Title

Regional Market Integration and Household Welfare:
Spatial Evidence from the East African Community

Author

Frederik Wild†[1]
University of Bayreuth

February 2024

Early draft. Please do not cite or distribute.

See most recent version here.

Abstract

The distributional consequences of trade liberalization in Africa are under-researched. In this paper, I investigate the differential impact of the East African Community (EAC) on household welfare using three distinct sets of longitudinal, geo-referenced household-level surveys from the three founding members Kenya, Tanzania and Uganda. I thereby treat the re-establishment of the EAC in 2001 – and the expansion to a customs union and common market in 2005 and 2010, respectively – as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the countries, a prediction I derive formally from a canonical New Economic Geography (NEG) model, i.e. from a quantitative spatial equilibrium with heterogenous intra-national space. To test this hypothesis, I employ a difference-in-differences specification with treatment intensity given by households' road distance to internal EAC border crossings, effectively comparing outcomes between 'interior' and 'border' households (first difference) before and after the intervention (second difference). Results reveal that households located closer to the internal EAC border did not experience positive welfare effects following the re-establishment. Rather, the results hint at the concentration of economic activity, as measured by increased consumption as well as extensive and intensive labor market opportunities in pre-existing agglomerations.

JEL Classification: F14, F15, R12, O15, O55

Keywords: FTA, East African Community, New Economic Geography, Difference-in-Differences

† University of Bayreuth, Bayreuth, Germany. frederik.wild@uni-bayreuth.de. www.frederikwild.com.

I. INTRODUCTION

Regional economic integration is widely regarded as a welfare-enhancing policy and has been a specifically popular intervention in developing economies (Schiff and Winters 2003). Particularly in Africa, deepened cooperation and trade has been suggested as ways to alleviate several barriers to development such as landlockedness, fragmented national markets as well as poor transport- and communications infrastructure (United Nations Development Programme 2011; World Bank 2020). Research on the aggregate effects of trade and integration generally supports such sentiments and point to largely positive (long-run) effects of trade liberalization (see e.g. Frankel and Romer 1999; Feyrer 2019). However, donor agencies have long emphasized the potentially inequality-enhancing impact of trade within countries (e.g. World Bank 2009), and there now exists a well-established literature which studies these distributional concerns and provides evidence for them (Pavcnik 2017). One aspect which has received particular attention is the spatial consequence of trade liberalization, i.e. the question what happens to countries' internal economic geography in response to external trade liberalization (for an overview see Brülhart 2011; Redding 2022). Heterogeneities may also form along factors such as the composition of labor markets (e.g. import-competing vs. export-oriented), income and consumption patterns of households, worker and capital mobility, and the nature of the distortions affected (Winters et al. 2004; Goldberg and Pavcnik 2007; Winters and Martuscelli 2014).

Regarding developing economies, the evidence on such distributive effects mainly stem from liberalization experiences in Asia or the Americas, with Mexico and India forming prominent country-cases (for an overview see Pavcnik 2017; Barros and Martínez-Zarzoso 2022). In Africa, similar assessments have only been explored recently, and are split along analyzing either household level outcomes using differential exposure to tariff cuts by sector (see Erten et al. 2019; McCaig and McMillan 2020; Giovannetti et al. 2022)[1] or, for a spatial analysis, on the use of economic proxies such as light emitted by night (e.g. Cadot et al. 2015; Brülhart et al. 2017; Eberhard-Ruiz and Moradi 2019).

In this paper, using a distinct set of geo-referenced household-level surveys, I provide novel evidence on the distributional effects of regional trade liberalization in Africa by combining the spatial considerations of market integration with a household-level analysis. I thereby treat the re-establishment of the East African Community (EAC) in 2001 – and the expansion to a customs union and common market in 2005 and 2010, respectively – as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the

[1] Here, exposure is typically defined at an administrative boundary and differentiated by the relative composition of specific industries within these regions.

countries. I derive this prediction from a New Economic Geography (NEG) model with heterogeneous intra-national space, i.e. a quantitative spatial equilibrium which is constructed to fit the East African Community's spatial layout. The results of the model show that as trade costs among member countries decrease, internal EAC border regions become relatively more attractive. And given that all three countries have long hosted preeminent economic agglomerations in the "interior"[2] of their countries, i.e. in Nairobi, Dar es Salaam, and Kampala, regional market integration in the EAC is predicted to act as a dispersion force and to decrease previous spatial inequalities. These predictions are brought to an empirical test using a distinct set of geo-referenced household surveys before and after trade liberalization.

The empirical results show that households living closer to the internal EAC border did not experience positive welfare effects following the establishment of the EAC, as measured by an array of consumption measures as well as intensive and extensive labor market outcomes. Rather, I observe that regional market accession within the EAC had a statistically significant and economically relevant effect on households living in the preeminent interior economic hubs. For instance, households surveyed in these agglomerated outlays increased the consumption of consumer durables by 13% compared to those from all other regions in the periods before the EAC. Further, they have a 11% increase in working in more skill-intensive occupations and show relevant decreased in the occurrence of basic consumption goods such as food, water and medicine by 32%. Corresponding to this increased agglomeration force, I additionally document a strong increase in agglomeration, as measure by population density, in these urban hubs in the years thereafter. As such, my findings go against the general prediction of the theoretical simulations and also against the hypothesis prominently outlined in Krugman & Elizondo (1996), who were the first to predict a dispersion of the formerly concentrated economic activity of developing countries following liberalization. My results are also in contrast to other recent empirical findings, which have regularly documented regions closer to the new market (potential) to profit from the less costly access to them (for an overview see Brülhart 2011). While at odds with the general predictions of the model, the theory as constructed allows a more differentiated insight into potential theoretical reasons for these empirical results. More specifically, the endogenization of the foreign country which is also outlined by heterogenous intra-national space renders foreign economic inequality as a non-negligible moderating force. In particular, the presence of economically dominating interiors in the EAC potentially may weaken the draw to the border to the point where a core-periphery pattern as existent prior to the EAC remains a possible long-run stable equilibrium even after regional trade liberalization.

[2] As seen from their relative position against their respective EAC partner countries.

II. RELATED LITERATURE

The paper relates to the body of research investigating the impact of trade on households and welfare on the one hand, and the literature analyzing the spatial consequences of trade liberalization on the other.

Trade and Household Welfare

Increased availability of detailed survey data has aided the growth of the literature assessing the link between trade liberalization and household welfare (for an overview see Goldberg and Pavcnik 2007; Winters and Martuscelli 2014; Pavcnik 2017; Barros and Martínez-Zarzoso 2022). This body of research confirms the notion that trade does not unequivocally increase the welfare of all households within a country, i.e. produces winners and losers. An analytical starting point in thinking about these heterogenous effects is given by a stylized production-consumption schedule of households and may be encapsulated by $\Delta W = (q_i - c_i)\Delta p_i$, whereby welfare changes are explicitly moderated by (trade-induced) price changes (see Deaton 1997; Winters et al. 2004). Depending on whether the household is a net consumer (c_i) or producer (q_i) of product i, a given price change Δp_i will either lead to net benefits or net losses. In his seminal paper Porto (2006) extends such partial equilibrium statics to a general equilibrium model of trade, taking account the simultaneous changes prices of non-traded goods, and subsequently, second-round effects resultant of altered factor- rewards and intensities in specific industries.[3] Evidently, these dynamics are highly relevant in cases where specific sectors are facing increased import-competition from international exporters or where export-oriented produces are drawing increased demand from abroad. As such, one and same trade policy may render very different results depending on the goods affected, households' production and consumption schedule, and subsequent general equilibrium effects.

Porto's approach has been subsequently extended and employed to study trade effects in various contexts, including Mexico (Nicita 2009), Brazil (Borraz et al. 2013), India (Marchand 2012; Ural Marchand 2019), Tunisia (Martínez-Zarzoso et al. 2016), as well as in six African countries (Nicita et al. 2014). These studies typically employ changes in (non-)traded goods prices together with income-consumption shares reported in household surveys. To assess the overall welfare impact, these changes are then compared across the (income or expenditure) distribution to assess the pro-poor or pro-rich character of a trade-policy. Most of these studies provide evidence of a pro-poor effect of

[3] Evidently, Porto's approach depends largely on the parametrization of wage and (cross-)price elasticities as well as the pass-through rate of the border price (Goldberg and Pavcnik 2007).

trade, some of them showing mixed results, and Nicita (2009) being the only exception in showing a clear "rich-only" impact of trade in Mexico.

A second branch of the literature on trade and household welfare has relied on "Bartik-style" shift-share instruments to identify trade effects.[4] Here, exposure is typically defined at an aggregate level, such as at a particular administrative unit (e.g. districts or regions). The intensity of trade on households living within a specific region is then differentiated by the pre-liberalization concentration of a industries and the respective tariff cuts (see Goldberg and Pavcnik 2007; Winters and Martuscelli 2014). For instance, McCaig (2011) shows that the U.S.-Vietnam Bilateral Trade Agreement accelerated poverty decline, as export growth due to tariff removal was largest in the low-skilled labor-intensive apparel and clothing sectors. On the other hand, Topalova (2010) provides evidence that India's trade liberalization of 1991 actually slowed poverty decline in the most affected regions, i.e. the ones intensive in agriculture, given that such sectors faced increased import-competition. Related studies have looked at similar issues in Brazil (Castilho et al. 2012), China (Emran and Hou 2013), India (Edmonds et al. 2010), Indonesia (Kis-Katos and Sparrow 2015), and Vietnam (Fukase 2013; Vo and Nguyen 2020). These studies are mixed in finding both decreases as well as increases in relative poverty. In Africa, the evidence on liberalization experiences in this literature is almost universally negative. For instance, drawing on South Africa's trade liberalization of the 1990s, Erten et al. (2019) find decreased formal as well as informal employment for more affected regions and no effects on wages for those remaining employed. Relatedly, McCaig and McMillan (2020) find neither a contraction nor an expansion of industries in neighboring Botswana, which was affected by the same liberalization schedule.[5] Rather, they report higher likelihoods of being employed informally for more intensely affected regions. In the same vein, evidence from Ethiopia suggests increased unemployment levels in regions more exposed to trade liberalization and import competition in light of the Structural Adjustment Programs (SAP) of the early 1990s. One exemption to these findings is Giovannetti et al. (2021) who provide evidence of a negative effect of protective policies in Egypt shortly after the Spring Revolution. Interestingly, they find neither positive nor negative results of trade liberalization in the preceding decades.

To my knowledge, there exists no study analyzing household-level welfare concerns of trade liberalization in Africa from a spatial point of view.[6] One exception to this is Cali (2014), who assesses Uganda's progressive liberalization policy with Kenya in the 1990s on wage premia, i.e. changing

---

4 As introduced by Bartik (1991) as well as Blanchard and Katz (1992).
5 Botswana is a member of the South Africa Customs Union (SACU).
6 As a matter of fact, I was not able to identify a study exploiting the geo-referencing of survey locales to study these links on any continent for that matter.

returns to schooling.[7] However, the variation across space is given at a district level (GADM2), of which there are a total of 38 and 45 in the study across the two survey rounds, respectively. To compare, in this paper, households' location is defined by latitude-longitude combinations comparable to GADM3 or finer. As such, I draw from a minimum of 104 and a mean of 324 GPS locations per country per round, or a mean of 299, 326, 353 for Uganda, Tanzania and Kenya, respectively. As such, analyzing the spatial response of household welfare to trade liberalization (with higher precision) represents a research gap I aim to fill. Motivating differential trade effects across space requires an overview of the relevant theoretical and empirical findings in this regard, which is provided in the next section.

Spatial Effects of Trade

The second strand of literature to which I contribute investigates the spatial consequence of trade liberalization. This growing body of research has its roots in New Economic Geography (NEG) and has extended to an active field now better referred to as "quantitative spatial economics" (for an overview see Redding and Rossi-Hansberg 2017; Brakman et al. 2019; Redding 2022).

Krugman's (1991) seminal paper was a crucial expansion on earlier conceptualizations of spatial economic distribution, which mainly concentrated on allocations within cities, such as the von-Thünen model (1826), or the relative size of cities (Henderson 1974, 1982). The advantage of NEG in comparison to these earlier specifications lies in the fact that it can explain the spatial distribution of cities against each other such that there are not simply "floating islands" (Brakman et al. 2019; 3). Krugman's model is essentially based on new-trade-theory (Krugman 1979, 1980) and thereby combines monopolistic competition (Dixit and Stiglitz 1977) with increasing returns to scale. Most importantly, trade costs factor in between locations, regulating their spatial allocation against each other (Krugman 1991). Hence, the endogenous allocation of activity ultimately boils down to producer- and consumer problems, who optimize over given a set of preferences and production technology, while factoring in trade costs. Agglomeration is then a product of cost (forward) and demand (backward) linkages which produce centripetal forces, while dispersion is a product of increased competition, the costs of urban congestion, or immobile factors of production. For instance, because firms operate under increasing returns to scale and incur transport costs, they benefit from the increased demand in larger locations, i.e. move where demand is highest (demand linkage).[8] And given that consumers have a "love of variety" and will additionally save on higher price tag for

[7] The analysis is motivated by a Hecksher-Ohlin type trade effects, thereby suggesting to decrease wage inequality in a developing country who is labor abundant and human capital scarce.

[8] Note that in large markets, the additional presence of a firm increases demand mechanically, and by being able to pay higher wages, thereby further strengthening the backward linkage.

shipping, consumers prefer to locate close to (a large number) of producers (cost linkage). However, while large regions offers firms high demand and consumers lower prices, competition as well as costs of congestion (commuting, land rents) are increased which decreases agglomeration tendencies. In the long-run, an equilibrium is given by the balance of these forces, i.e. when the advantages and disadvantages of agglomeration or dispersion, expressed in real wages, are net zero. In this scenario, there exists no incentive for firms or workers to relocate.

This endogenization of the spatial allocation of economic activity has provided a workhorse model and spurred subsequent extensions and applications to questions on how spatial inequalities form and how they may be affected. Importantly, NEG allows the comparative statics examination of what happens the centrifugal and centripetal tensions in response to changes in *internal transportation costs* or, importantly, *external trade costs* (for a synthesis see Fujita et al. 2001).[9] Concerning the latter, both theoretical and empirical results vary in their prediction of whether liberalization increases or decreases spatial disparities within countries (for an overview see Brülhart 2011). Krugman and Elizondo's initial treatment (1996) famously predicted the dissolving of the "giant Third World metropolis" of developing countries in response to external trade liberalization. The model extends the stylized two-region case to a three-region-economy, with two regions situated in the home country, and one region ("rest of the world") posing as the international market to which trade costs are successively lowered (a 2+1 economy). Krugman and Elizondo (1996) sparked an array of refinements and extensions to this basic setup. Interestingly, however, the prediction from these theoretical advancements is far from uniform. While Behrens et al. (2003, 2007) confirm the original prediction, several adaptations arrive at the contrary result, i.e. that increased trade liberalization sparks intra-national agglomeration. For instance, in the same original 2+1 setup, Paluzie (2001) as well as Brülhart et al. (2004) and Crozet and Koenig (2004) provides evidence of increased agglomeration in response to external trade liberalization. Further studies have extended the setup to 2+2 economies, confirming these predictions (Monfort and Nicolini 2000; Monfort and van Ypersele 2003). The difference among all of these studies is how they chose key elements from the "menu of building blocks" (Redding and Rossi-Hansberg 2017; 25), i.e. how consumer preferences (CES or quasilinear) as well as dispersion forces (immobile workers vs. congestion) are modeled.[10] One particularly interesting adaptation of this literature is to allow for *heterogenous intra-national space*, i.e. regions (within-countries) to differ from

[9] While the core model of NEG is known for its "bang-bang" property for changes in transport costs, i.e. equilibria between complete spreading or agglomeration, subsequent adaptations have accommodated a wider range of equilibria, using stronger centrifugal (dispersion) forces such as interregional labor immobility (e.g. Krugman and Venables 1995), diminishing returns in the non-traded sector (e.g. Puga 1999), or housing (e.g. Helpman 1998).

[10] In addition, a full menu of is also outlined by choices on the production technology, trade costs, externalities, labor mobility, as well as endowment structure across regions.

one another ex-ante. For instance, in Mansori (2003), Brülhart et al. (2004), Crozet and Koenig (2004) and Behrens et al. (2006) they additionally test what happens to the prediction if one region has better access to the international market than the other, i.e. poses as a "border" or "gate" region. What these class of models show is that in almost all instances, external trade liberalization leads to increased "draw" to the border, i.e. to the region with the better foreign market access (Crozet and Koenig 2004b). However, depending on the relative size and the export intensity of the home and foreign markets, this draw to the border may be alleviated as the interior as it acts as a shield to foreign competition (Brülhart et al. 2004). These effects may be further mediated by varying intra-national transport costs which regulates the pass-through of changes in international trade costs towards the interior as well as the symmetry of the foreign country (Behrens et al. 2006). These initial refinements to an asymmetric regions were first steps into what is now more richly embodied in "quantitative spatial economics" whereby first-nature characteristics (e.g. local endowments such as productivity, amenities or floor space) are paired with the "classical" second-nature agglomeration and dispersion forces, which are produced by the endogenous relative position of agents against each other (see for a distinction Redding 2022).

The empirical evidence reflects the ambiguity shown across these models. While evidence from cross-country settings lean towards the convergence of economic activity in response to trade liberalization, within-country evidence has shown increasing inequalities for various settings (see for an overview Brülhart 2011). However, a rather robust empirical result across the empirical literature is that regions with relatively better access to foreign markets, often border regions or regions near the coast, generally stand to benefit comparatively more. This mirrors the theoretical results of the class of models with heterogenous intra-national space. Naturally, whether this leads to convergence or divergence of economic activity within countries naturally depends on the pre-liberalization diffusion of economic activity. For instance, convergence is found to occur in settings where market access is higher in the historically economically weaker border regions, as was the case in Austria (Brülhart et al. 2012) or Germany (Redding and Sturm 2008).[11] On the other hand, divergence is found somewhat more frequently, as documented by the increasing activity to the already industrialized U.S.-Mexican border following NAFTA (Hanson 1994, 1997), or in China, where trade has benefitted the already more developed coastal areas (Kanbur and Zhang 2005). Next to singular country cases, a growing field of literature employ large scale evidence employing satellite imagery, were lights emitted by night serve as a proxy for economic activity to assess spatial within-country

[11] Redding and Sturm (2008) show evidence for population movements west, increasing regional inequality in Germany. The reason for my conclusion is that the population movement was induced by market loss of border regions, rather than market gain, which would vice-versa lead to the opposite result.

inequality.[12] So far, much of the evidence has a tendency for trade to increase within-country inequality, and particularly so in developing regions (Ezcurra and Rodríguez-Pose 2014; Hirte et al. 2020; Ezcurra and Del Villar 2021).

Within-country evidence for Africa is scarce and is mostly conflated with these large-scale studies of all world regions. Particular country-case investigation in Africa so far has also exclusively relied on nighttime lights as a source of data across space. For instance, Cadot et al. (2015) who look at the influence of improved trade on the border shadow in sub-Saharan Africa. Similarly, Brülhart et al. (2017) estimate this border shadow for Uganda and Rwanda in specific. Lastly, similar to my study, Eberhard-Ruiz and Moradi (2019) who also analyze the impact of the East African Community on city growth within Kenya, Tanzania and Uganda.[13]

As noted in the previous section, there exists no study for the African continent which use geo-referenced household-level data analyze these dynamics. However, there are distinct benefits in using household level data to measure distributive effects of trade policy. First, it allows to analyze consumption of households, which is arguably a better measure for overall welfare at one specific point in time if one assumes intertemporal optimization/smoothing of consumption. Further, trade policies tend to alter prices in a non-uniform way which affects both income and consumption, better captured by consumption as an outcome of both. Second, household-level data allows us to additionally explore potential mechanisms regarding human capital, occupation, gender, as well as the household composition (production-consumption) which may drive these effects at the aggregate level. And lastly, nightlight data may not be as reliable in our setting. Recent research on the quality of nightlight data has cautioned practitioners of the quick application, particularly in developing countries. Results have suggested that precisely the areas relevant to development economists, i.e. low density, rural (agricultural) areas are due to non-negligible measurement issues (e.g. Bickenbach et al. 2016; Gibson et al. 2020, 2021). Relevant to our case, studies have shown that nightlight-to-GDP elasticities may differ largely between rural and urban areas, which may lead to conflating a systematic measurement error with policy impacts (Bluhm and McCord 2022).

The use of three, independently collected household surveys may potentially alleviate such concerns, and additionally, helps in exploring potential mechanisms regarding human capital, occupation, and household composition (production-consumption) which may drive the effects at the aggregate level.

---

[12] See e.g. Gibson et al. (2020) on the various uses of night light data in economics.

[13] This study focuses on assessing spatial city growth in the East African Community. The authors show that the effects of the EAC were highly localized. Increased city growth after the EAC was only observed relatively close to internal borders and also only short-lived.

Institutional Background

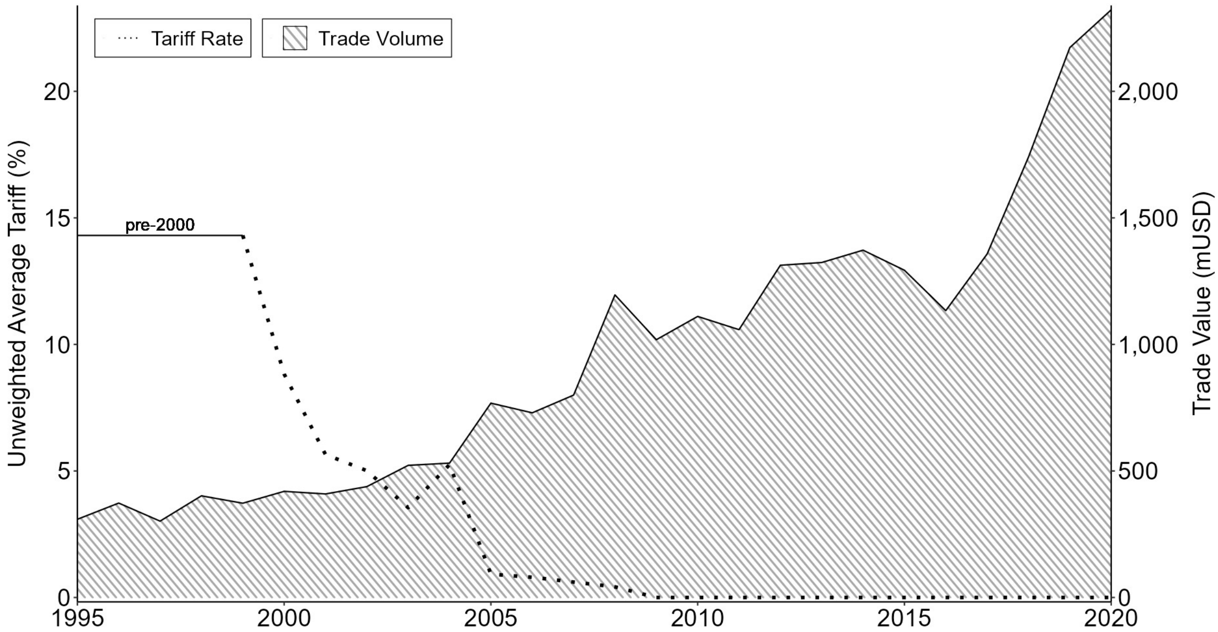
The East African Community (EAC) was originally found by the Republics of Kenya, Tanzania, and Uganda in 1967. Placed around Lake Victoria in East Africa, the three countries share two common borders each and economic and political cooperation between the countries has historical roots. In pre-independence periods, roughly from 1900-1960, they shared large infrastructure outlays such as railways, telecommunication, postal service and a common currency (Hazlewood 1979; EAC 1999). However, not soon after the first formal treaty towards the establishment of an East African Community was signed in 1967, questions on sovereignty, and particularly the "disproportionate sharing of benefits of the Community among the Partner States" arose (EAC 1999; 1). While attempts of redistribution of benefits were made, it was deemed as insufficient by the member states and trade restrictions were levied between them even while formally in union (Mugomba 1978; Hazlewood 1979). Next to a "lack of strong political will" (EAC 1999; 1), these are often cited reasons for the ultimate demise of the original EAC in 1977 when it was formally dissolved. However, the mutual interest of working together in a union was kept alive in the decades thereafter, as seen by the gradual move towards the modern EAC for instance by the establishment of the "Permanent Tripartite Commission for East African Cooperation" in 1993 or the "East African Cooperation Development Strategy" in 1997, which focused on the for closer co-ordination in economic, political, fiscal, immigration, infrastructural as well as social and cultural arenas (EAC 1999).

The institutional establishment of the modern day East African Community was initiated with the treaty of 1999, which was ratified on July 7th of 2000, and the new EAC began to operate as a free trade area on January 15th of 2001 (EAC 1999; Kaahwa 2003). Hence, it was not before 2001 after which the substantial lowering of tariff rates by member states was initiated.[14] The EAC consistently moved towards deeper integration in the years thereafter, with the protocol for a customs union operational from the 1st of January 2005 followed by a transitional period to a common market on the 1st of July in 2010. While member states have since ratified the move towards a monetary union in 2013, a common currency has not been implemented as of yet. Figure 1 depicts these developments quantitatively, by plotting the simple (unweighted) average tariffs among the EAC founding members together with their total merchandise trade in mUSD from 1995 to 2020 (UNCTAD 2022; UNSD 2022).[15]

---

[14] For instance, Tanzania postponed many substantial tariff line removals to the budgetary year beginning July 1st 2001 and for sugar even until July 1st of 2002, which was the 7th highest valued import in the years between 1996 and 2001 of all 96 chapters in the H1 nomenclature (UNCTAD 2022). See also Eberhard-Ruiz and Moradi (2019) for a more detailed account on the tariff structure around the implementation.

[15] The numbers reported reflect current dollar values of the respective year. We use import values as there are some gaps in the reporting of exports values.



Notes: The unweighted average tariff depicted on the dotted line represents the 'AHS simple average tariff' and comes from the United Nations Conference on Trade and Development (UNCTAD) Trade Analysis Information System (TRAINS). The values between 1995 and 2000 are irregularly reported and stem from a single country only, hencewhy the average is taken. The trade volume depicted by the shaded area represents the total USD value of all goods imported within the EAC member states and stems from the United Nations Commodity Trade Statistics Database. Both data are facilitated via the World Integrated Trade Solution (WITS) platform.

Figure 1: Tariffs and Trade in the East African Community (EAC)

The EAC has also expanded outwards to contiguous countries of the region, with the accession of Burundi and Rwanda in 2007, South Sudan in 2016, the Democratic Republic of Congo in 2022 and also Somalia most recently in 2023. However, the three founding members still account for the overwhelming majority of economic activity with over 70% of the EAC's total GDP in 2022 . As such and given that I aim to evaluate effects over the entire timeline of the modern EAC, this paper concentrates primarily on evidence drawn from the founding members Kenya, Tanzania, and Uganda. The section on robustness and extensions in Chapter IV includes additional insights from the two first accession countries Burundi and Rwanda.

Since the establishment of the EAC in 2001 the three economies have grown by a total of 111%, 86% and 47%, respectively. However, with a per capita GPD (PPP) of 2,624$ and 2,280$ two of the three countries, Tanzania and Uganda, are still categorized as "low-income-countries".[16] Only Kenya has graduated to a "lower-middle-income country" as of 2014, with a current GDP per capita of 4,882$. These low levels can be partly attributed to the substantial population growth within these

[16] GDP figures are expressed in constant 2017 international (PPP) USD.

countries over the same time period, roughly doubling from a 90 million in 2000 to 166m in 2022.[17] Concerning the economic structure of the countries, they are heavily reliant on agriculture and services. In Kenya, the service sector makes up a total of 48% of the GDP, followed by agriculture (38%) and manufacturing (9%). Services also dominate in Tanzania (40% of GDP), who hosts a large tourism sector, with agriculture making up 32%. Manufacturing is not as important in Tanzania with a contribution to total GDP of 6%. In Uganda, the respective figures total to 27%, 52% and 9% (WTO 2019).

Concerning trade, merchandise exports display a relevant contribution to the economies' GDP with shares of 25%, 28 %, and 29%, in Kenya, Tanzania and Uganda, respectively (World Bank 2022). However, regarding the direction trade, the large majority of merchandise is still sourced and exported to markets outside of the continent, with extra-African export and import shares of 40-50% and 80-90%, respectively. China, India, as well as markets in the Middle East and the EU have been the predominant trading partners within the last decade (WTO 2019).[18] As such, the share of intra-community ("intra-EAC") trade has been relatively low, hovering around 10% of total trade since its establishment in 2000 (UNU-CRIS 2019).[19] Some of the reasons for the relatively low volumes of intra-regional trade in the EAC are outlined by the complementarity of goods produced, several non-tariff barriers of trade, infrastructural shortcomings but also the importance of informal cross-border trade (WTO 2019). However, compared to the other eight officially recognized regional economic communities (RECs) on the continent, the EAC has the second highest intra-regional trade share, trailing only the Southern African Development Community (SADC) whose members' intra-regional trade account for 20% of total trade.[20] Further, there is a significant asymmetry in the pattern of trade, as only 6% of imports are sourced within the EAC, but 20% of countries' exports are directed to markets within the EAC (WTO 2019). The predominant type of goods traded within the original EAC members are comprised of primary products such as mineral fuels and oils, gemstones as well as cereals but also manufactured goods such as rolled iron, steel and steel products, vehicles and electrical machinery, plastic goods, processed food and beverages as well as pharmaceuticals (UNCTAD 2022).

