporting the IRS case, but suitable data on trade costs, traffic flows, and road upgrades along African highways could produce such evidence.

Figure 40: Local Utilitarian Welfare Gains Under Increasing Returns



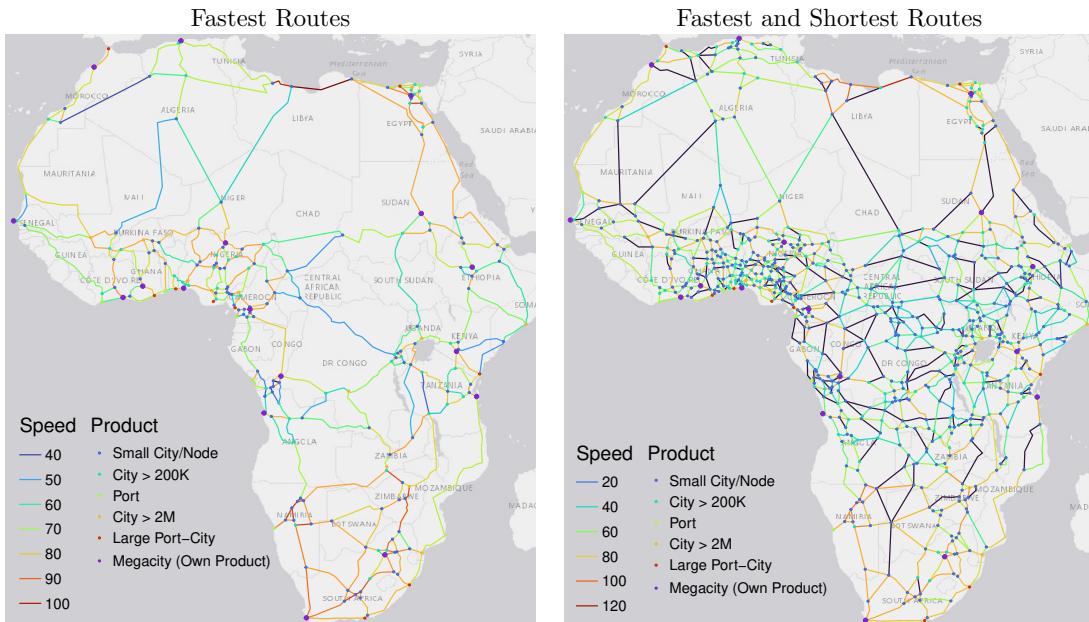
Notes: Figure shows welfare gains from IRS allocations with/without an inequality averse planner and border frictions.

<sup>37</sup>This result, verified by re-solving the model under both allocations and frictions with the same parameters, is purely the outcome of 'minor' strategic differences in road investments as shown in the bottom half of Figure 39.

## 6.2 Optimal Trans-African Network Investments

Having examined optimal regional network investments by letting cities of different sizes and ports trade with each other, I now examine optimal investments in trans-African roads connecting the largest 47 (port-)cities on the reduced networks developed in Section 4.4. To create strong incentives for trans-African trade, I let 17 large and strategically located megacities produce their own goods. Other cities are classified into large port cities (population >1M and outflows >1M TEU in 2020Q1), large cities (population >2M), ports (outflows >0), medium sizes cities (population >200K) and small cities otherwise. Figure 41 shows the classified networks.

Figure 41: Trans-African Network Connecting Large Cities: Parameterization



*Notes:* Figure shows networks connecting 47 large (port-)cities with link speeds and a classification of traded products.

These networks are consistent with the larger network considered so far and parameterized using the same routing data. Edges are contracted by summing distances and travel times and recalculating travel speeds on the new, longer edges. Appendix Figure A21 provides an equivalent representation of the fastest-routes network with real roads along the original edges, illustrating that some of these 'long edges' are a patchwork of different roads at heterogeneous travel speeds.

### 6.2.1 Fastest Route Investments under Decreasing Returns

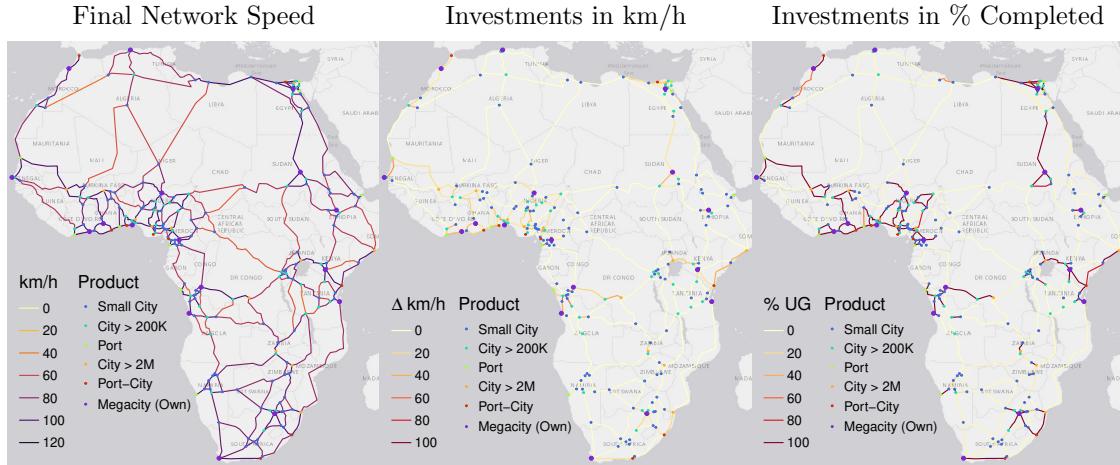
I begin by considering just upgrades along the existing fastest routes. These sum to a total cost of \$33B. I assume the planner has \$10B to invest in trans-African roads - about 1/3 of the total needs. The parameterization otherwise follows the previous subsection, with added international productivity in port cities to account for imports. I use  $\sigma = 3.8$  as benchmark elasticity following Bajzik et al. (2020) and Armington (1969), but also run simulations with  $\sigma = 5$  and  $\sigma = 2$ .

Figure 42 shows that under DRS ( $\gamma < \beta$ ), the \$10B planner mainly invests in connections between large cities, megacities, and ports. This includes better connecting Khartoum to upper Egypt, Kampala to Mombasa, Kinshasa to Mbaji-Mayi, and improving connectivity in West Africa, particularly along the extended coastline between Abidjan and Yaounde. In total the planner completely upgrades 40,164km of roads. The welfare gains of only 0.04% are very small and nonredistributional at  $r = 0.13$  correlation with the IWI. Similar to the  $\sigma = 4$  cases considered above, the equilibrium is close to autarky: small cities consume >95% and megacities >99.7% of their own produce. Appendix Figure A22 visualizes the flow of goods.

Unlike the four good case on the full network, optimal investments are quite sensitive to the value of  $\sigma$ . Appendix Figure A23 provides a comparison. While  $\sigma$  values of 3.8 and 5 yield relatively similar results (under  $\sigma = 5$  the planner interestingly better connects Kinshasa to

Bukavu/Rwanda), the  $\sigma = 2$  planner spends much more on roads crossing central Africa. Notably, a significant fraction of the budget is devoted to upgrading an expensive road across the DRC connecting Bukavu/Rwanda to Bangui in the Central African Republic. (S)he also better connects Khartoum to West Africa and upgrades a segment in Algeria common to two trans-Saharan roads, instead of better connecting Khartoum to Egypt, Dakar to Bamako, and Mombasa to Mogadishu.

Figure 42: Optimal Trans-African Network Investments



*Notes:* Figure shows the optimal allocation of a \$10B social planner investing in the trans-African fastest routes network.

In the model,  $\sigma$  governs consumers' love for trans-African variety, but we may also reinterpret  $\sigma$  from the production side as a propensity to form trans-African regional value chains (RVCs). Thus, the results signify that, under decreasing returns to infrastructure, significant efforts to upgrade central African roads - the worst on the continent - are only sensible if consumers' love for variety or the propensity to form RVCs is relatively strong. Ignoring imports again shifts optimal investments slightly to interior roads, as the upper middle panel of Appendix Figure A23 shows.

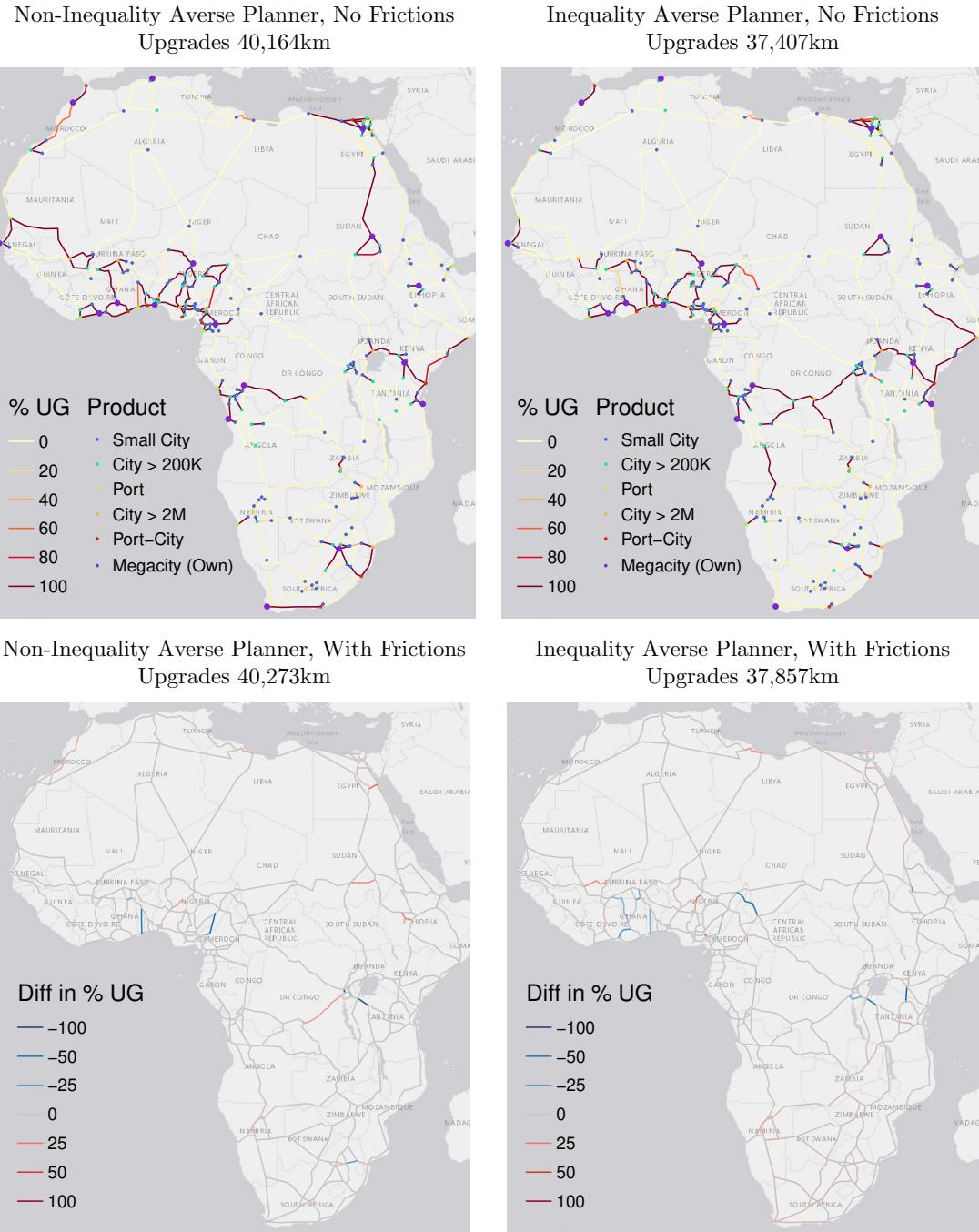
With inequality aversion ( $\rho = 2$ ), the optimal allocation, reported in Figure 43 and further in Appendix Figure A24 for different  $\sigma$  values, changes slightly. The planner invests more in poorer countries such as Angola, Congo, and Somalia/Ethiopia, and less in other regions. Welfare gains are again small at maximally 0.5% under  $\sigma = 2$  (and 0.03% under the benchmark  $\sigma = 3.8$ ) and thus not reported in further detail. But their distribution remains modestly non-redistributive at  $r = 0.13$ . Consumption remains relatively autark, with megacities consuming >99.7% of their own produce under the benchmark  $\sigma = 3.8$ , and >98.7% under  $\sigma = 2$ .

The optimal investment allocation is also largely unaffected by the inclusion of border frictions into  $\delta_{jk}^\tau$  (Eq. 12), as the bottom panel of Figure 43 and the lower middle panels of Appendix Figures A23 and A24 show. Since in PE (Section 5), frictions had a significant impact on optimal allocations, this suggests that the iceberg formulation (Eq. 11) with logarithmic distance (Eq. 12) may not be the most appropriate way to include them. Better accounting for border frictions in quantitative spatial models with endogenous infrastructure thus invites further research.

### 6.2.2 Fastest Route Investments under Increasing Returns

Surprisingly, with increasing returns to infrastructure ( $\gamma > \beta$ ), the optimal allocation, reported in Figure 44 and, in more detail, Appendix Figure A26, hardly changes. Only locally there are minor differences to Figure 42 in terms of which segments are upgraded. This suggests that the returns to infrastructure are less important when restricting the planners' choice to important transport routes. The propensity to trade long distances, governed by  $\sigma$ , is the key parameter governing optimal allocations. However, interestingly, trade flows still increase considerably under IRS, as Appendix Figure A25 shows. The total welfare gains from the investment at 0.16% are still small but considerably larger than the 0.04% under decreasing returns. But they remain non-redistributional ( $r = 0.14$ ), and megacities consume more than 98.2% of their own produce. Appendix Figure A27 provides sensitivity analysis with very similar results to the DRS case.

Figure 43: Optimal \$10B Trans-African Network Investments: Different Scenarios



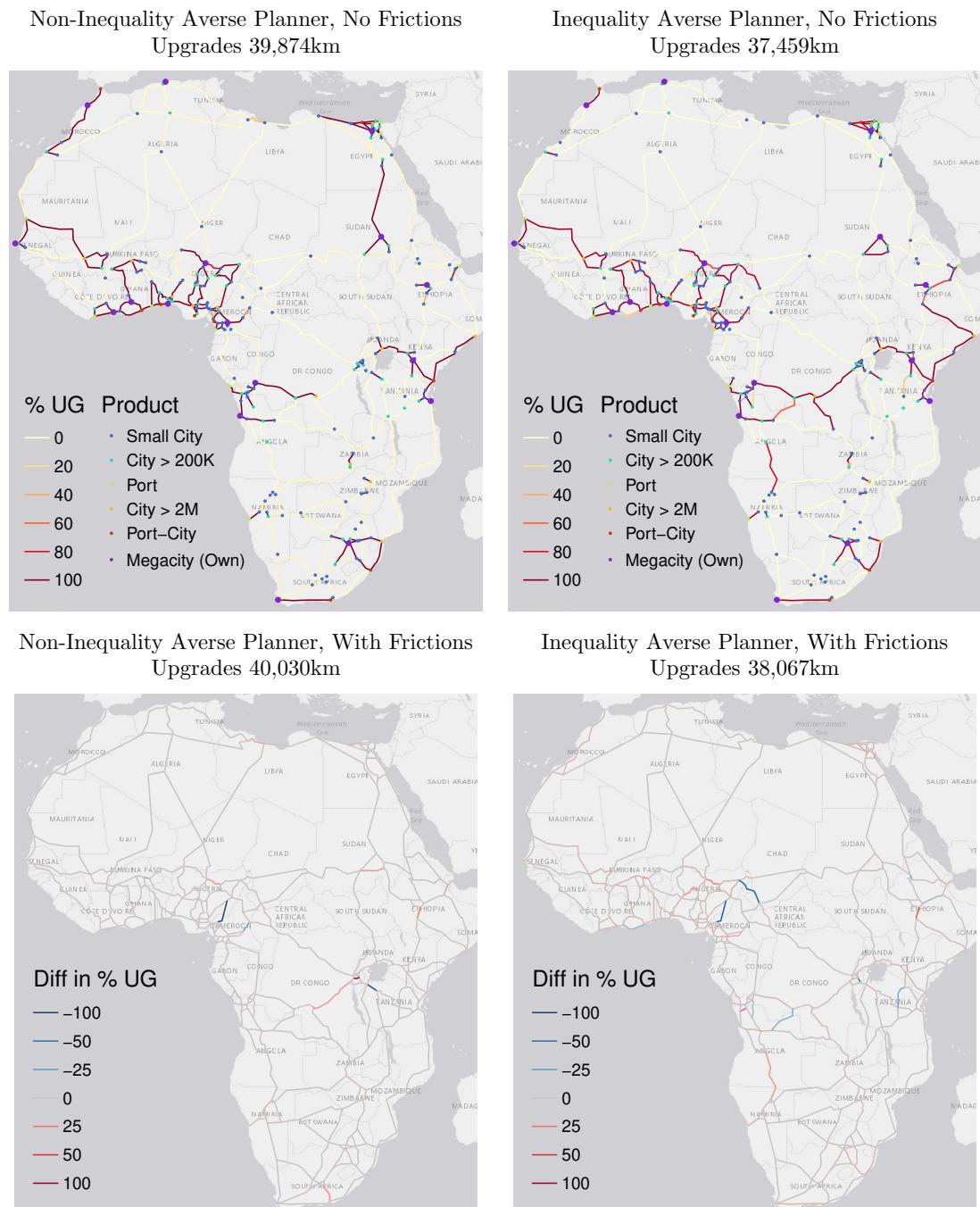
*Notes:* Figure compares optimal allocations (in % upgraded) with/without inequality averse planner and border frictions.

As under DRS, making the planner inequality averse ( $\rho = 2$ ) redistributes investments to poorer regions, particularly in DR Congo, Angola, and Sudan/Ethiopia. The allocation is again relatively insensitive to including border frictions, the most notable difference being cross-border roads into Nigeria from Cameroon and Chad, which are not upgraded under frictions. The average cost of crossing the Nigerian border is estimated equivalent to 1944km of road distance.

Thus, the distribution of optimal investments in trans-African roads depends on the elasticity of substitution ( $\sigma$ ) and the planner's inequality aversion ( $\rho$ ). The central African infrastructure gap evident from Figures 2 and 4 is particularly interesting. This vast area is largely unpopulated; hence, the optimal regional planner of Section 6.1 does not place much infrastructure there. The non-averse trans-African planner in this section does the same under standard assumptions of  $\sigma = 3.8$  or  $\sigma = 5$ . With inequality aversion ( $\rho = 2$ ) (s)he fully upgrades the road from Kinshasa

to Bukavu and builds roads in Angola and Somalia. However, significant efforts to close the gap in northern central Africa require elasticities as low as  $\sigma = 2$  (see Appendix Figures A23, A24, A27, and A28). In other words, it requires optimism about the development of trans-African trade and RVCs to prioritize upgrading these roads. Presently, trade between ECOWAS and the EAC + IGAD combined makes up 0.2% of total inner-African trade and 0.7% of inner-African trade between different regional economic communities. Closing the central African infrastructure gap is likely to increase these shares significantly, but, as discussed, only a policy priority if  $\sigma$  is lower than typically assumed. Another priority is the reduction of border frictions, which, as Figures 8-9 show, are extremely high along Congolese borders. Appendix Figures A23 and A24 show that under frictions and  $\sigma = 2$ , the planner reduces investments on the Bukavu-Bangui link by around 25% - and likely by more if frictions were incorporated into the framework in a better way. Thus, this intriguing link is not a reasonable present policy option for a \$10B social planner.