One particularly pertinent aspect of the three countries is their economic geography, which is outlined by exceedingly high levels of urban economic primacy. While the large majority of the population is dependent on agriculture and lives in rural environments (70% in 2022, down from 80%

[17] The population of the three countries grew from 31m to 54m in Kenya, 34m to 65m in Tanzania and from 24m to 47m in Uganda.

[18] The main goods exported are primary products (mainly agriculture produce) which make up 60%, 61%, and 43% of total export value across the three countries respectively, and those declared manufactures 28%, 18% and 18% (WTO 2019).

[19] Between 10 to 20% when including the trade in services (IMF 2023).

[20] The eight RECs have an average intra-regional trade share of 6% (UNU-CRIS 2019).

in 2000), the majority of the countries' economic activity is concentrated in the geographically confined hubs Nairobi, Dar Es Salaam, and Kampala, respectively (World Bank 2022). For instance, around the time of the EAC's establishment in 2001, Dar es Salaam hosted only 7% of the country's population but 51% of formal employment of the private sector and contributed to over 57% of the total wage bill (Tanzania National Bureau of Statistics 2004, 2006).[21] Considering that the administrative region of Dar es Salaam makes up a mere 0.16% of Tanzania's total land area, this describes a large intra-national discrepancy in economic activity. To compare, the next largest contributors to the wage bill in 2001 were Kilimanjaro, Arusha and Dodoma with 6%, 5% and 5%, respectively, and land shares of 1.5% and 4% and 5%. This pattern has continued to persist and is particularly pronounced in the high value-added manufacturing sector. For instance, in 2008, Dar es Salaam hosted 55% of manufacturing establishments (30% of the manufacturing labor force) while contributing over 51% of the country's total value added (Tanzania National Bureau of Statistics 2010). In the latest available survey of 2016, Dar es Salaam still contributed to over 41% to total value added, albeit hosted a smaller share 27% of all manufacturing establishments, and 32% of the manufacturing workforce, which is however, well over twice the amount the next largest region Morogoro.[22] The structure of the EAC partner countries evince the same spatial pattern. Concerning Kenya, Nairobi accounted for 46% of (formal) wage employment in 2001 and for 51% of the total wage bill among main towns (Kenya Central Bureau of Statistics 2003; Kenya National Bureau of Statistics 2011).[23] In 2009, almost a decade later, these figures were virtually unaltered. Together with the second largest industrial hub, Mombasa, these figures increase to over 63% and 69% in 2009 for the employment and wage bill, respectively. Again, to compare, Nairobi makes up only 0.12% of the total land area and 7 (8%) of the population as per the census of 1999 (2009) (Kenya Central Bureau of Statistics 2001; Kenya National Bureau of Statistics 2011). Concerning the industrial structure, the main sectors clustered in Nairobi are manufacturing, construction, and financial services, and in 2009 Nairobi hosted 49% of all manufacturing employment and 51% of the total manufacturing wage bill (Kenya National Bureau of Statistics 2011, 2013).[24] And lastly, in Uganda, Kampala hosted 45% of all formal businesses establishments in 2001 and 2006, followed by Mbarara and Wakiso as the second

---

[21] For instance, around the time of the EAC's establishment in 2002, Dar es Salaam hosted only 7% of the country's mainland population but contributed to over 40% of the total wage bill and hosted 57% of total employment in the private sector (Tanzania National Bureau of Statistics 2006, 2007).

[22] Dar es Salaam is also a hub for large firms, hosting over 33% of all firms sized over 100 employees and 13 out of the 44 firms over 500 employees. The second largest region Pwani, which encloses Dar es Salaam geographically, hosts a mere 7% of such (Tanzania National Bureau of Statistics 2018).

[23] Earnings in informal sector and rural small scale agriculture as well as pastoralists activities are excluded (Kenya National Bureau of Statistics 2011; 236).

[24] The respective figures for construction and financial services are 75% and 64% (Kenya National Bureau of Statistics 2011).

largest industrial cities with a share of 5% each (Uganda Bureau of Statistics 2003, 2007). If one includes the "Central" region of Uganda which encloses Kampala geographically, the figure increases to 63% in 2001 and 65% in 2006.[25] Similar to Nairobi and Dar Es Salaam, Kampala contains the majority of the high value-added manufacturing sector with 42% of all firms operating in Kampala and 61% together with the central region in 2006. As such, the Kamapala region contributed to 47% of value added in 2006 and over 77% when including the central district (Uganda Bureau of Statistics 2006).[26] Similar to Nairobi and Dar es Salaam, Kampala only makes up 0.09% of the total land area and 5% (4%) of the population in 2002 (2014) (Uganda Bureau of Statistics 2016).[27]

III. A Four Region Economy

To lay the theoretical groundwork on which to analyze the exposure of regional market integration in the EAC across space, this chapter builds a canonical, four-region quantitative spatial equilibrium model, which combines aspects from the models discussed in the previous chapter. The model is built on Krugman's (1991) core fundamentals while adding an external economy as introduced by Krugman and Elizondo (1996). However, rather than the 2+1 cases in which the external economy acts as one region (e.g. Paluzie 2001), I extend the foreign economy to two regions as in Monfort and Nicolini (2000) and Zeng and Zhao (2010). Finally, the model is rendered unique as I tweak the structure of intra-national transport costs borrowing from the 2+1 models of Crozet and Koenig (2004) and Brülhart et al. (2012) such that the regions within the two countries are outlined by differential access to foreign markets, i.e. the model encapsulates heterogenous intra-national space. The model thereby allows to additionally analyze the potential implications of *foreign* economic (in)equality on the *domestic* distribution of activity, particularly in the context of increasing regional integration. I refrain from computationally more involved multi-region approaches, as the 2+2 case nicely encapsulates the stylized facts of the EAC in terms of its spatial layout and keeps the model tractable.

As such, we have a four-region world economy consisting of R locations denoted by $r = \{1, 2, 3, 4\}$. We define regions 1 and 2 to be in the "home" country and refer to regions 3 and 4 as situated in the "foreign country". Note that most of the analysis conducted in the subsequent chapter refers to effects on regions 1 and 2. However, by symmetry, this readily translates into a view from

---

25 Establishments with 5 employees or more. If one includes informal businesses, Kampala has contained 30% and 29% of all businesses in 2001 and 2011 and 60% and 59% when including the central region, respectively (Uganda Bureau of Statistics 2003, 2012).

26 Kampala also hosts the majority of large firms with 40% of firms with 100 employees or more in 2006 (Uganda Bureau of Statistics 2007). The central region also had the largest increase in manufacturing businesses, with a 40% increase between 2001 to 2006.

27 Together with the Central region, this increases to 20%.

the other regions, i.e. from the "foreign" country, also. Moving on with the model, there are two sectors in the economy, manufacturing, and agriculture. The latter sector is characterized by perfect competition and produces the homogenous agricultural good "food" under constant returns to scale using the immobile, inelastically supplied input "farmers". The modern manufacturing sector is characterized by monopolistic competition and thereby produces a variety of differentiated goods, "manufactures", using the input factor "workers". Farmers and workers within each country are drawn from a total population mass L of which $L^M = L \cdot \delta$ are engaged in manufacturing and the rest $L^F = (1 - \delta)L$ in agriculture, hence $0 < \delta < 1$. Manufacturing workers are mobile between regions but not across sectors or countries, i.e. only mobile between regions 1 and 2 or 3 and 4, respectively. As such, the total manufacturing workforce within countries is fixed, but workers allocate themselves endogenously across regions over time in response to real wage differentials. The respective shares of manufacturing of each region are given by λ_r, which satisfies $\sum_{i=r}^{R} \lambda_r = 1$. We make the simplification that $(\lambda_1 + \lambda_2) = (\lambda_3 + \lambda_4)$, such that the total manufacturing workforce of the two countries is equal, albeit with the potential to be unequally distributed within. The distribution of the immobile agricultural farmers is exogenously fixed and spread evenly across all regions such that their respective shares across regions are given by $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 0.25$. To ease notification, we set the total population mass of the economy L to $L = 2$ and assume countries to be of equal size, i.e. $L_H = L_F = 1$. As we will see later, this allows us to express the share of manufacturing workforce for each region in a country by a λ which is between zero and one. This facilitates the interpretation of λ as a measure of the relative economic disparity within a country and eases interpretation down the line.

a) Consumer Preferences & Behavior

As in classical NEG models, a consumer decides how to spend her income Y with a preference assumed to be of Cobb-Douglas type. In fact, all consumers have a preference representation of Cobb-Douglas which combines a utility derived from the consumption of the agricultural good, F, as well as a Dixit-Stiglitz (Dixit and Stiglitz 1977) Constant-Elasticity-of-Substitution (CES) sub-utility for manufactures, M:

$$U = F^{1-\delta} \cdot M^{\delta} \tag{1}$$

$$M = \left[\sum_{i=1}^{n} c_i^{\rho} \right]^{\frac{1}{\rho}} \tag{2}$$

$$with \quad 0 < \delta < 1 \qquad and \qquad 0 < \rho < 1$$

Whereby δ denotes the share of income spent on consumption of the manufacturing variety such that the share of income not spent on manufactures $(1 - \delta)$ is spent on the consumption of food. c_i specifies the level of consumption of manufacturing variety i of a total of n varieties, among which the consumer chooses with elasticity ρ. ρ is chosen to be constrained between 0 and 1 such that varieties are substitutable but not perfect substitutes. Often ρ is set to $\varepsilon = \frac{1}{1-\rho}$ such that epsilon represents the elasticity of substitution. From (2) it is immediate that M is increasing more strongly in n than in c which reflects the well-known "love of variety" property, the strength of which regulated by ε. The consumer problem is then given by maximizing utility U subject to the budget constraint which is given by income Y from working either in agriculture or manufacturing:

$$Y = p^F F + \sum_{i=1}^{n} p_i c_i \tag{3}$$

Solving the consumer problem thereby involves first finding an optimal allocation of income Y on F and M, and then, maximizing the sub-utility derived from consumption of the composite index M subject to the budget constraint for such manufacturing varieties which follows from the first optimization problem. Hence, our first optimization problem is given by:

$$max\, U = F^{1-\delta} \cdot M^{\delta}$$

$$s.t.\ Y = p^F F + \sum_{i=1}^{n} p_i c_i$$

Some algebra leads to the well-known result that consumers spend share δ of income Y on manufactures, and $(1 - \delta)Y$ on food:

$$p^F F = (1 - \delta)Y \tag{4}$$

$$\sum_{i=1}^{n} p_i c_i = \delta Y \tag{5}$$

The next step involves finding the optimal spending *among* manufacturing varieties n, which is encapsulated by the following optimization problem:

$$\max M = \left[\sum_{i=1}^{n} c_i^{\rho}\right]^{\frac{1}{\rho}}$$

$$s.t. \sum_{i=1}^{n} p_i c_i = \delta Y$$

Taking the ratio of first order conditions for a pair of varieties, the maximization problem yields the equality of marginal rates of substitution to price ratios:

$$\frac{c_i^{p-1}}{c_j^{p-1}} = \frac{p_i}{p_j}$$

$$\text{or } c_i = p_i^{-\varepsilon} \cdot p_j^{\varepsilon} c_j \tag{6}$$

Once we substitute this result into the budget constraint for manufactures (5) we get:

$$\sum_{i=1}^{n} p_i c_i = \sum_{i=1}^{n} p_i \cdot \left(p_i^{-\varepsilon} \cdot p_j^{\varepsilon} c_j\right) = p_j^{\varepsilon} c_j \cdot \sum_{i=1}^{n} p_i^{1-\varepsilon} = c_j = p_j^{-\varepsilon} \cdot I^{\varepsilon-1} \cdot \delta Y \tag{7}$$

$$using\ I \equiv \left[\sum_{i=1}^{n} p_i^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \tag{8}$$

Hence, the expenditure needed to attain M is:

$$M = \left[\sum_{j=1}^{n} c_j^{\rho}\right]^{\frac{1}{\rho}} = \left[\sum_{j=1}^{n} (p_j^{-\varepsilon} \cdot I^{\varepsilon-1} \cdot \delta Y)^{p}\right]^{\frac{1}{\rho}} = I^{\varepsilon-1} \cdot \delta Y \left[\sum_{j=1}^{n} \left(p_j^{-\varepsilon p}\right)\right]^{\frac{1}{\rho}},$$

$$M = I^{\varepsilon-1} \cdot \delta Y \cdot I^{-\varepsilon} \tag{9}$$

Where we made use of that $-\varepsilon p = 1 - \varepsilon$, and $\frac{1}{p} = \frac{-\varepsilon}{1-\varepsilon}$, given that $\varepsilon = \frac{1}{1-p}$. Given that I multiplied by the quantity composite manufacturing consumption M is equal to expenditure δY, I is also known as the price index, which measures the minimum cost of purchasing manufacturing goods bundle M. Consumer demand functions are thereby:

$$F = \frac{(1-\delta)Y}{p^F} \tag{10}$$

$$M = \frac{\delta Y}{I} \tag{11}$$

Plugging these utility-maximizing consumption levels of F and M into (1) leads to the indirect utility function:

$$U = \delta^{\delta}(1-\delta)^{1-\delta} \cdot Y \cdot I^{-\delta}(p^F)^{-(1-\delta)} \tag{12}$$

Hence, the maximum attainable welfare is a function of the income Y weighted by the cost of living as given by price indices I and p^F together with their relative consumption shares δ and $1 - \delta$.

b) Transport Costs & Heterogenous intra-national Space

All manufacturing varieties can be consumed in each home or foreign location. However, evidently, a variety locally consumed but not produced needs to be imported, which entails transport costs. As is standard in NEG models, these transport costs are encapsulated by the *Samuelson-Von Thünen* iceberg-type, which envisions only a fraction of the goods to arrive at a destination, i.e. goods "melting" in transit (von Thünen 1826; Samuelson 1952). Thereby, a producer located in region 1 has to dispatch an additional amount together with the demanded amount, summing to T, for $1/T$ to arrive at the destination. For instance, if 20% of the dispatched goods regularly melt away en-route between regions i and j, iceberg transport costs are given by $T_{ij} = 1.25$. In other words, for one-unit of a good produced in region i to arrive at region j, suppliers located in region 1 have to dispatch 1.25 units of the good. Note at this point that we assume food to be transported costlessly across all national and international regions.

As anticipated above, the present model is outlined by heterogenous intra-national space, which is operationalized by specific transport cost structure. The reason for this adjustment is, of course, added realism on the one hand, but more importantly, because the spatial layout of the EAC as anticipated in Chapter II lends itself naturally to this modification. More precisely, note that all three urban centers (Nairobi in Kenya, Dar es Salaam in Tanzania and Kampala in Uganda) are geographically tucked away from the common borders connecting the respective EAC partner state(s). In the data, the average road distance to EAC border crossings for the three cities is 395km, 922km, and 269km, respectively.[28] The travel time over road is particularly relevant for intra-EAC trade, as over 95% of the regional trade in the area is transported via the road network, and only 5% via rail (Nathan Associates 2011). To operationalize this specific spatial layout of intra-EAC trade in the model, I assume that among the two regions within each country, one of the regions has better access to the foreign market, i.e. is a "border" or "gated" region (Behrens et al. 2006). As such, shipping goods from a non-border region to a foreign location means transiting through this region, i.e. higher trade costs. This effectively places the four regions on a line with regions 1 and 4 at the end of the spectrum and regions 2 and 3 connecting the two home and two foreign countries. As expected, regions 1 and 4 represent the economic hubs of the countries, i.e. Nairobi, Dar es Salaam and Kampala, and are denoted as "interior" or "core" regions. As in Brülhart et al. (2012), I formalize this transport structure by simply accumulating all transport costs which accrue throughout the transit, i.e. multiply all types iceberg transport costs T which lie between the origin and the destination region. For instance, for region 1, which is an interior regions, sending (importing) goods to (from) regions 2, 3, and 4 entails total iceberg transport costs of $T_{12} = T_{12}$, $T_{13} = T_{12} \cdot T_{23}$, and $T_{14} = T_{12} \cdot T_{23} \cdot T_{34}$. I additionally assume that intra-national transport costs in the home and foreign country are identical and that transport costs are symmetric, such that $T_{12} = T_{34} = T_{D(omestic)}$, $T_{23} = T_{32} = T_{F(oreign)}$ and $T_{ij} = T_{ji}$. Finally, note that transport costs are zero when consuming a variety produced within the same region, i.e. $T_{ij} = 1$, for all $i = j$. As such, the transport costs of trading goods between the four sending regions $i = \{1, 2, 3, 4\}$ and arrival regions $j = \{1, 2, 3, 4\}$ can be summarized by the following five types of total trade costs across regions.

$$1 = T_{ij} \begin{cases} for\ i = i\ and\ j = i \\ for\ i = j\ and\ j = j \end{cases}$$

$$T_D = T_{ij} \begin{cases} for\ i = 1\ and\ j = 2 \\ for\ i = 3\ and\ j = 4 \end{cases}$$

$$T_F = T_{ij} \begin{cases} for\ i = 2\ and\ j = 3 \\ for\ i = 3\ and\ j = 2 \end{cases}$$

$$T_{DF} \equiv T_D \cdot T_F = T_{ij} \begin{cases} for\ i = 1\ and\ j = 3 \\ for\ i = 3\ and\ j = 1 \end{cases}$$

[28] The minimum distances to the nearest EAC border crossings are 152km, 389km and 185km, respectively.

$$T_{DFD} \equiv T_D \cdot T_F \cdot T_D = T_{ij} \begin{cases} for\ i = 1\ and\ j = 4 \\ for\ i = 4\ and\ j = 1 \end{cases}$$

Figure 2 depicts this spatial cost structure for the 2+2 model illustratively using Uganda and Kenya as a stylized example. The dashed line in the countries depicts the main trade route between the countries called the "northern corridor" (Nathan Associates 2011).[29] The vertically dotted line illustrates the border.

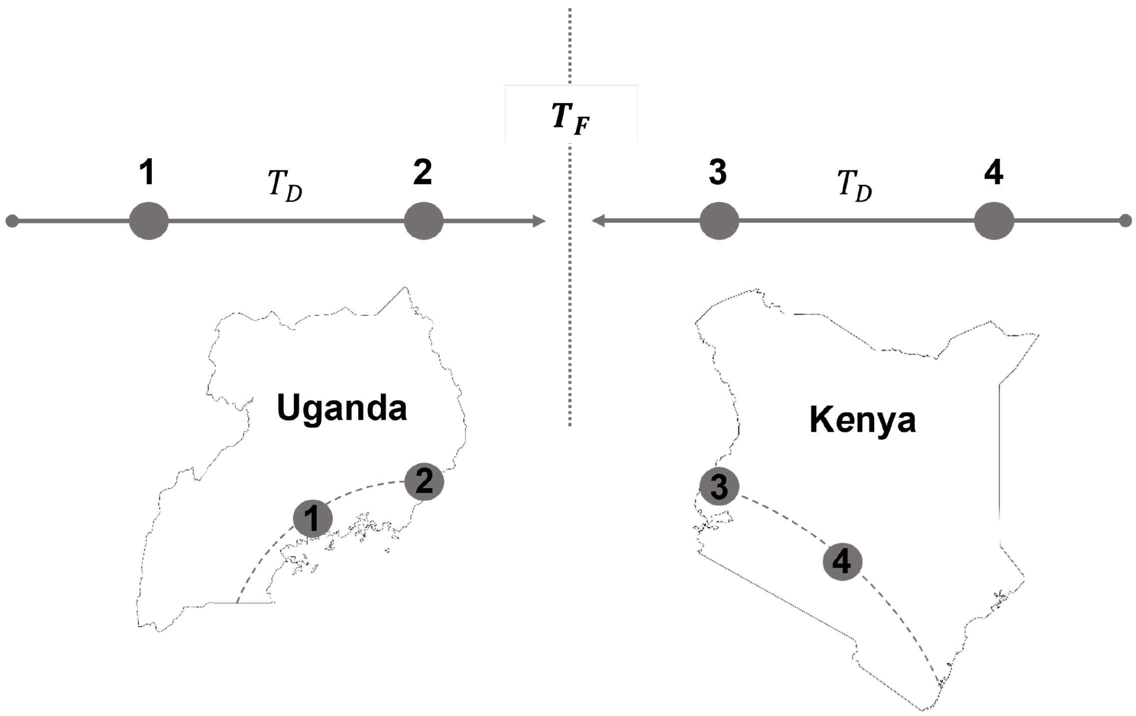


Figure 2: Transport Cost Structure in the four-region Economy

As is depicted in Figure 2, regions 1 and 4 represent the economic hubs Kampala and Nairobi, respectively, with 2 and 3 posing as the "border" regions.[30] Note that Tanzania borders the depicted countries to the south, respectively, and given the position of Dar es Salaam, creates a similar spatial pattern.[31]As such, this transport cost structure is assumed to be symmetric and thereby extends to the two other trade pairs, Tanzania-Uganda and Tanzania-Kenya analogously. Granted that this a simplification of the spatial realities on the ground, including varying absolute and relative distances, differing processing times etc., this transport cost structure is nonetheless useful because it easily lets

[29] Note that the figure is not drawn up to scale and serves as a stylized model of the spatial trade structure, only (see Appendix A2.1 for a more accurate depiction of the geography as well as a depiction of the "central corridor" which connects the countries via Tanzania).
[30] Malaba is the main border-crossing connecting these countries (Nathan Associates 2011).
[31] See detailed maps of these routes, i.e. the "northern corridor" as well as the "central corridor" connecting the larger region in Figure A.2 in the Appendix from Nathan Associates (2011).

us operationalize the comparative statics of a change in regional market integration and the subsequent effect on (pre-existing) regional disparities by solely altering the costs of moving goods between regions 2 and 3, i.e. by altering T_F.

Carrying on with the model, these transport costs imply that the delivered price is T_{ij} higher than the f.o.b. price.[32] A standard assumption I follow is that all transport costs are incurred by consumers such that the total cost of consuming one-unit of variety produced in i in region j increases to $p_i = p_i T_{ij}$. Note that given (7), the demand for a variety produced in region i, consumed in location j is now given by:

$$c_j = (p_i T_{ij})^{-\varepsilon} \cdot I_j^{\varepsilon-1} \cdot \delta Y_j \tag{13}$$

Note that this necessitates the simplifying assumption that one manufacturing variety is produced at one location only, which follows from internal economies of scale, and also, that all varieties n produced in this respective location are produced using the same technology and, therefore, price. The total price index I of region j is then given by:

$$I_j = \left[\sum_{i=1}^{R} n_i \cdot \left(p_i T_{ij}\right)^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}} \tag{14}$$

To arrive at the total sales of a given variety i, we sum demand for this variety over all regions R using (13), and note that the supply incurs shipping T_{ij} units of i. Hence we arrive at:

$$q_i = \delta \sum_{j=1}^{R} Y_j \cdot (p_i T_{ij})^{-\varepsilon} \cdot I_j^{\varepsilon-1} \cdot T_{ij} \tag{15}$$

This encapsulates that total demand of a variety q_i is decreasing in the price of the good p_i and the transport cost incurred T_{ij} for the respective importing region. Demand is increasing in income Y_j and price index I_j of regions as well as in the share spent on manufactures δ.

---

32 The "mill" or "f.o.b.", free on board, price, is the price charged at the "mill", the production location, not incurring shipping costs.

c) Producer Behavior

As defined previously, food is produced under constant returns to scale as well as under the assumption of perfect competition. Given that we have just assumed food to be traded costlessly across all regions, the price of food is equal everywhere and so is the wage given that farmers are paid their marginal product. We then set the technology coefficient of food production to 1 such that $w^F = p^F = 1$ and the agricultural good acts as the *numeraire* throughout the analysis. In the manufacturing sector, production technology is of increasing returns to scale. It thereby involves a fixed cost of production F and marginal costs per unit c. Given that labor is our only input factor, the production of a quantity q of a variety i produced in location i is given by labor input requirement:

$$l = F + cq \tag{16}$$

and this is assumed to be the same technology for all varieties. Given increasing returns to scale, consumer preference for variety, firms will choose to produce a variety, not produced by any other firm such that a variety is produced only in one location by one firm.[33] This has the result that the number of available varieties is equal to the number of firms. The profit of a specific firm producing at location i with a given wage rate w_i, and an f.o.b. price p_i is:

$$\pi_i = p_i q_i - w_i(F + cq_i) \tag{17}$$

Making the simplification $q = Bp_i{}^{-\varepsilon}$ (see Brakman et al. 2020) and differentiating w.r.t. price and setting equal to zero leads to the $f.o.c.$:

$$(1-\varepsilon)Bp_i{}^{1-\varepsilon} + \varepsilon w_i \cdot cBp_i{}^{-\varepsilon-1} = 0 \tag{18}$$

Rearranging leads us to the well-known result that prices are a combination of f.o.b. price, which are given by marginal costs $w_i c$, and a mark-up, determined by the elasticity of substitution ε:

$$p_i\left(1 - \frac{1}{\varepsilon}\right) = w_i c, \text{ or}$$

33 Where an additional assumption is that the number of varieties goes to infinity,

$$p_i = \frac{cw_i}{\rho} \tag{19}$$

Given that we assume free entry and exit, profits are driven to zero. Using the new pricing rule (19) in the profit function (17) and setting to zero leads:

$$\pi_i = \frac{cw_i}{\varepsilon - 1}(q_i - \frac{F(\varepsilon - 1)}{c}) \tag{20}$$

Hence, equilibrium output by any active firm i is the constant:

$$q^* = \frac{F}{c}(\varepsilon - 1) \tag{21}$$

And the required labor input producing this amount is then given by plugging (21) into the production technology used (16):

$$l^* = F + c\left(\frac{F}{c}(\varepsilon - 1)\right), or$$

$$l^* = F\varepsilon \tag{22}$$

Which carries the result that the number of varieties n produced in a location i, and thereby the number of manufacturing firms, is directly proportional to the manufacturing population at this location, $\lambda_i \delta L$:

$$n_i = \frac{\lambda_i \delta L}{F\varepsilon} \tag{23}$$

d) Short-run Equilibrium

In equilibrium, output of firms must match demand by consumers. Using (14) we have:

$$q_i^* = \delta \sum_{j=1}^{R} Y_j \cdot p_i^{-\varepsilon} \cdot {T_{ij}}^{1-\varepsilon} \cdot {I_j}^{\varepsilon-1} \tag{24}$$

In other words, firms break even if the price they charge equals:

$$p_i^{\varepsilon} = \frac{\delta}{q_i^*} \sum_{j=1}^{R} Y_j \cdot T_{ij}{}^{1-\varepsilon} \cdot I_j{}^{\varepsilon-1} \tag{25}$$

Plugging in the pricing rule (19), leads to the well-known wage equation:

$$w_i = \left(\frac{\varepsilon - 1}{\varepsilon c}\right)\left(\frac{\delta}{q_i^*} \sum_{j=1}^{R} Y_j \cdot T_{ij}{}^{1-\varepsilon} \cdot I_j{}^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{26}$$

To arrive at real wages, ω, we simply have to divide nominal wages (26) by the cost of living, which is a combination of the manufacturing price index of the region (14) and food prices:

$$\omega_i = w_i \cdot I^{-\varepsilon} \cdot (p^F)^{-(1-\delta)} \tag{27}$$

It is convenient to use some normalizations to simplify analysis (Fujita et al. 2001). Hence, we redefined the marginal labor requirement is:

$$c = \frac{\varepsilon - 1}{\varepsilon} = \rho \tag{28}$$

Then, (19) turns to:

$$p_i = w_i \tag{29}$$

Also, we set a unit of measurement for the number of firms n, such that the fixed input requirement F is given by:

$$F = \frac{\delta}{\varepsilon} \tag{30}$$

Remember that the number of firms in each location is directly proportional to the manufacturing labor force in this location $\lambda_i \delta L$, such that (23) reduces to:

$$n_i = \frac{\lambda_i \delta L}{F\varepsilon} = \lambda_i L \tag{31}$$

From this, the price index (14) as well as the wage equation can be simply expressed as:

$$I_j = \left[\sum_{i=1}^{R} \lambda_i L \left(w_i \cdot T_{ij}\right)^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}} \tag{32}$$

$$w_i = \left(\sum_{j=1}^{R} Y_j \cdot \, T_{ij}{}^{1-\varepsilon} \cdot I_j{}^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{33}$$

These constitute the first two of three equations that characterize the short-run equilibrium. What is missing is the income-determining equation, which is easily defined by the sum of wage income from manufacturing workers in the region $\lambda_i \delta L$ as well as from farm workers $\phi_i(1-\delta)L$.[34] Hence, the income of a region i is given $Y_i = \lambda_i \cdot w_i \cdot \delta L + \phi_i(1-\delta)L$. Taking into account our initial simplifications, namely that the manufacturing workforce is immobile across countries and exogenously set to $\phi = 0.25$, that the distributions of the manufacturing workforce is given by $\sum_{i=r}^{4} \lambda_r = 1$, that the total mass of population is set to $L = 2$ and that the two countries are of equal size lets us write the income equation in our four region case as:

$$Y_i = \lambda_i \cdot w_i \cdot \delta + \frac{(1-\delta)}{2}, \quad 0 \leq \lambda_i \leq 1 \tag{34}$$

Where we additional use of our assumption $\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4$ which enables us to set $2\lambda_i = \lambda_i$. And similarly, the price index as simplifies to:

$$I_j = \left[\sum_{i=1}^{R} \lambda_i \left(w_i \cdot T_{ij}\right)^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}}, \quad 0 \leq \lambda_i \leq 1 \tag{35}$$

[34] Note that given constant returns to scale and perfect competition, the wages for agricultural labor are equal everywhere is set as the numeraire.

Given that manufactures can be traded across all regions, and our economy is made up of four regions in total, the short-run equilibrium relationship is expressed by 12 equations (3 for each region) given in (36) through (47):

$$Y_1 = \lambda_1 \cdot w_1 \cdot \delta + \frac{(1-\delta)}{2} \tag{36}$$

$$Y_2 = \lambda_2 \cdot w_2 \cdot \delta + \frac{(1-\delta)}{2} \tag{37}$$

$$Y_3 = \lambda_3 \cdot w_3 \cdot \delta + \frac{(1-\delta)}{2} \tag{38}$$

$$Y_4 = \lambda_4 \cdot w_4 \cdot \delta + \frac{(1-\delta)}{2} \tag{39}$$

$$I_1 = \left[\lambda_1 \cdot {w_1}^{1-\varepsilon} + \lambda_2 (w_2 \cdot T_D)^{(1-\varepsilon)} + \lambda_3 (w_3 \cdot T_{DF})^{(1-\varepsilon)} + \lambda_4 (w_4 \cdot T_{DFD})^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}} \tag{40}$$

$$I_2 = \left[\, \lambda_1 (w_1 \cdot T_D)^{(1-\varepsilon)} + \lambda_2 \cdot {w_2}^{1-\varepsilon} + \lambda_3 (w_3 \cdot T_F)^{(1-\varepsilon)} \lambda_4 + (w_4 \cdot T_{DF})^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}} \tag{41}$$

$$I_3 = \left[\, \lambda_1 (w_1 \cdot T_{FD})^{(1-\varepsilon)} + \lambda_2 (w_2 \cdot T_F)^{(1-\varepsilon)} + \lambda_3 \cdot {w_3}^{1-\varepsilon} + \lambda_4 (w_4 \cdot T_D)^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}} \tag{42}$$

$$I_4 = \left[\, \lambda_1 (w_1 \cdot T_{DFD})^{(1-\varepsilon)} + \lambda_2 (w_2 \cdot T_{DF})^{(1-\varepsilon)} + \lambda_3 (w_3 \cdot T_D)^{(1-\varepsilon)} + \lambda_4 \cdot w_4^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \tag{43}$$

$$w_1 = \left(Y_1 \cdot {I_1}^{\varepsilon-1} + Y_2 \cdot {T_D}^{1-\varepsilon} \cdot I_2^{\varepsilon-1} + Y_3 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_3^{\varepsilon-1} + Y_4 \cdot {T_{DFD}}^{1-\varepsilon} \cdot I_4^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{44}$$

$$w_2 = \left(Y_1 \cdot {T_D}^{1-\varepsilon} \cdot I_1^{\varepsilon-1} + Y_2 \cdot I_2^{\varepsilon-1} + Y_3 \cdot {T_F}^{1-\varepsilon} \cdot I_3^{\varepsilon-1} + Y_4 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_4^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{45}$$

$$w_3 = \left(Y_1 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_1^{\varepsilon-1} + Y_2 \cdot {T_F}^{1-\varepsilon} \cdot I_2^{\varepsilon-1} + Y_3 \cdot I_3^{\varepsilon-1} + Y_4 \cdot {T_D}^{1-\varepsilon} \cdot I_4^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{46}$$

$$w_4 = \left(Y_1 \cdot {T_{DFD}}^{1-\varepsilon} \cdot I_1^{\varepsilon-1} + Y_2 \cdot {T_{DF}}^{1-\varepsilon} \cdot I_2^{\varepsilon-1} + Y_3 \cdot {T_D}^{1-\varepsilon} \cdot I_3^{\varepsilon-1} + Y_4 \cdot {I_4}^{\varepsilon-1}\right)^{\frac{1}{\varepsilon}} \tag{47}$$

These 12 equilibrium conditions formalize the notion of centripetal (demand and cost linkages) as well as centrifugal forces (competition) anticipated in Chapter II. Take first, the price index given in equations (40) through (43). Consumer prices at one particular location can be seen as a weighted average of all source location sizes (λ) and prizes (which, given (19) are directly proportional to the wage rate w) with weights given by the distance to these exporting locations (T), respectively.[35] As such, the price index is lower in those regions, where a higher share of demand is sourced from large (high λ), low wage (low w) and importantly, nearby locations (low T); and of course, most cheaply sourced locally, i.e. when $T = 1$. In other words, locations with large shares of own or close by manufacturing employment have lower price indices given that a smaller share of the total consumption needs to be imported; this is the "price index effect" analytically derived in Fujita et al. (2001). These dynamics describe the cost (forward) linkage described in Chapter II, whereby a larger home market provides lower consumer prices. As such, moving to a region, i.e. making it larger, thereby displays a self-reinforcing centripetal force.

The wage equations given in (44) through (47) can be interpreted similarly. In essence, wages are higher in regions where income Y, and thereby expenditure, is high or in regions where these larger markets are more proximate (low T). Put simply, firms are able to pay higher wages if they have better access to large markets. This describes the demand (backward) linkage anticipated before and indicates that a larger number of workers, and thereby, consumers, increase the local demand which increases the wage firms are able to pay. Similarly, as for the cost linkage, this attracts more workers to this region, and also firms, thereby acting as a self-reinforcing centripetal force. This is described by the "home market effect" (for the full derivation, see Fujita et al. 2001). Importantly, the wage equation also encapsulates a centrifugal force which is given by its positive dependence on the price index I. As just established, the price index is lower in larger regions, i.e. those with a higher number of manufacturing varieties. And given that the number of manufactures is regulated not by output per firm, but by the number of firms, a lower price index automatically indicates a larger number of competing firms, which exerts a downward pressure on the wages a firm is able to pay.[36] As a results, firms may seek to relocate in order to shelter from competition allowing them to pay higher wages, which may also draw workers.

In the end, the relative strength of these centrifugal and centripetal forces can be handily manifested in real wage differentials across regions, which combine the effects on nominal wages and prices. Formally, the real wages ω of regions are given by dividing the total wage income w by the consumer price index of both manufactures I and food F together with their relative consumption

[35] Consumer prices at one particular location can be seen as a weighted average of all source locations and their prizes (which, given (19) are directly proportional to the wage rate w) as well as distance T with weights given by the relative size of these locations λ.

[36] This also be validated in (8) or (15), where demand of an individual firm is inversely related to the price index.

shares δ, hence $\omega_i = w_i \cdot I^{-\delta}(p^F)^{-(1-\delta)}$. Note that we are able to dismiss the component of the agricultural good, as it is set as the numeraire. Real wages of all four regions are then expressed by:

$$\omega_1 = w_1 \cdot I^{-\delta} \tag{48}$$

$$\omega_2 = w_2 \cdot I^{-\delta} \tag{49}$$

$$\omega_3 = w_3 \cdot I^{-\delta} \tag{50}$$

$$\omega_4 = w_4 \cdot I^{-\delta} \tag{51}$$

Where the values of the right-hand side are given by the simultaneous solution to the 12 short run equilibrium conditions (36) and (47). In the long run, we assume that workers respond to the real wage differential across regions by migrating such that the share of manufacturing workers within the two home and foreign economies, λ_1 and λ_2, as well as λ_3 and λ_4, are endogenously determined. I assume workers to move between regions with the following dynamics:

$$\frac{d\lambda}{dt} = \gamma \begin{cases} \dfrac{\omega_i}{\omega_j} - 1 & if\ 0 < \lambda < 1 \\ min\left\{0, \dfrac{\omega_i}{\omega_j} - 1\right\} & if\ \lambda = 1 \\ max\left\{0, \dfrac{\omega_i}{\omega_j} - 1\right\} & if\ \lambda = 0 \end{cases} \tag{52}$$

Hence, for a given real wage differential and spatial configuration λ, workers move between regions across regions with a particular speed γ. We now have all the ingredients we need to define a *long-run* equilibrium. By (52) the first type of long-run equilibrium can be described by a spatial configuration for which real wages across regions are equalized, i.e. a situation in which workers have no incentive to move. Formally, this is given by a $\lambda \epsilon [0,1]$ for which $\omega_i/\omega_j = 1$. One specific case of such is the equal spreading of workers, i.e. for our four-region model $\lambda_{1,2} = \lambda_{3,4} = 0.5$ and $\omega_{1,3}/\omega_{2,4} = 1$. This is also called the "symmetric" or spreading" equilibrium. The model also admits a second type of a long-run equilibrium, one in which real wages are not equalized. In these cases, all of the manufacturing workforce is agglomerated in one of the regions, which represents a corner solution. Formally, such an equilibrium is given by $\lambda_{1,4} = 1$ and $\lambda_{2,3} = 0$, and often referred to as an "agglomerated" or "core-periphery" equilibrium. To complete the discussion on long-run equilibria, one important distinction to make is whether such an equilibrium is also a *stable* one. In general, the stability of an equilibria depends on whether a small perturbation in the manufacturing workforce at this spatial configuration triggers dynamics which reinstates the just left allocation of workers or not. For the first type of equilibria, the stability is thereby defined by a second condition which is that the derivative of the real wage differential w.r.t. an infinitesimal change in the manufacturing workforce is smaller or equal to zero, i.e. formally whether $d(\omega_i/\omega_j)/d\lambda_i \leq 0$. Put simply, if migrating from region j to region i increases the real wage differential ω_i/ω_j, then the previous equilibrium was not a stable one. In the second type of equilibria, the stability condition entails that the real wage differential is skewed in favor of the agglomerated region, such that for $\lambda_1 = 1$, $\frac{\omega_1}{\omega_2} \geq 1$ and for $\lambda_1 = 0$ and $\frac{\omega_1}{\omega_2} \leq 1$. I analyze the stability of these two types of equilibria more thoroughly in Appendix A.1.

Spatial Equilibria and Regional Trade Liberalization

The four-region model and the long-run equilibrium conditions just developed lends itself to the comparative static examination of what happens to the forces inducing agglomeration or dispersion once trade across countries is liberalized. In particular, we can use the solutions to the simultaneous equilibrium conditions (36) through (47) as inputs to compute the real-wage differential which dictates the dynamic process towards a stable long-run equilibrium described in (52). The analysis in this section entails tracking what happens to the real wage differential across regions inside the countries once the costs connecting the two economies T_F are lowered from a former prohibitive level (i.e. autarky) down to levels which mirror those incurred within the respective countries, i.e. $T_F = T_D$.[37] This will effectively allow us to analyze how the process of trade liberalization affects the (stability of) specific long-run equilibrium allocations of workers across regions. Note, however, given that the real wage differential ω_i/ω_j depends on twelve simultaneous non-linear equations, the real wage differential is not a simple function of λ_i. As such, and as is common in the NEG literature, I will analyze the dynamics of the spatial equilibria mainly via numerical simulations. This is most efficiently done by plotting the real wage differential ω_1/ω_2 across the full range of potential manufacturing distributions $\lambda \in [0,1]$ which may be realized at any point in time. To nonetheless provide some analytical insights into the numerical results, Appendix A.1 provides a "sustain" and "break" analysis in the vein of Fujita et al. (2001), which revolves around assessing the stability of the two specific types of equilibria described above, i.e. "agglomeration" and "spreading". Some of the key results of this analysis are discussed in this section as well.

[37] In this scenario, the cost associated with trading goods across borders mirrors those incurred when shipping goods intra-nationally, i.e. $T_F = T_D$.

As a final remark on the approach of this section's analysis, it turns out to be instructive to compare the results of the model to a more general version of it. To be specific, I will conduct the simulations additionally for a four-region model with *homogenous* (or *asymmetric*) intra-national space. This model mirrors the one described by equations (36) through (47), but with a tweak regarding the transport cost structure. This is done by simply setting the three different types of external iceberg trade costs equal such that $T_F = T_{DF} = T_{DFD}$, and the two home regions have identical international trade costs to both foreign regions. In the vein of Figure 2, one can think of regions 1 and 2 as well as 3 and 4 in this adjusted model as placed on a line parallel, rather than perpendicular to the border with roads diagonally connecting the home and foreign regions, respectively. Note that this model thereby reduces to the one studied in Monfort and Nicolini (2000), and their conclusions apply analogously. However, comparing the results from our model to this one provides an intuitive reference and helps in evaluating the role *heterogenous* (or a*symmetric*) intra-national space and thereby unequal access to the newly integrated foreign markets plays, which is ultimately how trade plays out in the East African Community as we have established previously.

Figure 3 initiates our analysis and plots the real wage differential between regions 1 and 2 across the full range of possible manufacturing distributions $\lambda \in [0,1]$ as well as for three levels of international trade costs T_F, respectively. Note that solving for this set necessitates a choice on the exogenous parameter values given by δ and ε, and the intra-national trade costs T_D. I use values commonly employed in the literature which are given below the figures. Table A.1 in the Appendix additionally provides a sensibility test for a wider range of values and the main interpretations remain. Notice that although our model entails four endogenous parameters, λ_1, λ_2, λ_3, and λ_4, the plots in Figure 3 only depict two at a time, i.e. is two-dimensional. This is done by setting $\lambda_3 = \lambda_4 = 0.5$, i.e. by assuming an equal distribution of manufacturing in the foreign country. This assumption is relaxed later in Figure 5, when we assess the moderating influence of foreign economic inequality. Note that as anticipated earlier, we focus our view on the home country, i.e. regions 1 and 2, but the results and intuitions apply identically, given symmetry. Figure A2.1 of the Appendix also provide the full three-dimensional plots, which effectively combine the results depicted in Figure 3 and 5.

The numerical simulations depicted in Figure 3 provide the main insights into the process of trade liberalization across a four-region economy, i.e. lowering the intra-national trade costs T_F. It depicts the results for two spatial setups (symmetric and asymmetric intra-national space) for two sets of parameter values. More specifically, Panels A and C represents the symmetric case for values of the elasticity of substitution $\varepsilon = 6$ and $\varepsilon = 4$. And Panels A and D present the results analogously for the asymmetric case.

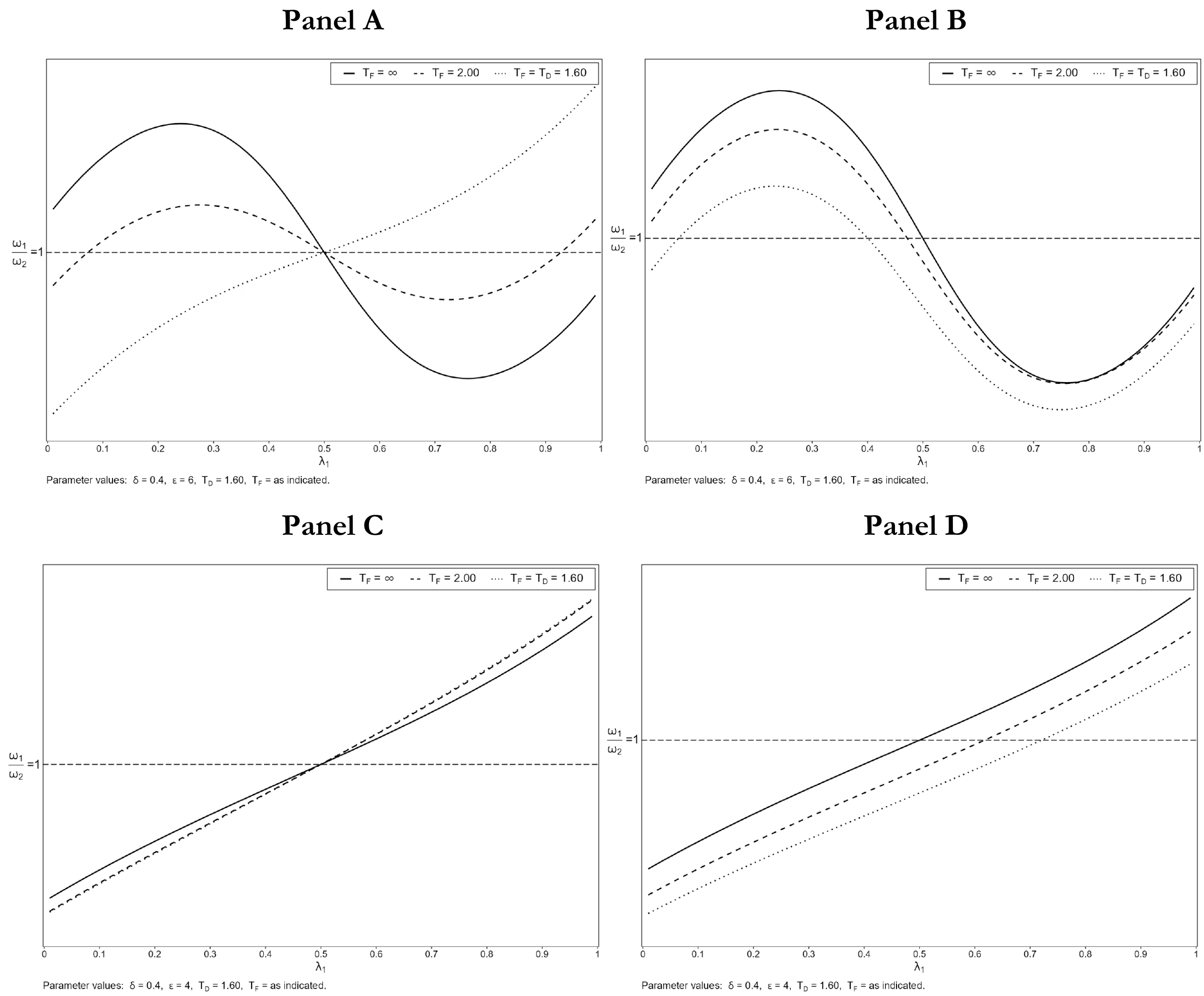


Figure 3: Trade Liberalization and Spatial Equilibria

We focus first on the real wage differentials in autarky, i.e. where international trade costs are prohibitively high $T_F = \infty$, as depicted by the solid line. In the case of low product differentiation ($\varepsilon = 6$), Panels A and B, we notice that there exists a long-run stable symmetric equilibrium where the workforce is equally spread across the two home regions for both models, as can be seen by the negative slope passing-through real wage parity. While this equilibrium also exists for the case where product differentiation is high (Panel C and D), this equilibrium is not stable anymore, as can be depicted by the positive slope through the point where $\lambda_1 = 0.50$. What happens to this type of equilibria in the home country when the external trade costs to regions 3 and 4 are lowered? This is depicted by the new equilibrium real wage differentials given by the dashed ($T_F = 2.00$) and dotted lines ($T_F = T_D = 1.60$).[38] As a first pass through the Panels, and as shown in previous results, lowering the costs to trade with an external market increases agglomerating tendencies, i.e. increases intra-national inequality (e.g. Monfort and Nicolini 2000; Paluzie 2001). This can be seen by a general attenuation of the slopes passing through the symmetric equilibria. Most starkly, in Panel A, the slope concludes a full rotation from negative to positive values from autarky to free trade. Hence, when the countries are liberalized, the former stable equilibrium for equal distribution of manufacturing activity turns out to be instable. As we defined in the previous section, this is so because an infinitesimal small shock (increase) to the manufacturing workers in any direction would also cause a higher real wage skewed towards this region, which would not induce workers to move back to the symmetric equilibrium. As such, once trade is liberalized, the strength of the force holding together the equal spreading, i.e. the costs of serving remote markets, weakens. This may therefore set in motion a cumulative causation for a small increase of consumers in region 1, leading to full agglomeration in region 1, and vice versa, for region 2 if initially moved in the opposite direction. However, this effect on the slope is generally not as pronounced in the model with heterogenous intra-national space. For instance, in Panel B, while the slope is reduced for higher values of trade liberalization, there still exists a stable equilibrium not leading to a full core-periphery pattern as it would in Panel A. Remarkably, this long-run stable equilibrium is brought about at an unequal distribution of the workforce within the home country. That is, we observe a shift of the curve which cuts the constant parity line parallel to the left. This effectively indicates a stable equilibrium at an unequal distribution across the home regions. Hence, there now exists an increased draw to the border, given that λ_1 reduces from 0.50 in autarky to around 0.40 in the free trade scenario, which indicates that now over half of the manufacturing is operating at the border. This is similar to the result provided in (Crozet and Koenig 2004b), albeit in a 2+1 setup.[39]

If we move our view to the results in Panel C and D, this result is further corroborated. In this scenario, the centripetal forces are accentuated as can be seen by positive values of the slope of ω_1/ω_2 throughout. Given that the only amendment is a lower elasticity of substitution ε, it seems that higher product differentiation causes the strength of scale economies to increase. Notice, from (40) through (47) how the strength of the centripetal and centrifugal forces depends on the parameter ε. For one, in the price indices, a lower elasticity of substitution (ε) increases the strength of the love of variety, such that for any increase in the low-cost access of goods (high λ, as well as a low w and

[38] We thereby implicitly assume a change in the ad valorem tariff of crossing international borders down to 25% and 0%.

[39] Note that setting $\lambda_3 = \lambda_4 = 0.5$ is not equal to the case of one region in the foreign country.

T), the price index is lower than for higher values of ε.[40] Intuitively speaking, the higher the differentiation between varieties, the higher the added utility gain of (increased availability of) a further variety to consumers. Hence, lowering ε causes the forward (cost) linkage to intensify. Note however, as established above, that this also automatically also leads to stiffer product market competition among varieties, as I is reduced.[41] In the wage equation, this means that lower values of ε has a negative effect on the wage firms are able to afford, which displaying a centrifugal force. However, as by the exponent of I, this negative pressure is less intensive in environments of high product differentiation, which is intuitively plausible. And secondly, this also means that any increase in market access (high Y and low T) also increases the wages firms are able to afford. Hence, lowering ε also seems to cause the backward (demand) linkage to intensify.

As seen by the comparison between top and bottom Panels, the forward and backward linkages are strengthened, i.e. centripetal forces dominate the centrifugal forces caused by a decrease in ε.[42] As such, a core-periphery pattern is more likely at any level of intra- or international trade costs. This is seen by positive slopes in both Panel C and D. Again, while decreases in T_F causes only minor changes in the slope for Panel C, it significantly alters the equilibrium configuration in the model with heterogenous intra-national space. However, different from the case in B, the cutpoint has now shifted to the right. This is easily reconciled, by interpreting this change as essentially decreasing the basin of attraction which would lead to full agglomeration in region 1. Put simply, in autarky, it suffices to increase the manufacturing workforce in region 1 an infinitesimal amount over 50%, for dynamics to unfold which lead to full agglomeration in 1. In the free trade scenario, over 70% of the manufacturing workforce would have to be in region 1 for this cumulative causation mechanism to kick in. Again, Crozet and Koenig (2004) as well as Brülhart et al. (2004) show qualitatively similar results, albeit for a 2+1 setting.[43] This is the second noteworthy departure from the model with symmetric intra-national space and highlights that relevant different conclusions arise when the access to foreign regions is unequal.

Hence, these results paint two consistent insights. For one, liberalizing trade across the two countries model increases internal agglomeration tendencies. And secondly, that this agglomeration is

---

[40] To validate this, note that the negative exponent of the entire bracket in (40) through (44) gets larger, while the negative exponents of w and T get smaller. From (8) and noting that varieties are produced with the same technology in all locations, which renders the price index as $I = p \cdot n^{1/1-\varepsilon}$. It is easily seen that I is more strongly decreasing in n (varieties) for lower values of ε.

[41] This can also be confirmed in (7), i.e. $c_j = p_j^{-\varepsilon} \cdot I^{\varepsilon-1} \cdot \delta Y$. As established in footnote 40, increasing n decreases I which thereby lowers the demand for any variety. This is also seen in (21) and (23), whereby a decrease in ε causes equilibrium output per firm q^* to decrease with an accompanied increase in varieties n at each location.

[42] Table A.1. shows that this is the case for all tested parameter configurations.

[43] Brülhart et al. (2004) additionally departs from CES and uses a quasilinear consumer utility.

more likely to occur in the region bordering the newly accessed markets. For additional insights into these dynamics, Figure 4 reproduces part of the analytics carried out in Appendix A.1. As such, it plots the results of the "sustain" analysis, which essentially evaluates the stability of the agglomerated equilibrium, i.e. depicts the range of intra-national transport costs T_D for which the agglomerated equilibrium in region 1 proves sustainable. Remember that the stability condition of this equilibrium at $\lambda_1 = 1$ requires $\omega_1/\omega_2 \geq 1$. Given that we have seen an increased draw to the border (region 2), we are interested for which range of values a sustainable agglomeration in the interior (region 1) can be upheld. Again, Panels A and C provides the case for the symmetric case, while Panels B and D conducts the analysis for our main model again for the three levels of international transport costs T_F.

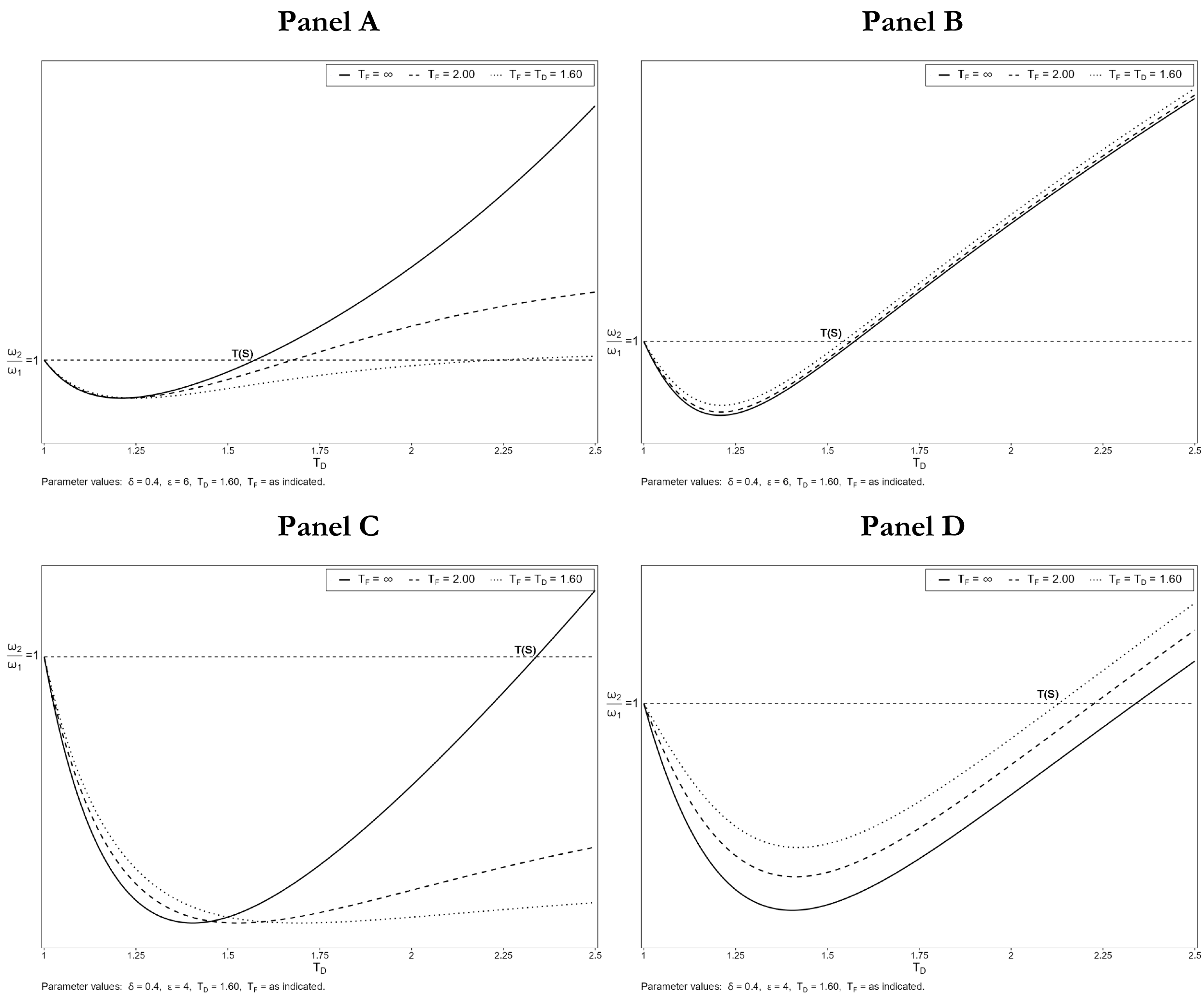


Figure 4: Internal transport costs and sustainable agglomeration

The point where the line crosses the real-wage differential from below is called the "sustain" point, $T(S)$ and describes the maximum level of transport costs for which agglomeration is still sustainable, i.e. for which $\omega_2/\omega_1 \leq 1$ (note the reversal). This equates to a real wage differential at λ_1 in Figure

3 which stays above the parity line. Beyond this point, agglomeration is not sustainable anymore, i.e. a case where the line is below parity in Figure 3. What happens when trade is liberalized? In Panel A, the sustain point shifts to the right, which indicates that agglomeration is able to be upheld for a wider range of domestic transport costs T_D. By design, this result mirrors the one in Monfort and Nicolini (2000) and also what we have seen in Figure 3 Panel A, i.e. that decreased cross-border trade costs increase the agglomeration forces. Note that this is mainly due to a decrease in the slope of the ascending part of the plotted lines. At these levels, an infinitesimally small increase in T_D increases ω_2/ω_1 and liberalization thereby seems to influence the centrifugal forces (competition) to a larger degree than the centripetal forces (cost and demand linkages), as seen by the positive slope (Crozet and Koenig 2004a). As such, a decrease in international trade costs mainly modulates the strength of the dispersion forces. In Panel C, these centripetal strengths dominate by reasons given above, such that the core-periphery pattern is upheld for a larger range of intra-national transport costs. So much so, that in full trade scenario, there exists no sustain point and agglomeration in region 1 is the never broken.