Figure 44: Optimal \$10B Trans-African Network Investments under IRS: Different Scenarios



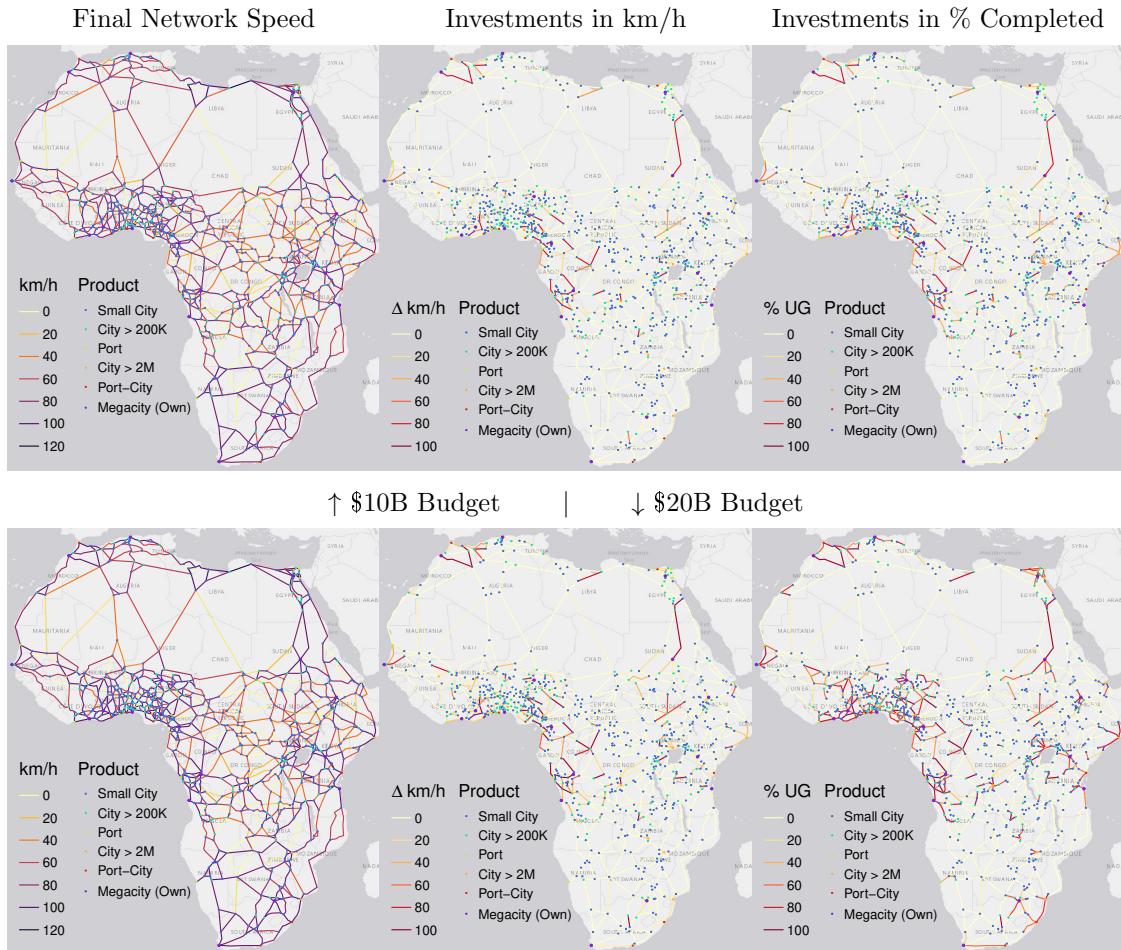
Notes: Figure compares optimal allocations (in % upgraded) under IRS with/without inequality aversion and frictions.

### 6.2.3 Fastest and Shortest Route Investments

It remains to consider the trans-African network comprising both fastest and shortest routes. With 707 nodes and 1274 edges, it gives the planner considerably more choices to improve trans-African connectivity. The total potential investment volume into this network is \$93B. Therefore, a planning budget of only \$10B imposes a tight resource allocation problem. To loosen the resource constraint a bit, I additionally consider a \$20B planner. With 22 goods, the model takes modern non-linear optimizers to their limits. I can only solve it by assuming strong duality, which, following Fajgelbaum & Schaal (2020), requires  $\gamma < \beta$  and  $\beta \leq 1$ . Hence, I set  $\beta = 1$ .

Figure 45 shows both planner's allocations. Ostensibly, they are more similar to the optimal regional investments from Section 6.1 than those in fastest routes. Both planners better connect Khartoum to Egypt, build/upgrade roads surrounding Juba, Kampala, Kinshasa, Luanda, Mbandaka, Dakar, and many West African cities, and better connect Morocco to Algeria. Thus, even with a richer trade characterization incentivizing trans-African trade, investments in major trans-African roads are not a priority if planners can divert away from them, at least under  $\sigma = 3.8$ .

Figure 45: Optimal Trans-African Network Investments: Shortest Routes



Notes: Figure shows optimal allocations of \$10B & \$20B planners investing in the trans-African shortest routes network.

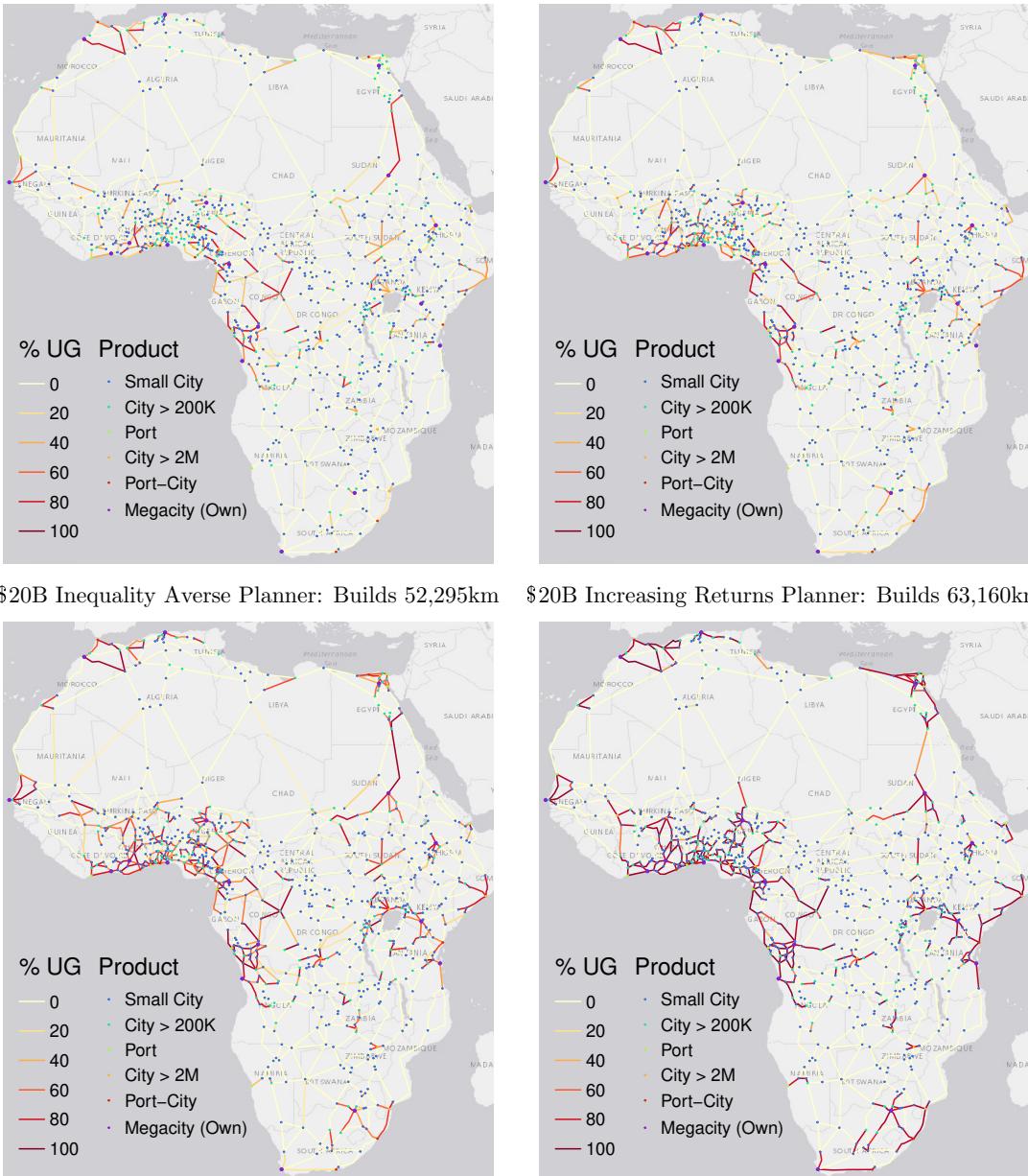
Appendix Figure A30 provides results for different  $\sigma$  values, including the impact of border frictions. Under  $\sigma = 2$  the planners again become more cosmopolitan and, notably, invest mildly in new trans-Saharan links (not in upgrading existing ones), links better connecting Cape Town to Botswana and further north, and significantly improve the connectivity of Luanda, Kinshasa and other Congolese cities to West Africa. Building new trans-Saharan links rather than upgrading existing ones is a particularly interesting choice. Likely these links are significantly more expensive than modelled taking into account just terrain and population, thus this allocation is not cost-efficient. But it clearly shows the planners desire to enable more route-efficient travel between North- and Sub-Saharan Africa. Compared to the network with only fastest routes, the flow of

goods is much more focused on nearby cities. The welfare gains under  $\sigma = 3.8$  are small at 0.12%/0.21% for the \$10B/\$20B planner and mildly redistributive ( $r = -0.08 / -0.1$ ). When lowering  $\sigma$  to 2, the welfare gains increase to 1.34%/2.11%. Appendix Figure A31 provides details.

To cut short the discussion of different scenarios, Figure 46 reports results for inequality averse planners ( $\rho = 2$ ) and IRS planners ( $\gamma > \beta$ ). Averse planners again invest more in central Africa, Angola and Somalia. IRS planners expend more efforts connecting Morocco to Algeria and the populated regions around Kinshasa and Luanda to West Africa, while reducing connectivity from Khartoum to Egypt and around major cities such as Juba. Aggregate welfare gains again drop slightly to 0.1%/1.7% under averse planners, and increase to 0.52%/0.91% under the IRS planners. Appendix Figures A32 and A33 analyze their sensitivity to different  $\sigma$  values.

Figure 46: Optimal Trans-African Network Investments: Shortest Routes: Different Scenarios

\$10B Inequality Averse Planner: Builds 21,491km    \$10B Increasing Returns Planner: Builds 26,848km



*Notes:* Figure illustrates optimal \$10B and \$20B allocations under IRS ( $\gamma > \beta$ ) or inequality averse planners.

Overall, the results even under  $\sigma = 2$  are quite regionally focused. I thus also run simulations for  $\sigma = 1.5$ , generating much larger trade flows and incentives for pan-African investments. Appendix Figures A34 and A35 reports these results for different planners. The investments are indeed much more pan-African. Research on Armington elasticities between African cities could disambiguate.

## 7 Summary and Conclusion

Using detailed routing data, large and accurate network graphs, and a rich economic geography including population, wealth/gdp, trade through ports, border frictions, road construction and upgrading costs, this paper rigorously examines Africa's full continental road network and characterizes optimal investments into it in partial and general equilibrium (PE and GE).

Presently, Africa has 1.6 million km of non-residential roads, around 30% of which are paved. This fraction differs significantly by country and road type. Particularly in central Africa, the Sahara, and the Horn of Africa region, most roads are unpaved and rough, severely limiting trans-African connectivity. The network's overall route efficiency (RE) of 0.68 is acceptable but considerably lower than that of the US network at 0.84. The time efficiency and average travel speed along many roads is significantly worse than in other world regions. Particularly the aforementioned regions can only be traversed at speeds around 40km/h. In addition, cross-border frictions substantially impede continental market access (MA). Estimates suggest that crossing the median border in Africa is as costly as 1043km of road distance, and the time commitment is equivalent to 3.2 days of travel time. These frictions induce an aggregate MA loss between 26% and 78%. Locations across the border from major markets such as Nigeria suffer MA losses of up to 80%.

Given a detailed network and economic geography, optimal road investments still depend on the geographic scale and magnitude of the investment and the planner's objective (MA or welfare maximization with/without inequality aversion). Structural parameters such as the trade elasticity/elasticity of substitution and returns to scale in infrastructure also noticeably impact optimal spatial allocations, as do heterogeneous road construction costs and border frictions. The paper navigates this complexity via a graduated approach, starting with high-resolution small-scale PE investments and gradually progressing towards lower-resolution GE simulations where social planners spend large amounts. This requires different network representations and methods.

Considering first local network investments surrounding 50.3km equally spaced grid cells, I find that increasing connectivity in densely populated high-activity areas such as West Africa around Nigeria, greater Johannesburg, East Africa between Bukavu and Nairobi, Sudan around Khartoum, Upper Egypt, Kinshasa and Luanda, and in corridors connecting these economic areas where transport roads are already present, yields the greatest marginal MA gains. Figures 17 and A11 map out such corridors of high marginal gains, synthesizing rich data on the road network and the distribution of economic activity via an optimally weighted graph.

To increase the scale of investments to entire roads or collections of segments, I build a detailed representation of the transport network connecting 447 important cities and 52 ports along the fastest routes. My network graph has 1379 nodes and 2344 edges (links), summing to 315K km. I apply a network finding algorithm and find 481 potential new links summing to 104K km. Building them all would raise RE by 7.1%, from 0.68 to 0.72, and MA (denominated in road distance) by 5.5%. The most valuable links increasing RE/MA by more than 0.1% are new trans-Saharan roads through Mauritania, Mali, and Chad and reopening the closed Morocco-Algeria border. Some new roads in central and southern Africa would also significantly increase RE/MA. When border frictions are activated, the MA gain drops from 5.5% to 3.9%, with particularly high losses on remote and border-crossing links. The median link yields a 50% lower MA gain under frictions.

According to the World Bank ROCKS database, the median 2-lane highway in Africa costs \$611K 2015 USD per km. Using a heterogeneous road costs prediction model considering road length, terrain ruggedness, and population density, I estimate the cost of the existing transport network at \$186B USD'15, and the proposed extension of \$104K km at \$56B, yielding an average MA return of \$1.4/km per \$ invested in new links, which drops to ~\$0.7/km/\$ if frictions are added.

Examining road quality measured by travel speed, I find that upgrading all links to  $\geq 100$ km/h yields a time-denominated MA gain of 42.1%. New links yield a 14.1% MA gain if built at 100km/h on top of the existing (slow) network. Upgrading the entire network to fast links and also building the proposed ones could yield a 50% MA gain. There is significant heterogeneity in the value of upgrading links, with trans-Saharan links yielding the greatest MA gains. Applying border frictions

lets the total MA gain from upgrading all links drop to 27%. Again, particularly peripheral and transnational links lose value. I estimate the cost of upgrading the entire network at \$106B, comprising \$67.8B full upgrades, \$36.5B mixed works, and \$12.3B asphalt resurfacing. Upgrading all roads yields average gains of \$6/min per \$ spent, which drops to \$2.4/min/\$ under frictions.

To raise average returns, I consider investment packages targeting high marginal gain links (both upgrades and new roads). I propose three packages with links at marginal gains  $>1$ ,  $2$ , or  $4$  \$/min/\$. They cost \$60.9B/\$36.6B/\$17B, yield frictionless MA gains of 38.1%/30.6%/20%, and 26.8%/24%/19.5% under frictions. These packages are globally macroeconomically feasible under frictions if they can raise Africa's aggregate growth rate (4.1%) by 0.33%/0.2%/0.09%. For the  $>\$1$  (\$60.9B) package, this would require that growth in MA translates into economic growth at a rate of  $\geq 81:1$ . This is plausible in light of empirical results suggesting economic returns to MA gains as high as 2:1, but there are several factors that could raise the cost of these packages considerably and lower the macroeconomic returns, especially if they are financed through debt.

Calibrating a general equilibrium framework by [Fajgelbaum & Schaal \(2020\)](#) permitting global optimization over the space of networks, I also derive optimal investments in spatial equilibrium, letting a welfare-maximizing social planner spend a fixed infrastructure budget. My calibration accounts for trade in multiple goods and through international ports. I consider cases with/without inequality aversion, imports, cross-border frictions, and increasing returns to infrastructure. I also simulate with different values for the elasticity of substitution ( $\sigma$ ). Due to the computational complexity of the problem, I split the simulations into two parts. In Part 1, a regional planner improves connectivity between small/medium/large cities and ports. In Part 2, a trans-African planner optimally invests in continental transport roads connecting 47 major (port-)cities.

The regional planner with a budget of \$50B optimally connects large cities and ports with each other and surrounding smaller cities, focusing on populated/productive areas. An inequality-averse planner invests slightly more in the poorer central African and the Horn of Africa regions. With increasing returns to infrastructure, the investments become less concentrated around large cities, and the planner upgrades more roads (building few new ones). This increases welfare gains to 7.4%, up from 5.8% under decreasing returns. With inequality aversion, the gains become more equitable at minor aggregate losses, and the planner upgrades more roads in central Africa. Due to concave utility of consumption, remote cities gain more despite receiving less infrastructure.

The trans-African planner, endowed with \$10B, also mostly improves regional connectivity under standard assumptions of  $\sigma$  between 3.8 and 5. Much is invested along the West African coastline, but also in better connecting Dakar to Bamako, Khartoum to Egypt, Kampala to Mombasa, and Kinshasa to Mbaji Mayi. An inequality-averse planner fully connects Kinshasa to Bukavu/Rwanda and further to Mombasa, Mogadishu, and Addis Ababa. Thus, the planner upgrades a corridor crossing populated central Africa. At low  $\sigma = 2$ , both planners also close the large infrastructure gap in sparsely populated northern central Africa by upgrading roads from Rwanda/Bukavu and Khartoum to West Africa. Thus, optimal trans-African investments depend critically on the propensity to trade. Only at low  $\sigma$  do major infrastructure gaps in central Africa become a policy priority. High border frictions reduce equilibrium investments along affected links.

In conclusion, the paper characterizes optimality for road network investments at different spatial scales and economic objectives from a pan-African perspective. It thus provides broad guidance for policymakers interested in improving Africa's road infrastructure. An overarching finding is that trans-African roads are neither MA nor welfare maximizing compared to benefits from improved local connectivity in populated areas or between nearby cities and coastal ports. This is especially true under border frictions, which reduce the value of trans-African links. Yet, these links are essential for trans-African trade under AfCFTA to pick up. A general policy recommendation is thus to prioritize regional connectivity improvements, reduce border frictions, and invest in trans-African roads along specific high-yield links/corridors. The graphical results provided in the paper and appendix should be helpful in prioritizing continental and regional investments, with detailed data and results available upon request. The multi-level approach combining routing engines, PE, and GE analysis can also be adapted to further examine specific regions/investments. Towards this end, a new [Julia library](#) for general equilibrium analysis is utile.

## References

- Abbasi, M., Lebrand, M. S. M., Mongoue, A. B., Pongou, R., & Zhang, F. (2022, March). *Roads, Electricity, and Jobs: Evidence of Infrastructure Complementarity in Sub-Saharan Africa* (Policy Research Working Paper Series No. 9976). The World Bank. Retrieved from <https://ideas.repec.org/p/wbk/wbrwps/9976.html>
- ADB. (2010). *African development report 2010: Ports, logistics and trade in africa*. African Development Bank. Retrieved from <https://www.afdb.org/en/documents/document/african-development-report-2010-27559>
- African Development Bank. (2014). *Study on road infrastructure costs: Analysis of unit costs and cost overruns of road infrastructure projects in africa* (Tech. Rep.). Abidjan, Côte d'Ivoire: African Development Bank Group. Retrieved 2023-05-23, from [https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/Study\\_on\\_Road\\_Infrastructure\\_Costs-\\_Analysis\\_of\\_Unit\\_Costs\\_and\\_Cost\\_Overruns\\_of\\_Road\\_Infrastructure\\_Projects\\_in\\_Africa.pdf](https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/Study_on_Road_Infrastructure_Costs-_Analysis_of_Unit_Costs_and_Cost_Overruns_of_Road_Infrastructure_Projects_in_Africa.pdf)
- Alder, S. (2016). Chinese roads in india: The effect of transport infrastructure on economic development. *SSRN Working Paper*. Retrieved from <https://ssrn.com/abstract=2856050> doi: 10.2139/ssrn.2856050
- Armington, P. S. (1969). A theory of demand for products distinguished by place of production (une théorie de la demande de produits différenciés d'après leur origine)(una teoría de la demanda de productos distinguiéndolos según el lugar de producción). *Staff Papers-International Monetary Fund*, 159–178.
- Asher, S., & Novosad, P. (2020, March). Rural roads and local economic development. *American Economic Review*, 110(3), 797-823. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/aer.20180268> doi: 10.1257/aer.20180268
- Atkin, D., & Donaldson, D. (2015). *Who's getting globalized?: The size and implications of intra-national trade costs* (No. w21439). National Bureau of Economic Research Cambridge, MA.
- Atkin, D., & Donaldson, D. (2022). The role of trade in economic development. In *Handbook of International Economics* (Vol. 5, pp. 1–59). Elsevier.
- Bajzik, J., Havranek, T., Irsova, Z., & Schwarz, J. (2020). Estimating the armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, 127, 103383.
- Barnes, R. (2020). dggridr: Discrete global grids [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=dggridR> (R package version 2.0.4)
- Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115, 103124.
- Bonfatti, R., Gu, Y., & Poelhekk, S. (2019). *Priority roads: The political economy of Africa's interior-to-coast roads* (Discussion Papers No. 2019-04). University of Nottingham, GEP. Retrieved from <https://ideas.repec.org/p/not/notgep/2019-04.html>
- Bosio, E., Arlet, J., Comas, A. A. N., & Leger, N. A. (2018). Road costs knowledge system (rocks)–update. *The World Bank, Washington, DC*.
- Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., & Padró i Miquel, G. (2015). The value of democracy: evidence from road building in kenya. *American Economic Review*, 105(6), 1817–1851.
- Collier, P., Kirchberger, M., & Söderbom, M. (2016). The cost of road infrastructure in low-and middle-income countries. *The World Bank Economic Review*, 30(3), 522–548.
- Couture, V., Duranton, G., & Turner, M. A. (2018, 10). Speed. *The Review of Economics and Statistics*, 100(4), 725-739. Retrieved from [https://doi.org/10.1162/rest\\_a\\_00744](https://doi.org/10.1162/rest_a_00744) doi: 10.1162/rest\_a\_00744