Again, the results for our main model with heterogenous intra-national space provide different conclusions. While the range of transport costs for which a core-periphery pattern is upheld also decreases in ε (compare Panels B and D), trade liberalization works towards the opposite, i.e. puts negative pressure on the full agglomeration in region 1, as seen by the negative shift of the sustain point $T(S)$ to the left. As such, regional market integration decreases the range of values for which agglomeration away from the border region can be upheld. This time, the change in the slope occurs mainly for the descending part, indicating that trade liberalization decreases centripetal forces for region 1.[44] These results essentially confirm analytically what is depicted numerically Figure 3, i.e. that there is an increased draw to the border with increased trade liberalization. However, one interesting aspect is that, for any level of regional trade integration T_F, agglomeration in region 1 is more likely to be upheld in the case of high product differentiation. Hence, it seems that the increased competitive pressures of an increased number of firms from abroad push activity into the interior, where they are sheltered (Crozet and Koenig 2004b). See also the discussion of Appendix A.1.

So far, we have confirmed the results of previous symmetric $2+2$ settings (Monfort and Nicolini 2000) and extended those from asymmetric $2+1$ layouts to an economy with four regions. What is left to investigate in our unique $2+2$ setting is the role *foreign* economic inequality, given that

[44] Note importantly, that given the unequal intra-national space, a reversal of the analysis, i.e. evaluating the stability of an agglomerated equilibrium at the border with $\lambda_1 = 0$ and $\omega_1/\omega_2 \leq 1$ would render the shift of the lines as in Panels A and C, i.e. would move the sustain point to the right.

we have previously set $\lambda_3 = \lambda_4 = 0.5$. We now relax this assumption and discuss additional results for the full range of spatial configurations λ ϵ [0,1] in the foreign country. Given this added dimension, Figure 3 turns three-dimensional which makes it a bit cumbersome to evaluate at first sight (see Figure A.2.1 of the Appendix). To make it more accessible, for the moment, we restrain ourselves to assessing the influence of a varying foreign manufacturing distribution on our two types of long-run equilibria depicted in Figure 3. As such, Figure 5 plots the combinations of λ_1 and λ_4 where the real wage differential ω_1/ω_2 is equalized, i.e. give the contour lines of the plane spanned by the two endogenous variables as given in Figure A.2.1.[45]

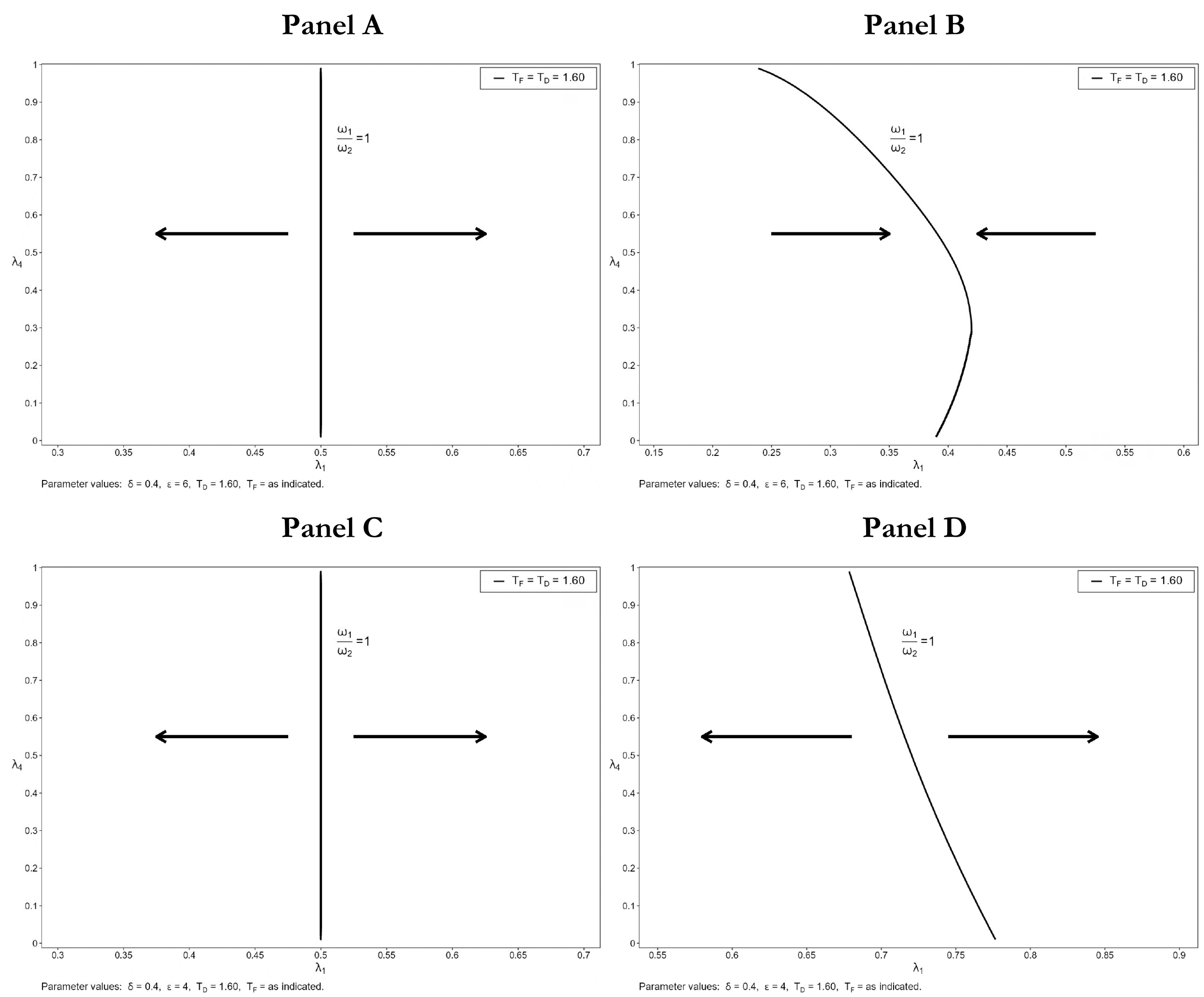


Figure 5: Foreign economic inequality and spatial equilibria

As is now common, Panel A and C shows the result for the homogenous $2 + 2$ model whereas Panel B and D depicts our asymmetric case. Given that the foreign spatial configuration can only exert

45 Remember that $\lambda_2 = (1 - \lambda_1)$ and $\lambda_3 = (1 - \lambda_4)$.

influence when trade costs are not prohibitive, we analyze the case for $T_F = 1.60$. To no surprise, Panel A and C shows a vertical line at $\lambda_1 = 0.50$. This is because when the home country is equally spread, and both regions have equal access to the foreign market via T_F, there is now difference in the relative real wages of the two regions. Hence, shifting shares of the workforce in the foreign regions doesn't affect the existence of this equilibrium.[46] Panels B and D paint a wholly different picture. For the case where a long-run stable equilibrium exists ($\varepsilon = 6$), we see that a changing share of workers in the interior of the foreign country modulates the domestic allocation for which this equilibrium is reached.[47] Panel D corroborates this view for the case where trade liberalization has led to a core-periphery pattern as the only stable equilibrium. In general, the higher the foreign spatial distribution is skewed towards the interior (a higher λ_4), the lower the share of workers in region 1 needed for both types of equilibria depicted in Panel B and D in Figure 3. What is the implication of this result? In the case of less intensive scale effects ($\varepsilon = 6$), the draw to the border is further increased (Panel B). On the other hand, when scale effects are large and full agglomeration is the only stable equilibrium ($\varepsilon = 4$), this result is reversed (Panel D). While there is still an increased draw to the domestic border in free trade, as can be seen by values for λ_1 above 0.50 on the x-axis, this draw is decreasing in λ_4. Intuitively speaking, given a stark regional inequality in the foreign country which is outlined by an economically strong interior (high λ_4), the basin of attraction leading to domestic agglomeration at the border is decreased when product differentiation is high. This shows that in the case of two equally sized countries, sheltering from increased competition in the interior is not as relevant, as a higher share of foreign activity at the border (low λ_4) decreases the relative real wage of region 1.

In sum, the quantitative spatial equilibrium model developed in this section and the counterfactual exercise of an increased market integration performed through it hold three main insights. First, given heterogenous intra-national space, progressive trade liberalization draws economic activity to the border, i.e. the real wages at border regions are relatively higher than in the interior when compared to autarky. Second, agglomeration in one of the regions is a more likely outcome the freer trade is, although agglomeration is more likely to occur at the border. And third, foreign spatial inequality has non-negligible impacts on these domestic effects, such that a core-periphery pattern in the foreign economy may attenuate or reinforce the first and second results. As we will see in the empirical results of the next chapter, this interpretation holds relevant insights the

46 In fact, Monfort and Nicolini (2000) show that there is one special case for which the *stability* of the spreading equilibrium depends on the foreign distribution.

47 The arrows in the graphs indicate the stability of the equilibrium, not depictable in contour lines. Arrows pointing towards the line indicate a stable equilibrium, given that economic forces (the real wage differential) lead consumers back to the original allocation, and vice versa for arrows pointing away from the line.

case of the East African Community, where the countries integrating are in fact all outlined by large interior hubs which host most of the economic activity.

IV. EMPIRICAL STRATEGY & DATA

Empirical Strategy

The theoretical exercise of the previous chapter motivates the empirical strategy. As is seen from the simulations, lowering trade costs among EAC members is predicted to increase the draw to the border, i.e. to the region with better access to the new markets. As discussed, this result is corroborated by previous theoretical models as well as by empirical evidence from both developed and developing settings.[48] However, what we also saw in the simulations is that this draw may be attenuated or reinforced depending on the economic (in)equality present in the foreign economy. Given the particular spatial layout of all EAC member countries, trade liberalization among them presents a fitting empirical case on which to study these dynamics, i.e. testing whether the re-establishment of the EAC did increase the relative attractiveness of border regions in comparison to the preeminent economic hubs, and also provides a setting with which to identify household-level effects of trade integration.

Note that so far, we have measured the attractiveness of border regions with real wages, and particularly, real wage differentials. However, we can readily translate these real wages into household welfare, as first established using indirect utility in (12), by simply noting that food was set as the numeraire.[49] Thereby, the real wages discussed broadly in Chapter III encapsulate what we envision as household welfare in the simplest form, which is, the income consumers earn and the prices they face (see e.g. Deaton 1997; Fujita et al. 2001; Winters 2002; Brülhart et al. 2012). The comparative statics tested theoretically thereby translate naturally to the empirics and revolve around assessing what happens to households' welfare (indirect utility) across space following a change in the international trade costs from a former prohibitive level down to levels of trade costs that mirror those of the type within the domestic country, i.e. only given by the geographic distance between locations; while holding trade costs between regions 1 and 2 as well as 3 and 4 constant throughout. To operationalize this, I employ a difference-in-differences (DiD) specification comparing the changes in welfare of

[48] Note also that a pure reference to new trade theory is not strictly necessary to render an increased impact of trade closer to borders. It has been shown in other developing settings that price pass through is highest directly at the border and decays perpendicular to it (e.g. Nicita 2009; Cali 2014; Atkin and Donaldson 2015).

[49] For a given set of exogenous parameter values, the indirect utility function (12) then reduces to a function of income, which is varying across space only in the nominal manufacturing wages w (see (36) through (39)), and consumer prices for manufactures I.

households living *relatively* closer to borders, ω_2, with those of households living *relatively* closer to the interior agglomerations ω_1, before and after the establishment of the EAC. To flexibly allow for treatment across space, I model this relationship nonparametrically, employing a continuous treatment intensity instead of dummies for the respective regions, which is captured by households' geographic (road) distances to EAC border crossings. The estimating equation therefore reads:

$$Y_{ict} = \alpha + \beta_1 Dist_i^{EAC} + \beta_2 CoreAgglom_i + \sum_{t=\{EAC,\}}^{\{CM\}} \beta_{3,t}(\gamma_t \cdot Dist_i^{EAC}) + \sum_{t=\{EAC,\}}^{\{CM\}} \beta_{4,t}(\gamma_t \cdot CoreAgglom_i) + X'_i + \delta_{ct/i/h} + e_{ict} \quad (1)$$

$Y_{i,c,t}$ represents the respective welfare indicator of individual i living in country c, surveyed at survey-sampling period t. $Dist_i^{EAC}$ is the inverse, relative within-country distance to the nearest EAC border crossing (0-1), such that a value of 1 indicate individuals in the sample living closest to the border in the sample, and value of 0 those furthest away. Given this definition, in the interpretation of results, we refer to this variable as $EAC\ Border$ (0-1). $CoreAgglom_i$ is a dummy (0/1) indicating individuals living within 50 kilometers of the three preeminent interior agglomerations, namely Nairobi in Kenya, Dar es Salaam in Tanzania and Kampala in Uganda.[50] γ_t is an indicator for the respective integration period i.e. switching to 1 for the free trade period (EAC) between 2001 and 2004, the customs union period (CU) between 2005 and 2009, as well as the common market period (CM) after 2010, respectively. Therefore, under specific assumptions discussed in the section on robustness and validity, β_3 and β_4 represent estimates of the effect of the EAC on border regions as well as on interior agglomerations. Specifically, they give estimates of the differential effect of households living at the border compared to those living furthest away, and those living in interior agglomerations to those living in the auxiliary, compared before and after the EAC was established. β_3 and β_4 can thus be seen as a test on the theoretical predictions, i.e. whether the EAC led to larger relative increases in welfare in border regions, i.e. $\Delta\omega_1/\Delta\omega_2 < 1$, given by a $\beta_3 \neq 0$ which also satisfies $\beta_3 > \beta_4$, rather than the opposite, i.e. in preexisting interior agglomerations, $\Delta\omega_1/\Delta\omega_2 > 1$, given by $\beta_4 \neq 0$ for which $\beta_3 < \beta_4$.[51] These estimates therefore also indicate if we should expect *dispersion* of the previously concentrated economic activity rather than *concentration* as proposed by the endogenous adjustment process in (52). $\boldsymbol{X}$ represents a matrix of individual-level control variables which allows us to account

[50] We also test the lower distance thresholds 25km and 10km in the robustness tests of section V.

[51] Of course, this statement is true only for the maximum effect (change) of border distance, i.e. going from the largest distance to the smallest distance in the sample. In-sample prediction, such as an interquartile range bound between 0 and 1 is arguably more appropriate as a comparison, where β_3 must be more than infinitesimally larger than β_4.

for all influences potentially conflating the relationship between access to (new) markets and household welfare. I include country-time fixed effects such that identification comes from variation within individual member countries in specific survey-periods in time. Standard errors are constructed by allowing for spatial correlation of errors, i.e. Conley standard errors are used (Conley 1999, 2010), and I additionally check for the clustering of errors at the level of the survey enumeration area, i.e. at the survey cluster level.[52] Binary dependent variables are estimated with a simple Linear Probability Model (LPM) specification.[53]

Data

I employ a distinct set of longitudinal, geo-referenced household-level surveys that were sampled in all three founding members of the EAC. First, I make use of the complete set of available Demographic and Health Surveys (DHS). DHS are cross-sectional, household-based surveys which are representative at both the national- as well as regional level, and collecting a broad array of information on topics such as demographics, education, employment and occupation, as well as fertility and family planning (Croft et al. 2018).[54] The main respondents are women of reproductive age (15-49), but DHS also provides information on men and children living in the sampled households, as well as household-specific information such as consumer durables and wealth assets in possession. To increase the variable size and sample space, I pool these additional data from the Men- and Household recodes also, leading to a main sample of 141,879 individuals living in 88,196 households located across 5,110 survey locales and interviewed between 1999 and 2016. Later extensions and robustness checks expand this sample to include non-GPS survey rounds sampled from 1989 onwards, and further add special survey rounds such as the AIDS Indicator Survey (AIS), the Malaria Indicator Survey (MIS), as well as the Knowledge, Attitudes and Practices Survey (KAP), to gather a higher frequency of survey years.[55] This extended (full) sample consists of a total of 332,725 individuals living in 203,150 households across 7,962 survey locales interviewed between 1989 and 2020.

---

52 The cut-off for Conley standard errors is chosen by the function, i.e. ensures a large enough sample size within a certain distance cutoff and additionally robust to sub-sampling. I provide sensitivity checks of the Conley standard errors in Table A of the Appendix.

53 Results for binary dependent variables estimated via *Probit* yields qualitatively identical and quantitatively similar marginal effects.

54 More precisely, the first level administrative subdivision, most often referred to as regions, districts, provinces or states.

55 The surveys were sampled in Kenya in 1989, 1993, 1998, 2003, 2008-09, 2014, 2015 (MIS), and 2020 (MIS), in Tanzania in 1991-92, 1994 (KAP) 1995, 1996, 1999, 2003-04 (AIS), 2004-05, 2007-08 (AIS), 2010, 2011-12 (AIS), 2015-16, and 2017, and in Uganda in 1988-89, 1995-96, 1995, 2000-01, 2006, 2009 (MIS), 2011, 2011 (AIS), 2014-15 (MIS) 2016 and 2018-19 (MIS).

Second, I make use the geo-referenced Afrobarometer (AFB) survey rounds, which span a timeframe of 18 years (from 1999 to 2017) across seven survey waves, i.e. rounds 1 through 7 (Afrobarometer 2019).[56] Afrobarometer surveys are representative at the national level, and the main respondents are adults of the sampled households. They carry individual- and household level information on basic characteristics, socio-demographics as well as own (economic) living conditions, household assets, and additionally, provide information on individuals' sentiments as well as opinions towards the economy, democracy, governance and society. Afrobarometer fits geo-coordinates (latitude and longitude) to respondents at the level of their respective enumeration area, and the sampling procedure aims for eight individuals/households per EA (BenYishay et al. 2017). The Afrobarometer adds information on 38,644 individuals (households) living in 3,414 geo-referenced localities across Kenya, Uganda and Tanzania to the sample, and additionally, provides the opportunity to test specific sentiments and attitudes towards free trade, which we return to in the section on robustness and extensions.

Lastly, I supplant the analysis with information from the Kagera Health and Development Survey (KHDS) (World Bank and University of Dar es Salaam 1994, 2004, 2010). The KHDS is a representative panel originally sampled from Kagera, a GADM-1 administrative region of Tanzania bordering Uganda in the northwest. The panel collected detailed information on households' and individuals' wealth and poverty dynamics, such as (self-)employment, salary, (non-) durable assets as well as food- and non-food consumption, all for which values in constant (deflated) Tanzanian Shilling are provided (Beegle et al. 2006; De Weerdt et al. 2010). The KHDS also includes information on migration decisions of individuals as well as community-level variables such as the price of commodities in local markets. The KHDS set out by interviewing 6,356 individuals living in 915 households spread across 51 sampling clusters between the first four yearly survey waves between 1991 and 1994 (round 1).[57] All of the initially sampled households (rather, the individuals living within those households) were sought to be re-contacted in the succeeding two survey rounds in 2004 and 2010, respectively. The tracking of individuals was highly successful, the sample evincing re-contact rates of 80% for individuals and over 90% for singular households.[58] Importantly, the number of administered households (and individuals) grew significantly over the sample timespan, as all members residing in the (new) households of original respondents were to be fully included in the survey in the

[56] Surveys were sampled in sampled in 2000-2001 (only Tanzania and Uganda), 2002-03, 2005, 2008, 2011-12, 2014-15 and 2017.

[57] Not all of the households were interviewed in all of the first four waves.

[58] And over 90% for those cases where at least one of the original household members was aimed to be re-interviewed.

later survey years as well.[59] For the present paper, the KHDS is able to add information on 21,696 distinct individuals – interviewed a minimum of one-, and a maximum of six times – whose households are spread across 2,019 survey locales in Tanzania and Uganda. Importantly, out of the 6,356 original survey respondents sampled between 1991-1994, 4,430 individuals were successfully (re)-interviewed in 2004 and 3,848 were able to be contacted in all three survey rounds, including 2010.

Figure 6 visually depicts this distinct set of geo-referenced data by plotting the sample enumeration areas of households from each of these three sources across East Africa. Notice that the map also depicts enumeration areas of contiguous EAC-accession as well as non-accession countries. These data will be employed in the extensions and robustness tests of the following chapter.

a) Dependent Variables

As anticipated above, real wages are proxied by household welfare in its simplest form, i.e. indicating the (sources of) income consumers earn and the prices they face (Deaton 1997; Winters 2002). Given the common data restrictions of household surveys, i.e. a lack of precise wage and price data, I capture these dynamics along a set of intensive and extensive labor market outcomes (work, employment, and income) as well as of consumption measures (food and non-food consumption, durable as well as non-durable assets).

To measure the levels of consumption in the Demographic and Health Surveys, I make use of the 1) *Wealth Index (1-5),* which is a DHS-constructed index which places households on a relative scale of wealth within their respective sample.[60] I countercheck these results employing the *Comparative Wealth Index* established by the DHS, which facilitates the comparison of the wealth scores underlying the wealth indices across countries and samples (Rutstein and Staveteig 2014). I additionally construct the *International Wealth Index (IWI)* as established in Smits and Steendijk (2015) as a further attempt to make household's wealth ranking more comparable across surveys.[61] Concerning labor market outcomes as the second dimension, I test the variable 2) *Employed Work (0/1)*, which indicates whether the respondent worked for someone outside of the household (conditionally on having work). In later extensions for these labor market results, I also test *Worked in last Year (0/1)*, which is the baseline measurement indicating whether survey respondents were pursuing some activity aside from housework within the last calendar year, on which 2) is conditioned on, and test whether this activity was also remunerated, i.e. *Paid in Cash (0/1).*

[59] As such, the number of singular households contained in the survey expanded from 915 in the first round (1991 to 1994) to 2,719 in 2004 and 3,314 in 2010 (De Weerdt et al. 2010).

[60] The construction of the index is based on an array of consumer durables, such as the construction of dwelling, sanitation facilities and as well as possessions such as a TV, motor vehicles etc. (Rutstein and Johnson 2004).

[61] The IWI exploits information from the entire universe of developing countries household surveys to construct factor loadings of specific household wealth items.

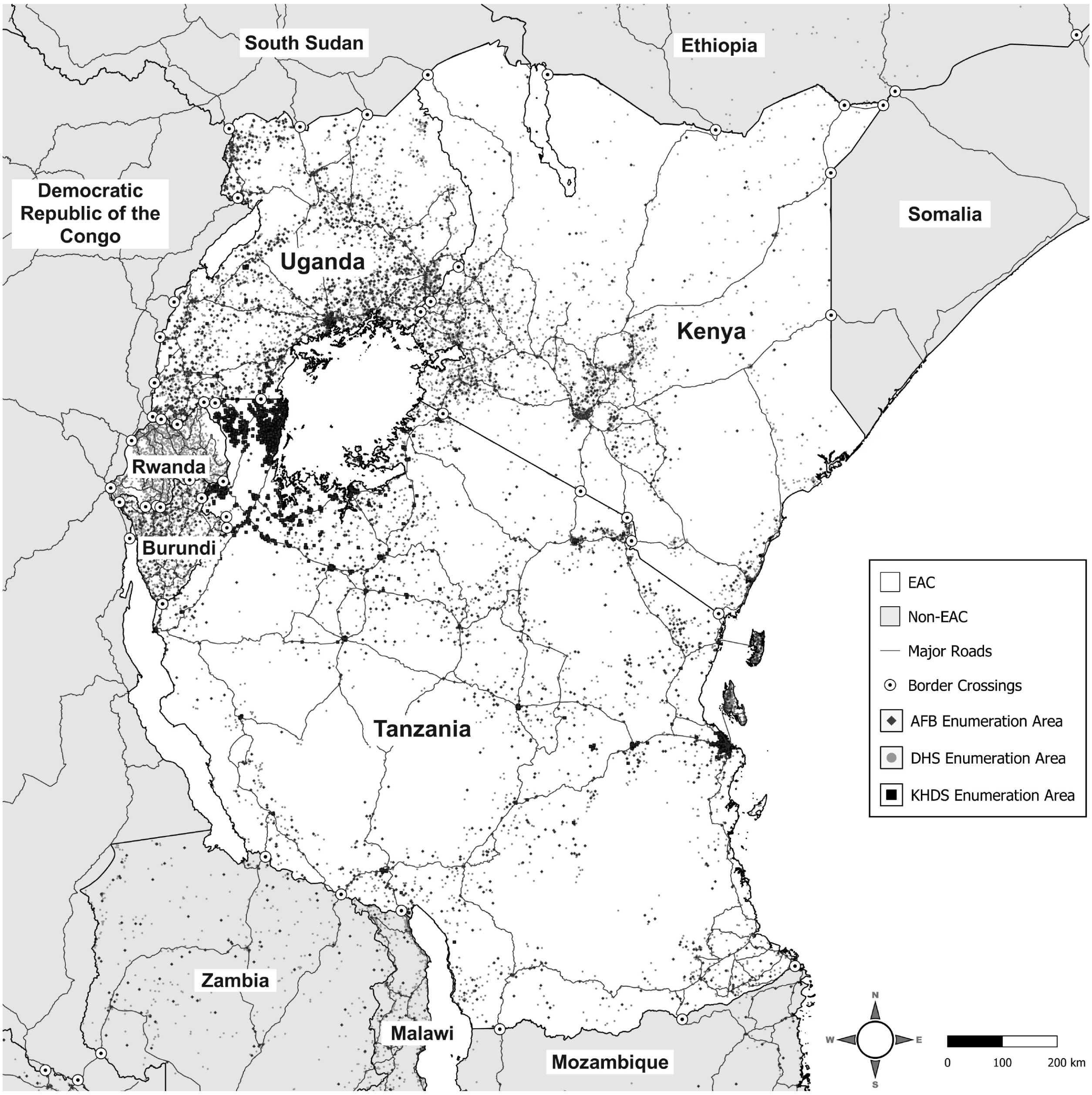


Figure 6: Sample Coverage

Lastly, I test the *Occupational Type (1-3)* of work, which places all activities categorized within the survey schedule from 'agrarian' (1), 'worker' (2) and 'professional' (3) activities.[62] This may also be regarded as a test on the skill-intensity of these occupations.[63]Concerning the Afrobarometer, the level of (basic) household consumption is measured by the variable 1) *Frequency gone without: [Water / Food /*

[62] For instance, 'workers' are occupations such as traders, artisans, or unskilled manual labor. 'Professional' is comprised of lawyers, accounts and teachers.

[63] A simple regression of *Occupation Level (1-3)* on the individual characteristics age, age squared, years of education as well as a female dummy and country-time fixed effects shows that each year of education increases the index by 0.06 units and a standard error of 0.0045.

Medical Care] (0-4) which is constructed by averaging individuals' responses in these three items. The three separate questions read: "Over the past year, how often, if ever, have you or anyone in your family gone without: Enough clean water for home use" / "[…]: Enough food to eat" / "[…]: Medicines or medical treatment?". The response values range from 'never' (0), 'just once or twice' (1), 'several times' (2), 'many times' (3) and 'always' (4). Similar to the DHS, the primary test on individuals' labor market outcome is measured via 2) *Employed Work (0/1)*. Again, I test whether individuals *Worked in Last Year (0/1)*, as well as their *Occupation Type* (1-3) in the section on extensions of Chapter V. One of the key characteristics of the Afrobarometer surveys is the component containing opinions, attitudes, and sentiments on individual, political, as well as domestic and international economic topics. As such, I test the variable *Support for: Regional Integration (1-5)* which evaluates the strength of the support for free movement and trade.[64] Further, I test whether individuals living closer or further from the border assess the *Ease of Crossing Borders (1-4)* as more or less difficult, how much they evaluate the *EAC* as well as the *African Union (AU)* in helping their country *(0-3)*, the latter as the preeminent supranational trade facilitators of the continent, and evaluate whether they would like having an *Immigrant as a Neighbor* (1-5). I supplant these variables with subjective assessments of the *Present vs. Past: Life Satisfaction (1-4)*, as well as their *Present vs. Past: Living Standards of People (1-4)*.

The Kagera Health and Development Survey (KHDS) allows us more detailed access into the consumption and income dimensions of households and the individuals therein. First, I test changes in the 1) *Annual per capita Household Consumption* which expresses aggregate food and non-food consumption in constant, i.e. deflated, 2010 Tanzanian Shilling ('000 TZS).[65] Food items are constituted of both purchased as well as home-produced food, non-food items are comprised of expenses from items such as clothing, schooling, services like haircuts, or utilities.[66] I also test for *Food-* and *Non-Food Consumption* in the extensions separately, to identify potential systematic differences across the two and to countercheck the results on food consumption droughts measured in the Afrobarometer. Secondly, I test for changes in household wealth similar to the DHS using the 2) *Value of Durable Assets* as well as the *Value of the occupied Dwelling*.[67] Both measures are given in deflated, 2004 Tanzanian Shilling ('000 TZS) as that this component was last administered in the survey wave of

[64] The question asked in Round 6 of the Afrobarometer probes the support by agreeing with either of the two following mutually exclusive statements. Statement 1: "People living in the sub-region should be able to move freely across international borders to trade or work". Statement 2: "because foreign workers take away jobs, and foreign traders sell their goods at very low prices, governments should protect their own citizens and limit the cross-border movement of people and goods"(Afrobarometer 2019)

[65] The total annual household consumption is distributed equally across all household members.

[66] For the full information on the construction of the aggregate consumption, see https://microdata.worldbank.org/index.php/catalog/2251/download/34035.

[67] Durable goods include e.g. Radios, Refrigerators, Telephones. The figure for the value of the occupied dwelling represents an estimate of the head of the household.

2004. Concerning income, I proceed in similar to the previous two surveys and ask whether the respondent has 3) *Employed Work (0/1)*, whether this work is 4) *Salaried Work (0/1)* and the 5) *Occupation Type (1-3)*. In extensions, I also test whether the overall likelihood of having any kind of work via *Worked last Year (0/1)*, as well as employed individuals' *Monthly Salary* in deflated 2004 Tanzanian Shilling ('000 TZS). Similarly to the Afrobarometer, the KHDS contains an array of subjective assessments and evaluations. As such, I test whether respondents' are more likely to have *Ever Migrated (0/1)*, whether the *Main reason for Migration was Economic* and whether movers found *Paid (formal) Employment (0/1)* right after migrating to the current location. Again, these measures are supplanted by assessments of their S*ubjective HH. Wealth (1-5)* both at the survey time (2004) as well as 10 years before (1994) and the general *Life Satisfaction on a Ladder (1-9)*.

As a last empirical investigation into the spatial corollaries anticipated in Chapter III, I test for changes in the extent of agglomeration at respondents' geographic locations (enumeration areas) across all three surveys. To do this, I merge granular population data provided by the Gridded Population of the World (CIESIN 2017) to evaluate the EAC's effect on *Population Density (sdz.)*, a measure of the total number of persons per square kilometer at the specific geography.[68] Testing changes in population density across space before and after the EAC is a direct test of the long-run dynamics induced by the agglomeration vs. dispersion forces discussed in chapter III and proxied by the household welfare components just established. In another vein, population density may be useful as a more general indicator capturing "underlying differences in productivity and quality of life" (Breinlich et al. 2014; 733), which are relevant consequences for policy intervention such as market integration.[69]

b) Independent Variables

The main explanatory variable of interest $Dist_i^{EAC}$, or respectively, $EAC\ Border$ (0/1), is measured by calculating the shortest road distance from each respondent's enumeration area to the nearest (within country) internal EAC border crossing (depicted in Figure 6).[70] To circumvent endogeneity in the construction of roads, I only use major roads, i.e. motorways, trunk- and primary roads as provided by OpenStreetMap (OSM 2022), which can be tracked back to the pre-EAC era. Border crossings are defined as points where these major roads connect to both sides of the border. To assess the sensibility of the results to the specific distance calculation, I also construct beeline (as the crow flies) distances from all enumeration areas to both the border crossings as well as to the nearest possible point on the

[68] To facilitate comparison, I standardize the value at a mean of 0 and a standard deviation of one.
[69] See for instance Rappaport and Sachs (2003).
[70] I measure distances using the projection of coordinates along the earth's ellipsoid (using WGS 84, EPSG 7030).

entire borderline spanned by two EAC country pairs. Shapefile data for country administrative areas, i.e. the boundaries of which I use come from the Center for Spatial Sciences at the University of California (GADM 2020). The empirical counterpart to $EAC\,Border$ (0/1), is given by $CoreAgglom$ (0/1) which represents an indicator of individuals living in core agglomerations, defined as households located within 50km of the country's respective economic hub (i.e. Dar es Salaam, Nairobi, and Kampala).[71] To also provide a test on the general tendency of agglomeration derived the Chapter III, i.e. irrespective of border distance and excluding the preeminent economic hubs, I construct *Agglomeration (0/1)*, a dummy indicating whether individuals live in environments that had a population density of 100,000 inhabitants per sq. km or more in the year 2000 (CIESIN 2017), and make us of the surveys' *Urban (0/1)* indicator, which is a sample-specific assessment of the level of urbanization at the specific location.[72]

To control for influences which may conflate the relationship between household welfare and trade-related aspects, I include the individual-level covariates *Age*, *Age squared*, a dichotomous indicator of gender, *Female (0/1)*, as well as individuals' *Educational Attainment* in completed levels or schooling years depending on the sample. I additionally account for potentially correlated geographic influences of development across distance and closely follow Henderson et al. (2018) with a set of important physical geographic features. I therefore include the location's *Elevation* (Farr et al. 2007), *Ruggedness* (Nunn and Puga 2012) as well as agricultural characteristics such as the number of *Growing Days* (Ramankutty et al. 2002) as well as average long-term *Monthly Temperature* and *Monthly Rainfall* (Fick and Hijmans 2017). In later robustness checks, I additionally control for the location's *Malaria Ecology* (Sachs et al. 2004), *Absolute Latitude*, as well as household's distance to *Navigable Rivers*, as well as *Major Lakes* and *Major Harbors.*[73] Lastly, I add country-year fixed effects to control for time-specific influences as well as country-specific influences at specific points in time, such as the Kenyan Post-Election Crisis of 2007-2008, and additionally include household-, respectively, individual-fixed effects for estimations using the Kagera Health and Development Survey.

[71] Later robustness tests relax this distance cutoff and also test areas within 25km and 10km, respectively. Notice that for the KHDS survey, there is no data for households living in these hubs in the pre-EAC era, as the survey was initially sampled in Kagera only. For these cases, the dummy switches to 1 for individuals living in 'Bukoba' the urban capital of Kagera.

[72] Urban stems either from country census information (DHS) or on the assessment of sample enumerators (Afrobarometer). See e.g. https://www.idhsdata.org/idhs-action/variables/URBAN#comparability_section and https://www.afrobarometer.org/wp-content/uploads/2022/07/AB_R9.-Survey-Manual_eng_FINAL_20jul22.pdf.

[73] The criteria for rivers' "navigability" as well as the importance of lakes ('major') is defined as in Henderson et al. (2018), i.e. I select all natural rivers within size categories 1-5 (scale 1-7) as defined in Natural Earth (2019) and lakes with a surface area of over 5,000 sq. kilometers (Lehner and Döll 2004). Concerning harbors, I define all large and medium sized ports listed in the World Port Index (WPI) as 'major harbors' (NGA 2019) as in (Wild and Stadelmann 2022), who provide recent individual-level evidence that access to such harbors is a robust predictor of living standards and household welfare.

Table B2.1 of the Appendix provides summary statistics for the dependent- and independent variables grouped by sample source. The table also provides first evidence into the distribution of outcome variables across space by grouping values into (border) distance quartiles and reporting a separate mean of outcomes within the three core agglomerations (see panel b). As expected, the countries are highly polarized. For instance, employed work and capital-intensive sectors are predominantly found in the core agglomerations. These are also the regions where most of the educated workforce lives and where consumption is highest.

V. RESULTS

Table 1 presents the first set of results estimated via regression equation (1). Importantly, these results set out the analysis by reporting a simple "pre vs. post" effect estimated by setting $\gamma_t = 1$ for all responses collected after mid-2001. The effects differentiated across the specific post-integration periods, i.e. the test on the temporal evolution of the EAC towards a Customs Union (CU) as well as a Common Market (CM), are established from Table 2 and onwards. Note also, that the results reported in all tables of this section are restricted to the reporting of the two difference-in-differences estimates β_3 and β_4, only. This is because we are primarily concerned with the relative development of border vs. interior agglomerations, and because the identification strategy is ultimately bound to estimate differential, i.e. distributional effects of regional market integration only.[74]

We first focus on results produced from the two nationally representative household-surveys, the Afrobarometer and the DHS for which the table depicts estimates on three specific outcomes, each representing a distinct dimension of and contributing to spatial inequality. To facilitate the comparison of measures across samples, the dependent variables are consistently grouped into consumption, income and agglomeration categories. Concerning the consumption dimension reported in columns (1) and (4), the results paint a stark picture. Both the AFB and the DHS provide no evidence of increased household welfare in regions closer to internal EAC borders following the establishment of the EAC. In fact, the results show that consumption was differentially affected in the negative direction, by a statistically significant, and economically relevant shift of relative household wealth of 0.128 (=0.558*0.23) units comparing household located at the median distance in the sample (222 km) with households living closest the border. This represents a reduction in wealth quintiles by 4% relative to the overall sample mean.

[74] Given that the control group is merely treated less intensely, however effectively located in the same intervention jurisdiction.

Table 1: Aggregate Difference-in-Differences Estimate

	Dependent Variable					
	Afrobarometer			DHS		
	Consumption	Income	Agglomeration	Consumption	Income	Agglomeration
	Freq. gone without: [Wat./Food/Med.] (0-4)	Employed Work (0/1)	Population Density (sdz.)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[3.21]	[0.19]	[0.00]
EAC Border (0-1) * EAC 1[t ≥ 2001]	0.114	0.076*	0.084	-0.558**	-0.060	0.110
	(0.278)	(0.041)	(0.110)	(0.242)	(0.041)	(0.124)
Core Agglom. (0/1) * EAC 1[t ≥ 2001]	-0.362***	-0.045	0.712***	0.402***	0.023	0.883***
	(0.069)	(0.062)	(0.220)	(0.108)	(0.040)	(0.160)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	38,234	104,483	71,738	104,467
R-Squared	0.13	0.14	0.24	0.37	0.16	0.29
R-Squared -Within	0.10	0.10	0.21	0.35	0.11	0.28

Notes: The results in each column are produced by a separate regression. In columns (1) through (3), data come from the Kenya, Uganda and Tanzania Afrobarometer surveys rounds 1 through 7 sampled between 2000 and 2017. In columns (4) through (6), data come from the Kenya, Uganda and Tanzania Demographic and Health surveys (DHS) sampled between 1999 and 2020. For all DHS variables measured at the household level, (4) and (6), the answers from the main survey respondents (women) are used. The sample mean of the respective dependent variable is given in brackets above the estimates. EAC Border (0-1) is the inverse, relative within country distance to the nearest EAC border crossing. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). EAC 1[t ≥ 2001] switches to one for individuals sampled from the second half of 2001 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Concerning the complementary DiD estimate for core agglomerations, we see the opposite effect. Households experienced a large, statistically significant increase in (basic) consumption measured by the (AFB) DHS. Relative to the overall sample mean, households living in Nairobi, Dar es Salaam or Kampala increased their position on the wealth quintile by 0.402 units (13%) and reduced the occurrence of basic consumption droughts by 0.362 (32%). Concerning income, the results do not provide evidence for a change in the likelihood of having employed work, only showing a marginally significant increase (at the 10% level) by 1.9 percentage points (8%) for border households in the Afrobarometer. Turning to the evidence on agglomeration patterns, we see the results reported in the consumption modules corroborated, as the population density of preexisting core agglomerations further intensifies after the establishment of the EAC by almost one standard deviation, compared to household locations in the auxiliary.

In sum, this first set of results does not provide evidence in favor of the first prediction of Chapter III, i.e. that relative household welfare (real wages) is increased in border regions. Households and individuals living closer to borders did not experience (greater) relative welfare gains after the EAC was established, compared to individuals living further away. This result is consistent throughout the outcome variables tested in both the Afrobarometer as well as the DHS samples. Inasmuch as trade liberalization weakens agglomeration tendencies, which was the second prediction of Chapter III, the findings strongly negate this notion, as we can observe large differential increases in household consumption and population density in the core agglomerations predating the EAC.

I now turn to a more nuanced assessment of these aggregate effects and estimate the full set of period-specific difference-in-differences estimates anticipated in equation (1). Table 2 reports these results, which effectively expand the simple DiD effect of Table 1 to three separate estimates for border and interior regions, which compare outcomes across space in the initial free trade regiment (EAC), the customs union (CU), and the common market era (CM) to the same pre-EAC period. As such, the estimates are directly comparable and allow an insight into the temporal dynamics of the mean increases in spatial inequality across the EAC shown in Table 1. While these temporally differentiated effects overall confirm the average effect shown in Table 1, the results nonetheless provide three interesting insights. First, border regions did not seem to have benefitted differentially more than more distant region in any of the EAC's time periods. Importantly, this result is true even in the early years following the re-establishment, which goes against Eberhard-Ruiz and Moradi (2019) who show that growth of lights emitted by night was differentially higher for cities in border regions of the EAC precisely, and only, in the initial periods. In fact, the present results show that it is more likely that regions closer to the border experienced relative welfare reductions following re-establishment. Three out of the four significant estimates show reductions in the wealth index as well as employed work opportunities (columns 4 and 5). The only positive effect depictable is a 2.4 percentage point increase in employed work in the CM period, which may or not be suggestive evidence of a common market protocol, i.e. free movement of labor potentially benefitting individuals closer to new markets. Second, while the positive effect for households living in the core agglomerations is present for all years, the effect is non-increasing, falling in absolute terms over the deepening of the EAC, and even vanishes for one of the estimates (column 5). The estimates for changes in population density show the opposite dynamic, which constitutes the third noticeable insight of Table 2. While in the EAC period, the results either show non-identifiable, or negative differences (column 6 and 3, respectively) in population density for survey locales' close to the interior hubs, the estimates on later periods provide evidence of growing agglomeration tendencies, and thereby, increasing spatial disparity over time.

Table 2: Difference-in-Differences across three Integration Periods

	Dependent Variable					
	Afrobarometer			DHS		
	Consumption	Income	Agglomeration	Consumption	Income	Agglomeration
	Freq. gone without: [Wat./Food/Med.] (0-4)	Employed Work (0/1)	Population Density (sdz.)	*Wealth Index* (1-5)	Employed Work (0/1)	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[3.21]	[0.19]	[0.00]
EAC Border (0-1) * EAC 1[2001-2004]	-0.083	0.050	0.080	-1.006***	-0.129	0.516
	(0.296)	(0.058)	(0.102)	(0.366)	(0.089)	(0.355)
EAC Border (0-1) * CU 1[2005-2009]	0.172	0.060	0.030	-0.389	-0.138**	0.190
	(0.292)	(0.065)	(0.094)	(0.286)	(0.054)	(0.136)
EAC Border (0-1) * CM 1[t ≥ 2010]	0.142	0.098**	0.103	-0.578**	-0.034	0.087
	(0.282)	(0.039)	(0.118)	(0.250)	(0.040)	(0.118)
Core Agglom. (0/1) * EAC 1[2001-2004]	-0.261***	-0.013	-0.204**	0.536***	0.073*	-0.140
	(0.077)	(0.050)	(0.095)	(0.118)	(0.040)	(0.320)
Core Agglom. (0/1) * CU 1[2005-2009]	-0.470***	-0.006	0.657***	0.379***	0.037	0.876***
	(0.097)	(0.045)	(0.247)	(0.104)	(0.039)	(0.340)
Core Agglom. (0/1) * CM 1[t ≥ 2010]	-0.338***	-0.076	0.978***	0.383***	0.006	1.109***
	(0.065)	(0.069)	(0.269)	(0.114)	(0.041)	(0.172)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	36,042	25,465	38,234	104,483	71,738	104,467
R-Squared	0.13	0.14	0.26	0.37	0.16	0.30
R-Squared -Within	0.11	0.10	0.23	0.35	0.11	0.29

Notes: The results in each column are produced by a separate regression. In columns (1) through (3), data come from the Kenya, Uganda and Tanzania Afrobarometer surveys rounds 1 through 7 sampled between 2000 and 2017. In columns (4) through (6), data come from the Kenya, Uganda and Tanzania Demographic and Health surveys (DHS) sampled between 1999 and 2020. The sample mean of the respective dependent variable is given in brackets above the estimates. For all DHS variables measured at the household level, (4) and (6), the answers from the main survey respondents (women) are used. EAC Border (0-1) is the inverse, relative within country distance to the nearest EAC border crossing. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). EAC 1[2001-2004] switches to one for individuals sampled from the second half of 2001 to and including 2004, CU 1[2005-2009] for individuals sampled from 2005 and including 2009, and CM 1[t ≥ 2010] for individuals sampled from 2010 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Expressed quantitatively, and in comparison, to the rest of the country, the agglomeration in survey locals in close vicinity to the capital cities grew more strongly by a magnitude of one standard deviation when measuring differences in the common market period to years preceding the EAC. Reconciling this observation with the positive welfare effects of agglomerations (columns 1 and 4) which consistently predate the positive responses of population density in time, the finding provides supportive evidence for individuals (labor) responding with the dynamic proposed in (52) of Chapter

III, i.e. respond to positive welfare differentials with migration inflows.[75] Given that estimated welfare differences decrease or stagnate in the periods after these large population inflows, these findings also support the notion that by moving, individuals (labor) themselves endogenously regulate welfare differences (downward). The loss of significance on the positive employed work outcome (column 5) from the CU period onwards may provide some (weak) evidence for this in the form of an increase in the elasticity of labor supply due to migration.

As a last note on this set of results, and in an attempt to reconcile them with the theoretical predictions of chapter III, the dynamics just described hint towards three broad possibilities. The first trivial one is that the cross-border activity triggered by regional market integration of the EAC did not induce a large enough shock to render proximity to new markets as relevant enough to break a general agglomeration tendency towards interior hubs.[76] Second, the interior economic hubs display too strong of an attraction that even in the event of a fully integrated market, agglomeration in Nairobi, Dar Es Salaam and Kampala is sustained. This latter scenario can be intuitively displayed by referring to Panel D in Figure 3, where a large enough share of either previous manufacturing distribution or population inflows renders a fully agglomerated equilibrium of interiors possible even for the scenario of full trade liberalization, i.e. a $T_F = T_D = 1.60$. Depending on the degree of integration, the necessary share of manufacturing positioned in hubs prior to liberalization setting in motion such an equilibrium has a minimum of 50% and a maximum of 70%. Importantly, this is the case only for the foreign market which is equally balanced. As is established in Panel D of Figure 5, this necessary labor share is decreasing in foreign spatial inequality, where 78% depicts the scenario when all foreign activity is situated at the border and less than 70% when the foreign economy is fully agglomerated in the interior.[77] As such, given the significant polarization of all member countries before the establishment of the EAC (see Chapter II) one can imagine a scenario in which liberalization does not suffice to break agglomeration, because countries open up to similarly spatially unequal foreign markets.

Thirdly, the dynamics displayed in Table 2 may also point towards a transitory shift to a new welfare equalizing equilibrium e.g. as indicated by the dwindling differences of AFB and DHS consumption indices over time. The mechanics can be displayed in Panel B of Figure 3 and are as follows. With a starting point of equal labor distribution, trade liberalization sets in, which lowers the real wage differential of core agglomerations. Labor immediately responds with outflow as seen in

[75] The effects on consumption are strongest in the EAC and CU years (columns 1 and 4), the effects on population density in the periods directly succeeding these eras (columns 3 and 6).

[76] Of course, only if in this setting the crucial assumptions of the difference-in-differences hold. We explore this point in the section on robustness & extensions.

[77] For results in all three trade scenarios, see Figure A.2.1.

column (3) of Table 2. However, these outflows are too large such that they quickly offshoot the long-run stable equilibria between 40% and 50% and render real wages in interior hubs as exceeding those at borders. The subsequent response is a migration inflow back into core agglomerations until welfare equalization at a new (lower) equilibrium share of labor is reached. Three factors go against this possibility, however: First, an equal or less than equal spatial distribution of labor and welfare in favor of region 2 was not the initial position of all three countries before the EAC. Second, population inflows into border regions are not observed in Table 2. And third, most importantly, while differential change in welfare and population decrease across time, they remain positive in all periods post liberalization, which is only possible in a move towards a fully agglomerated equilibrium. One last possibility, of course, is that there is no practical spatial heterogeneity, i.e. border regions do not have better access to the markets.[78] In such a scenario, depicted in Panel A and C in Figures 3, the result is fully robust with trade setting in motion full agglomeration, as depicted by the rotating lines in all parameter configurations. In any case, what is not supported by the results is a dissolving of previous metropolis or 'urban giants' as prominently anticipated for developing settings (Ades and Glaeser 1995; Krugman and Elizondo 1996).

Extended Labor Market Results. The results on the remaining set of labor market outcomes introduced in the Data section are presented in Table B.5 of the Appendix. The table shows evidence on differential development of having any type of activity outside of household work within the last year (columns 1 and 4), whether conditionally on such an activity, one is paid in cash compared to in-kind or no compensation, and, at which level of the occupational skill dimension, i.e. *Occupation Type (1-3)* the work is situated.

Overall, the results across variables and samples provide no systematic evidence in favor of or against spatial inequalities following the establishment of the EAC. The most robust result is a positive change in the likelihood of having work outside of household activities. However, the effect is decreasing in absolute terms and in statistical significance over time (column 4). Also, evidence on the type or remuneration of this work shows that the likelihood of it being low-skilled and unpaid work is larger (compare columns 5 and 6). Concerning core agglomerations, there is some weak evidence on having work (column 4) and being remunerated for it, household's occurrence of experiencing monetary droughts increased in the CU period (column 3), which mirrors the increased likelihood of being paid in cash (column 6). However, given the inconsistency of these results, further interpretation

[78] One may think of shifting transport types i.e. using airports or harbors rather than roads.

of these results would require more evidence. The next section looks at these issues arguably more fittingly, using the panel dataset of the sample.

Kagera Health and Development Survey. Table 3 provides the final set of main results, reporting estimates produced from the Kagera Health and Development Survey (KHDS). A few notes towards the interpretation of results in comparison to the Afrobarometer and the DHS. First, given the timeframe of the survey, I am unable to provide longer term evidence as the third and final sampling round was conducted in 2010 and Table 3 thereby reports estimates on the effect of the EAC and CU periods only. Second, some survey items were not administered in 2010, such that values cannot be estimated for these time periods either. Third, concerning the interpretation of the *Core Agglom. (0/1)* dummy. Given the spatial limitations of the KHDS survey in the first waves between 1991-1994, I am unable to measure outcomes for households in Dar es Salaam prior to 2004. Hence, the DiD estimate on *Core Agglom. (0/1)* is given by comparing differences of individuals over time living in Kagera's urban capital 'Bukoba' in addition to those individuals who have later moved to (or were initially sampled in) Dar es Salaam or Kampala in the second and third rounds. The robustness checks in section removes the latter group and thereby provide a test on differential spatial sorting and on the general tendency of agglomeration predicted in Chapter III.[79] Fourth and last, as a panel, the KHDS allows the inclusion of household-, respectively, individual fixed effects, and identification thereby stems from changes of within households or individuals across time.[80]

Columns (1) and (2) depict the first set of results which test for differential changes in per capita consumption of food and non-food items as well as the value of durable items if the household. Both figures are expressed in deflated Tanzanian Shilling (TZS) priced in constant 2010 and 2004 levels, respectively. Households living in core agglomerations increased their consumption following the establishment of the EAC, while households living closer to the regions bordering the new markets do not evince statistically significant differences in either integration period. Quantitatively, each individual living in households in core agglomerations consumes an extra of over 170,320 TZS worth of food and non-food items more than in the pre-EAC period, compared to the development of the rest of the country over the same timespan. In relative terms, this indicates an increase of roughly 31% compared to the average of the sample mean. This figure grows to almost 50% in the post-EAC years. Households also increase the consumption of consumer durables, owning new stock of goods in value of over 706,607 TZS higher than the comparison group.

[79] I pick up concerns about spatial sorting in the section on Robustness Checks, Validity Tests, and Extensions.
[80] Depending on whether questions are administered at the household- or individual-level.

Table 3: Difference-in-Differences using the Kagera Health and Development Survey (KHDS)

	Dependent Variable					
	Kagera Health and Development Survey (KHDS)					
	Consumption		Income			Agglomeration
	Annual p.c. Consumption (in 2010 TZS '000)	Value of dur. Assets (in 2004 TZS '000)	Employed Work (0/1)	Salaried Work (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[553.78]	[112.23]	[0.12]	[0.01]	[1.20]	[0.48]
EAC Border (0-1) * EAC 1[2004]	-853.376	187.152	0.581*	0.071	0.107	0.251
	(1072.010)	(754.481)	(0.331)	(0.049)	(0.261)	(0.703)
EAC Border (0-1) * CU 1[2010]	-1317.845		0.432		-0.010	-0.697
	(1154.386)		(0.305)		(0.404)	(0.438)
Core Agglom. (0/1) * EAC 1[2004]	170.320***	706.607***	-0.013	0.026***	0.134***	0.907***
	(42.328)	(192.323)	(0.023)	(0.008)	(0.021)	(0.090)
Core Agglom. (0/1) * CU 1[2010]	275.036***		-0.039		0.069	1.020***
	(57.368)		(0.025)		(0.067)	(0.074)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	YES	YES	YES	YES
Household Fixed Effects	YES	YES	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,492	2,695	24,972	14,254	15,685	7,366
Observations - Fixed Effects	3,816	2,363	12,747	6,988	6,253	3,933
R-Squared	0.88	0.97	0.64	0.72	0.82	0.97
R-Squared -Within	0.12	0.09	0.04	0.01	0.14	0.49

Notes: The results in each column are produced by a separate regression. Data come from the Kagera Health and Development Surveys (KHDS) collected in four waves across 1991-1994, as well as one wave in 2004 and 2010, respectively. In columns (1) through (3) outcome variables represent aggregate household information provided by the head of the household provided by the head of the household, in columns (4) through (6) they are administered on an individual level. Certain indicators were not sampled in the survey wave of 2010, which is why there is no estimate given for these columns. The sample mean of the respective dependent variable is given in brackets above the estimates. EAC Border (0-1) is the inverse, relative within country distance to the nearest EAC border crossing. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). For the initial KHDS survey waves 'Bukoba' - the capital of Kagera representes the core agglomeration. EAC 1[2004] switches on for individuals (re-)sampled in 2004. CU 1[2010], switches on for individuals (re-)sampled in 2010, the second re-interview period of the KHDS. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include an indicator whether the household is living in proximity to (former) refugee camps. The regressions testing household-level outcomes, columns (1) through (3), include household fixed effects, the regressions testing individual-level outcomes, columns (4) through (6), include individual fixed effects. All regressions include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

While I cannot identify positive effects on employed work in agglomerations, there is a weakly significant effect at borders. A change from the median distance to borders in the sample to the minimum distance leads to an increase in the likelihood of employed work by 5 (=0.581*0.086) percentage points. However, the remaining labor market outcomes paint a more consistent compared to the cross-sectional results. That is, individuals in core agglomeration profit differentially more from the establishment of the EAC both in terms of extensive and intensive labor market outcomes. Living in core agglomerations compared to the periphery increases the likelihood of having salaried work by 2.6 percentage points and raises the occupation type in which the individual works by 0.134 units. In other words, the skill gap in work done by individuals in capital cities versus peripheries increased by another 11% relative to the sample mean. As a further exploration into these dynamics Table B6 of the Appendix provides the findings on the remaining labor market outcomes introduced in the Data section. The findings show no differential effects across space for the likelihood of having worked in the past year or showing an increase in monthly salary.[81] Table B6 also tests an expanded set of consumption components, i.e. tests for differences in food vs. non-food consumption and tests the value of the occupied dwelling of the household. With the exception of row 3 of column (2), all results confirm the evidence shown in Table 3, namely that households proximate to borders do not gain differentially more than interior regions but for the core agglomerations, which show strong, positive effects on consumption and wealth.

Concerning the last dimension of and contributing to spatial inequality, population density, we see quantitatively highly similar and qualitatively identical results compared to the cross-sectional samples. The disparity in population density is growing, by up to one standard deviation as measured in 2010. Evincing strengthened agglomeration patterns.[82] In sum, the results from the individual-level support the evidence produced from the cross-sectional samples, The results in Table 3 confirm my previous results from m. As was shown in Tables 1 and 2, the positive effect on population density emerges and is again, increasing over time.

Robustness Checks, Validity Tests and Extensions

Robustness Checks. As a first insight into the stability of the presented results, I conducted an array of robustness tests summarized in Tables B2 through B4 of Appendix B and briefly discussed here. As before, all of the results are produced using regression specification (1), unless otherwise indicated. The tables also report the baseline coefficients from Table 2 in the top rows with which to compare the results of the adapted estimations. Given the lack of effects for border regions throughout the samples and variables tested, the table confines the reporting of the three difference-in-difference coefficients on *Core Agglom. (0/1)*. The full results including the DiD estimates for *EAC Border (0-1)* can be accessed in the source tables which are referred to in each of the respective tests.

[81] Interestingly, regressions relaxing the strict individual fixed effects show a strong differential increase of wages in agglomeration and a decrease of them at borders (significant at the 10 and 5 percent level).

[82] An explanation of the larger and statistically more significant effect in the initial EAC period is that this estimate measure changes in the tail end of the EAC period (i.e. in 2004), in contrast to the earlier samples, which average changes between 2001 and 2004.