- Djankov, S., Freund, C., & Pham, C. S. (2010). Trading on time. *The Review of Economics and Statistics*, 92(1), 166–173.
- Donaldson, D., & Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2), 799–858.
- Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3), 1046–1070.
- Fajgelbaum, P. D., & Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, 88(4), 1411–1452.
- Fiorini, M., Sanfilippo, M., & Sundaram, A. (2021). Trade liberalization, roads and firm productivity. *Journal of Development Economics*, 153, 102712.
- Foster, V., & Briceño-Garmendia, C. (2010). *Africa's infrastructure: a time for transformation*. World Bank.
- Gennaioli, N., La Porta, R., Lopez-de Silanes, F., & Shleifer, A. (2013). Human capital and regional development. *The Quarterly Journal of Economics*, 128(1), 105–164.
- Ghodsi, M., Grüberl, J., Reiter, O., & Stehrer, R. (2017). *The evolution of non-tariff measures and their diverse effects on trade* (wiw Research Report No. 419). Vienna. Retrieved from <https://hdl.handle.net/10419/204191>
- Gorton, N. E., & Ianovichina, E. (2021, November). *Trade Networks in Latin America : Spatial Inefficiencies and Optimal Expansions* (Policy Research Working Paper Series No. 9843). The World Bank. Retrieved from <https://ideas.repec.org/p/wbk/wbrwps/9843.html>
- Graff, T. (2024). Spatial inefficiencies in africa's trade network. *Journal of Development Economics*, 103319.
- Jedwab, R., Kerby, E., & Moradi, A. (2017). History, path dependence and development: Evidence from colonial railways, settlers and cities in kenya. *The Economic Journal*, 127(603), 1467–1494.
- Jedwab, R., & Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: evidence from africa. *Review of Economics and Statistics*, 98(2), 268–284.
- Jedwab, R., & Storeygard, A. (2022). The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960–2010. *Journal of the European Economic Association*, 20(1), 1–38.
- Krantz, S. (2023). Mapping africa's infrastructure potential with geospatial big data and causal ml. *SSRN Working Paper*. Retrieved from <https://ssrn.com/abstract=4537867> doi: 10.2139/ssrn.4537867
- Kummu, M., Taka, M., & Guillaume, J. H. (2018). Gridded global datasets for gross domestic product and human development index over 1990–2015. *Scientific Data*, 5(1), 1–15.
- Lamarque, H., & Nugent, P. (2022). *Transport corridors in africa*. Boydell & Brewer.
- Lee, K., & Braithwaite, J. (2022). High-resolution poverty maps in sub-saharan africa. *World Development*, 159, 106028.
- Mayer, M. (2023). missranger: Fast imputation of missing values [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=missRanger> (R package version 2.2.1)
- Moneke, N. (2020). Can big push infrastructure unlock development? evidence from ethiopia. *STEG Theme*, 3, 14–15.
- Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1), 20–36.
- Peng, C., & Chen, W. (2021). *Roads to development? Examining the Zambian context using AI-Sat* [Working Paper]. Retrieved from <https://www.congpeng.org/>

- Poot, J., Alimi, O., Cameron, M. P., & Maré, D. C. (2016). The gravity model of migration: the successful comeback of an ageing superstar in regional science. *Investigaciones Regionales-Journal of Regional Research*(36), 63–86.
- 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.
- Porteous, O. (2022). Reverse dutch disease with trade costs: Prospects for agriculture in africa's oil-rich economies. *Journal of International Economics*, 138, 103651.
- Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics*, 101, 148–167.
- Sahr, K. (2022). User documentation for discrete global grid generation software. *Southern Oregon Univ., Ashland, OR, USA, Tech. Rep. Dggrid version, 7.5*.
- Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, 30(2), 121–134.
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest: non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.
- Teravaninthorn, S., & Raballand, G. (2009). *Transport prices and costs in africa: A review of the international corridors* (No. 6610). The World Bank Group. Retrieved from <https://ideas.repec.org/b/wbk/wbpubs/6610.html>
- Wang, H., Li, J., Chen, Q.-Y., & Ni, D. (2011). Logistic modeling of the equilibrium speed–density relationship. *Transportation Research Part A: Policy and Practice*, 45(6), 554–566.
- Wolfram, C. (2021, April). *Efficiency of road networks*. Blog post. Retrieved from <https://christopherwolfram.com/projects/efficiency-of-road-networks/> (Accessed on 27th March 2024)
- Zipf, G. K. (1946). The  $p_1^*p_2/d$  hypothesis: on the intercity movement of persons. *American Sociological Review*, 11(6), 677–686.

## Appendix

Figure A1: Average Network Speed and WorldPop Population Estimate

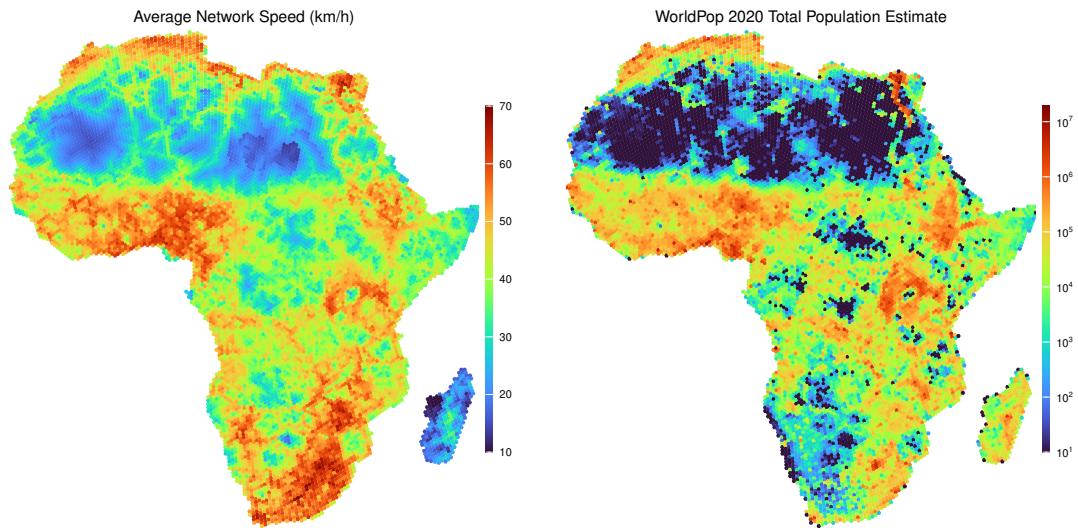


Figure A2: Gridded GDP and International Wealth Index

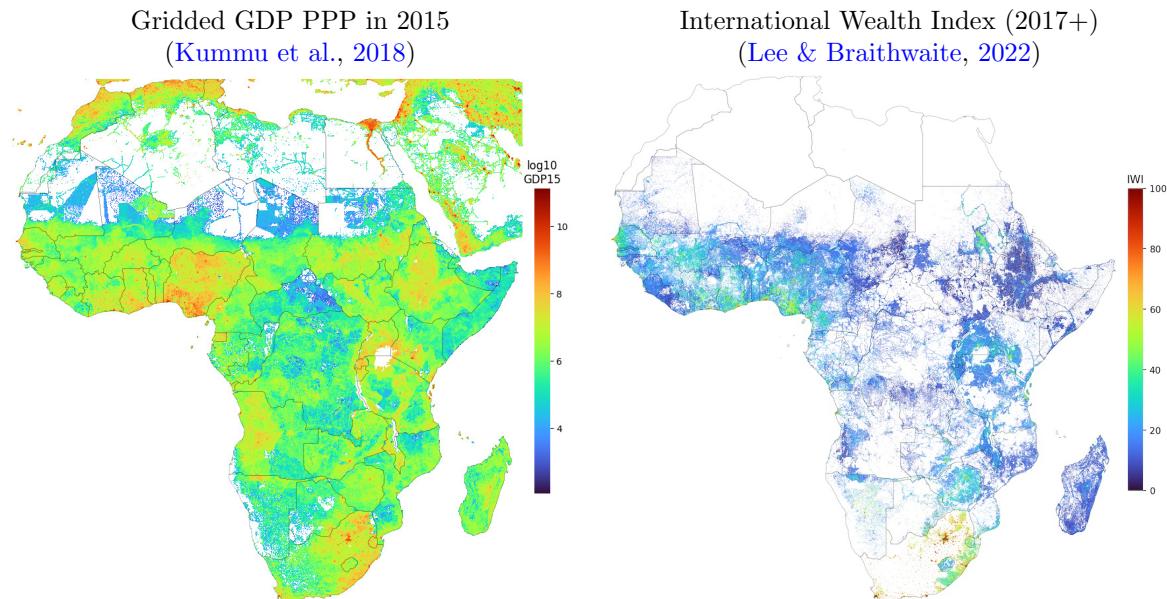
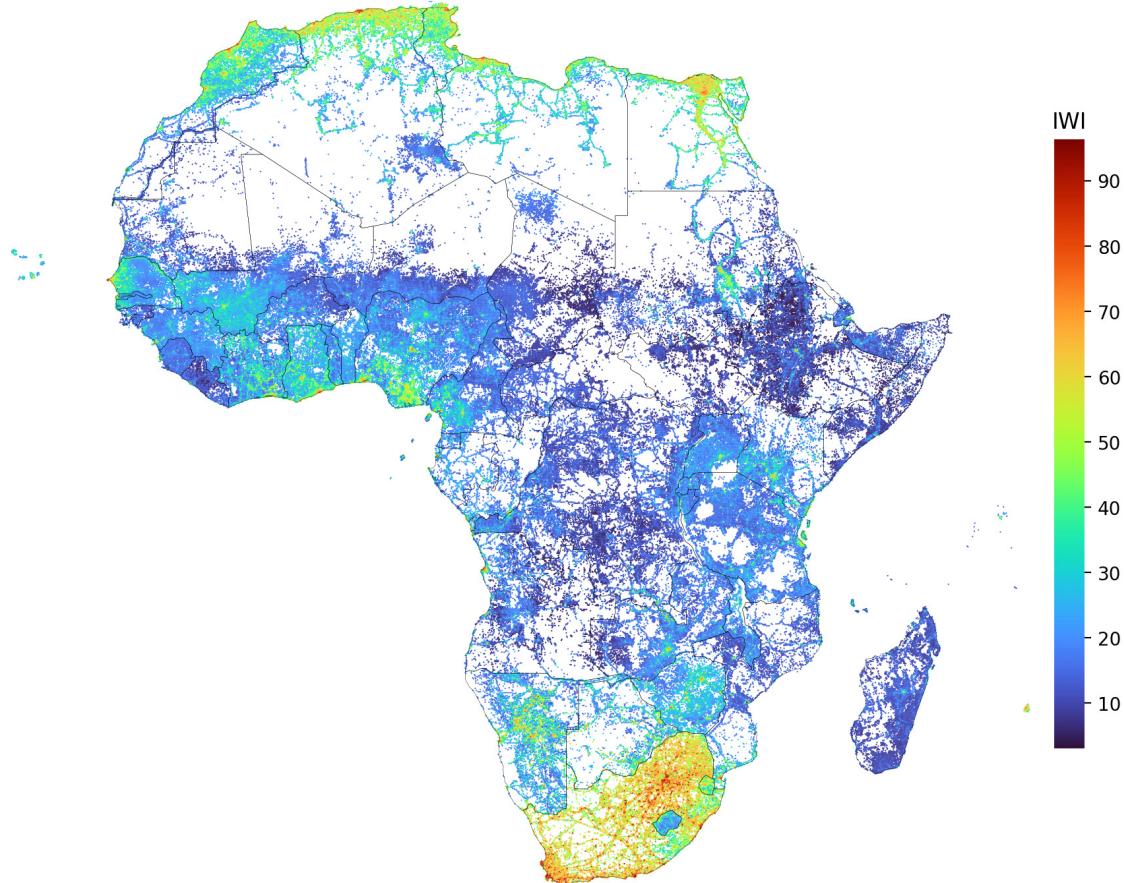


Figure A3: Imputed International Wealth Index using Database of [Krantz \(2023\)](#)

Predicted International Wealth Index: Imputed With MissForest, R-Squared = 0.970



**Notes:** The high  $R^2$  is not surprising as the estimates of [Lee & Braithwaite \(2022\)](#) also utilize many features from OSM in their methodology and Random Forests is very flexible. A concern with the imputation is that the relationship between infrastructure and wealth may be different south of the Sahara. [Lee & Braithwaite \(2022\)](#) mention the absence of (recent) DHS surveys in North Africa as an obstacle to extending their methodology to them. Their success in estimating cross-country models for all of SSA, including South Africa, however suggests that these models - approximated by *MissForest* - should also provide acceptable predictions in North Africa.

Figure A4: Market Access Maps using Total IWI-Based Wealth

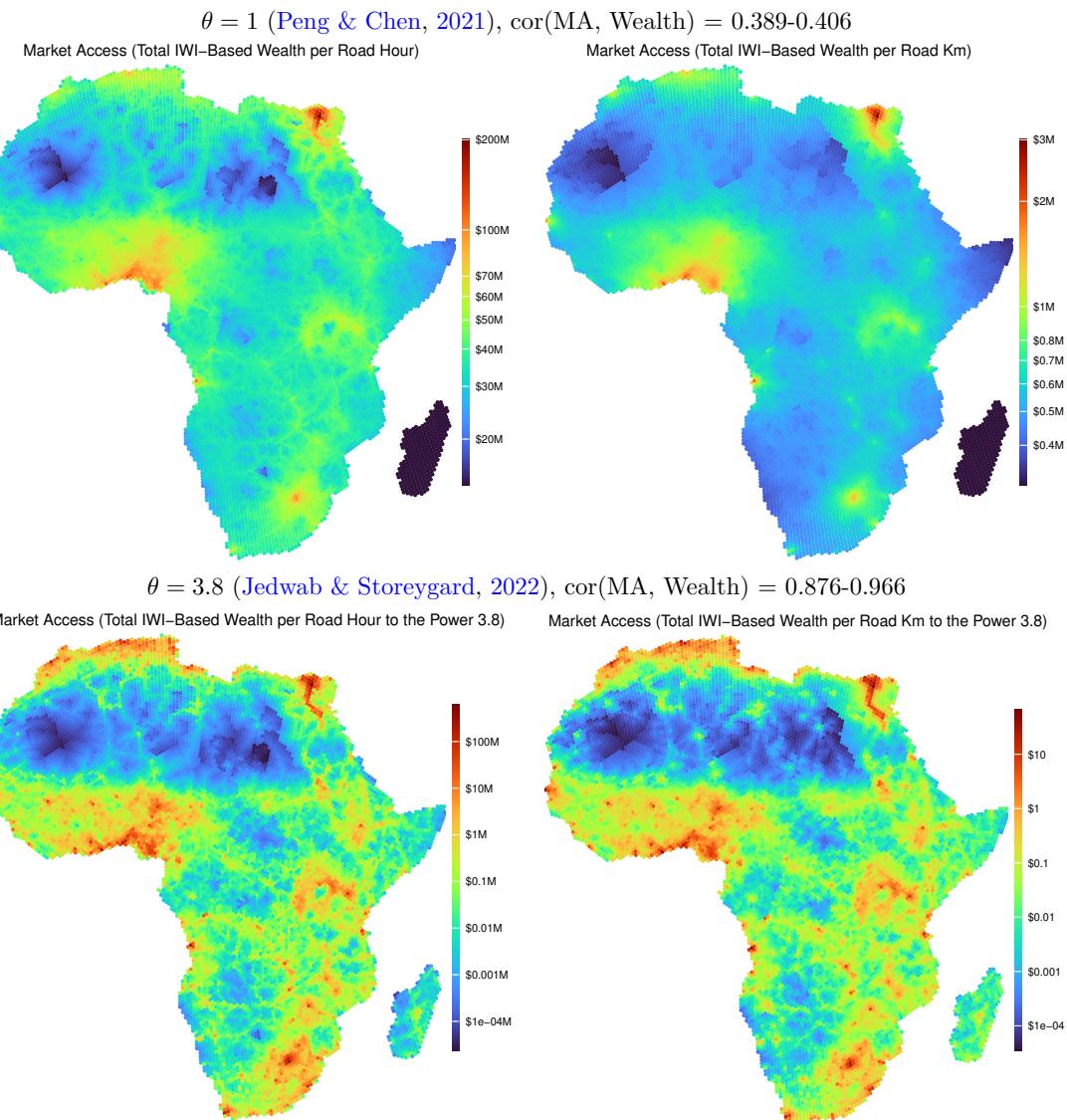


Figure A5: Market Access Maps using GDP in 2015 USD PPP: With Frictions

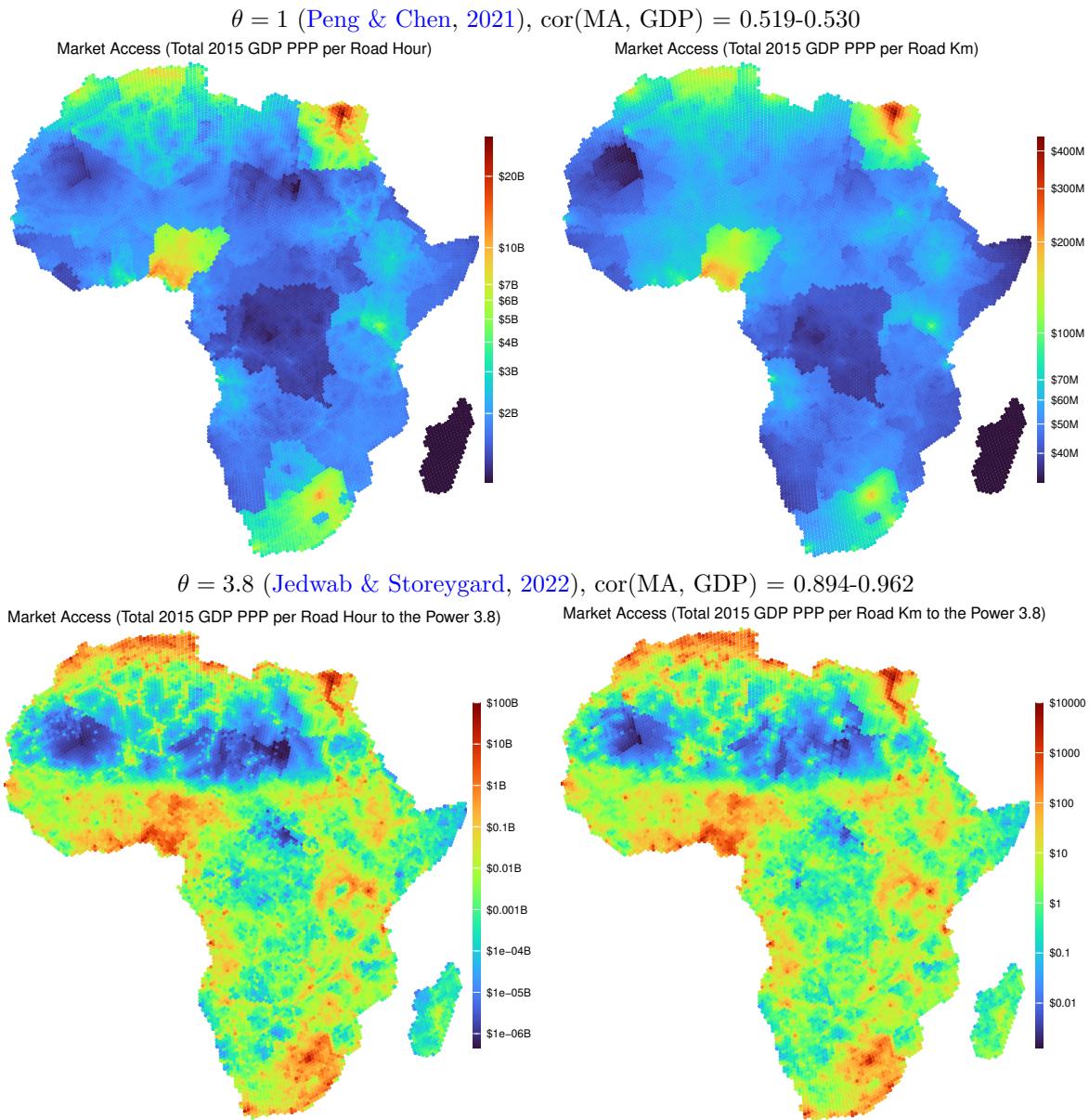


Figure A6: Market Access Maps using Total IWI-Based Wealth: With Frictions

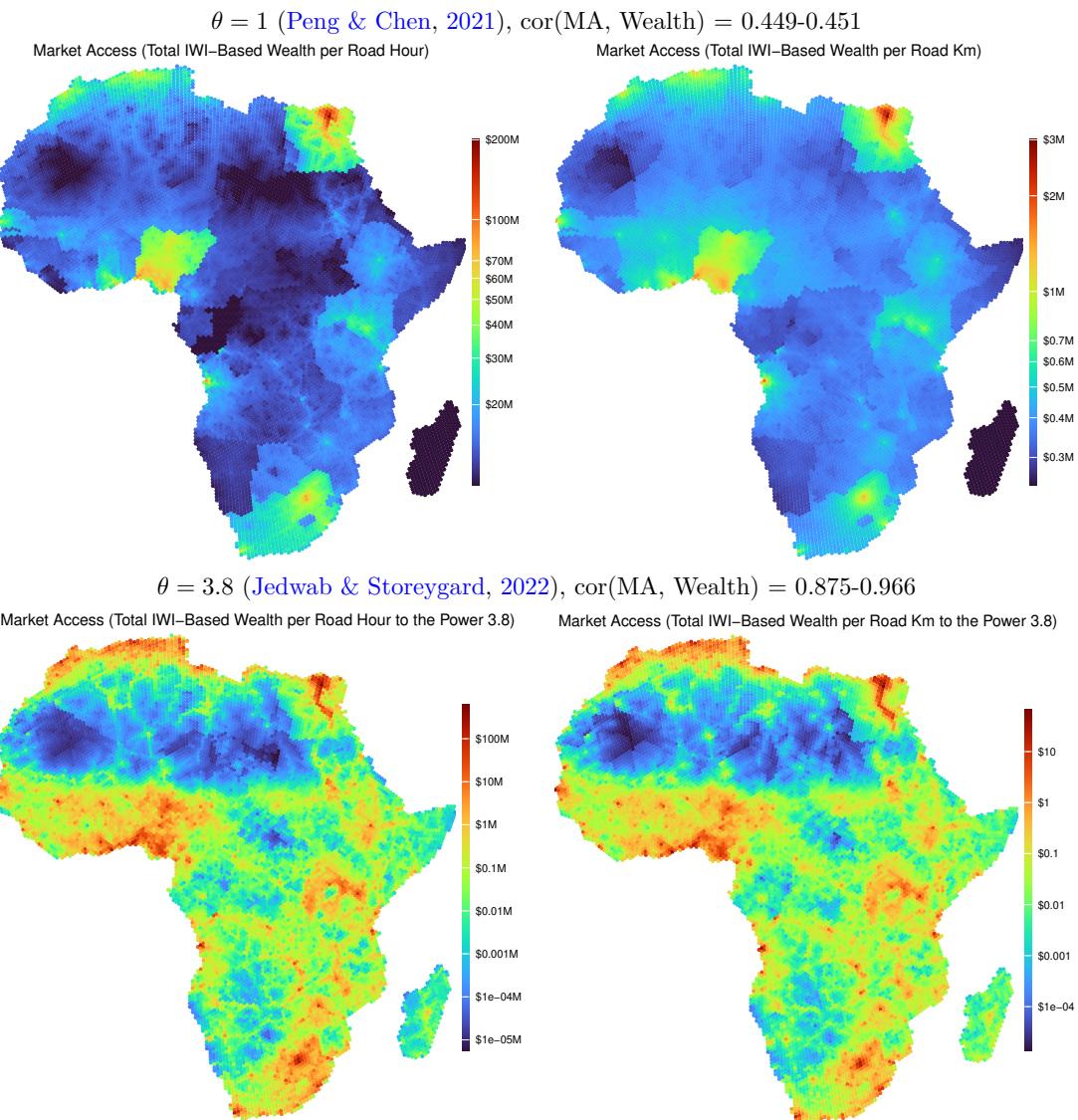


Figure A7: Market Access Loss (%) from Border Frictions using Total IWI-Based Wealth

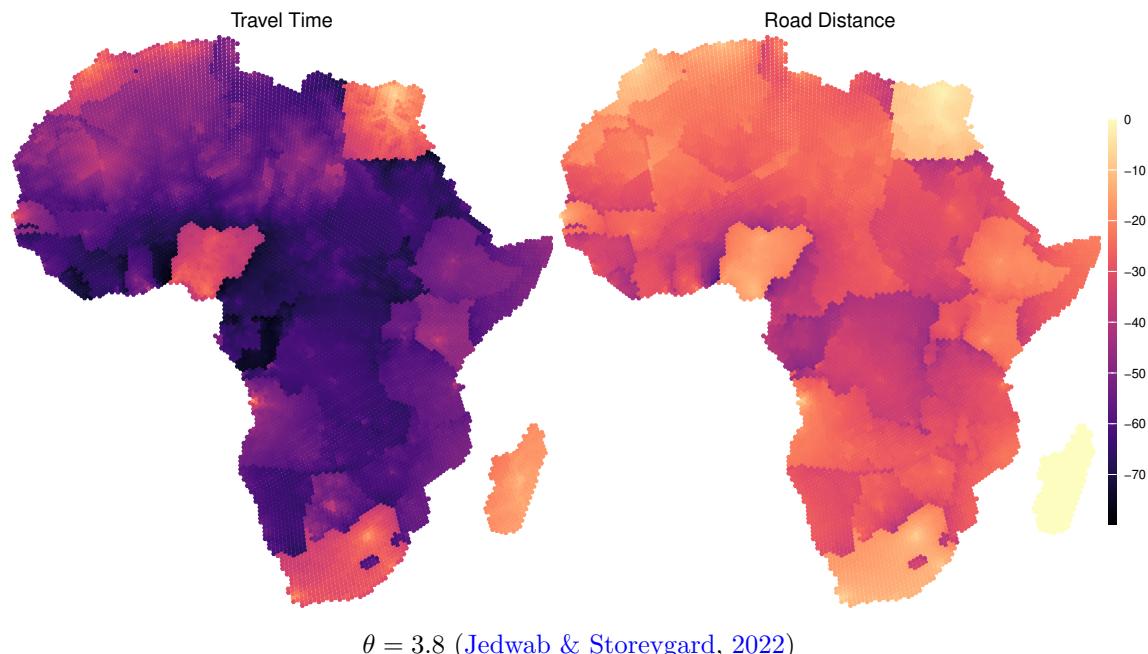
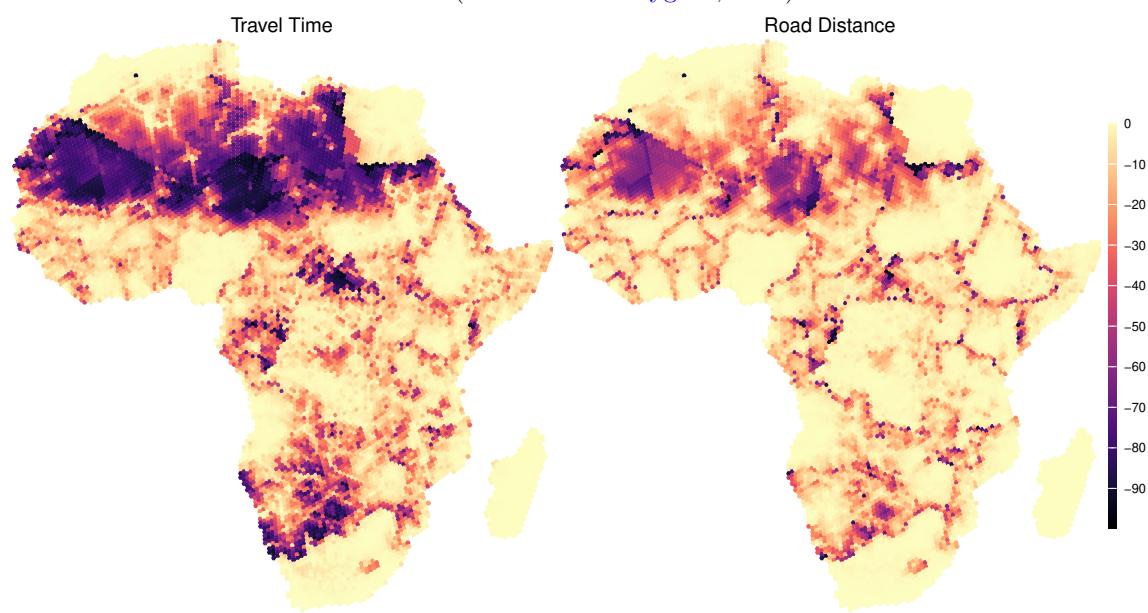
 $\theta = 1$  (Peng & Chen, 2021) $\theta = 3.8$  (Jedwab & Storeygard, 2022)

Figure A8: Optimized Full Network Graph: Duration Weighted Edges

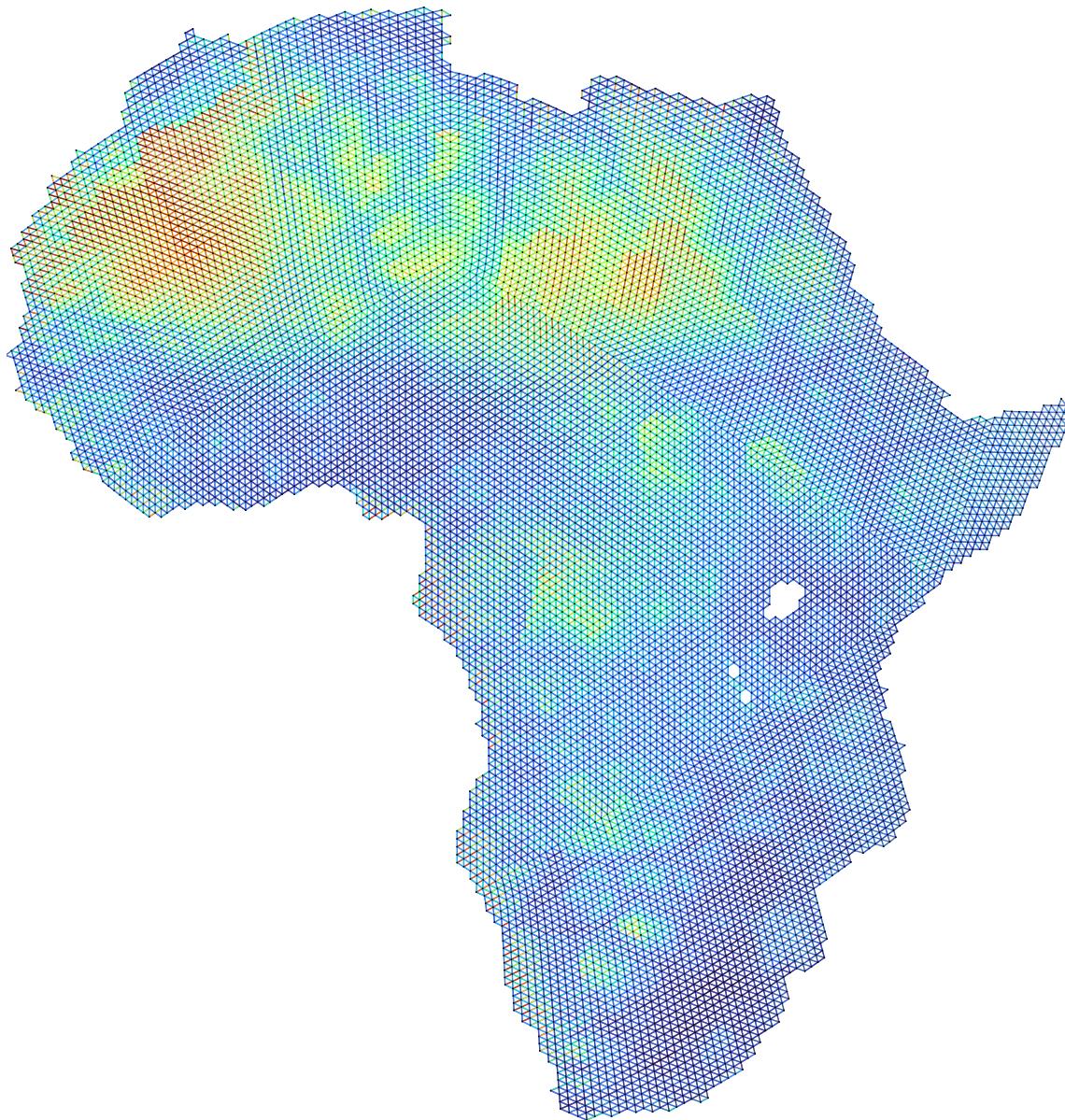


Figure A9: Discretized Trans-African Network Plus Original Routes

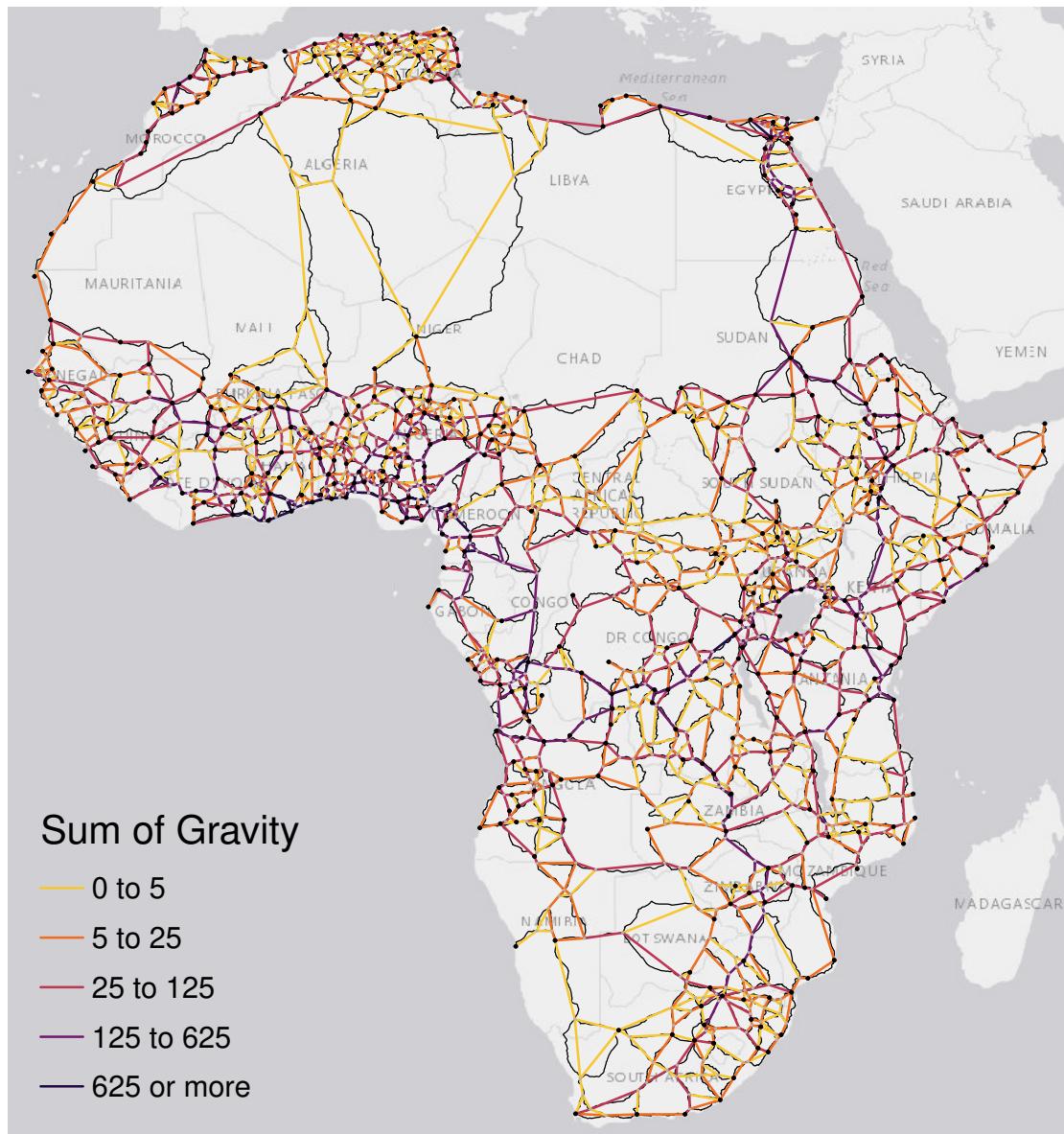


Figure A10: Discretized Trans-African Network and New Links for US Route Efficiency

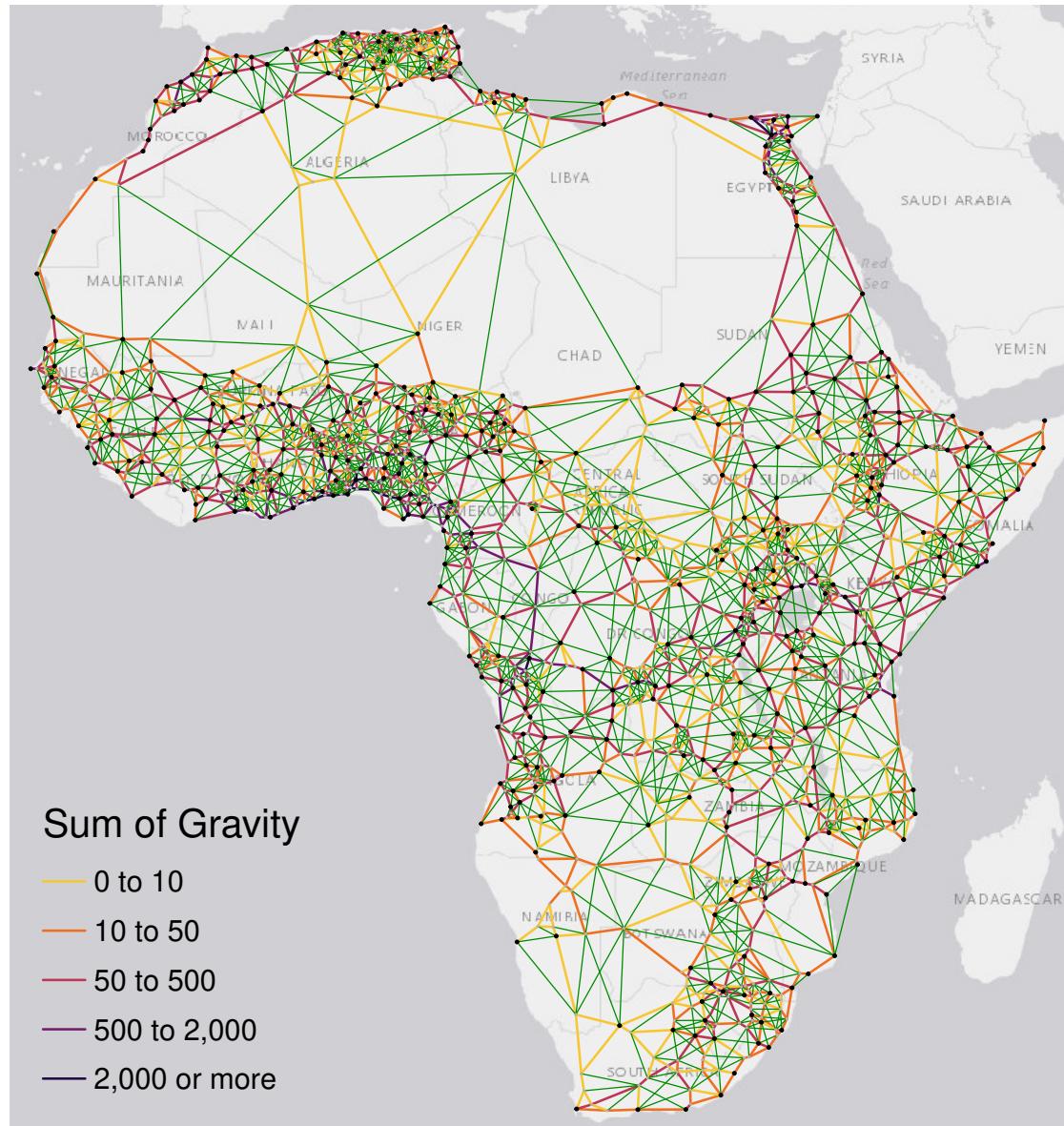


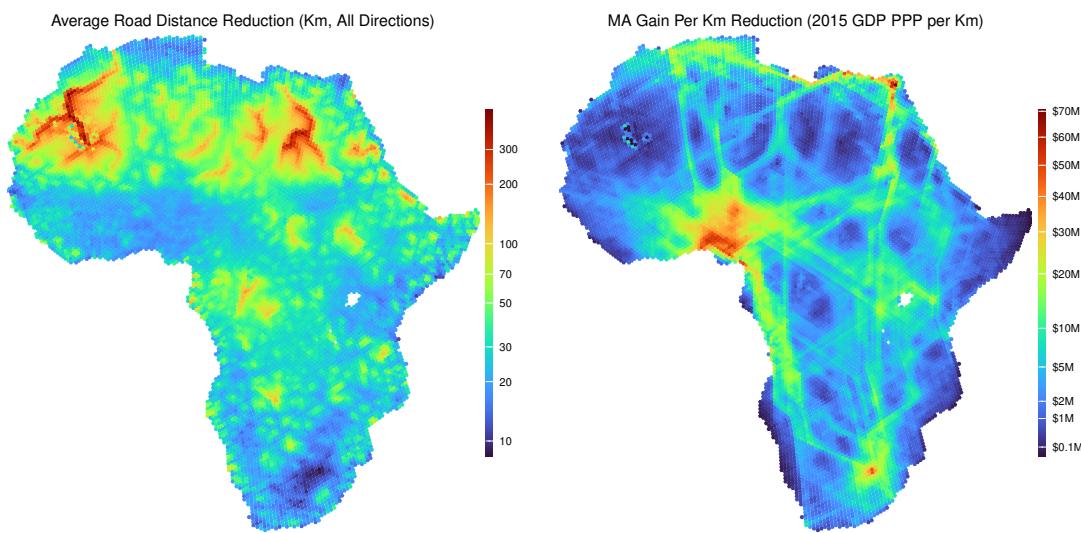
Figure A11: Local Road Distance Reduction:  $\min(\kappa_{ik}^r) = 0.85$ 