To begin, (a) allows for the clustering of standard errors at the enumeration area level instead of implementing Conley standard errors. (b) Removes all individual and geographic controls from the regression.[83] (c) Adds the extended set of geographic controls anticipated in chapter IV, namely locations' *Absolute Latitude*, *Malaria Ecology* (Sachs et al. 2004), *Navigable Rivers*, *Major Lakes* and *Major Harbors* to control for other trade-related influences and adds the dummy *non-EAC ≤ 100km (0/1)* as well as the interaction of it with all period dummies to net out effects potentially stemming from a change (loss) in market access at non-EAC borders. (d) Employs the sample survey weights provided by the Afrobarometer and DHS, accounting for the pooling across countries and years by standardizing the weights for each country-survey round pair.[84] (e) Excludes low-precision localities. For the Afrobarometer, this is implemented by dropping all observations for which the AidData precision code is above 2 (AidData 2017).[85] In the DHS survey, I drop all observations for which coordinates are generated from a GPS receiver used by the fieldworker.[86] This test cannot be conducted for the KHDS as there is no distinction in the precision of GPS locales. (f) Replaces the *Core Agglom. (0/1)* dummy in specification (1) with *Agglomeration (0/1)* which switches to 1 for households living within 50km of an 'urban center' demarcated as such in the year 2000. The list of urban centers I use are provided by the European Commission (2019).[87] This can be seen as a general test on the agglomeration tendency of the results in Chapter III, irrespective of border distance or proximity to the capital cities. As a further test on this, I also try the *Urban (0/1)* dummy attached to the surveys.[88] (g) Reduces the spatial cut-off criteria for living in core agglomerations to 25km and (h) to 10km, respectively. (i) Tests a flexible distance specification in identifying differential effects across distance by interacting both the continuous distance as well as the squared continuous distance to border crossings as well as hubs with the dummy *EAC 1[t ≥ 2001]*. (j) Splits the CM period into a two dummies, namely *CM 1[2010-2014]* and add a post-CM time period *post-CM 1[t ≥ 2015]* to provide a test on the hypothesized transitory shift to a welfare-equalizing equilibrium as discussed in the previous section.[89] (k) Excludes the individuals in the sample which did not live in the survey location at least three years before the establishment of the EAC (before 1999), i.e. excludes "post-

[83] For the KHDS survey, I cluster observations at a specific geographic delineation, which is based on 2 decimal places of latitude-longitude combination, i.e. raster of slightly larger than 1 square km. at the equator.

[84] I.e. transform the weights such that they sum to 1 for each pair.

[85] The scale ranges from 1-8. Using precision code 1 leads to a loss of data in the range of

[86] See https://dhsprogram.com/Methodology/upload/MEASURE-DHS-GPS-Data-Format.pdf.

[87] The definition reads: "The spatially-generalized high-density clusters of contiguous grid cells of 1 km2 with a density of at least 1,500 inhabitants per km2 of land surface or at least 50% built -up surface share per km2 of land surface, and a minimum population of 50,000." (European Commission 2019; 13). For my purposes, I use a minimum population threshold of 100,000.

[88] Results can be requested from the author.

[89] Notice that 'post-CM' has no further meaning other than nomenclature. The common market of the EAC has continued to persist.

EAC Migrants".[90] (i) Uses the entire universe of data, including AIS, KAP and MIS surveys to fill the sample space of the main DHS surveys. (m) Takes the logarithm of outcome variables expressed in constant Tanzanian Shilling.

While the general upshot of all these sensitivity checks is that all previous conclusions and interpretations hold, there are some interesting takeaways for two specific tests which have bearings on the theoretical results. For one, by splitting up the CM years into two periods (j), we notice that the convergence anticipated in Table 2 is at least partly corroborated. While the Afrobarometer shows seemingly unaltered results compared to the years 2010-2014, estimates produced with the DHS survey, which include three more survey years from 2017 to and including 2020, drop both in size as well as significance. Most notably, the coefficient on population density is halved compared to the 'early' CM period. And secondly, when exchanging the *Core Agglom. (0/1)* dummy with a wider selection of urbanities (f), the effects are either non-significant, significantly weaker, or point in the opposite direction. Particularly interesting are the effects on population density which are insignificant for the EAC and CU period, and lower in magnitude by an order of 5-10 in the CM period compared to the developments in the core agglomeration. Only for the KHDS are results similar in magnitude and significance. This may be reconciled by the fact that the dummy is a test on 'urbanities' rather than economic hubs existent in 2000, and that 'urbanities' are endogenous to the outcome. However, given that endogenous formation of agglomeration is precisely what theory dictates, I take this observation seriously and evaluate this effect in the context of the main theoretical model, i.e. with heterogenous intra-national space. I do so by estimating a triple-difference specification which tests for a differential effect for (endogenous) agglomerations at border regions, which is effectively done by interacting the previous treatment interaction *EAC Border (0-1) * EAC* with the *Urban (0/1)* dummy provided in the Afrobarometer and DHS. The results are shown in Table B7 of the Appendix. The combined effect in row three provides no convincing evidence that agglomeration occurs at borders, and that relative welfare increases at those agglomerations are differentially higher. While there is weak evidence of increased population density for the DHS sample (significant at the 10 percent level), the wealth index, in turn, was reduced. Panel b) tests the extended set of labor market outcomes. Again, most results are insignificant or strongly negative (columns 5 and 6). This confirms the previous results which did not show a draw to the border.

Validity Tests. The fundamental assumption behind the employed difference-in-differences design requires that absent policy change, the spatial disparity in households' welfare within the EAC

---

[90] Because of missing migration information, this test is only possible for the DHS and KHDS sample.

countries would have evolved 'in parallel' i.e. continued their relative pre-intervention trajectories. In other words, for our estimates to represent a causal relationship, nothing other than the policy of regional market integration should have induced a differential welfare change across space in the timespan (shortly) the before and after the re-establishment. While in practice ultimately never verifiable, I provide three distinct pieces of evidence which may strengthen our confidence that this assumption holds.

Before the tests are discussed, there needs to be a check on other policy measures with the potential to influence the economic geography of the respective countries. In the timespan between 1995 and 2010, the most relevant policies I was able to identify mainly concentrated on trade facilitation. For instance, the Northern Corridor Transport Improvement Plan (NCIP) of 2004 aimed to improve transport infrastructure to facilitate trade integration. I do not deem this investment as undermining the results, as the completion of project goals aimed within the NCIP were temporally lagging the main results presented here (see World Bank 2016). A related concern is that the facilitation of "one-stop-border-posts" (OSBPs) may lead to differential success of integration across border. However, as in the case of the NCIP most OSBPs were erected many years after the large increases first set in in our results (Cadot et al. 2015; EAC 2015).[91]Another initial concern may be displayed by "Export Processing Zones" (EPZs) or "Special economic Zones" (SEZs) in or near the core agglomerations defined in the paper. While all of the three member countries actively promote SEZs, the timing as well as the spatial pattern relief concerns of an effect entirely attributable to such developments. For instance, in Kenya, the majority of EPZs are outside of Nairobi, many of them in the port of Mombasa not included in the *Core Agglom.* Definition.[92] In Tanzania, the operation of EPZs is possible since the ratification of the EPZ act of 2002 and the SEZ act of 2006, respectively. However, while data on firms operating under such licenses is untransparent with several contradicting reports on the absolute number, they agree that the general impact to industrialization was small (Andreoni et al. 2022). Concerning the spatial dimension, most recent data suggests that the majority of firms operating under an SEZ license are outside of Dar es Salaam (Kinyondo et al. 2016; Andreoni et al. 2022).[93] Lastly, the distribution of EPZs in Uganda does in fact evince a stark regional disparity skewed towards the Central region enclosing Kampala (UFZA 2022). However, SEZ only began

[91] I also estimate a potential heterogeneity across border regions in Table B12 of the Appendix, which is discussed at the end of this section. The findings provide no evidence of such an influence.

[92] While many of them are in the Machakos county, next to Nairobi, road and beeline distances are above the commonly used threshold of 50km. Also note that the results are robust to the narrower spatial delineation (i.e. 25 and 10km), if one assumes laborers to commute to these EPZs.

[93] Even though the largest of the literal SE zones are located in the Dar es Salaam-Bagamoyo corridor. Bagamoyo is similar distance to Dar es Salaam as Machakos to Nairobi in Kenya.

operating after 2014, which is the year the Ugandan Free Zones Act was ratified, rendering our results robust to this development.

Placebo Tests. The first formal test against the difference-in-differences assumption is presented in Table 4. The results shown are produced in the identical way as they were in Table 1, but use data on contiguous "placebo countries", i.e. estimate a differential change across pre and post EAC time periods for non-EAC countries bordering Kenya, Tanzania and Uganda. The countries available in the data are Malawi, Mozambique, and Zambia in the Afrobarometer. The DHS expands the countries to include Ethiopia and Rwanda.[94] In these countries *EAC-Border (0-1)* represents the inverse, relative within-country road distance to the nearest major road crossings with an EAC country and *Core Agglom. (0/1)* identifies individuals living within the respective core agglomeration (i.e. Addis Abeba in Ethiopia, Kigali in Rwanda, Lilongwe in Malawi, Lusaka in Zambia and Maputo in Mozambique). The results provide suggestive evidence that the differential welfare change in economic hubs was not a larger, regional trend within the timeframe under evaluation. More generally, the results do not show any consistent evidence of spatial inequalities increasing or decreasing across time. While there are positive effects on population density in the Afrobarometer, the arguably more precise spatial estimates produced from the DHS does not confirm this finding.[95] Further, while there some improvements in wealth at borders (column 4 of panel a), column 3 in panel b) shows that the occurrence of cash income droughts was similarly strongly decreased in core agglomerations.[96] The rest of the results show no hint towards a differential development following the EAC, which strengthens our confidence in attributing our estimates to the EAC. For completeness, B8 of the Appendix shows the companion results which splits the *EAC* $1[t \geq 2001]$ dummy into the three integration periods. The results from this more nuanced regression do not provide reasons for an interpretations different than the ones just discussed. One interesting aspect, however is that, in the DHS sample, many of the significant effects for core agglomerations are only apparent in the EAC period. Given that they almost universally imply a negative impact and show no accompanied increase at borders, I deem these results as non-indicative of other region-wide processes influencing the results from the EAC countries.[97]

---

[94] We include data from Rwanda only until 2005, given that the country joined the EAC in 2007.

[95] The size of the DHS sample is larger by a magnitude of 8-9.

[96] Going from the median distance of the sample to the border (912km) renders an effect size of -0.287 (=0.667*0.43).

[97] Four out of five estimates imply negative welfare changes at core agglomerations.

Table 4: Placebo Tests – DiD in contiguous Countries

Panel a)	Afrobarometer: Consumption	Afrobarometer: Income	Afrobarometer: Agglomeration	DHS: Consumption	DHS: Income	DHS: Agglomeration
	Freq. gone without: [Wat./Food/Med.] (0-4)	Employed Work (0/1)	Population Density (sdz.)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.25]	[0.24]	[0.00]	[3.20]	[0.12]	[0.00]
EAC Border (0-1) * EAC 1[t ≥ 2001]	0.179	0.062	-0.202	0.844***	0.007	-0.102
	(0.143)	(0.075)	(0.249)	(0.132)	(0.037)	(0.140)
Core Agglom. (0/1) * EAC 1[t ≥ 2001]	-0.142	-0.040	0.725**	0.122	-0.110	-0.249
	(0.103)	(0.041)	(0.292)	(0.164)	(0.077)	(0.287)
Observations	28,541	18,994	28,573	200,133	110,361	233,841
R-Squared	0.09	0.19	0.22	0.29	0.14	0.33
R-Squared -Within	0.05	0.14	0.20	0.28	0.13	0.31

Panel b)	Afrobarometer (Income)	Afrobarometer (Income)	Afrobarometer (Income)	DHS (Income)	DHS (Income)	DHS (Income)
	Worked in last Year (0/1)	Occupation Type (Agr.-Worker-Prof.) (1-3)	*Freq. gone without:* [Cash Income] (0-4)	Worked in last Year (0/1)	Occupation Type (Agr.-Worker-Prof.) (1-3)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.49]	[1.62]	[2.12]	[0.64]	[1.36]	[0.37]
EAC Border (0-1) * EAC 1[t ≥ 2001]	0.022	0.027	-0.579***	-0.067	-0.011	-0.042
	0.074	0.138	(0.141)	0.094	0.112	0.115
Core Agglom. (0/1) * EAC 1[t ≥ 2001]	-0.038	-0.016	-0.361***	0.125**	-0.016	-0.048
	0.055	0.068	(0.073)	0.061	0.082	0.113
Observations	19,191	12,715	28,274	190,646	109,283	110,361
R-Squared	0.18	0.31	0.11	0.11	0.27	0.17
R-Squared -Within	0.10	0.27	0.06	0.07	0.23	0.13
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table conducts a 'placebo' analysis by testing for a spatially differentiated effect across contiguous, non-EAC countries within the time frame of the EAC's establishment and expansion. As such, in columns (1) through (3), data come from the Malawi, Mozambique and Zambia Afrobarometer surveys rounds 1 through 7 sampled between 1999 and 2018. In columns (4) through (6), data come from the Ethiopia, Malawi, Mozambique, Rwanda, and Zambia Demographic and Health surveys (DHS) sampled between 2000 and 2019. The sample mean of the respective dependent variable is given in brackets above the estimates. EAC Border (0-1) is the inverse, relative within country distance to the nearest border crossing of a contiguous EAC country. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Addis Abeba, Kigali, Lilongwe, Lusaka, Maputo). EAC 1[t ≥ 2001] switches to one for individuals sampled from the second half of 2001 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The results in each column and panel are produced by a separate regression. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Pre-Trends. One weakness of the datasets employed is the narrow timeframe of pre-integration periods. While a difference-in-difference approach implemented on data shortly before and shortly after may alleviate concerns of other concomitant policies driving the results, it excludes the possibility to net out potentially long-term, unit-specific time trends from the identified treatment effects. Also, it does not allow a test on the pre-intervention evolvement of relative welfare within the countries. Given that this is a crucial test concerning the validity of the results, I circumvent this data restriction by drawing on region-based estimates from the country, i.e. include non-GPS coded survey rounds of the DHS which extend back to 1988 (see section on Data sources). While there is no finely gridded information on the location of respondents available, DHS provides regional-based information, which identifies individuals' residence on the GADM-1 level.[98] I use this information and construct *Core Region (0/1)*, a dummy indicating whether individuals live in the capital city region. Note that while many of the surveys in pre-GPS years do not provide granularity reflecting the GADM-1 level, given their political and economic importance, Nairobi and Dar es Salaam were nonetheless demarcated as their own region at finer levels.[99] As such, the use of this *Core Region (0/1)* dummy likely captures much of what is also measured by the GPS-based *Core Agglom. (0/1)* dummy used in the main estimations.[100] Concerning the definition of border regions, the matter is not as straightforward. Many of the regions defined in the DHS, particularly in the pre-GPS years, could be considered both border as well as interior regions given their vast extent to the inland of countries. Hence a dummy categorization as used for capital cities will not likely suffice to capture true border households. I therefore try to improve upon a simple dummy with the following steps. First, I assign households in all survey rounds a regional correspondence for which boundaries are consistently available from as early on as possible. This yields 7, 20 and 4, regions for the years 1988 until 2020 for Kenya, Tanzania and Uganda respectively.[101] Second, using all available GPS samples, I retrieve the mean, modal as well as median values of road distances within these boundaries as central tendencies of the distribution within them. Third, I assign these values to all households nested within a specific region. Finally, I encode those households living in regions ranking in the 10th percentile of these boundary-based distances within their country as a 1 in the dummy *Border Region (0/1)*. Under the assumption that the distribution within regions has not dramatically changed between 1988 and the latter two decades, this arguably allows a more precise ordering of border to non-border regions within the sample.

Table 5 presents the result using these two region-based indicators, using the mean as the central tendency for the *Border Region (0/1)* dummy.[102] Panel a) relates to Table 1 and shows the aggregate difference-in-differences estimates before and after the establishment of the EAC, with the omitted

---

[98] See https://spatialdata.dhsprogram.com/boundaries/.
[99] For Kampala in Uganda, this was only done in the 1988-1989 survey for non-GPS surveys.
[100] For instance, until 2002, Uganda was located in the "central" region from which point on it was its own district.
[101] For one round, the 1991-92 Tanzania DHS, I am confined to 6 regions.
[102] The results are robust to using the median also.

time period, i.e. the reference group from 1988-2000. Panel b) explicitly tests for pre-trends by introducing a *pre-EAC 1[1996-2000]* dummy which tests for differential changes in border-, respectively, capital city regions in the across to period shortly before the establishment of the EAC.

Table 5: Region-Based Estimates & Pre-Tests

	Dependent Variable					
	DHS (Region-based)					
	Consumption	Income	Agglomeration	Income		
Panel a)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.22]	[0.19]	[0.00]	[0.73]	[1.44]	[0.57]
Border Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1988-2000]	0.021	-0.025	-0.055	-0.082	-0.030	-0.114*
	(0.170)	(0.019)	(0.105)	(0.063)	(0.045)	(0.057)
Core Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1988-2000]	0.386***	0.117***	2.268*	-0.037	0.137***	0.062*
	(0.122)	(0.021)	(1.235)	(0.054)	(0.042)	(0.033)
Observations	258,820	104,440	282,866	236,646	142,478	136,163
R-Squared	0.28	0.16	0.51	0.21	0.25	0.17
R-Squared -Within	0.27	0.10	0.49	0.18	0.16	0.07
Panel b)						
Border Region (0/1) * pre-EAC 1[1996-2000] pre-EAC 0[1988-1995]	0.098	0.040	0.126	-0.296**	0.037	-0.013
	(0.125)	(0.029)	(0.105)	(0.121)	(0.057)	(0.078)
Border Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1988-1995]	0.088	-0.001	-0.003	-0.304**	-0.008	-0.120
	(0.200)	(0.025)	(0.089)	(0.134)	(0.067)	(0.083)
Core Region (0/1) * pre-EAC 1[1996-2000] pre-EAC 0[1988-1995]	0.440*	0.066***	0.275	-0.245*	0.167*	-0.049
	(0.223)	(0.022)	(0.259)	(0.121)	(0.088)	(0.064)
Core Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1988-1995]	0.652***	0.146***	2.365*	-0.211	0.214***	0.042
	(0.229)	(0.016)	(1.314)	(0.137)	(0.062)	(0.055)
Observations	258,820	104,440	282,866	236,646	142,478	136,163
R-Squared	0.28	0.16	0.60	0.21	0.25	0.17
R-Squared -Within	0.27	0.10	0.58	0.18	0.16	0.07
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table makes use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999 and additionally conducts 'pre-tests' towards the difference-in-differences approach. The data thereby come from the full sample of Kenya, Tanzania, and Uganda DHS surveys sampled between 1989 and 2004, making use of AIS, KAP and MIS rounds as well. The sample mean of the respective dependent variable is given in brackets above the estimates. Border Region (0/1) switches to one for individuals living in a region with a median road distance to EAC border cossings below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region (0/1) is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). EAC 1[t ≥ 2001] switches to one for individuals sampled from the second half of 2001. Pre-EAC 1[1996-2000] switches to one for individuals sampled in survey years between 1996 and including 2000. As such, in panel a), the reference group of the estimates are comprised of individuals sampled in the full pre-EAC period, i.e. from 1991 to 2000, while in panel b), the reference group is formed by individuals sampled between 1991 and including 1995. Hence, the DiD estimate on 'pre-EAC' in panel b) represents the pre-test. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for clustering at the 'region' level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

The reference group for all estimates shown in this panel has therefore changed to pre-EAC 0[1988-1995]. Panel a) confirms the result shown in Table 1 from the GPS-based measurements. There are no indications of a positive relative household welfare change across all tested outcomes.[103] And further, we see the same strongly positive effects of welfare for households living in the core agglomerations.[104] Turning to the test of parallel trends in panel b), we see weak indication of an unequal trend of core vs. peripheral regions. However, only one of the estimates is significant at a conventional 5%-levels (column 2), and more than half in size compared to the effect in the EAC period. I conduct the same estimations with the placebo-countries analyzed in Table 4, and the results are reported in Table B9. As expected from the earlier findings of the contiguous countries, the results do not provide consistent evidence of a growing spatial inequality, which further corroborates the notion of a growing trend towards economic inequality in the larger East African region. Appendix Tables B10 and B11 also provide the results on the temporally disaggregated DiD effects. All of the previous conclusions remain.

Spatial Sorting. The last validity test offered addresses concerns about potential spatial sorting of individuals within the intervention years, which would systematically influence the measured treatment. Specifically, there exists the possibility that skilled individuals positively select, i.e. move into border-, respectively, capital city regions within treatment years in attempt to profit from the policy. This would lead to an upward bias of the results. Given that most of the results are drawn from cross-sectional households, entirely excluding this possibility is not possible. However, I offer three-part evidence that this may be unlikely. First, using regression specification (1), I test for the duration of having lived in the current residence KHDS survey to test for differential migration into border or capital city regions following the establishment of the EAC as well as the common market (CM) protocol. The results are shown in column 6, panel b) of Table B3 in the Appendix. There is no evidence on differential migration across regions induced by the EAC. Second, as seen in the robustness checks, the results are (highly) robust to excluding "post-EAC migrants", i.e. individuals that moved to the respective region of residence less than 3 years before the EAC was operational. Importantly, this is the case for the results of the KHDS panel survey. Secondly, using the KHDS panel survey once more, I test for spatial sorting of high-skilled individuals into the capital cities of Dar es Salaam and Kampala by regressing an indicator of having moved to these cities onto personal characteristics such as education, gender and age as well as country-time fixed effects. I also regress a

[103] Indeed, column 6 shows evidence of a negative effect of cash employment.

[104] Note that *Population Density*, as geography-based outcome variable, is constructed in the same fashion as the region-based distances.

dummy of living in capital cities on this set of covariates. In both cases, I find no systematic evidence of selection into economic hubs, with only the coefficient on education marginally significant (p values of [0.097] and [0.078], respectively) and small increases in the likelihood of 0.5 and 0.9 percentage points for each year of schooling.[105] And third, note that all of our main results include individual-level controls which may correlate with such a decision, i.e. education, age and gender, which do not alter the results dramatically (compare the baseline coefficients to panel (b) in Tables B2 through B4, respectively). Lastly, in the extensions discussed in the next paragraph, column (1) tests for the main motivation of migrating to the current place of residence for KHDS sample respondents. The effect does not show evidence for more or less economically motivated migration.

In sum, these results do not provide evidence in favor of a negative selection out of, and a positive selection into border- as well as capital city regions, respectively, which could drive the results identified in the main results.[106] And while neither of the validity tests can completely eradicate concerns about a potential ongoing trend, they do not undermine the findings to a degree which casts concerns about the nature of the main results, nor the validity of the identification strategy.

Extensions. I provide three extensions to the existing set of results, which provide potential insights into the nature of the findings. First, I test for a heterogeneous effect across the three EAC borders. I do so by replacing the continuous measure of border distance *EAC Border (0-1)* with dummies switching to 1 for households living within 100km of border distance to the three country-border pairs from both country directions. As such, I again estimate regression equation (1), albeit with separate difference-in-differences effects for *TZA-UGA*, *TZA-KEN*, and *KEN-UGA*, together with the *Core Agglom. (0/1)*. Note that this also constitutes a test on a parametric specification and identification of effects. The results are presented in Table B12. The findings provide no systematic evidence of heterogeneity, nor a prior misspecification of treatment. Except for the negative differential effect of having employed work at the Tanzania-Kenya border (column 9 and 10), as well as large negative effect of population density at the Tanzania-Uganda border (columns 11 and 12). Given the non-results produced from the KHDS sample, the latter indicates a potential negative effect rom the Ugandan side of this border.

The second extension makes use of the opinion polling of the Afrobarometer and KHDS surveys to evaluate (potentially altered) sentiments towards free trade and individuals' subjective wellbeing. Table B13 presents these results showing the component of the Afrobarometer in panel a)

---

105 Results can be obtained from the author.

106 Of course, many of these tests rely on the recall of individuals, which constitutes a general weakness of household surveys.

and the results drawn from the KHDS survey in panel b). Concerning subjective wellbeing, there is no uniform direction of differential effects across space. While individuals at borders evaluate their life satisfaction as worse than 5 years prior, they assess the general standard of living of people as higher.[107] The rest of the variables tested are usually found in only one of the survey rounds, which is why there is no DiD estimate possible. In these cases, we test simple differences between border- and core agglomeration regions to the rest of the country, having arguably benefited more from trade integration. While individuals deem it as easier to cross international borders in order to work and trade in foreign countries, the rest of the tested opinions show for non-significant differences in border regions.[108] Most importantly, individuals situated at borders, or in capital cities at that, do not support the free movement of labor across countries (column 3), nor do they assess the EAC or the African Union (AU) as vehicles promoting trade as more or less helpful to their country (column 3).[109] And lastly, individuals in capital cities show for a higher tendency to dislike immigrants or foreign workers as neighbors.[110] This may be evidence for increased experience with competition on the labor market, but given the general insignificance of the other results, this result is not further interpreted. Moving to the results on simple differences in the Kagera Health and Development Survey, we see

[107] Specifically, it sets a prior to compare the current living standards to the former military rule. The survey question reads: "We are going to compare our present system of government with the former system of military rule. Please tell me if the following things are better or worse now than they used to be. People have an adequate standard of living." The response values range from 'Much worse' (1), 'Somewhat worse' (2), 'No change' (3), 'Somewhat better' (4) and 'Much better' (5). I remove the observations valued 'Don't know'. Regarding the question in column (1), the survey question reads: "When you look at your life today, how satisfied do you feel compared with five years ago?". The response values range from 'Much less satisfied' (1), 'Slightly less satisfied' (2), 'About the same' (3), 'Slightly more satisfied' (4) and 'Much more satisfied' (5). I remove the observations valued 'Don't know'. In Tanzania, this question is asking the respondents to compare their life to one year ago.

[108] The survey question reads: "In your opinion, how easy or difficult is it for people in [West/South/East/North/Central] Africa to cross international borders in order to work or trade in other countries, or haven't you heard enough to say?". The response values range from 'Very difficult' (1), 'Difficult' (2), 'Easy' (3), 'Very Easy' (4) and 'Never try' (7). I remove the observations valued 'Never Try'.

[109] The survey question of column (3) reads: "Which of the following statements is closest to your view? Choose Statement 1 or Statement 2. Statement 1: People living in [West/South/East/North/Central] Africa should be able to move freely across international borders in order to trade or work in other countries. Statement 2: Because foreign migrants take away jobs, and foreign traders sell their goods at very cheap prices, governments should protect their own citizens and limit the cross-border movement of people and goods." The response values range from 'Agree very strongly with Statement 1' (1), 'Agree with Statement 1' (2), 'Agree with Statement 2' (3), 'Agree very strongly with Statement 2' (4), 'Agree with Neither' (5) and 'Don't know' (7). I recode 5 to represent the median value. I remove the observations valued 'Don't know'.

The survey question on the variable in column (4) and (5) reads: In your opinion, how much do each of the following do to help your country, or haven't you heard enough to say? [EAC/African Union]. The response values range from 'Don't help' (0), 'Help a little' (1), 'Help somewhat' (2), 'Help a lot' (3) and 'Don't know' (7). I remove the observations valued 'Don't know'.

[110] The survey question reads: "For each of the following types of people, please tell me whether you would like having people from this group as neighbors, dislike it, or not care: Immigrants or foreign workers." 'Strongly dislike' (1), 'Somewhat dislike' (2), 'Would not care' (3) and 'Somewhat like' (4), 'Strongly like' (5) and 'Don't know' (9). I remove the observations valued 'Don't know'.

some indication of what is found in the main results.[111] That is, individuals in border regions have a lower likelihood of having had paid (formal) employment after migrating there. Going from the median border distance in the sample () to respondents directly at the border decreases the likelihood of paid (formal) employment by 4.9 (3.5) percentage points The rest of the results indicating higher or lower subjective household life satisfaction or wealth are generally insignificant with the exception of a positive effect on the current assessment of the households as rich rather than poor.

The third and last extension, which to the main results is a last indirect test which tries to shed light on the non-effects shown across border distance. Without reference to NEG theory, studies have shown that following trade, the pass-through of price changes following trade liberalization decays strongly in border distance (see e.g. Nicita 2009 for the case of Mexico in NAFTA). To test for such a distance penalty in price pass through, I make use of the price questionnaire in the Kagera Health and Development Survey of rounds 1 and 2 which provide market prices of various food items. The prices are expressed per unit (e.g. per kilogram) and averaged across rainy and dry seasons to control periodic fluctuations. To test for border pass through I estimate regression equation (1), and assess the logged price of four homogenous, heavily consumed and, of course, traded goods across the three countries and check for nominal price differences before and after the establishment of the EAC with regards to border distance, and as always, core agglomerations. The results are presented in Appendix Table B14. As is seen, there are almost no differential effects of price changes across border distance. Only the price for Millet (a cereal grain), decreased slightly more at survey locales closer to borders. Finally, we see the now common effect for core agglomerations, which show statistically significant price decreases for all items tested. The relative size of the effects is in the range of 2-8%. This is suggestive evidence of what the NEG model predicts, i.e. a lower price index in agglomerations as a centripetal force. Why there is no positive price change at borders is up for debate. What can be said is that the results of similar tests on the integration of markets across (East) Africa is that borders remain a hinderance, even for country-pairs in a trade union (see Versailles 2012; Aker et al. 2014).

[111] There are weak findings for an increased likilhood of a any type of activity besides housework.

VI. CONCLUSION

This paper investigates the impact of the re-establishment of the East African Community (EAC) on household welfare using three distinct sets of longitudinal, geo-referenced household-level surveys from the three founding members Kenya, Tanzania and Uganda. I formally derive the potential impact of the EAC on households from a canonical New Economic Geography (NEG) model with heterogenous intra-national space, i.e. a quantitative spatial equilibrium, and test the predictions through a difference-in-differences specification with treatment intensity given by households' road distance to internal EAC border crossings. I therefore treat the re-establishment of the EAC in 2001 – and the expansion to a customs union and common market in 2005 and 2010, respectively – as a regional policy intervention having differential effects on individual households governed by their geo-spatial location within the countries.

My results show that households and individuals living closer to the internal EAC border did not experience a relative increase in welfare following the re-establishment, as measured by an array of consumption indices as well as intensive and extensive employment outcomes. Rather, the results hint at the strengthened concentration of economic activity, as evinced by the strong differential and economically relevant (short-run) increases in household welfare across the measured dimensions as well as subsequent inflows of population (density). Given the temporal persistence of the findings, and the subsequent population inflows, my results may be indicative of a "bang-bang" distribution following reductions in transport costs, whereby a long-run stable equilibrium is characterized by stark and growing regional inequalities in welfare. Whether this has set in or not, the theoretical as well as empirical findings do not support the hypothesis of Krugman & Elizondo (1996), who famously suggested that the reduction in trade costs will favor a dispersion of the formerly highly concentrated economic activity of developing countries.

VII. References

Ades, A. F., and E. L. Glaeser. 1995. 'Trade and Circuses: Explaining Urban Giants'. *The Quarterly Journal of Economics* 110 (1): 195–227.

Afrobarometer, [Kenya, Tanzania, Uganda] [Round 1, 2, 3, 4, 5, 6, 7]. 2019. 'Afrobarometer Data'.

AidData. 2017. 'Geocoding Methodology, Version 2.0.1.' Williamsburg, VA: AidData at William & Mary.: AidData Research and Evaluation Unit.

Aker, Jenny C., Michael W. Klein, Stephen A. O'Connell, and Muzhe Yang. 2014. 'Borders, Ethnicity and Trade'. *Journal of Development Economics* 107 (March): 1–16.

Andreoni, Antonio, Julian Boys, and Ole Therkildsen. 2022. 'The Political Economy of "Specialism" in Tanzania: How to Make Export Processing Zones Work via Conditional Special Licensing'. In . Anti-Corruption Evidence Working Paper 046. London, UK.: SOAS University of London. Anti-Corruption Evidence (ACE) Research Consortium.

Atkin, David, and Dave Donaldson. 2015. 'Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs'. w21439. Cambridge, MA: National Bureau of Economic Research.

Barros, Laura, and Inmaculada Martínez-Zarzoso. 2022. 'Systematic Literature Review on Trade Liberalization and Sustainable Development'. *Sustainable Production and Consumption* 33 (September): 921–31.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. 2006. 'Kagera Health and Development Survey 2004 Basic Information Document'. The World Bank, Washington, DC.

Behrens, Kristian, Carl Gaigné, Gianmarco I P Ottaviano, and Jacques-François Thisse. 2003. 'Inter-Regional and International Trade: Seventy Years After Ohlin'. In *CEPR Discussion Paper No. 4065*. Paris & London: CEPR Press.

Behrens, Kristian, Carl Gaigné, Gianmarco I.P. Ottaviano, and Jacques-François Thisse. 2006. 'Is Remoteness a Locational Disadvantage?' *Journal of Economic Geography* 6 (3): 347–68.

———. 2007. 'Countries, Regions and Trade: On the Welfare Impacts of Economic Integration'. *European Economic Review* 51 (5): 1277–1301.

Bickenbach, Frank, Eckhardt Bode, Peter Nunnenkamp, and Mareike Söder. 2016. 'Night Lights and Regional GDP'. *Review of World Economics* 152 (2): 425–47.

Blanchard, O., and Lawrence F. Katz. 1992. 'Regional Evolutions'. *Brooking Papers on Economic Activity* 23 (1): 1–76.

Bluhm, Richard, and Gordon C. McCord. 2022. 'What Can We Learn from Nighttime Lights for Small Geographies? Measurement Errors and Heterogeneous Elasticities'. *Remote Sensing* 14 (5): 1190.

Borraz, Fernando, Daniel Ferrés, and Máximo Rossi. 2013. 'Assessment of the Distributive Impact of National Trade Reforms in Brazil'. *The Journal of Economic Inequality* 11 (2): 215–35.

Brakman, Steven, Harry Garretsen, and Charles van Marrewijk. 2019. 'New Economic Geography After Thirty Years'. In *Handbook of Regional Science*, edited by Manfred M. Fischer and Peter Nijkamp, 1–23. Berlin, Heidelberg: Springer Berlin Heidelberg.

Brakman, Steven, Harry Garretsen, and Charles van Marrewijk. 2020. *An Introduction to Geographical and Urban Economics: A Spiky World*. Third edition. Cambridge, United Kingdom: Cambridge University Press.

Breinlich, Holger, Gianmarco I.P. Ottaviano, and Jonathan R.W. Temple. 2014. 'Regional Growth and Regional Decline'. In *Handbook of Economic Growth*, 2:683–779. Amsterdam: Elsevier.

Brülhart, Marius. 2011. 'The Spatial Effects of Trade Openness: A Survey'. *Review of World Economics* 147 (1): 59–83.

Brülhart, Marius, Olivier Cadot, and Alexander Himbert. 2017. 'Trade Integration and Spatially Balanced Development: Implications for Uganda and Rwanda'. *International Growth Centre Working Paper*, no. F-43406-CCP-1.

Brülhart, Marius, Céline Carrère, and Federico Trionfetti. 2012. 'How Wages and Employment Adjust to Trade Liberalization: Quasi-Experimental Evidence from Austria'. *Journal of International Economics* 86 (1): 68–81.

Brülhart, Marius, Matthieu Crozet, and Pamina Koenig. 2004. 'Enlargement and the EU Periphery: The Impact of Changing Market Potential'. *The World Economy* 27 (6): 853–75.

Cadot, Olivier, Alexander Himbert, and Marie-Agnès Jouanjean. 2015. 'Trade Facilitation and Concentration: Evidence from Sub-Saharan Africa'. *Shaping Policy for Development - ODI Research Reports*.

Cali, M. 2014. 'Trade Boom and Wage Inequality: Evidence from Ugandan Districts'. *Journal of Economic Geography* 14 (6): 1141–74.

Castilho, Marta, Marta Menéndez, and Aude Sztulman. 2012. 'Trade Liberalization, Inequality, and Poverty in Brazilian States'. *World Development* 40 (4): 821–35.

CIESIN. 2017. 'Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11'. Center For International Earth Science Information Network-CIESIN-Columbia University. Socioeconomic Data and Applications Center (SEDAC).

Conley, T.G. 1999. 'GMM Estimation with Cross Sectional Dependence'. *Journal of Econometrics* 92 (1): 1–45.

Conley, Timothy G. 2010. 'Spatial Econometrics'. In *Microeconometrics*, edited by Steven N. Durlauf and Lawrence E. Blume, 303–13. London: Palgrave Macmillan UK.

Crozet, Matthieu, and Pamina Koenig. 2004a. 'Trade Liberalization and the Internal Geography of Countries'. In *Multinational Firms' Location and the New Economic Geography*, by Jean-Louis Mucchielli and Thierry Mayer, 3311. Edward Elgar Publishing.

———. 2004b. 'EU Enlargement and the Internal Geography of Countries'. *Journal of Comparative Economics* 32 (2): 265–79.

De Weerdt, Joachim, Kathleen Beegle, Helene B. Lilleor, Stefan Dercon, Kalle Hirvonen, Martina Kirchberger, and Sofya Krutilov. 2010. 'Kagera Health and Development Survey 2010: Basic Information Document. Study Paper 46'. Copenhagen: Rockwool Foundation Working.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore, MD: Published for the World Bank [by] Johns Hopkins University Press.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. 'Monopolistic Competition and Optimum Product Diversity'. *The American Economic Review* 67 (3): 297–308.

EAC. 1999. 'Treaty for the Establishment of the East African Community'. *African Yearbook of International Law Online / Annuaire Africain de Droit International Online* 7 (1): 419–509.

———. 2015. 'Report of the Committee on Communication, Trade and Investments for the on-Spot Assessment on the EAC One Stop Border Posts (OSBPs) in EAC Partner States'. Arusha, Tanzania: East African Legislative Assembly.

Eberhard-Ruiz, Andreas, and Alexander Moradi. 2019. 'Regional Market Integration in East Africa: Local but No Regional Effects?' *Journal of Development Economics* 140 (September): 255–68.

Edmonds, Eric V, Nina Pavcnik, and Petia Topalova. 2010. 'Trade Adjustment and Human Capital Investments: Evidence from Indian Tariff Reform'. *American Economic Journal: Applied Economics* 2 (4): 42–75.

Emran, M. Shahe, and Zhaoyang Hou. 2013. ‘Access to Markets and Rural Poverty: Evidence from Household Consumption in China’. *Review of Economics and Statistics* 95 (2): 682–97.

Erten, Bilge, Jessica Leight, and Fiona Tregenna. 2019. ‘Trade Liberalization and Local Labor Market Adjustment in South Africa’. *Journal of International Economics* 118 (May): 448–67.

European Commission. 2019. *Description of the GHS Urban Centre Database 2015: Public Release 2019 : Version 1.0.* LU: European Commission Joint Research Centre. Publications Office.

Ezcurra, Roberto, and Alba Del Villar. 2021. ‘Globalization and Spatial Inequality: Does Economic Integration Affect Regional Disparities?’ *The Annals of Regional Science* 67 (2): 335–58.

Ezcurra, Roberto, and Andrés Rodríguez-Pose. 2014. ‘Trade Openness and Spatial Inequality in Emerging Countries’. *Spatial Economic Analysis* 9 (2): 162–82.

Farr, Tom G., Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, et al. 2007. ‘The Shuttle Radar Topography Mission’. *Reviews of Geophysics* 45 (2): RG2004.

Feyrer, James. 2019. ‘Trade and Income—Exploiting Time Series in Geography’. *American Economic Journal: Applied Economics* 11 (4): 1–35.

Fick, Stephen E., and Robert J. Hijmans. 2017. ‘WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas’. *International Journal of Climatology* 37 (12): 4302–15.

Frankel, Jeffrey A, and David Romer. 1999. ‘Does Trade Cause Growth?’ *American Economic Review* 89 (3): 379–99.

Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. 2001. *The Spatial Economy: Cities, Regions, and International Trade*. 1. paperback ed. Cambridge, Mass.: MIT Press.

Fukase, Emiko. 2013. ‘Export Liberalization, Job Creation, and the Skill Premium: Evidence from the US–Vietnam Bilateral Trade Agreement (BTA)’. *World Development* 41 (January): 317–37.

GADM. 2020. ‘Global Administrative Areas [Digital Geospatial Data]’. Center for Spatial Sciences at the University of California, Davis.

Gibson, John, Susan Olivia, and Geua Boe-Gibson. 2020. ‘NIGHT LIGHTS IN ECONOMICS: SOURCES AND USES [1]’. *Journal of Economic Surveys* 34 (5): 955–80.

Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li. 2021. ‘Which Night Lights Data Should We Use in Economics, and Where?’ *Journal of Development Economics* 149 (March): 102602.

Giovannetti, Giorgia, Enrico Marvasi, and Arianna Vivoli. 2021. ‘The Asymmetric Effects of 20 Years of Tariff Reforms on Egyptian Workers’. *Economia Politica* 38 (1): 89–130.

Giovannetti, Giorgia, Marco Sanfilippo, and Arianna Vivoli. 2022. ‘Import Tariff Liberalization, Employment, and Gender in Ethiopia’. *The Journal of Development Studies* 58 (11): 2366–92.

Goldberg, Pinelopi Koujianou, and Nina Pavcnik. 2007. ‘Distributional Effects of Globalization in Developing Countries’. *Journal of Economic Literature*, 75.

Hanson, Gordon. 1994. ‘Localization Economies, Vertical Organization and Trade’. *American Economic Review* 86 (5).

Hanson, Gordon H. 1997. ‘Increasing Returns, Trade and the Regional Structure of Wages’. *The Economic Journal* 107 (440): 113–33.

Hazlewood, Arthur. 1979. ‘THE END OF THE EAST AFRICAN COMMUNITY: WHAT ARE THE LESSONS FOR REGIONAL INTEGRATION SCHEMES ?’ *JCMS: Journal of Common Market Studies* 18 (1): 40–58.

Helpman, Elhanan. 1998. 'The Size of Regions'. *Topics in Public Economics: Theoretical and Applied Analysis*, 33–54.

Henderson, J. V. 1974. 'The Sizes and Types of Cities'. *The American Economic Review* 64 (4): 640–56.

Henderson, J.Vernon. 1982. 'Systems of Cities in Closed and Open Economies'. *Regional Science and Urban Economics* 12 (3): 325–50.

Henderson, Vernon J., Tim Squires, Adam Storeygard, and David Weil. 2018. 'The Global Distribution of Economic Activity: Nature, History, and the Role of Trade'. *The Quarterly Journal of Economics* 133 (1): 357–406.

Hirte, Georg, Christian Lessmann, and André Seidel. 2020. 'International Trade, Geographic Heterogeneity and Interregional Inequality'. *European Economic Review* 127 (August): 103427.

IMF. 2023. 'Direction of Trade Statistics (DOTS)'. 2023.

Kaahwa, Wilbert T. K. 2003. *East African Community: EAC Treaty and Challenges to the Community*. Occasional Papers (East African Community Secretariat), no. 3. Arusha, Tanzania: East African Community Secretariat in conjunction with the German Agency for Technical Co-operation.

Kanbur, Ravi, and Xiaobo Zhang. 2005. 'Fifty Years of Regional Inequality in China: A Journey Through Central Planning, Reform, and Openness: FIFTY YEARS OF REGIONAL INEQUALITY IN CHINA'. *Review of Development Economics* 9 (1): 87–106.

Kenya Central Bureau of Statistics. 2001. 'Statistical Abstract: 2001'. Nairobi, Kenya.: Kenya Central Bureau of Statistics, Ministry of Finance and Planning.

———. 2003. 'Statistical Abstract: 2003'. Nairobi, Kenya.: Kenya Central Bureau of Statistics, Ministry of Finance and Planning.

Kenya National Bureau of Statistics. 2011. 'Statistical Abstract: 2011'. Nairobi, Kenya.: Kenya National Bureau of Statistics.

———. 2013. 'Basic Report on The 2010 Census of Industrial Production (CIP)'. Nairobi, Kenya.: Kenya National Bureau of Statistics, Ministry of Industrialization.

Kinyondo, Abel, Carol Newman, Finn Tarp, and UNU-WIDER. 2016. *The Role and Effectiveness of Special Economic Zones in Tanzania*. 122nd ed. Vol. 2016. WIDER Working Paper. UNU-WIDER.

Kis-Katos, Krisztina, and Robert Sparrow. 2015. 'Poverty, Labor Markets and Trade Liberalization in Indonesia'. *Journal of Development Economics* 117 (November): 94–106.

Krugman, P., and A. J. Venables. 1995. 'Globalization and the Inequality of Nations'. *The Quarterly Journal of Economics* 110 (4): 857–80.

Krugman, Paul. 1980. 'Scale Economies, Product Differentiation, and the Pattern of Trade'. *The American Economic Review* 70 (5): 950–59.

———. 1991. 'Increasing Returns and Economic Geography'. *Journal of Political Economy* 99 (3): 483–99.

Krugman, Paul, and Raul Livas Elizondo. 1996. 'Trade Policy and the Third World Metropolis'. *Journal of Development Economics* 49: 137–50.

Krugman, Paul R. 1979. 'Increasing Returns, Monopolistic Competition, and International Trade'. *Journal of International Economics* 9 (4): 469–79.

Lehner, Bernhard, and Petra Döll. 2004. 'Development and Validation of a Global Database of Lakes, Reservoirs and Wetlands'. *Journal of Hydrology* 296 (1–4): 1–22.

Mansori, Kashif S. 2003. 'The Geographic Effects of Trade Liberalization with Increasing Returns in Transportation'. *Journal of Regional Science* 43 (2): 249–68.

Marchand, Beyza Ural. 2012. 'Tariff Pass-through and the Distributional Effects of Trade Liberalization'. *Journal of Development Economics* 99 (2): 265–81.

Martínez-Zarzoso, Inma, Leila Baghdadi, and Hendrik Wiard Kruse. 2016. 'Is Tunisian Trade Policy Pro-Poor?' In *1033*.

McCaig, Brian. 2011. 'Exporting out of Poverty: Provincial Poverty in Vietnam and U.S. Market Access'. *Journal of International Economics* 85 (1): 102–13.

McCaig, Brian, and Margaret S. McMillan. 2020. 'Trade Liberalisation and Labour Market Adjustment in Botswana'. *Journal of African Economies* 29 (3): 236–70.

Monfort, Philippe, and Rosella Nicolini. 2000. 'Regional Convergence and International Integration'. *Journal of Urban Economics* 48 (2): 286–306.

Monfort, Philippe, and Tanguy van Ypersele. 2003. 'Integration, Regional Agglomeration and International Trade'. In *CEPR Discussion Paper No. 3752*, 32. Paris & London: CEPR Press.

Mugomba, Agrippah T. 1978. 'Regional Organisations and African Underdevelopment: The Collapse of the East African Community'. *The Journal of Modern African Studies* 16 (2): 261–72.

Nathan Associates. 2011. 'Corridor Diagnostic Study of the Northern and Central Corridors of East Africa'. In *Volume 1: Main Report*. Alexandria, VA, USA.

Natural Earth. 2019. 'Natural Earth. Free Vector and Raster Map Data.'

NGA, Natioal Geospace-Intelligence Agency. 2019. 'World Port Index 27th Edition'. United States Government.

Nicita, Alessandro. 2009. 'The Price Effect of Tariff Liberalization: Measuring the Impact on Household Welfare'. *Journal of Development Economics* 89 (1): 19–27.

Nicita, Alessandro, Marcelo Olarreaga, and Guido Porto. 2014. 'Pro-Poor Trade Policy in Sub-Saharan Africa'. *Journal of International Economics* 92 (2): 252–65.

Nunn, Nathan, and Diego Puga. 2012. 'Ruggedness: The Blessing of Bad Geography in Africa'. *Review of Economics and Statistics* 94 (1): 20–36.

OSM. 2022. *OpenStreetMap Database*. OpenStreetMap contributors. Cambridge, United Kingdom: OpenStreetMap Foundation.

Paluzie, Elisenda. 2001. 'Trade Policy and Regional Inequalities*'. *Papers in Regional Science* 80 (1): 67–85.

Pavcnik, Nina. 2017. 'The Impact of Trade on Inequality in Developing Countries'. w23878. Cambridge, MA: National Bureau of Economic Research.

Porto, Guido G. 2006. 'Using Survey Data to Assess the Distributional Effects of Trade Policy'. *Journal of International Economics* 70 (1): 140–60.

Puga, Diego. 1999. 'The Rise and Fall of Regional Inequalities'. *European Economic Review*, 32.

Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney. 2002. 'The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change: *Cropland Sensitivity to Climate Change*'. *Global Ecology and Biogeography* 11 (5): 377–92.

Rappaport, Jordan, and Jeffrey D. Sachs. 2003. 'The United States as a Coastal Nation'. *Journal of Economic Growth* 8: 5–46.

Redding, Stephen J. 2022. 'Trade and Geography'. In *Handbook of International Economics*, 5:147–217. Elsevier.

Redding, Stephen J., and Esteban Rossi-Hansberg. 2017. 'Quantitative Spatial Economics'. *Annual Review of Economics* 9 (1): 21–58.

Redding, Stephen J, and Daniel M Sturm. 2008. 'The Costs of Remoteness: Evidence from German Division and Reunification'. *American Economic Review* 98 (5): 1766–97.

Rutstein, Shea O., and Kiersten Johnson. 2004. 'The DHS Wealth Index'. In *DHS Comparative Reports No. 6*. Calverton, Maryland, USA: Macro International.

Rutstein, Shea O., and Sarah Staveteig. 2014. 'Making the Demographic and Health Surveys Wealth Index Comparable'. ICF International Rockville, Maryland, USA.

Samuelson, Paul A. 1952. ‘The Transfer Problem and Transport Costs: The Terms of Trade When Impediments Are Absent’. *The Economic Journal* 62 (246): 278.
Schiff, Maurice W., and L. Alan Winters. 2003. *Regional Integration and Development*. Washington, DC : [s.l.]: World Bank ; Oxford University Press.
Smits, Jeroen, and Roel Steendijk. 2015. ‘The International Wealth Index (IWI)’. *Social Indicators Research* 122 (1): 65–85.
Tanzania National Bureau of Statistics. 2004. ‘Employment and Earnings Survey: 2001 Analytical Report’. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Planning, Economy and Empowerment. The United Republic of Tanzania.
———. 2006. ‘Population Census: 2002 Analytical Report’. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Planning, Economy and Empowerment. The United Republic of Tanzania.
———. 2007. ‘Employment and Earnings Survey: 2002 Analytical Report’. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Planning, Economy and Empowerment. The United Republic of Tanzania.
———. 2010. ‘Annual Survey of Industrial Production and Performance: 2008 Statistical Report’. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Industry, Trade and Marketing. The United Republic of Tanzania.
———. 2018. ‘Annual Survey of Industrial Production: 2016 Statistical Report’. Dar es Salaam, Tanzania: National Bureau of Statistics, Ministry of Industry, Trade and Investment. The United Republic of Tanzania.
Thünen, Johann H. von. 1826. *Der Isolierte Staat in Beziehung Auf Landwirtschaft Und Nationalökonomie*. Hamburg: Perthes.
Topalova, Petia. 2010. ‘Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India’. *American Economic Journal: Applied Economics* 2 (4): 1–41.
UFZA. 2022. ‘Uganda Free Zone Authority Annual Report 2021/2022’. Kampala, Uganda.
Uganda Bureau of Statistics. 2003. ‘Uganda Business Register: 2001/2002, Report’. Kampala, Uganda: Uganda Bureau of Statistics, Directorate of Business Industry, Agriculture and Energy Statistics.
———. 2006. ‘Uganda Business Inquiry: 2000/2001’. Entebbe, Uganda: Uganda Bureau of Statistics.
———. 2007. ‘Uganda Business Register: 2006/2007, Report’. Kampala, Uganda: Uganda Bureau of Statistics, Directorate of Business Industry, Agriculture and Energy Statistics.
———. 2012. ‘Statistical Abstract: 2012’. Kampala, Uganda: Uganda Bureau of Statistics.
———. 2016. ‘The National Population and Housing Census 2014 – Main Report’. Kampala, Uganda: Uganda Bureau of Statistics.
UNCTAD. 2022. ‘Trade Analysis and Information System (TRAINS)’. United Nations Conference on Trade and Development.
United Nations Development Programme. 2011. ‘Regional Integration and Human Development: A Pathway for Africa’. New York, NY: United Nations Development Programme.
UNSD. 2022. ‘International Merchandise Trade Statistics (COMTRADE)’. United Nations Statistics Division.
UNU-CRIS. 2019. ‘Regional Integration Knowledge System (RIKS)’.
Ural Marchand, Beyza. 2019. ‘Inequality and Trade Policy: The Pro-Poor Bias of Contemporary Trade Restrictions’. *Review of Income and Wealth* 65 (S1).
Versailles, Bruno. 2012. ‘Market Integration and Border Effects in Eastern Africa’, 48.
Vo, Thang T., and Dinh X. Nguyen. 2020. ‘Impact of Trade Liberalization on Household Welfare: An Analysis Using Household Exposure-to-Trade Indices’. *Social Indicators Research*, September.

Wild, Frederik, and David Stadelmann. 2022. ‘Coastal Proximity and Individual Living Standards: Econometric Evidence from Georeferenced Household Surveys in sub-Saharan Africa’. *Review of Development Economics* 26 (4): 1883–1901.

Winters, L Alan. 2002. ‘Trade Liberalisation and Poverty: What Are the Links?’, 29.

Winters, L. Alan, and Antonio Martuscelli. 2014. ‘Trade Liberalization and Poverty: What Have We Learned in a Decade?’ *Annual Review of Resource Economics* 6 (1): 493–512.

Winters, L Alan, Neil Mcculloch, and Andrew Mckay. 2004. ‘Trade Liberalization and Poverty: The Evidence So Far’. *Journal of Economic Literature*, 64.

World Bank. 2009. *World Development Report 2009: Reshaping Economic Geography*. 2009. Washington, D.C.: World Bank.

———. 2016. ‘Implementation Completion and Results Report (IDA-39300 IDA-45710) for the Northern Corridor Transport Improvement Project’. Report ICR00002771. Washington, D.C: The World Bank.

———. 2020. *The African Continental Free Trade Area: Economic and Distributional Effects*. Washington, DC: World Bank.

———. 2022. ‘Data.Worldbank.Org’. 2022.

World Bank, and University of Dar es Salaam. 1994. *Kagera Health and Development Survey 1991-1994 (Wave 1 to 4 Panel). Ref. TZA_1991_KHDS_v01_M. Dataset Downloaded from the World Bank Microdata Library (Www.Microdata.Worldbank.Org)*. World Bank and University of Dar es Salaam.

———. 2004. *Kagera Health and Development Survey 2004. Ref. TZA_2004_KHDS_v01_M.* World Bank and University of Dar es Salaam.

———. 2010. *Kagera Health and Development Survey 2010*. World Bank and University of Dar es Salaam.

WTO. 2019. ‘East African Community (EAC) Trade Policy Review: Report by the Secreteriat’. WT/TPR/S/384. Geneva, Switzerland: World Trade Organization.

Zeng, Dao-Zhi, and Laixun Zhao. 2010. ‘Globalization, Interregional and International Inequalities’. *Journal of Urban Economics* 67 (3): 352–61.

APPENDIX

"*Regional Market Integration and Household Welfare: Spatial Evidence from the East African Community*"

Appendix A.1

As anticipated in Chapter III, this section provides analytical insights into the simulation results of the main text. As in the NEG tradition, the analysis revolves around checking the stability of two specific equilibria of the model, namely the "spreading" equilibrium, i.e. where the real wage differential $\omega_1/\omega_2 = 1$ and $\lambda_1 = \lambda_2 = 0.50$, and the "agglomerated" equilibrium, where all manufacturing is concentrated in one of the regions such that $\lambda_1 = 1$ and $\omega_1/\omega_2 \geq 1$ or $\lambda_1 = 0$ and $\omega_1/\omega_2 \leq 1$. The analytical evaluation is thereby concerned with assessing the stability of both equilibria to varying internal transport costs, i.e. at which level of T_D spreading is broken and agglomeration sustainable, hence the name "sustain" and "break" analysis. As such, the treatment closely follows the exposition in Chapter 5 of Fujita et al. (2001), albeit for a modified spatial layout and focusing on the influence of external trade costs rather than internal ones only. Note that the analysis is constrained to these two cases because at these points, the system of non-linear equations reduces to a more tractable set which simplifies the analysis. However, given that λ varies all the way from 0 to 1, and choices have to be made on the other parameter values, (the stability of) further equilibria may depend on many such combinations, the main analysis of this paper relies on the numerical simulations of Chapter III in order to give a full picture of the long-run dynamics. As introduced in Chapter III, we compare the analytical results for the setting with *heterogenous* intra-national space to the ones drawn from a $2 + 2$ setting with *homogenous* intra-national space.

a) Symmetry Breaking

We start by analyzing the robustness of a symmetric equilibrium, that is the configuration in which $\lambda_1 = \lambda_2 = 0.50$ and $\omega_1/\omega_2 = 1$. From the discussion in Chapter III, and visually depictable in Figure 3, we know that this equilibrium is stable if migrating in either direction leads to a lower real wage in the destination region than in the origin. Stated more generally, for the symmetric equilibrium to be a stable one, the slope of the total differential with respect to $d\lambda_\mathrm{i}$ has to satisfy:

$$\frac{d\frac{\omega_i}{\omega_j}}{d\lambda_i} \leq 0. \tag{A.1.1}$$

Before we start deriving an expression for (A.1.1), notice that Figure 3, specifically, the differences across Panels A and B as well as C and D, already hold the insight insofar as a symmetric equilibrium can be upheld during trade liberalization within our heterogeneous $2 + 2$ setting. The simulations show that any move away from autarky ($\tau \neq \infty$) also entails a move away from the symmetric distribution of manufacturing as an equilibrium. This is observable by the shift of the cut point to the left (Panel B) and to the right (Panel D). Hence, contrary to Panels A and C, the relative share of manufacturing workforce across regions in the first type of equilibrium is dependent on external transport costs. As such, for the "symmetry breaking" analysis, we are limited to the $2 + 2$ setting with homogenous intra-national space. In this setup, the equal distribution is always a possible equilibrium, independent of the (external) transport costs. This is explicated in the following steps. First note that when $\lambda^{H(ome),F(oreign)} = 0.5$, income $Y^{H,F} = 0.5$ and from this, the wage reduces to $w^{H,F} = 1$, and this is true for all (home and foreign) regions, hence the drop of the indices.[1] We can confirm this by plugging these values into (36) through (44), and solving.

$$Y = \frac{\delta}{2} + \frac{(1-\delta)}{2} = 0.5 \qquad \text{(A.1.2)}$$

$$I = [0.5 + 0.5T_D^{1-\varepsilon} + T_F^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}} \qquad \text{(A.1.3)}$$

Plugging these two results into the wage equation leads to

$$w = \left(0.5I^{\varepsilon-1} + 0.5I^{\varepsilon-1}T_D{}^{1-\varepsilon} + I^{\varepsilon-1}T_F{}^{1-\varepsilon}\right)^{\frac{1}{\varepsilon}}, \text{ or}$$

$$w = \left(\frac{0.5 + 0.5T_D{}^{1-\varepsilon} + T_F{}^{1-\varepsilon}}{I^{1-\varepsilon}}\right)^{\frac{1}{\varepsilon}} = 1 \qquad \text{(A.1.4)}$$

$$\textit{given that } I^{1-\varepsilon} = 0.5 + 0.5T_D^{1-\varepsilon} + T_F^{1-\varepsilon}$$

Note that the subscripts can be dropped as income, price indices and therefore wages are equal in all regions at the symmetric equilibrium. Note that the existence of this equilibrium configuration does not depend on foreign trade costs T_F.[2] From this set of manipulations, specifically (A.1.3) and (A.1.4), it is easily seen that real wages across regions are equal. With these results in mind, we are able to proceed with a crucial simplification in the derivation of the total differential. Namely, that at this symmetric equilibrium, a change in one of the endogenous variables for one region requires the identical change for the other region in the opposite direction (Fujita et al. 2001). This can be

[1] Note that the superscripts H and F indicate identical values for all home (1 and 2) and foreign regions (3 and 4), respectively.