Figure A12: Estimated Network Building Cost per Kilometer

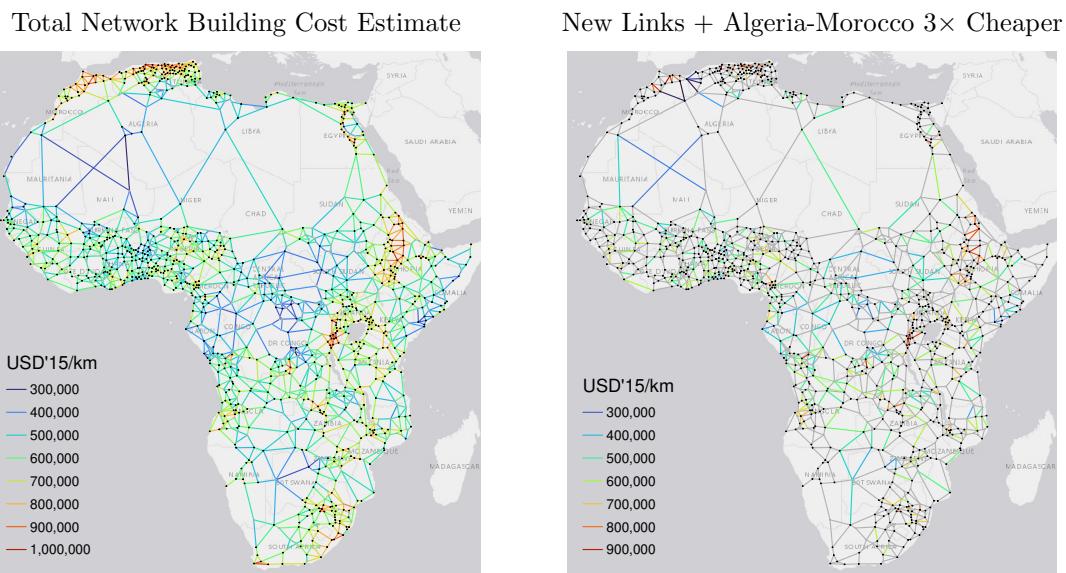


Figure A13: Nightlights vs. IWI to Measure City Productivity

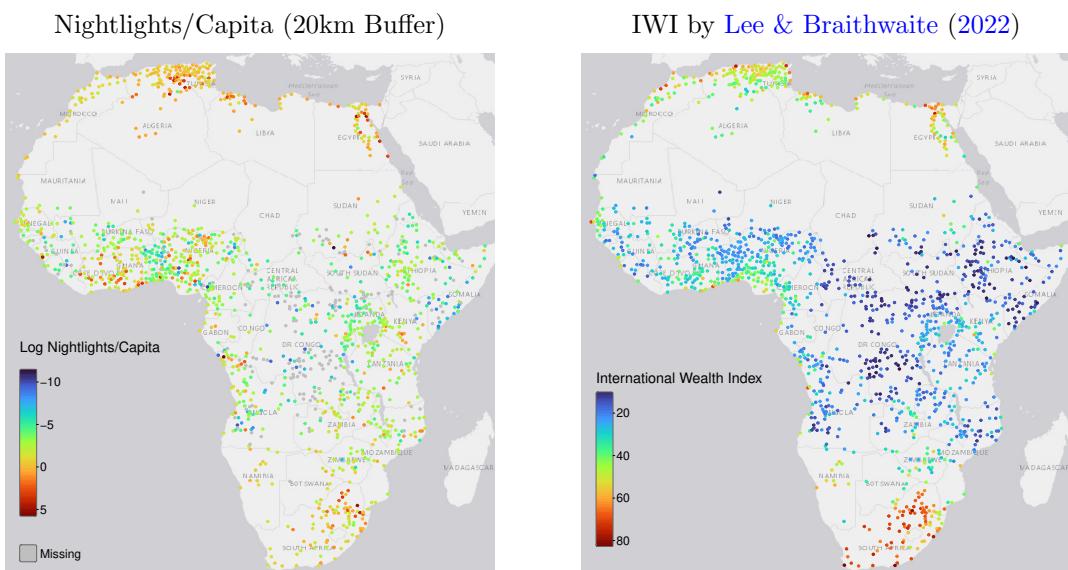


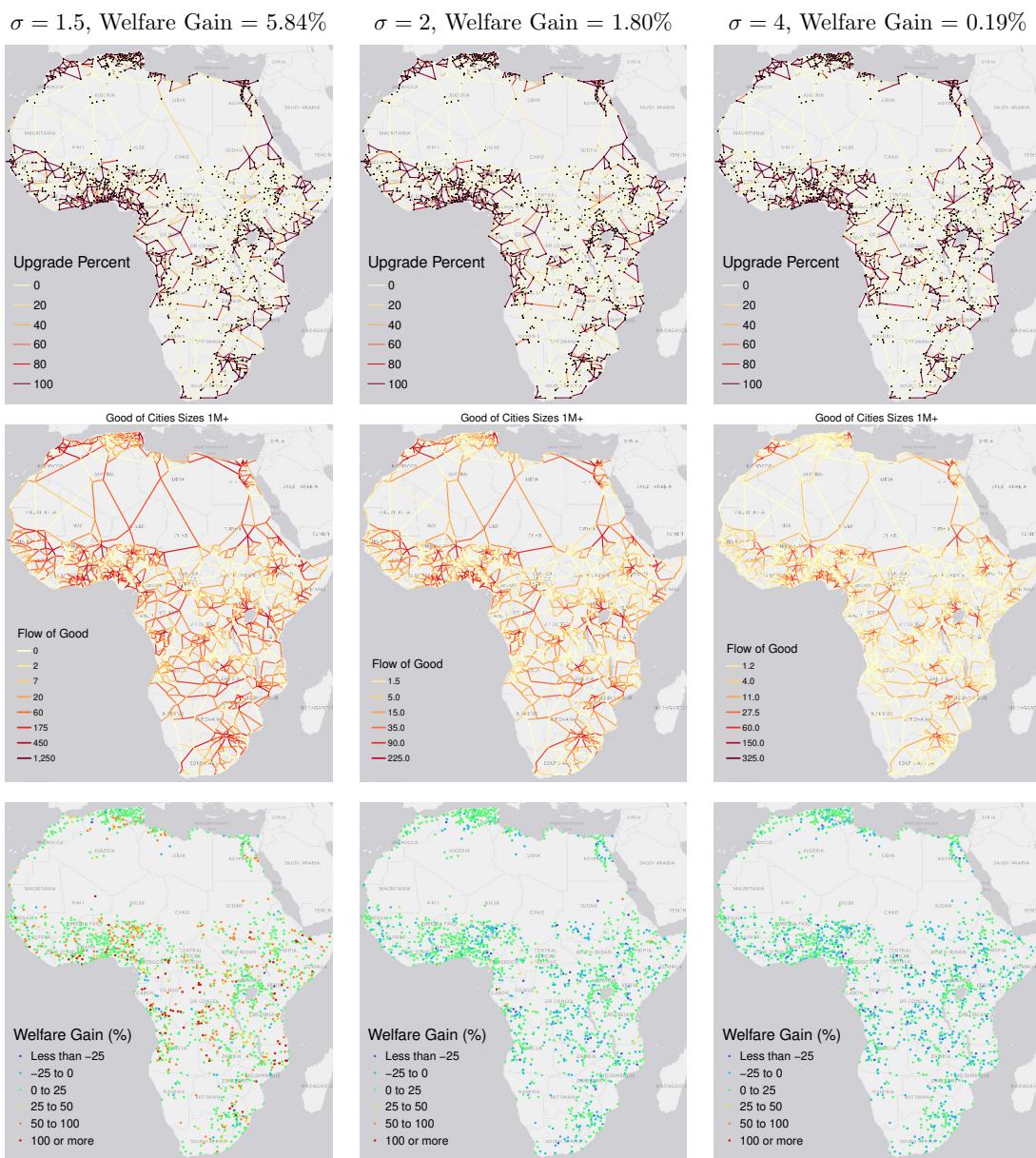
Figure A14: Optimal \$50B Investments, Trade, and Welfare by  $\sigma$ 

Figure A15: Optimal Infrastructure Allocation Without Imported Goods

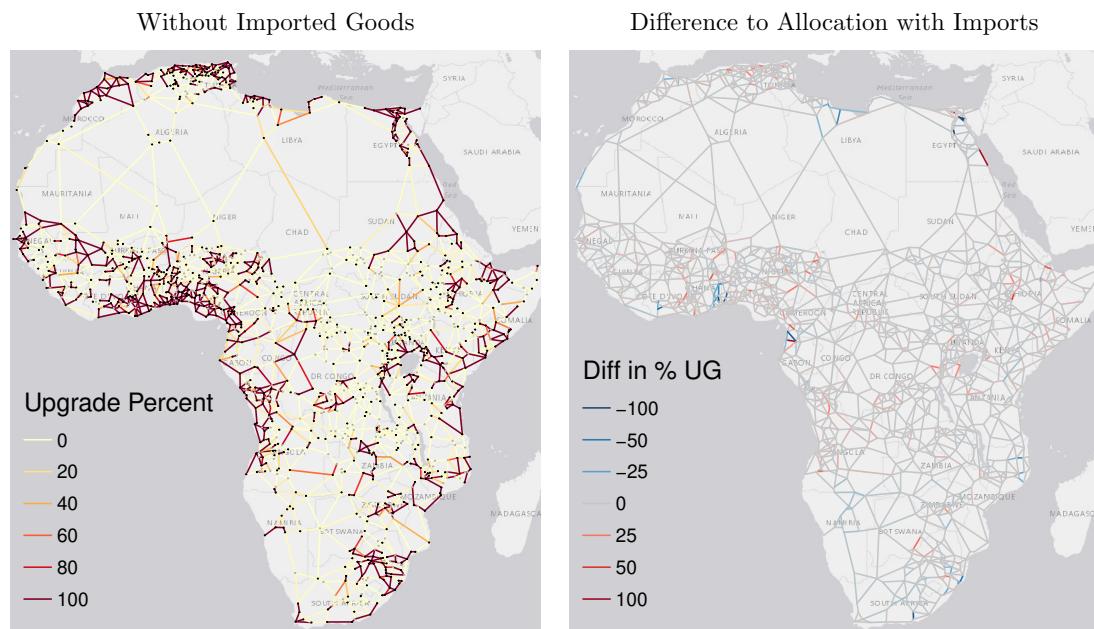


Figure A16: Ratio of Goods Flows under Border Frictions to Frictionless Flows

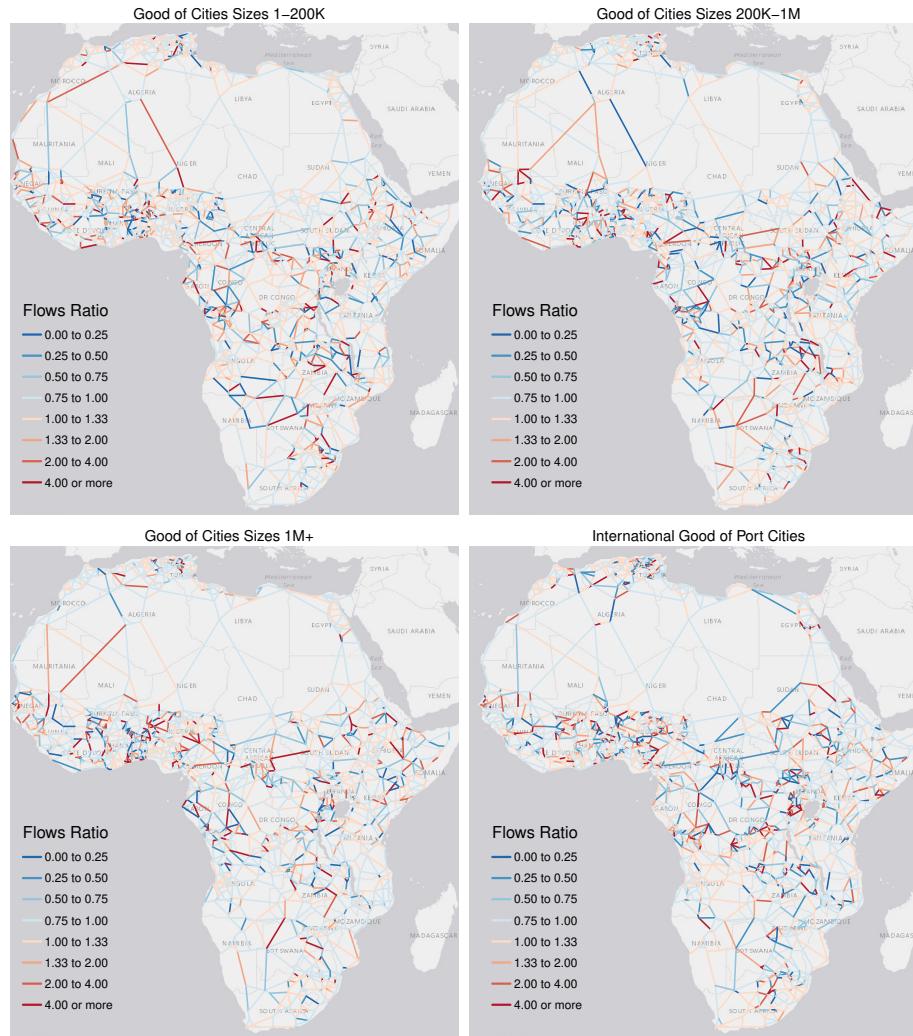


Figure A17: Flow of Goods and Local Consumption under IRS

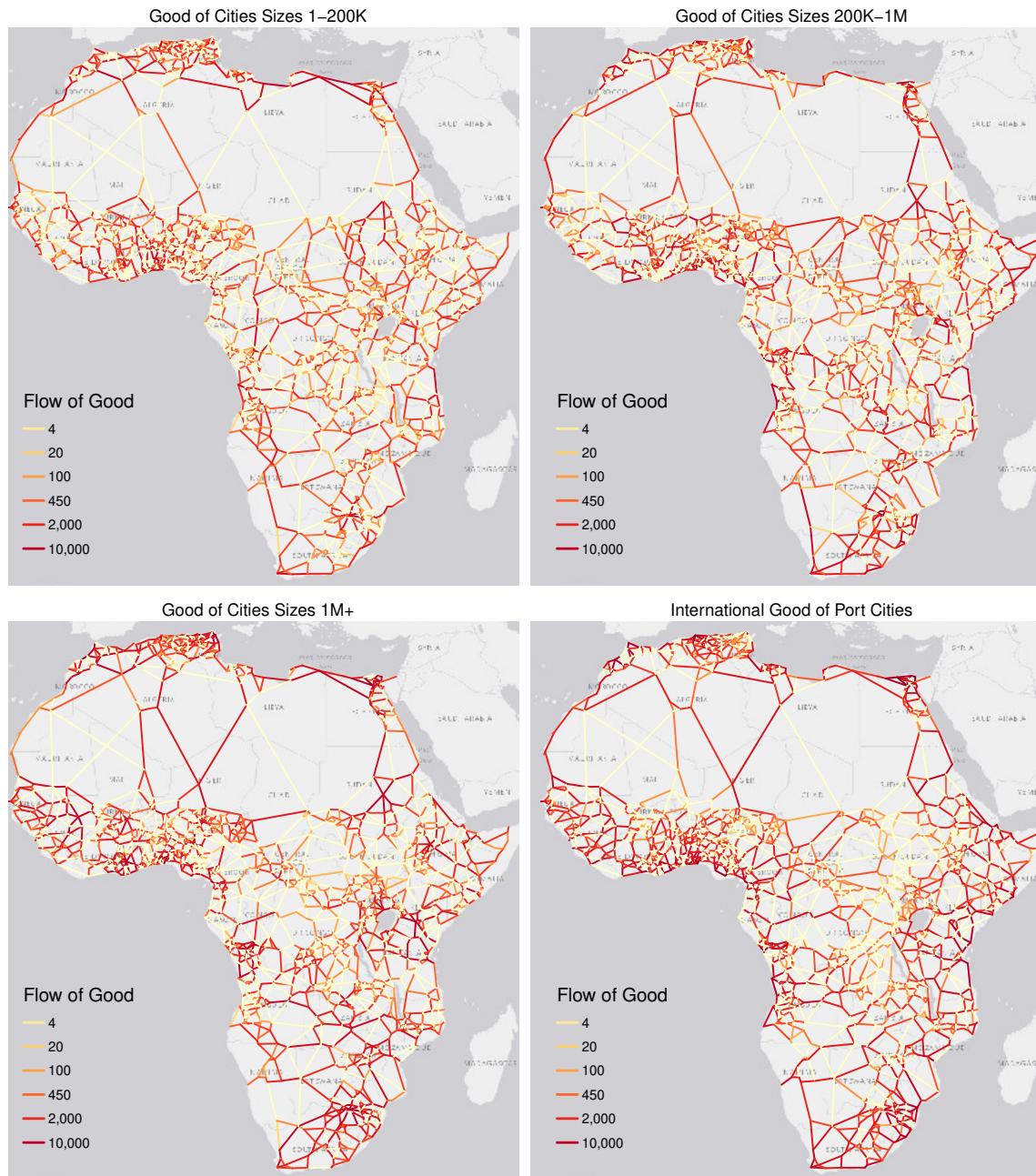


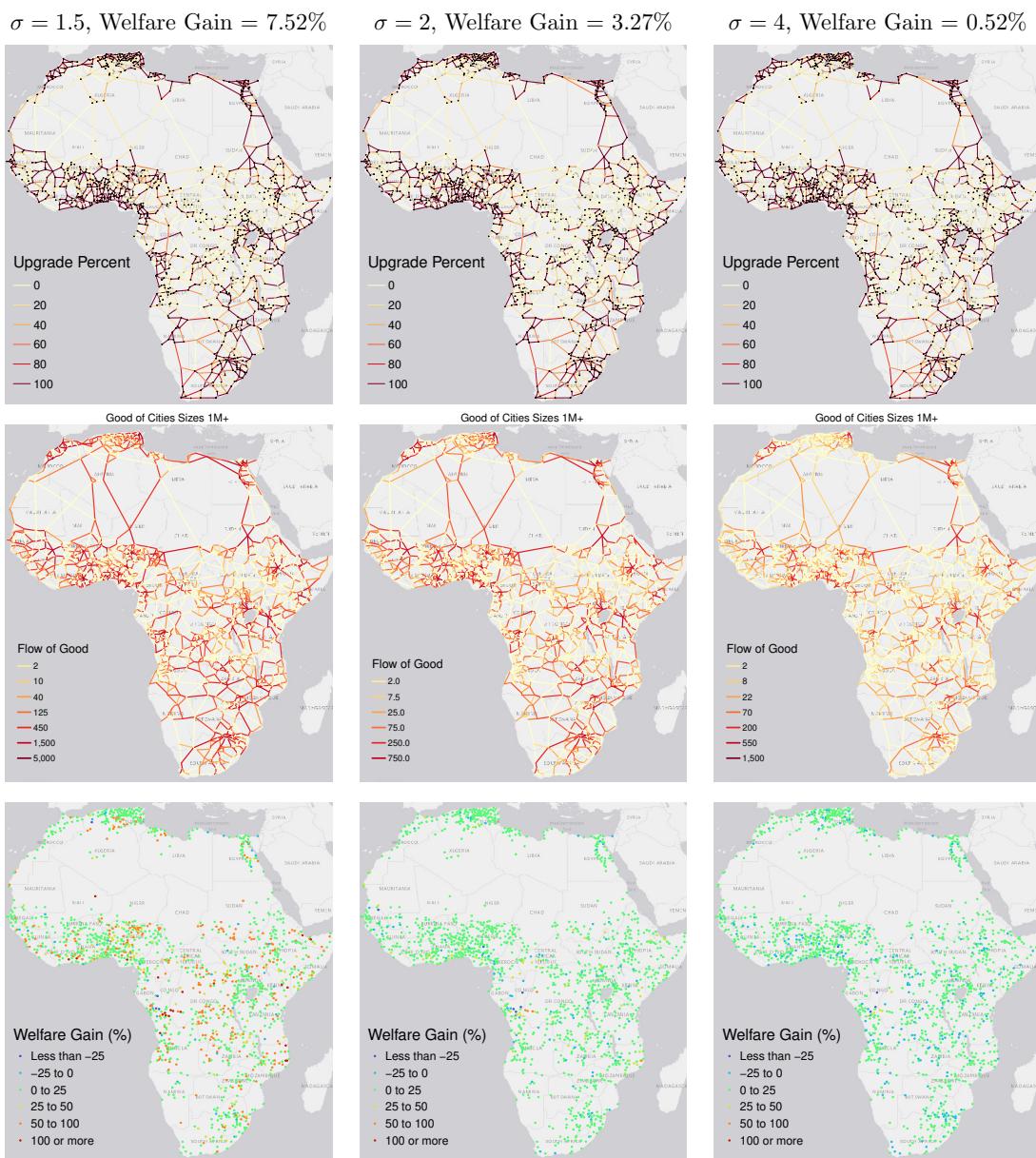
Figure A18: Optimal \$50B Investments, Trade, and Welfare under IRS by  $\sigma$ 

Figure A19: Optimal Infrastructure Allocation Without Imported Goods: IRS Case

Without Imported Goods

Difference to Allocation with Imports

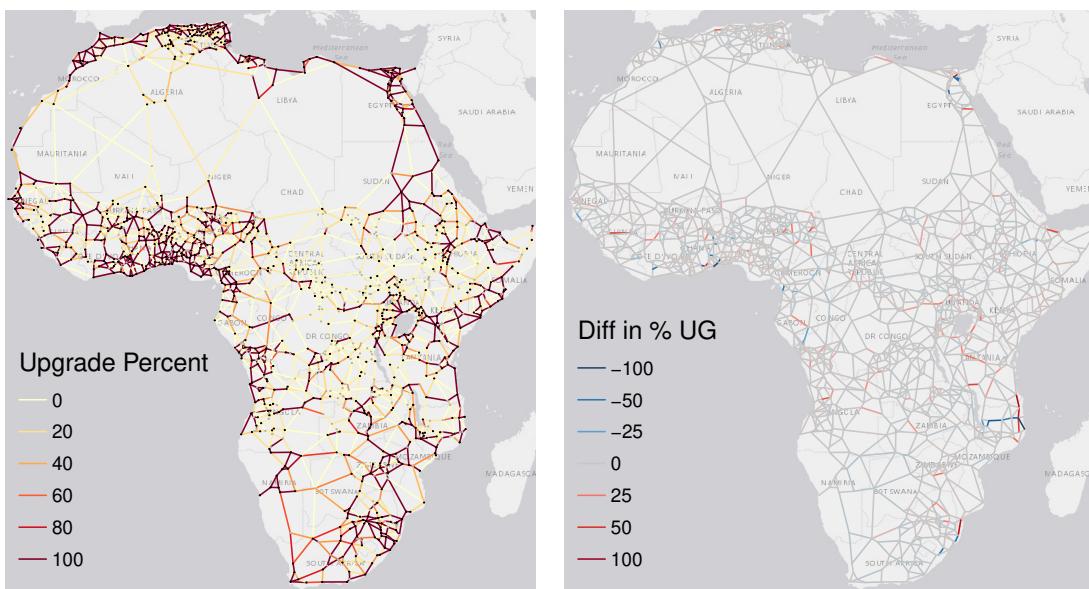


Figure A20: Ratio of Goods Flows under Border Frictions to Frictionless Flows: IRS Case

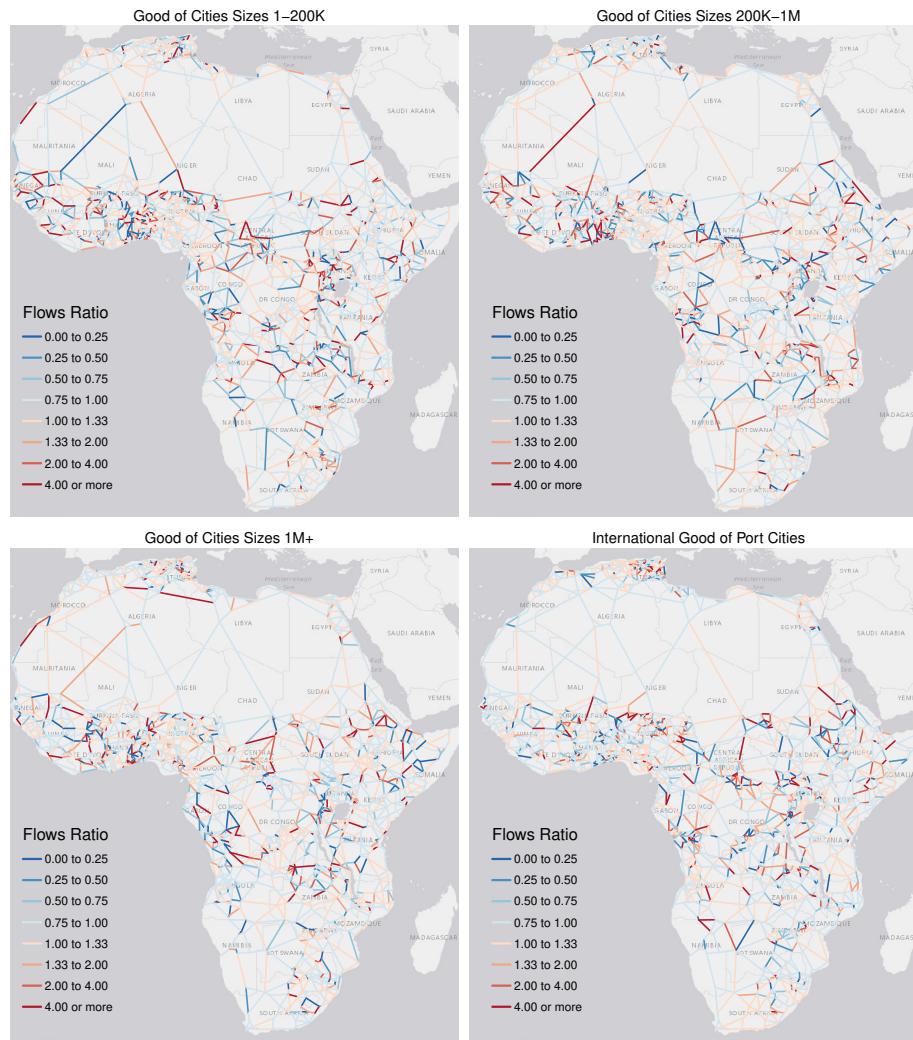


Figure A21: Trans-African Network Connecting Large Cities: Parameterization with Real Roads

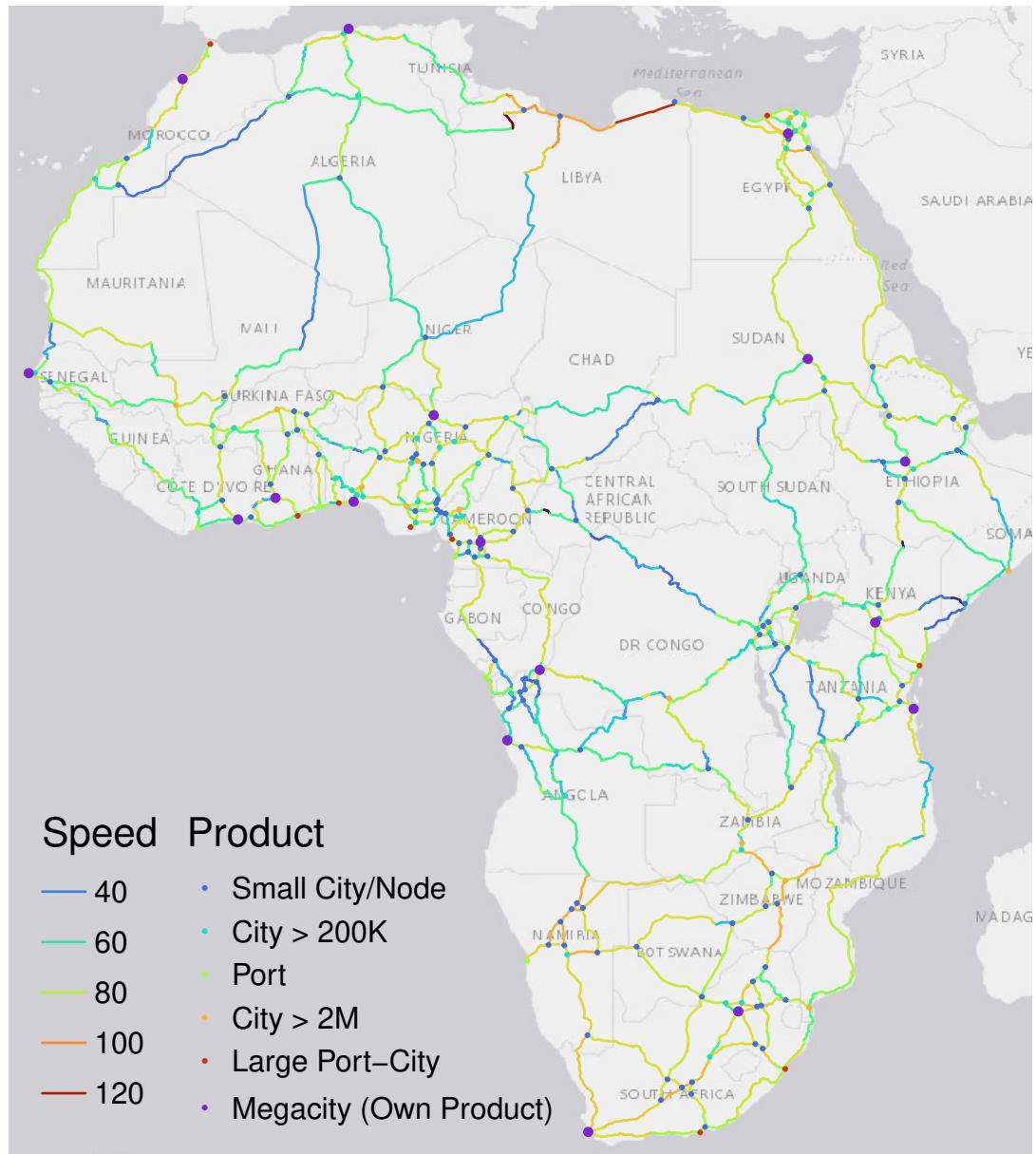


Figure A22: Flow of Goods Through Trans-African Network Connecting Large Cities

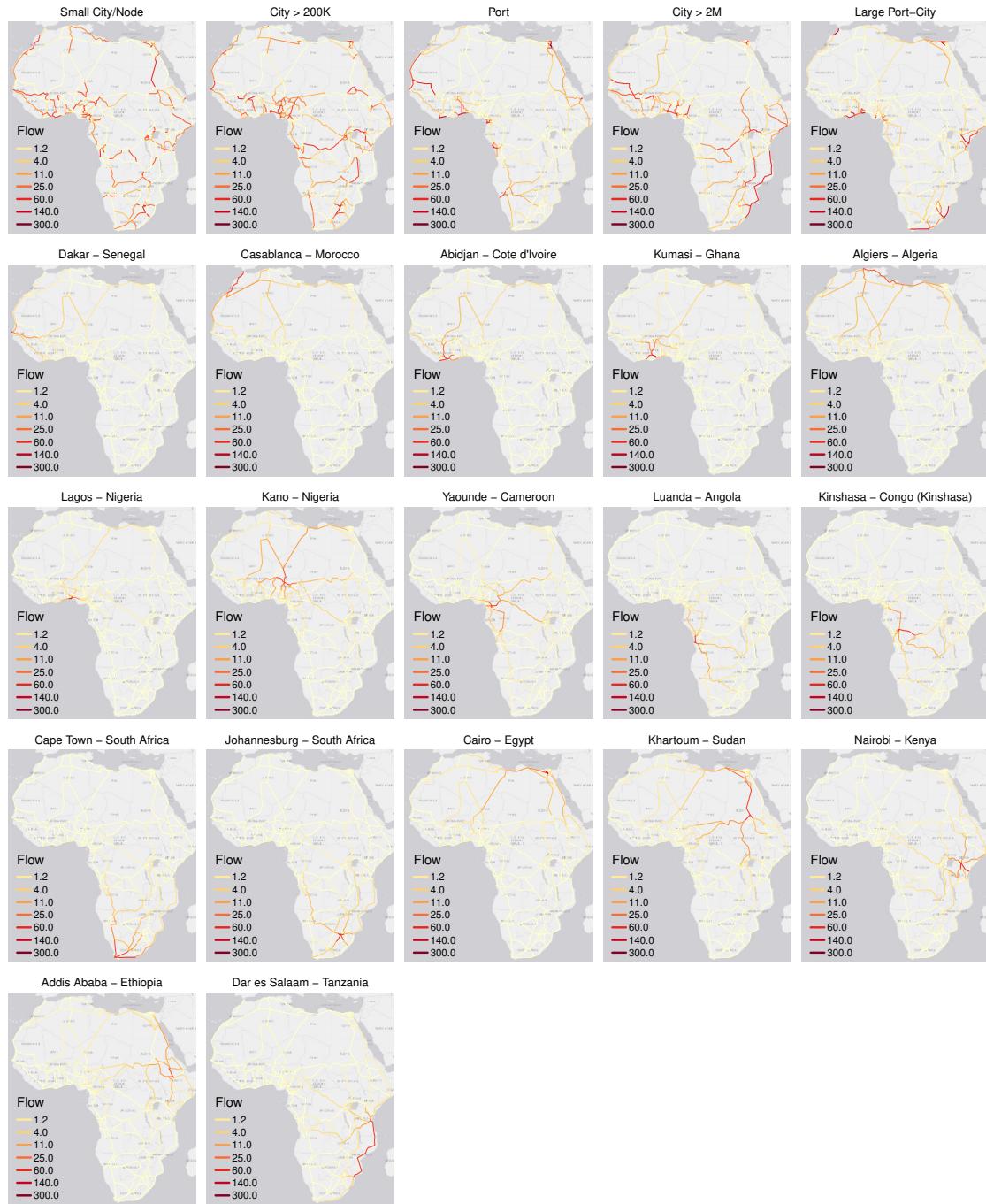


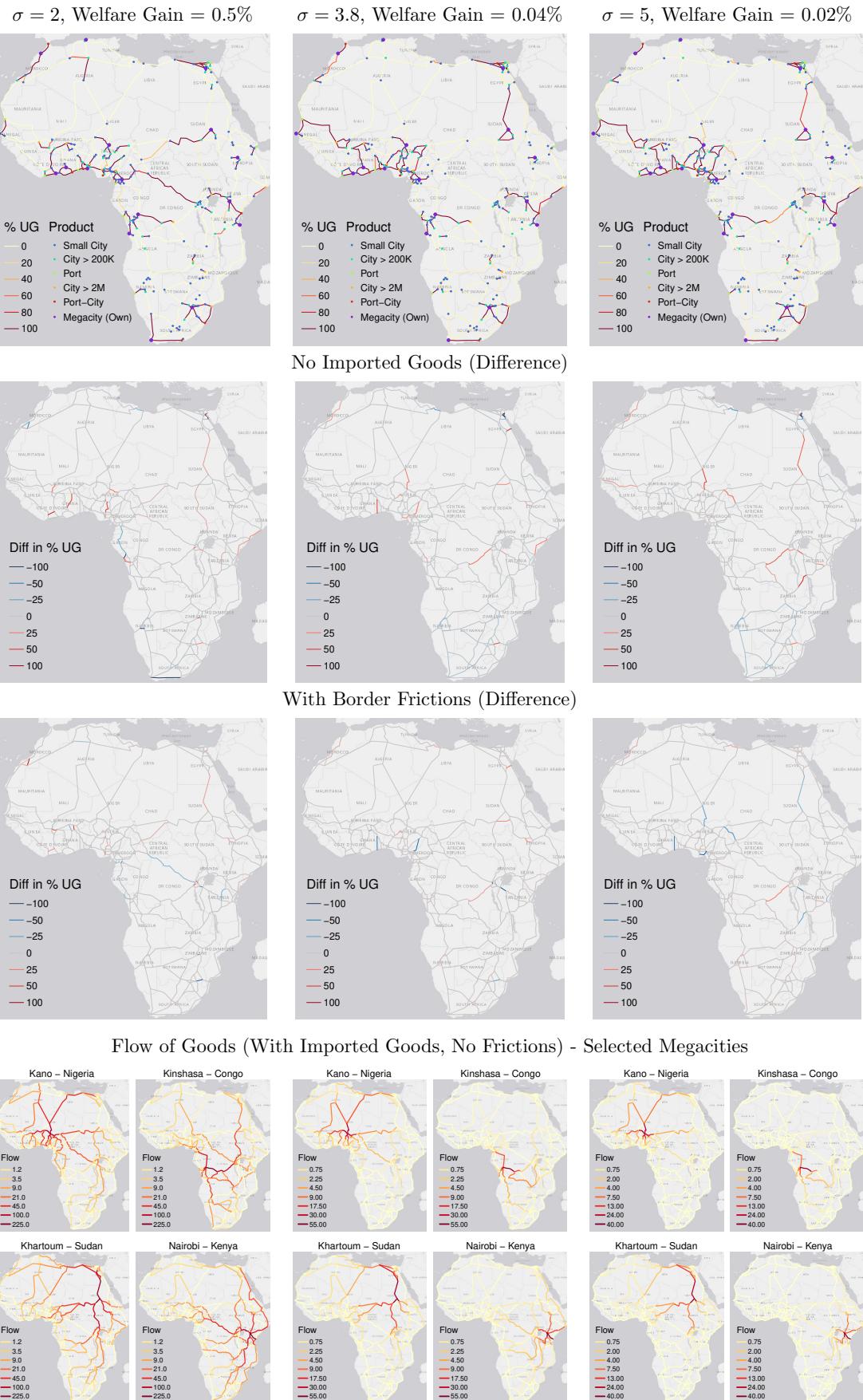
Figure A23: Optimal \$10B Trans-African Investments and Trade by  $\sigma$ 

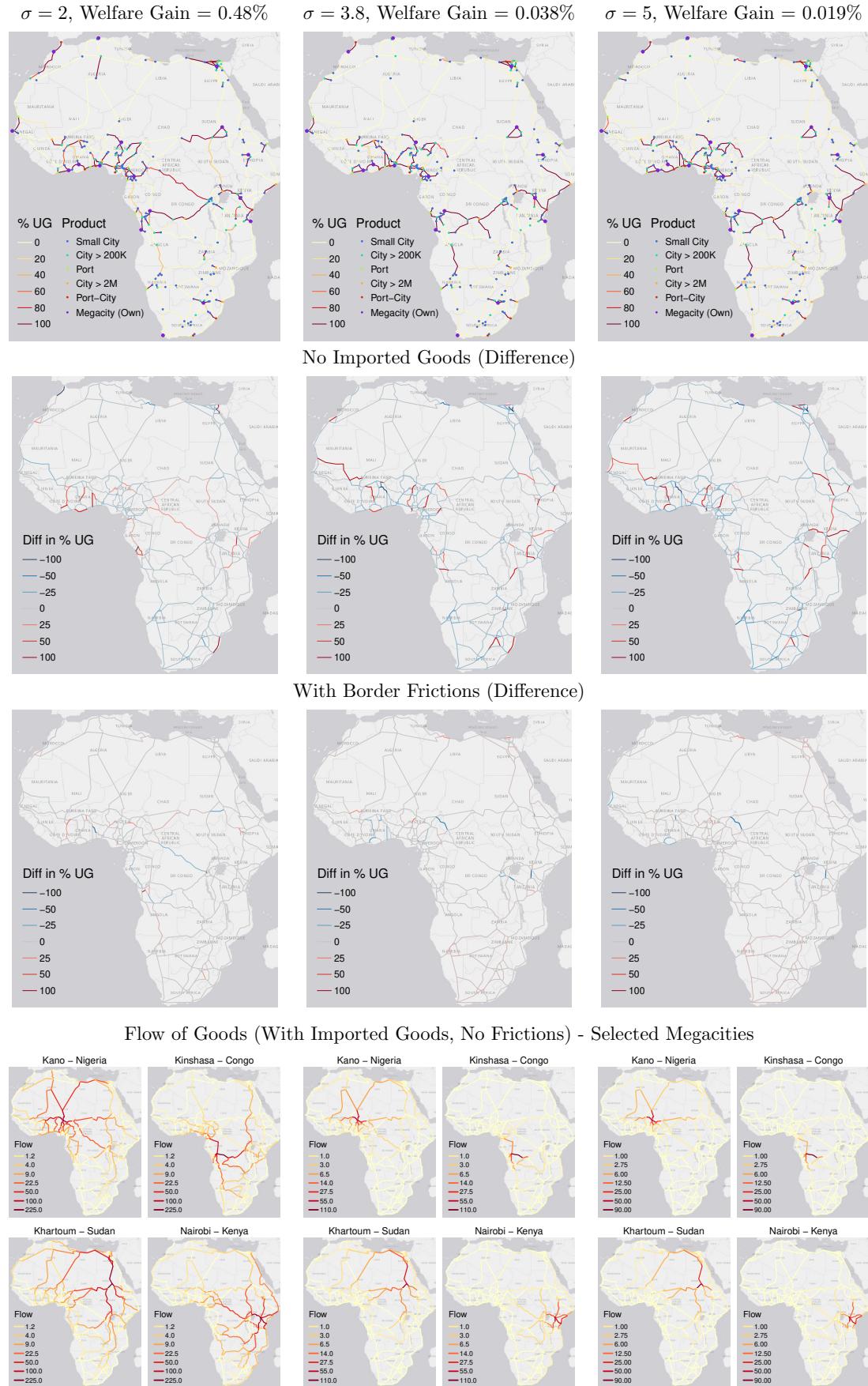
Figure A24: Optimal \$10B Trans-African Investments by  $\sigma$  with Inequality Aversion ( $\rho = 2$ )

Figure A25: Flow of Goods Through Trans-African Network Connecting Large Cities: IRS Case

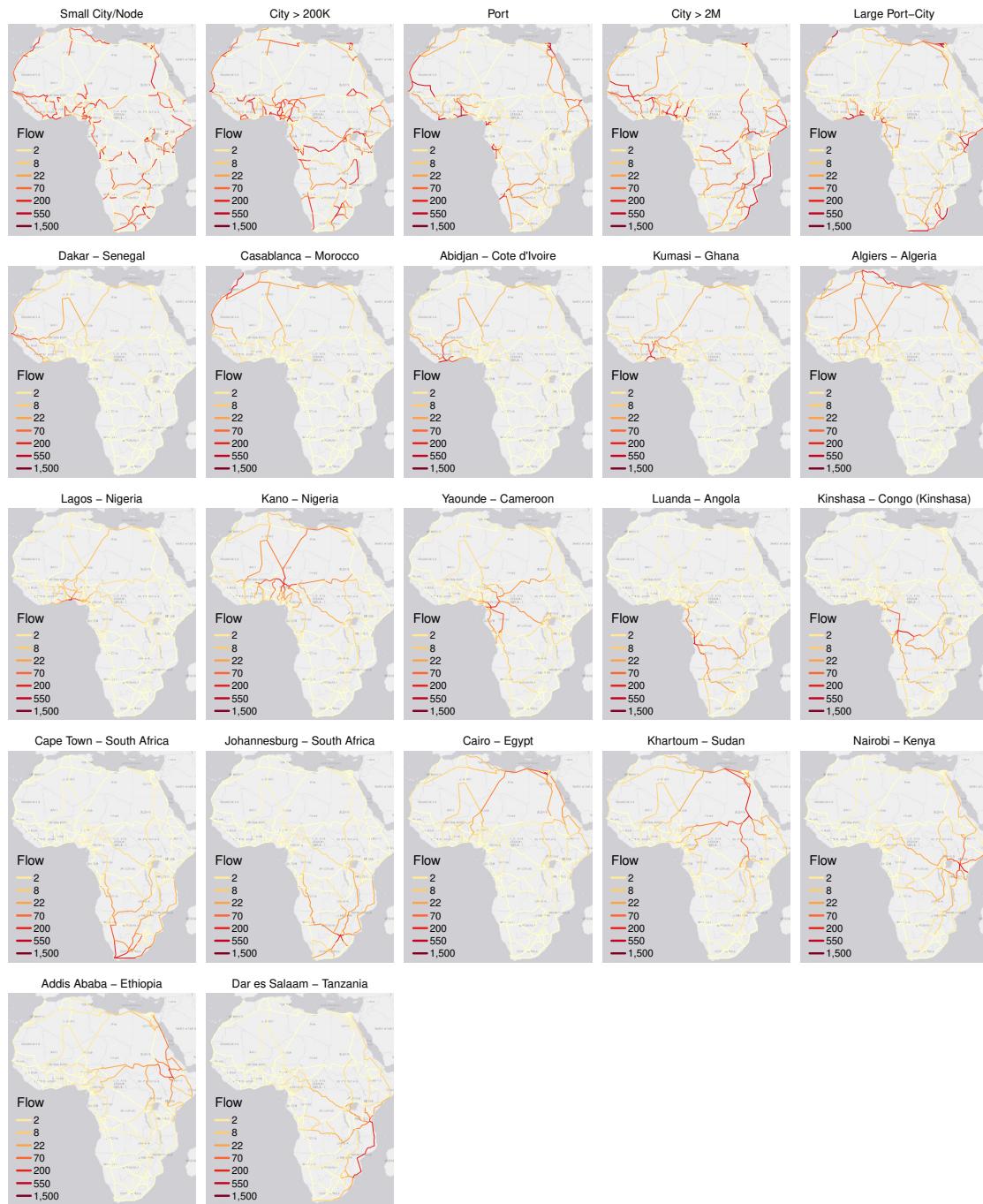


Figure A26: Optimal Trans-African Network Investments under Increasing Returns

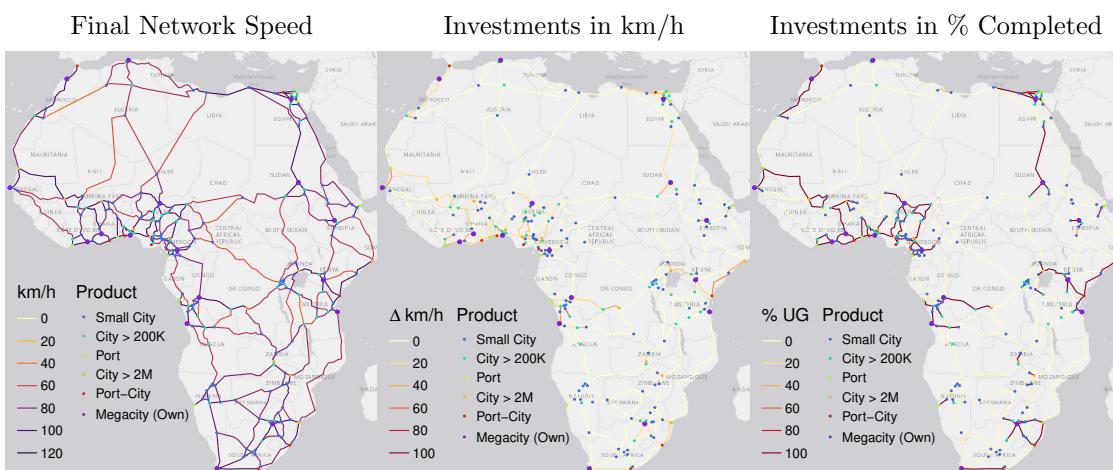


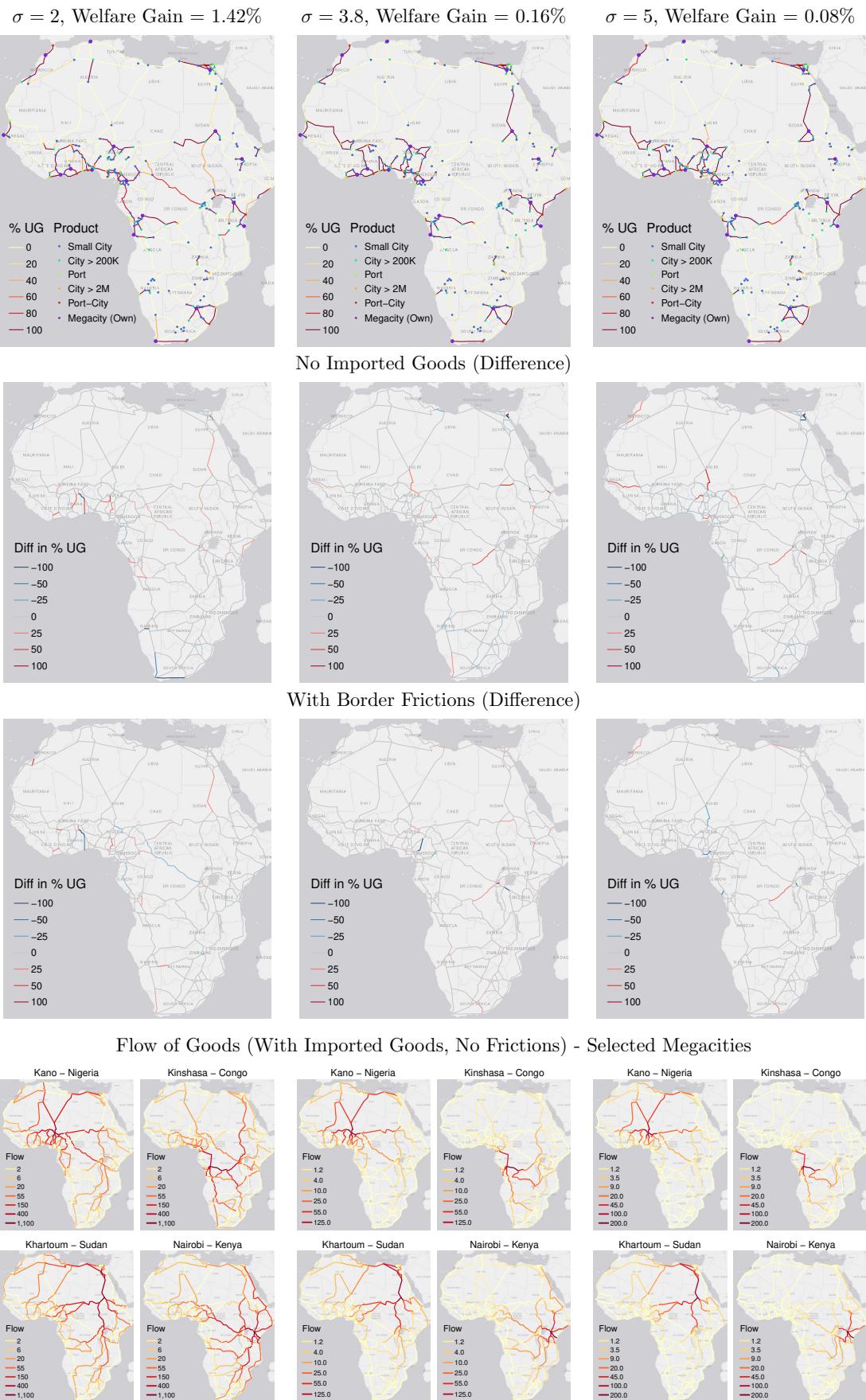
Figure A27: Optimal \$10B Trans-African Investments and Trade by  $\sigma$ : IRS Case

Figure A28: Optimal \$10B Trans-African Investments by  $\sigma$  with Inequality Aversion ( $\rho = 2$ ): IRS Case

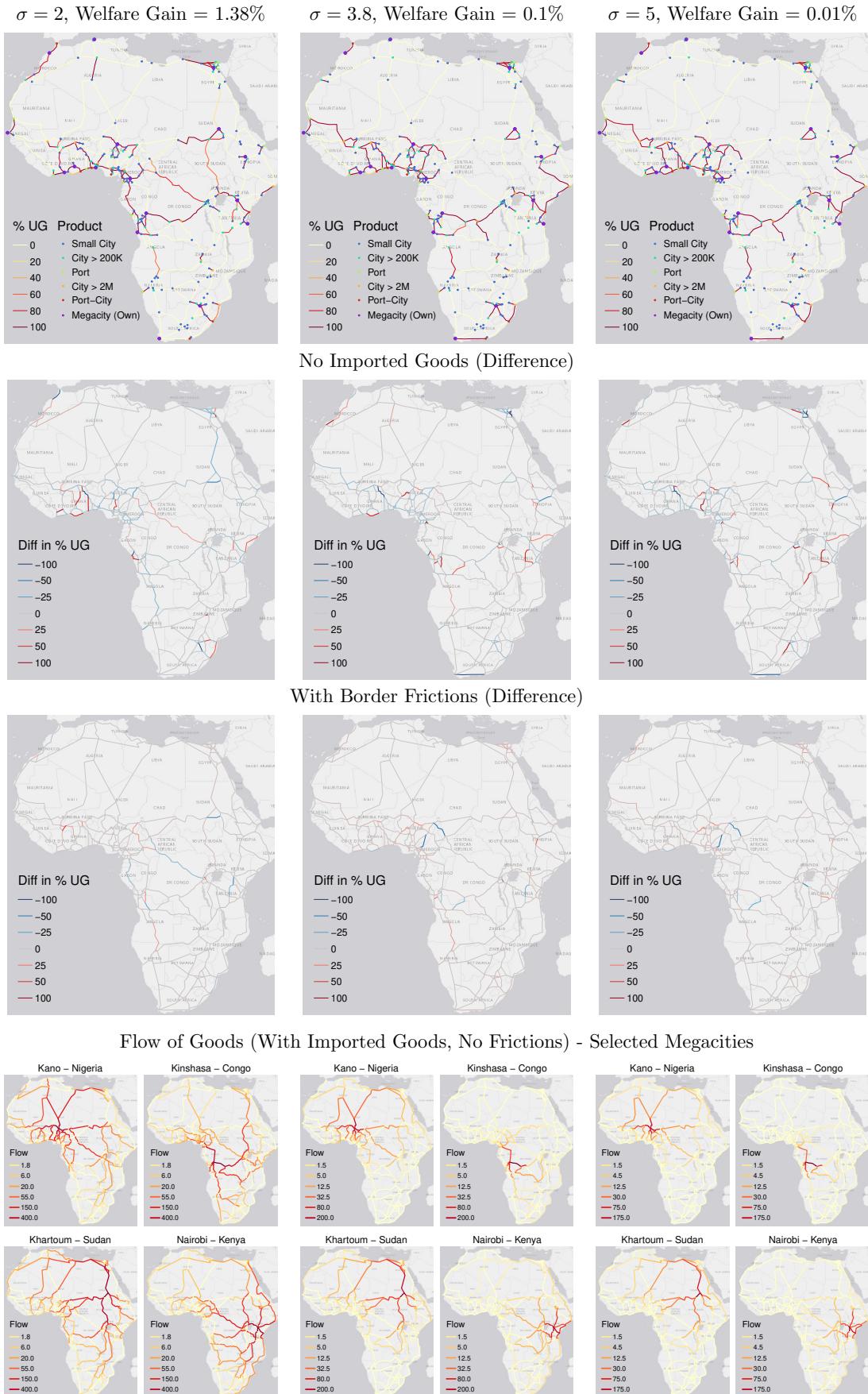


Figure A29: Flow of Goods Through Trans-African Network: Fastest and Shortest Routes

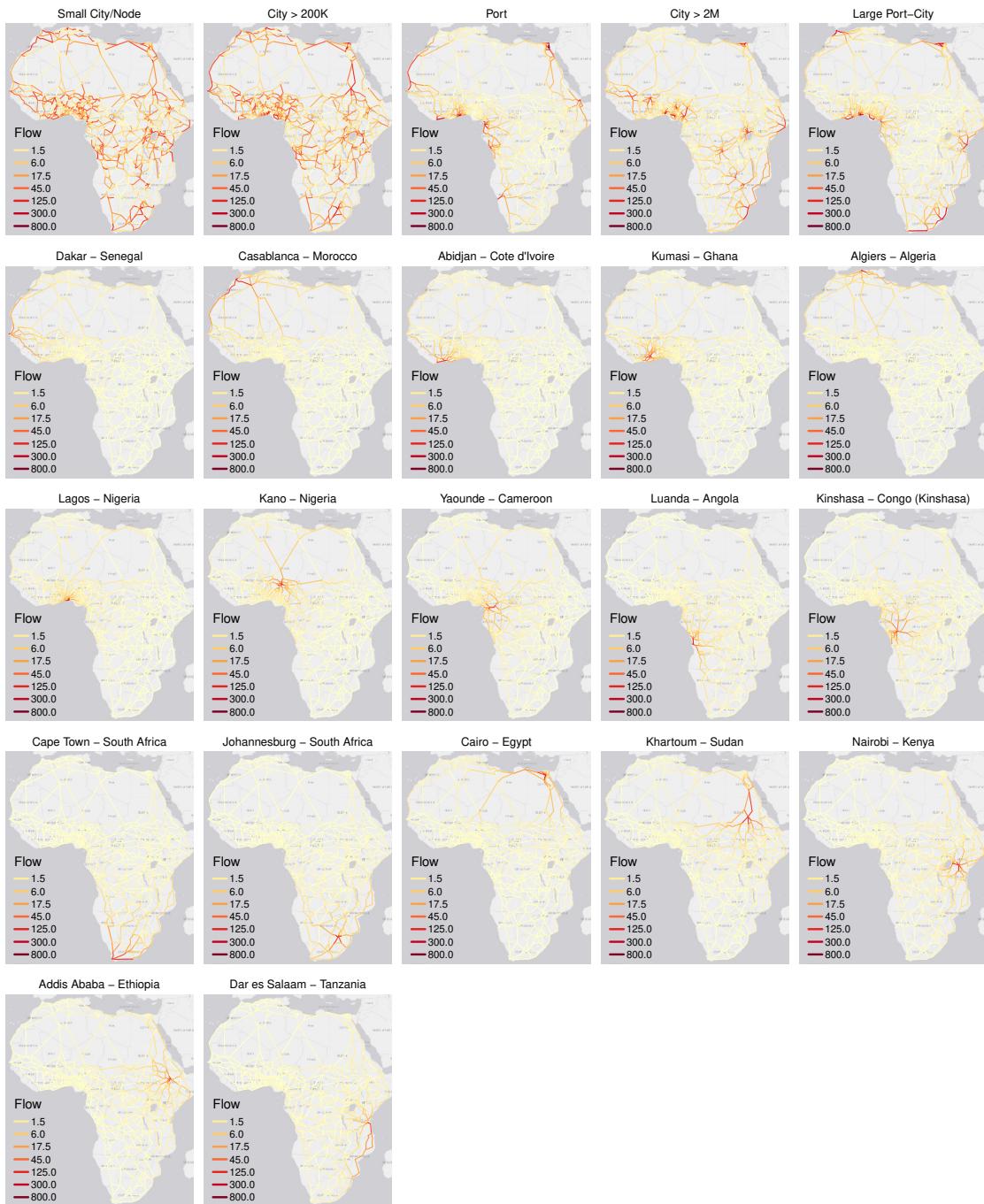


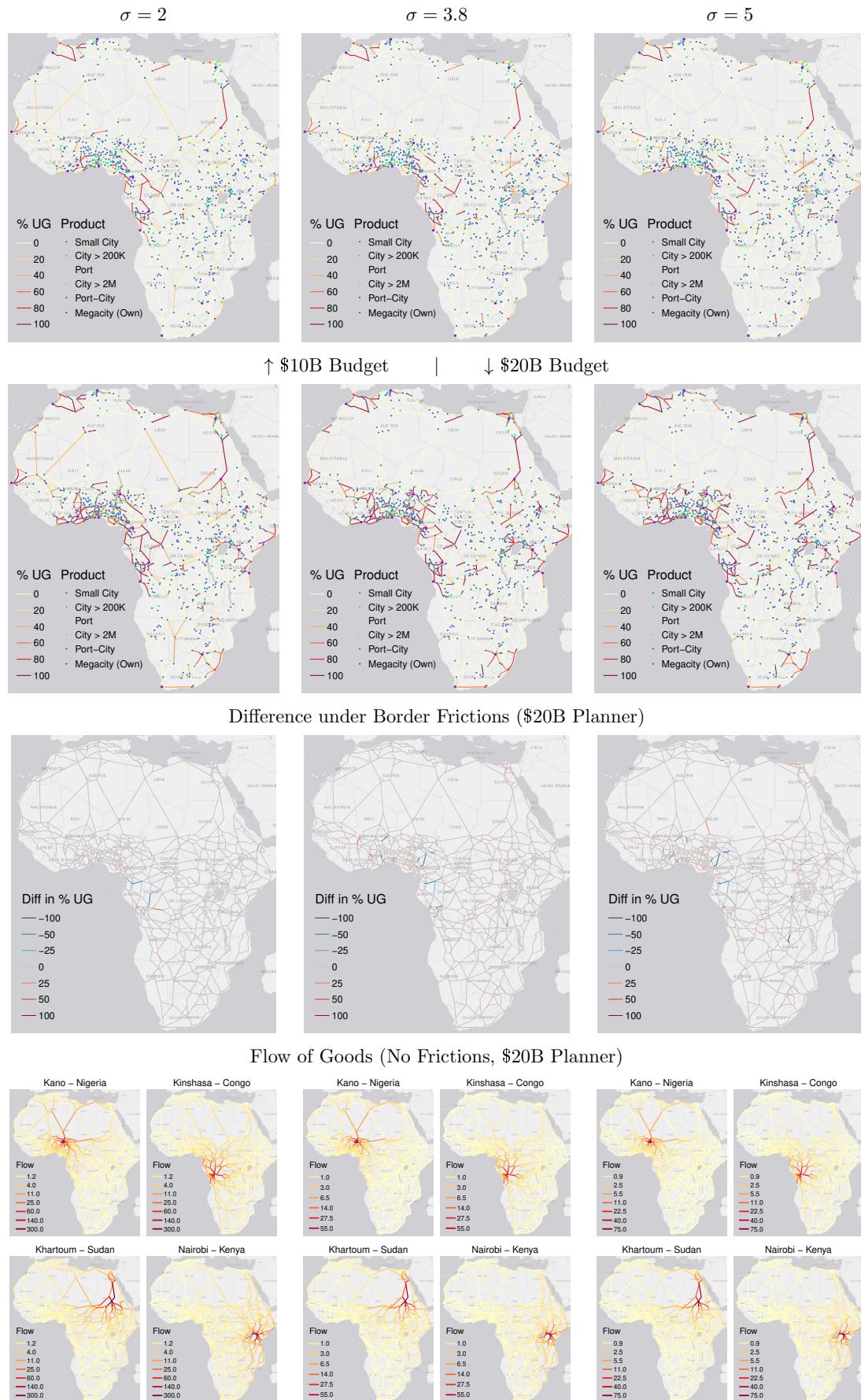
Figure A30: Optimal \$10B and \$20B Trans-African Investments by  $\sigma$  - Infrastructure

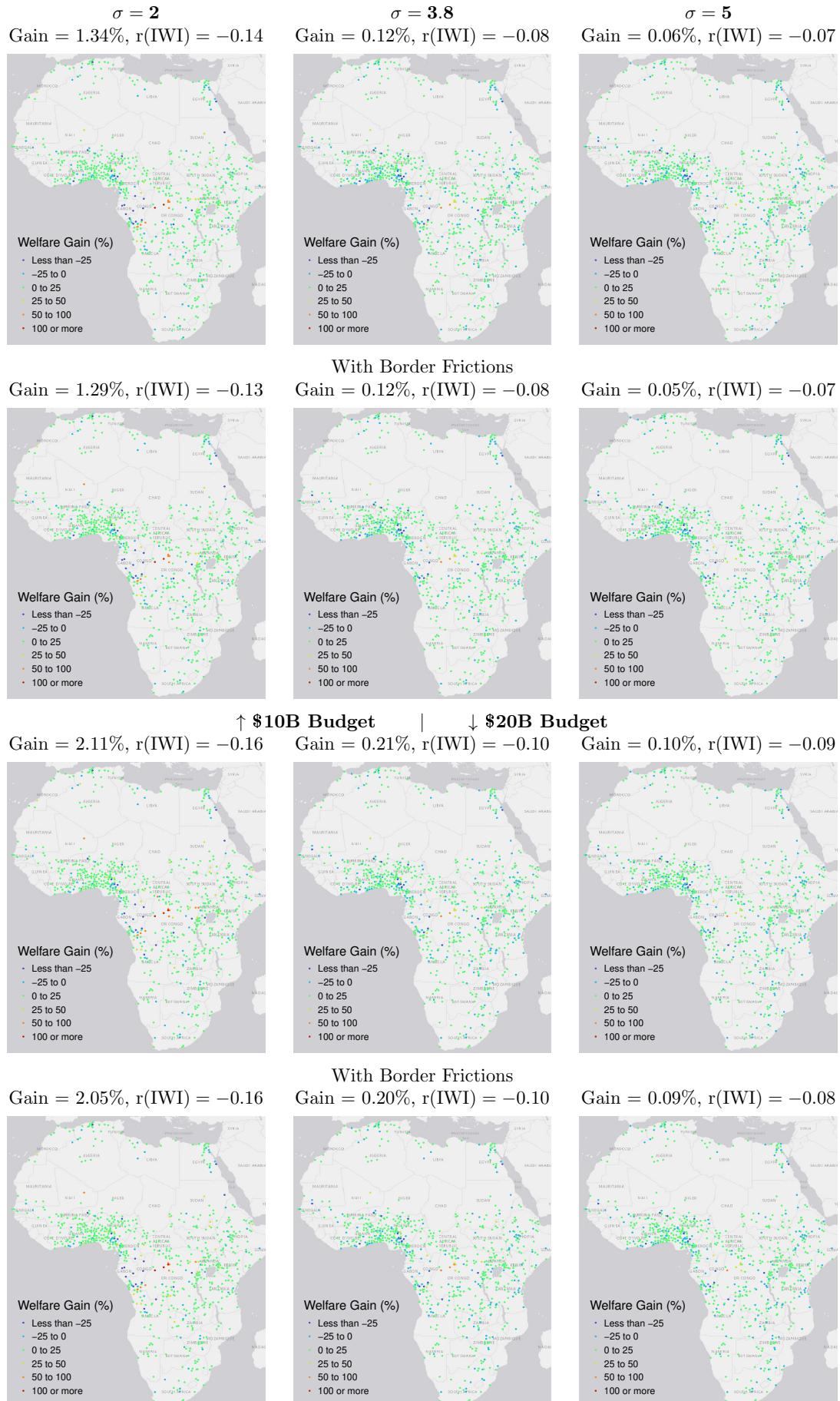
Figure A31: Welfare gains by  $\sigma$ : Standard and Frictions Scenarios

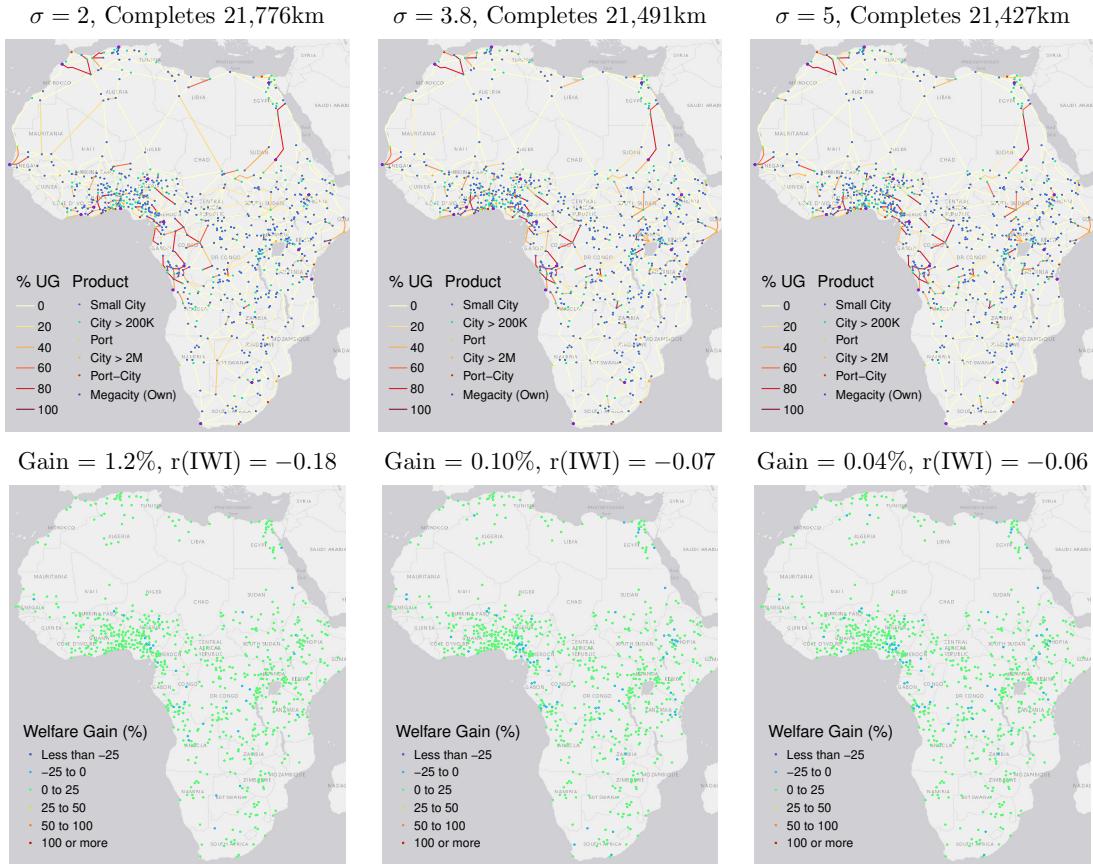
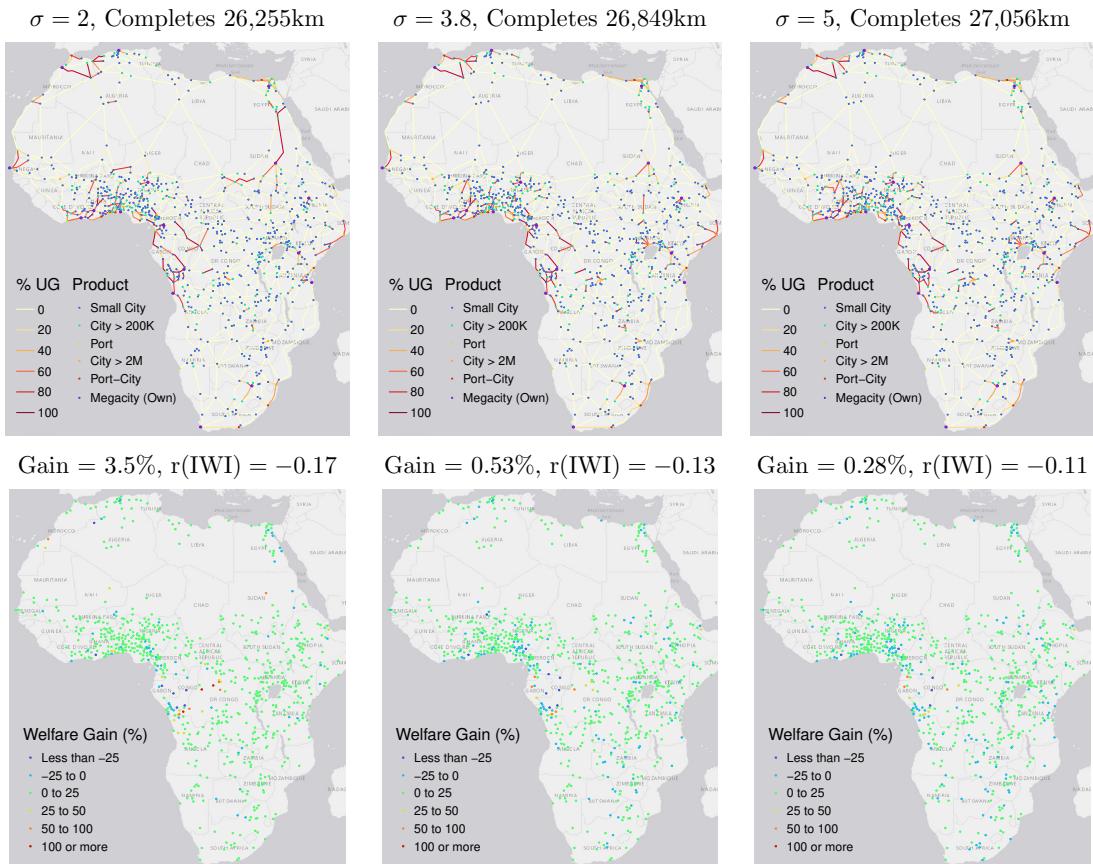
Figure A32: Optimal \$10B Trans-African Investments & Welfare with Inequality Aversion by  $\sigma$ Figure A33: Optimal \$10B Trans-African Investments & Welfare with Increasing Returns by  $\sigma$ 

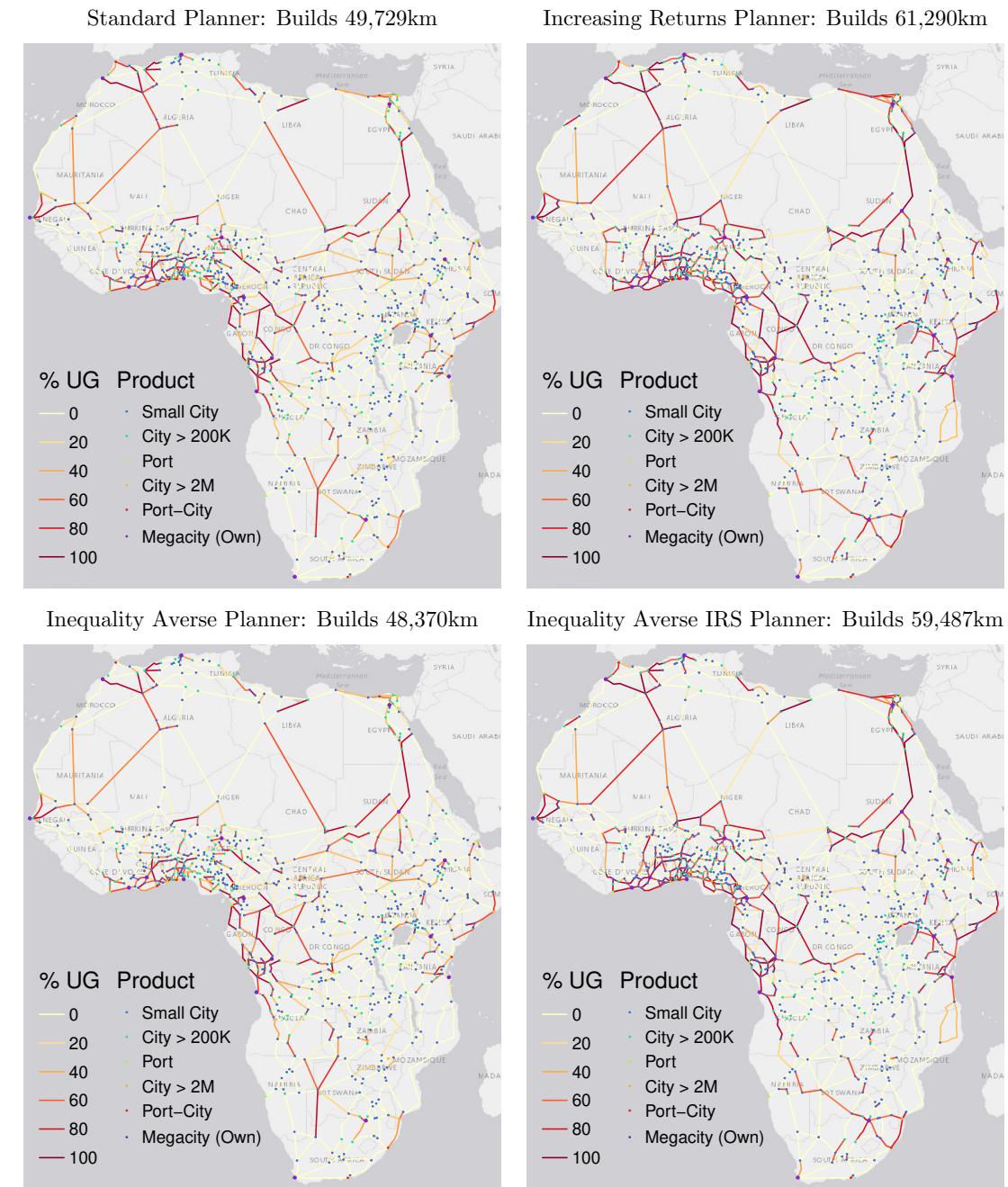
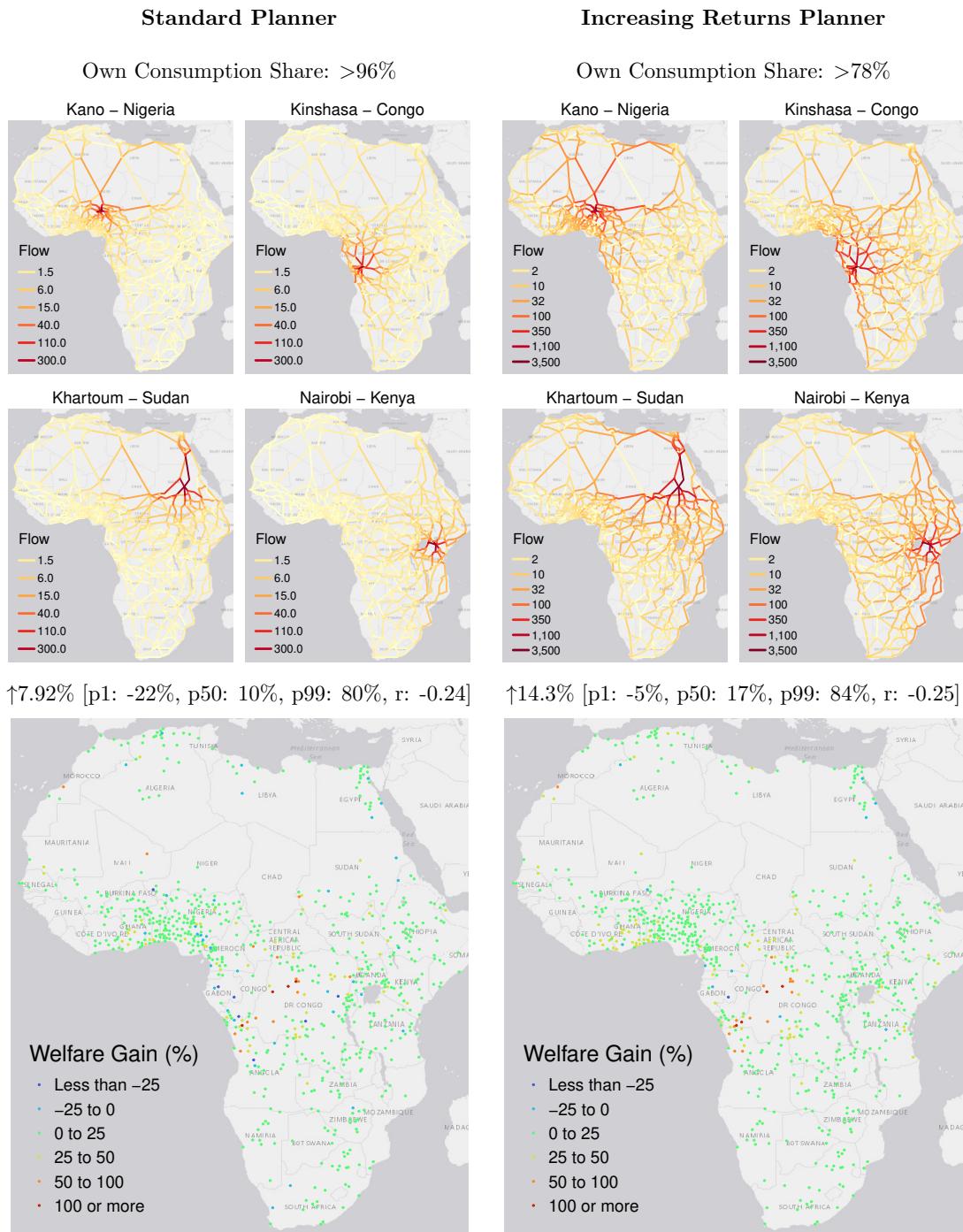
Figure A34: Optimal \$20B Trans-African Network Investments on Large Network with  $\sigma = 1.5$ 

Figure A35: Optimal \$20B Trans-African Flows & Welfare on Large Network with  $\sigma = 1.5$ 

## **Appendix A**

### **Acknowledgements**

I extend my greatest appreciation and gratitude to my supervisors, Prof. Tobias Heidland, Prof. Rainer Thiele, and Prof. Christoph Trebesch, for their invaluable input and support on the different papers of this dissertation and other research and organizational matters. In particular, Prof. Heidland has been enormously helpful in holding regular meetings to discuss my research, reading my drafts and providing useful comments, guiding me through publication processes, supporting research and conference visits abroad, and ensuring a rapid progression toward the PhD defense. I also thank Prof. Thiele for organizing conferences and activities within the Africa Initiative and providing valuable networking platforms and opportunities to present my research. In particular, the IDOS conference 2022 in Bonn, which led to a special issue on Africa's Regional and Global Integration, and the MIASA Policy Conference 2023 in Accra, Ghana, where I could present my work on infrastructure, were exceptional experiences. Prof. Thiele has also helped guide me through the (substantial) revision needed to publish my EAC paper in the Review of World Economics. Prof. Trebesch has been helpful in his strong endorsement of my pursuit to investigate Africa's infrastructure with Big Data, its effects on economic activity, and investment potentials. It has been a fruitful endeavor that led to this dissertation's most substantive empirical works (Chapter 4). Prof. Trebesch also endorsed the Africa Monitor, helped with marketing it (alongside Prof. Heidland), and giving it a prominent place on the Kiel Institute website.

I would also like to thank many other Kiel Institute colleagues who have inspired my research and provided valuable data and feedback. In particular, Hendrik Mahlkow, Sekou Metiki, Linda Maokomatanda, Julian Hinz, Frauke Steglich, Malte Becker, David Mihalyi, Sebastian Horn, Lukas Franz, Carsten Brockhaus, Alina Mulyukova and Prof. Jens-Boysen Hogrefe have assisted with feedback, data sharing, and inspirational thoughts and conversations.

Outside of the Kiel Institute I would like to thank Niclas Moneke (Oxford), Tillmann von Carnap (Stanford), Tilman Graff (Harvard), Giulio Schinaia (Chicago), Joshua Blumenstock (Berkeley), Woan Foong Wong (Oregon), Prof. Maximilian von Ehrlich (Bern), Joris Mueller (NUS) and anonymous referees for providing invaluable feedback and inspirational comments on my work.

I would also like to thank seminar/conference participants in Berlin, Stuttgart, Dresden, Bonn, Stellenbosch, Milan, Kampala, Accra, Berkeley, Oxford, the World Bank, and Kiel for helpful questions and comments.

Special thanks are due to Hylton Hollander (UCT), who invited me to Stellenbosch and facilitated my research visit there, my lovely hosts Jurie and Magdaleen Goosen during this visit, Prof. Jens-Boysen Hogrefe for giving me the opportunity and supporting me in consulting GIZ and the Beninese Ministry of Finance on government revenue forecasting, Joshua Blumenstock (Berkeley) for inviting me to present my work at the Global Policy Lab, Diana Beltekian and Woubet Kassa (World Bank) for inviting me to participate in a World Bank report on Africa's regional integration which I continue to enjoy, Prof. Stefan Kooths for providing vital inputs on the user interface of the Africa Monitor, the Kiel Institute IT department, in particular Christoph Schweickhardt, for procuring the server and facilitating the launch of the Africa Monitor and its later transition into the hands of Eisenschmidt Consulting, several Kiel Institute researchers for contributing data stories to the Africa Monitor, and Daan Steenkamp and Byron Botha from Codera Analytics for hosting and maintaining the South Africa Macroeconomic Database and the SA Nowcasting platform that I built accompanying my research visit.

Last but not least, I would like to thank my family and friends, particularly my parents and my lovely flatmates Sekou Metiki and Linda Maokomatanda, for providing social and emotional support during this, in many ways lonely and exhausting, journey of doctoral studies.

## **Appendix B**

### **Declarations**

**Declaration to confirm that the dissertation has been produced independently**

I hereby declare that I have produced my doctoral thesis „Africa’s Economic Transformation: A Big Data Economic Perspective“ independently and without external assistance and that I have specifically marked all passages taken verbatim from other authors, as well as the statements in my work that are closely based on the ideas of other authors, and that I have cited the sources according to the guidelines given to me.

---

Location, Date

Sebastian Krantz