[2] Accordingly, the existence of a symmetric equilibrium in the homogenous $2 + 2$ region case is independent of the foreign labor distribution.

confirmed, for instance, by computing the total derivatives of the two income equations for regions 1 and 2, plugging in the equilibrium values, and checking whether $dY_1 + dY_2 = 0$.

$$dY_1 = d\lambda_1 w_1 \delta + dw_1 \lambda_1 \delta \tag{A.1.5}$$

$$dY_2 = -d\lambda_1 w_2 \delta + dw_2 (1 - \lambda_1)\delta \tag{A.1.6}$$

Where we made us that $\lambda_2 = (1 - \lambda_1)$. If we now plug in the equilibrium values derived in (A.1.2) through (A.1.4) and assuming analogously that $dw_1 = -dw_2$, gives

$$dY_1 = d\lambda_1 \delta + dw_1 0.5\delta \; and \; dY_2 = -d\lambda_1 \delta - dw_1 0.5\delta \tag{A.1.7}$$

$$\text{which satisfies } dY_1 + dY_2 = 0. \tag{A.1.8}$$

This confirms that $dY^H \equiv dY_1 = dY_2$. Hence, the total derivate of the income at the symmetric equilibrium can be finally written as

$$dY^H = d\lambda\delta + \frac{\delta}{2}\, dw \tag{A.1.9}$$

And equally for the foreign country such that $dY \equiv dY^H = dY^F$. This operation can be confirmed for all other equilibrium equations, i.e. for price indices dI and wages dw.[3] Importantly, the same intuition applies to the total differential of the real wage equations also, such that it suffices to assess only the change in the real wage of one of the two foreign or home regions, and (A.1.1) effectively boils down to $d\omega_i^{H.F}/d\lambda_i^{H,F}$, e.g. given by

$$\frac{d\omega_i/d\omega_j}{d\lambda_i} \equiv \frac{d\omega}{d\lambda} = \frac{dw \cdot I^{-\delta} - dI \cdot w \cdot I^{-(1+\delta)} \cdot \delta}{d\lambda} \tag{A.1.10}$$

After some manipulations, which involves plugging in (A.1.9) into the equations for $dI^{H,F}$ and $dw^{H,F}$, and solving these four equations as a system, we arrive at expressions to plug into (A.1.10) which are solely dependent on the exogenous parameter values δ and ε as well the iceberg trade costs T_D and

[3] We do not show them here because they grow relatively large as they are additionally dependent on changes in the endogenous variables of the foreign country. For a full derivation, see the Maple replication script in the Online Appendix.

T_F.[4] As this expression hinges on the totals differentials from all four regions, the expressions is unwieldly compared to the core 2-region NEG model. Hence, to facilitate interpretation, the results from this "break" analysis are provided graphically in Figure A.1.1 for a given set of parameter values and the three levels of external transport costs T_F. Note that we are now also able to assess the stability of this equilibrium to a range of internal transport costs which is not simply set to $T_D = 1.60$ as previously. As in the NEG tradition, Table A.1. additionally provides the "break" values $T_D(B)$ together with the ones derived in b), i.e. the "sustain" values $T_D(S)$, for a range of parameter values δ and ε and the three levels of trade liberalization T_F.

Panel A **Panel B**

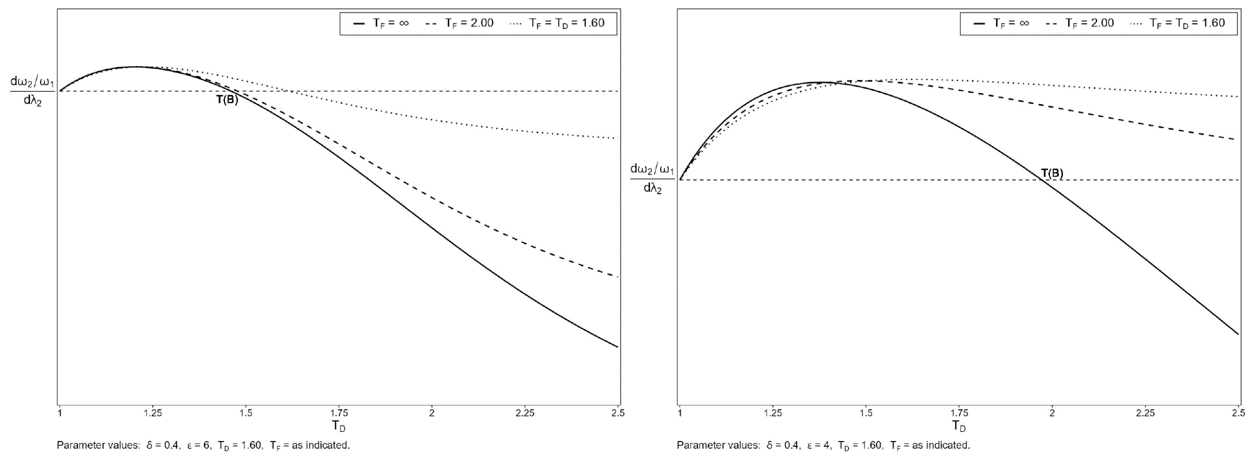


Figure A.1.1: Internal transport costs and symmetric equilibrium

Figure A.1.1 provides the results to the break analysis for the four-region model with homogenous intra-national space, plotting $(d\omega_2/d\omega_1)/d\lambda_2$ across an increasing value of intra-national trade costs T_D for three different values of international trade costs T_F, separately. Note that by symmetry, we are free to choose the direction of effects in any of the two home or foreign regions. We focus on the real wage differential of region 2, however, as it facilitates the comparison with the sustain analysis in b) and given that we have seen the increased draw towards border regions in the simulations discussed in Chapter III. Note first, that with zero transport costs, i.e. $T_D = 1$, there is no difference in the (real) wage across regions such that the total differential is zero at the origin.[5] Increasing the intra-national transport costs from this point on in Panel A shows that up until a level of internal transport costs of $T_D = 1.47$, the symmetric equilibrium is unstable, given that a move away from region 1 increases

[4] Also using $I = \left[0.5 + 0.5T_D^{(1-\varepsilon)} + T_F^{(1-\varepsilon)}\right]^{\frac{1}{1-\varepsilon}}$.

[5] The basic NEG setup hinges on positive transport costs, i.e. a $T_D \neq 1$ which essentially means that economies are not rendered identical in all respects.

real wages at the destination, i.e. in region 2, i.e. the total differential is positive.[6] In other words, in this scenario, the cost- and demand linkages of agglomerating are strong enough to render the cost of serving the demand of region 1 at a distance as profitable. For any increase in transport costs beyond this point, this is not true anymore and manufacturing activity spreads out. The dashed and dotted curves show the effect of trade liberalization, which is to increase the range of transport costs within which a symmetric equilibrium is unstable. Why is this the case? We need to analyze a couple of (countervailing) effects step by step in order to interpret the likely effects (e.g. Crozet and Koenig 2004a; Brülhart et al. 2004; Brülhart 2011). Note a first that liberalizing trade, i.e. decreasing T_F down from prohibitive levels, causes the components in both the price index (40-43) as well as the wage equation (44-47) that are dependent on external markets 3 and 4 to make up a larger component of the overall I and w at the respective location. This results in several dynamics. First, it lowers producers' need to locate close to consumers in the home country as a larger share of their sales come from abroad, i.e. the demand linkage is lowered. Secondly, it analogously decreases consumer's need to locate near producers in the home country as a larger share of their demand now stems from abroad, i.e. the cost linkage is also lowered (Crozet and Koenig 2004b). As such, lowering T_F essentially reduces agglomerating tendencies inside the domestic country which is the well-known result put forward in Krugman and Elizondo (1996). Note that this also means that the moderating force of these cost and demand linkages as given by the internal transport costs T_D is weakened also, as can be depicted by an attenuation of the slopes in Figure A.1.1. But why is agglomeration in this present model more likely then? There are two crucial differences to the model in Krugman and Elizondo (1996) which turn this result around. For one, as Crozet and Koenig (2004a) point out, Krugman and Elizondo not model an immobile agricultural sector the demands of which acting as a spreading force, and secondly, they explicitly model congestion costs of agglomerations (such as rent or commuting). These congestion costs are independent of trade costs, hence decreasing them does not lower the centrifugal tendency of them. On the contrary, the dispersion force of the type of model we employ, immobile farmers, is crucially dependent on the trade costs to serve them. Together with the key result of the original core NEG model (Krugman 1991), i.e. that the strength of the centrifugal force given by these farmers falls faster in (international) transport costs than the strength of the centripetal force, this may display one reason why the result is turned towards agglomeration in our case (see also the discussion in Brülhart 2011). However, there is one further potential reason why agglomeration tendencies may be increased by opening up to external markets, which is increased competition of firms from abroad (Crozet and Koenig 2004b). Remember from Chapter III that the dispersion force stems from increased competition given by the positive relationship between the price index I and the

[6] You can retrieve the precise value from Table A.1.

break-even wage rate firms are able to afford to pay, as seen in (44) through (47). Thereby, in a similar line of argument as above, from the point of producers, decreasing T_F lowers the relative importance of domestic competition, such that sheltering away from local firms is less important given the new competition foreign firms pose (Crozet and Koenig 2004a; Brülhart et al. 2004). By looking at our results in Figure 3 as well as A.1.1, it seems that this decreased competition effect dominates the decreased agglomeration forces. Additionally, when looking at Panel B in Figure A.1.1, we see this effect amplified up to a point where the does not even exist a break point anymore, i.e. $T_D = \emptyset$, and agglomeration is the only long-term stable equilibrium in the case of free trade. Note that the difference between Panel A and B makes intuitive sense, given that the only change between the two is the reduced elasticity of substitution, from $\varepsilon = 6$ to $\varepsilon = 4$. This change increases product differentiation i.e. lowers competition across varieties, and from the pricing rule (19), increases mark-ups. As such, sheltering from local firms is even less important now, which thereby acts as a further agglomerating force.

Notice how these results compare to the simulations in Figure 3, in which we have set $T_D = 1.60$. At this point on the x-Axis (Figure A.1.1), spreading is never sustainable for an $\varepsilon = 4$, as is confirmed by the positive slope in Figure 3 Panel B. And for an $\varepsilon = 6$ only stable when trade liberalization has not fully concluded yet, e.g. a value $2.00 \geq T_F > 1.60$ (Figure 3 Panel A). Table A.1.1 encapsulates these results at one glance. We see that a decrease in the elasticity of substitution consistently shifts the break point to the right, i.e. increases the range of values for which spreading is unsustainable. Notice, also that increase in δ, i.e. an increase in the share of income devoted to manufactures has the identical effect.

How do these deliberations compare to the model with heterogenous intra-national space? Panel B in Figure 3 shows that for the case with a lower product differentiation $(\varepsilon = 6)$, an equilibrium where economic activity is spread out is more likely at higher degrees of trade liberalization than in the homogenous case; albeit with higher shares of economic activity placed at the border. In this case, it seems that the competition effect from abroad does not yet seem to fully dominate the local one and its spreading tendency.[7] Intuitively, firms and consumers now also profit from increased agglomeration, but there is a bias towards agglomerating in the vicinity to the newly accessed markets, i.e. in region 2. This notion is further confirmed seen in Panel D of Figure 3, where, as in the case with homogenous intra-national space, decreased competitive pressures ($\varepsilon = 4$) fully reverses the curve to full agglomeration as the only stable, long-run equilibrium, but now this is more likely to happen at the border region 2, compared to the interior region 1.

[7] Although, notice the slope does in fact decline in Figure 4 Panel B slightly with progressing trade liberalization.

b) Sustainable Agglomeration

We now turn to the "sustain" analysis. Chapter III has already established that the stability of this equilibrium trivially depends on the condition $\omega_i/\omega_j \geq 1$ if $\lambda_i = 1$. As such, we need to derive an expression for the real wage differential at this point which, as in a), depends only on the parameter values δ, ε as well as, importantly, the different types of iceberg trade costs T_D, T_F, T_{DF}, T_{DFD}. For this analysis, we are able to derive analytical solutions for both spatial layouts, i.e. the homogenous as well as the heterogeneous layout of trade costs. As in a), the first step entails plugging in the equilibrium values for λ_i, i.e. $\lambda_1 = \lambda_4 = 1$ and correspondingly, $\lambda_2 = \lambda_3 = 0$, and noting that the wage equations of regions 1 and 4 reduce to $w_1 = w_4 = 1$.[8] To see this, note that the income equations (36) and (39) in this spatial configuration are given by

$$Y_{1,4} = \frac{(1+\delta)}{2} \tag{A.1.11}$$

$$Y_{2,3} = \frac{(1-\delta)}{2} \tag{A.1.12}$$

Note, from this set of four equations, income in region 1 is always higher, which represents the demand (backward linkage) introduced in Chapter II. Correspondingly, the price indices reduce to

$$I_{1,4} = [1 + T_F^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{A.1.13a}$$

$$I_{2,3} = [T_D^{1-\varepsilon} + T_F^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{A.1.14a}$$

And for the model with homogenous intra-national space:

$$I_{1,4} = [1 + T_{DFD}^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{A.1.13b}$$

$$I_{2,3} = [T_D^{1-\varepsilon} + T_{DF}^{1-\varepsilon}]^{\frac{1}{(1-\varepsilon)}} \tag{A.1.14b}$$

The first summand of the price index equations is equal for both spatial layouts (A.1.13a and A.1.13b). The difference lies in the second summand in the price indices. In the model with homogenous intra-

[8] This can be confirmed by guessing $w_1 = 1$, working out (36) through (43) using this value for W_1 and seeing that (44) is indeed 1. Notice that in our heterogeneous 2+2 case, this also entails assuming an equal distribution in the foreign country, i.e. full agglomeration as for instance given by $\lambda_4 = 1$ together with $W_4 = 1$.

nation space (A.1.13a and A.1.14a), the cost of living in the peripheral region is at best equal as that of the agglomerated region, but for transport which satisfy $T_D > 1$ are always higher than in the agglomerated region, given that $\varepsilon > 1$ which we assume by default given our CES utility structure. This is the cost (forward) linkage as described in Chapter I. However, in A.1.13b and A.1.14b, given that $T_{DF} < T_{DFD}$, we cannot readily determine whether cost of living is higher or lower in region 1. Moving on, we plug (A.1.11) through (A.1.14) into the wage equations:

$$w_{1,4} = \left[\frac{Y_{1,4}(1 + T_F^{1-\varepsilon})}{1 + T_F^{1-\varepsilon}} + \frac{Y_{2,3}(T_D^{1-\varepsilon} + T_F^{1-\varepsilon})}{T_D^{1-\varepsilon} + T_F^{1-\varepsilon}}\right]^{\frac{1}{\varepsilon}}, or$$

$$w_{1,4} = \left[\frac{(1+\delta)}{2} + \frac{(1-\delta)}{2}\right]^{\frac{1}{\varepsilon}} = 1 \tag{A.1.15a}$$

And similarly, for heterogeneous intra-national space

$$w_{1,4} = \left[\frac{Y_{1,4}(1 + T_{DFD}^{1-\varepsilon})}{1 + T_{DFD}^{1-\varepsilon}} + \frac{Y_{2,3}(T_D^{1-\varepsilon} + T_{DF}^{1-\varepsilon})}{T_D^{1-\varepsilon} + T_{DF}^{1-\varepsilon}}\right]^{\frac{1}{\varepsilon}}, or$$

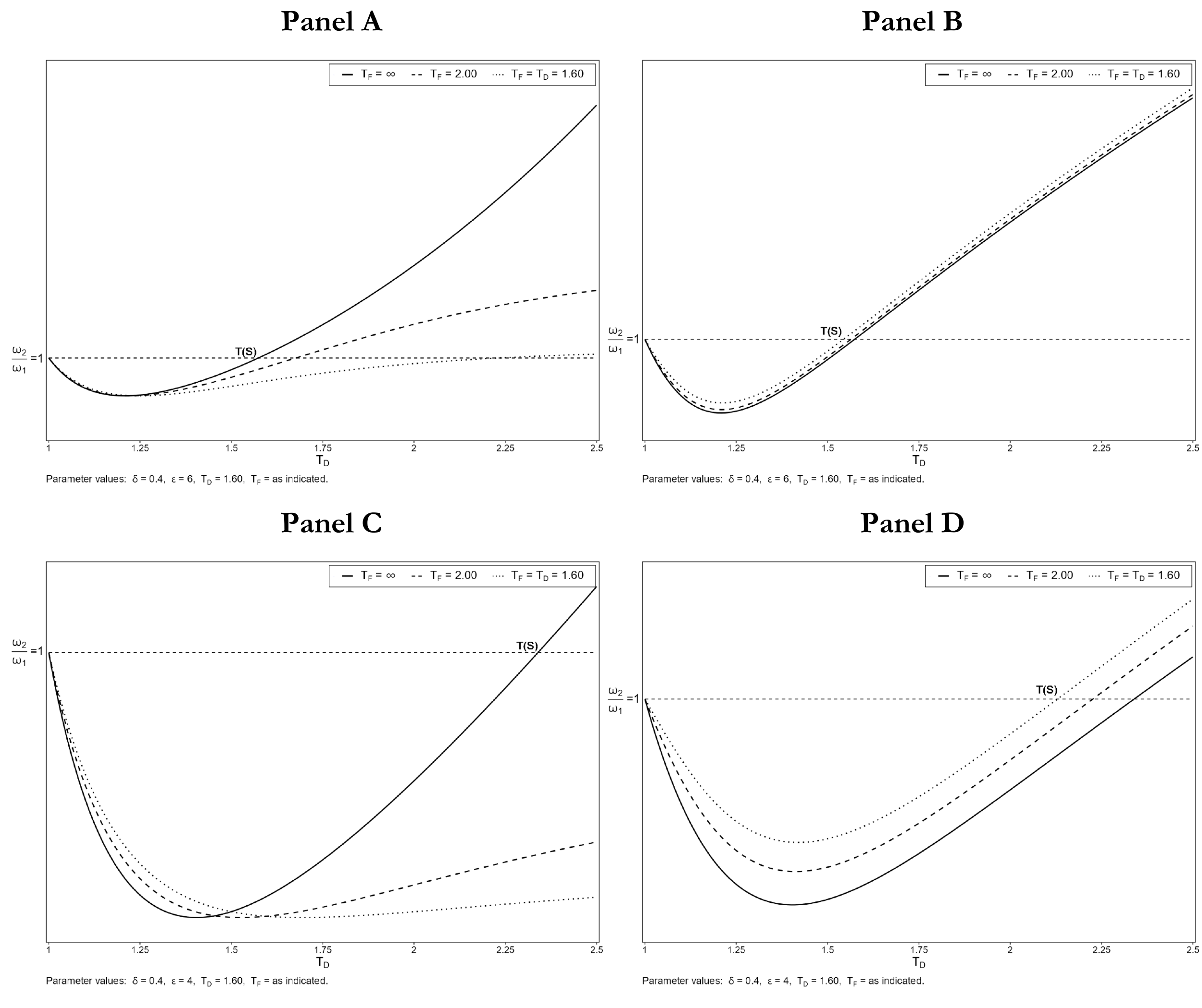
$$w_{1,4} = \left[\frac{(1+\delta)}{2} + \frac{(1-\delta)}{2}\right]^{\frac{1}{\varepsilon}} = 1 \tag{A.1.15b}$$

Where we made use of a similar manipulation as in a), i.e. that $I^{-1-\varepsilon} \equiv [I]^{-1}$. As such, wages in the interior are always 1. Now, real wage equations reduce to

$$\omega_{1,4} = \frac{1}{\left[(1 + T_F^{1-\varepsilon})^{\frac{1}{1-e}}\right]^{\delta}} \tag{A.1.16a}$$

$$\omega_{1,4} = \frac{1}{\left[(1 + T_{DFD}^{1-\varepsilon})^{\frac{1}{1-e}}\right]^{\delta}} \tag{A.1.16b}$$

Note that, technically, the (real) wage equations for regions 2 and 3 are only implied functions, as there is no actual manufacturing wage in this spatial configuration given by the absence of manufacturing workers, i.e. $\lambda_2 = \lambda_3 = 0$. One can think of these implied wages as the maximum wage that firms moving to this location would be able to pay (Fujita et al. 2001). The derivation of $\omega_{2,3}$ follow the same type of manipulations just made (A1.15 through A.1.16) and lead to expressions only dependent on the parameter values δ, ε and transport costs $T_D, T_F, T_{DF}, T_{DFD}$.[9] We now have all the ingredients for an expression of the real wage differential within the home or foreign economy, i.e. $\omega_i^{H,F}/\omega_j^{H,F}$. As in a), to assess the analytical results, we plot the real wage differential in the home economy against the intra-national transport costs T_D and for our three levels of external transport costs T_F. This is done in Figure A.1.2. Importantly, note the change in the y-axis; we now express the real wage differential from the point of view of the peripheral region, i.e. plot ω_2/ω_1 to facilitate a comparison to the break analysis in a). More precisely, the range of transport costs T_D for which this type of equilibrium is sustainable also lies below the constant. Again, we analyze the dynamics from the point of view of region 2, i.e. assess when full agglomeration in region 1 is unsustainable. In contrast to a) we are now also able to discuss peculiarities of the model with heterogeneous intra-national space analytically.



[9] Results for these equations are not reported here and relegated to the Maple code of the online Appendix.

Figure A.1.2: Internal transport costs and sustainable agglomeration

The results shown in Figure A.1.2 show similar tendencies as in a), that is, increased trade liberalization increases the range of values for which a fully agglomeration equilibrium is more likely (note the shift in the sustain point $T(S)$ to the right in Panel A). By construction, this mirrors the result in Monfort and Nicolini (2000). Also, by reasons given in Chapter III, a lower level of ε cause agglomeration forces to be strengthened, up to the point where a sustain point does not exist for low transport costs across countries (Panel C). Hence, for a level of $T_F \leq 2.00$ and $\varepsilon = 4$, there does not exist a level of internal transport costs for which agglomeration becomes unsustainable. In other words, the existence of external markets renders the costs of serving domestic markets from a distance negligible and it increasingly pays to agglomerate given reduced international trade costs. Notice, however, the stark difference to Panels B and D, i.e. in the case of heterogeneous intra-national space. Here, external trade liberalization causes a decrease in the range of intra-national transport costs T_D for which agglomeration in region 1 is sustainable, and for all parameters tested, there exists a level where agglomeration in region 2 is broken (see Table A.1). Whence the difference? We need to latch on to the discussion in a) where we discussed the relative influence of centrifugal and centripetal forces of a changing T_F. For heterogenous intra-national space, there is now an additional component which mediates the relative strength of these two forces, namely, the differential exposure to the external markets, initially shown by Crozet and Koenig 2004b). To the former, while agglomeration tendencies are lowered, one may expect an increased draw of firms and consumers to the border so as to benefit from the better access to new demand and supply, respectively. To the latter, there is the possibility that the dispersion force is further amplified which pushes economic activity towards region 1 given that its larger distance to the border provides an increased level of protection as given by new foreign competition. What are the implications for our present results? For all of the parameter configurations of ε and δ tested (see Table A.1), we see a falling range of intra-national transport costs T_D for which agglomeration in region 1 is sustainable. Hence, sheltering in the interior regions does not seem to happen to a larger degree than the draw to the border. This goes against Crozet and Koenig (2004b) where a push to the interior happens at intermediate international trade costs. The difference in these results most likely stem from the setup of the foreign economy and the moderating force of this. In their model, the foreign economy is larger than the domestic one which has arguably larger bearings on the competition effect just described.[10] However, what Crozet and Koenig (2004b) are not analyze

[10] Of course, this also depends on the structure of the economies, i.e. whether the two economies are complementary in their trade or whether one of the countries dominates in either imports or exports. These effects are analyzed in Brülhart et al. (2004), albeit for a different model set-up concerning the utility function and thereby not directly comparable to the one in this paper.

in their $2 + 1$ setup, is the influence of the relative size of the foreign economy on these dynamics. Figure 5 in Chapter III provides the main results of this analysis, where we have seen that the push to the border may be lower or higher when foreign economic activity agglomerates in the interior, but nonetheless exists. Hence, for two equally sized economies, foreign economic inequality cannot turn around our main results, which is that the draw to the border dominates any benefit by sheltering from the foreign competition. What we can say, however, is that this effect may be moderated by foreign economic inequality. As such, from the results in Figure 5, seems as if sheltering in the interior is more important when product differentiation is high ($\varepsilon = 4$) and less important when it is low ($\varepsilon = 6$). This may be easier understood when envisioning the scenario in which all foreign activity is agglomerated at the border, i.e. in region 3. Here, the need to shelter in the domestic interior is relatively more important when product differentiation is high, than when it is low. Note that this notion can be visually depicted in the graphs in Panels B and C as the changes for increasing liberalization stem from changes in the descending part of the slope, whereas in Panels B and D, this pattern is reversed. From the discussion in Chapter III, we know that the decreasing portion of ω_2/ω_1 along T_D i.e. at low levels of T_D, agglomeration forces dominate dispersion forces while for the positive slopes, dispersion forces dominate. The crucial difference therefore lies in the increased strength of dispersion forces at lower levels of T_D, which are strengthened by an increase in trade liberalization. This difference is driven by the position of region 2 vis-à-vis the foreign economy, because then, not only are the local dispersion forces lowered, but region 1 lowered sheltering from local competition is even less important. At full trade integration ($T_F = 1.60$), and for low levels of intra-national transport costs, each small increase in T_D causes trade costs.

Table A.1: Sustain and Break points across δ, ε and T_F

			Sustain and Break Values					
			$\delta = 0.4$		$\delta = 0.5$		$\delta = 0.6$	
			$T_D(B)$	$T_D(S)$	$T_D(B)$	$T_D(S)$	$T_D(B)$	$T_D(S)$
$T_F = \infty$	$\varepsilon = 4$	Symmetric	1.97	2.34	2.47	4.00	3.30	14.62
		Asymmetric	-	2.34	-	4.00	-	14.62
	$\varepsilon = 5$	Symmetric	1.63	1.81	1.90	2.52	2.30	5.00
		Asymmetric	-	1.81	-	2.52	-	5.00
	$\varepsilon = 6$	Symmetric	1.46	1.57	1.64	2.00	1.90	3.16
		Asymmetric	-	1.57 ∅	∅	2.00 ∅	∅	3.16 ∅
$T_F = 2$	$\varepsilon = 4$	Symmetric	3.58					
		Asymmetric	-	2.23	-	3.69 ∅	∅	12.50 ∅
	$\varepsilon = 5$	Symmetric	1.83	2.28	2.69			
		Asymmetric	-	1.78	-	2.45	-	4.78 ∅
	$\varepsilon = 6$	Symmetric	1.52	1.68	1.79	3.83	2.48	
		Asymmetric	∅	1.56 ∅	∅	3.83 ∅	∅	3.41 ∅
$T_F = T_D = 1.60$	$\varepsilon = 4$	Symmetric						
		Asymmetric	-	2.13 ∅	∅	3.45 ∅	∅	10.93 ∅
	$\varepsilon = 5$	Symmetric	3.02					
		Asymmetric	-	1.73	∅	2.37 ∅	∅	4.49 ∅
	$\varepsilon = 6$	Symmetric	1.70	2.23				
		Asymmetric	-	1.54	-	1.94	-	3.01

Notes: The values in this table represent the intra-national iceberg transport costs at which agglomeration turns "sustainable" [T(S)] and where the symmetric (spreading) equilibrium is "broken" [T(B)], i.e. at which real wages in the agglomeration exceed those in the periphery and a migration towards one of the regions leads to real wage gains, respectively. For more details on the derivation, see Appendix A.1.

Appendix A.2

As anticipated in Chapter II, this section provides the full set of simulations, of which selected results are presented and discussed in the main text. I thereby provide the three-dimensional depictions of the simulations which were discussed as simpler, two-dimensional illustrations before. This Figures A.2.1 and A.2.2 plot the plane of real wage differentials $\omega_1 \,/\, \omega_2$ spanned by all possible home and foreign spatial configurations, given by relative shares of the home and foreign workforces λ_1 and λ_4, respectively for a given set of parameter values. Additionally, I provide the full set of corresponding contour lines in Figures A.2.3 and A.2.4 which depict the changing influence of foreign economic inequality for stable and unstable equilibria. As established in Chapter II, I plot the results for all three levels of external transport costs and additionally, compare results from the main 2+2 setting against the more general 2+2 setting with homogeneous intra-national space.

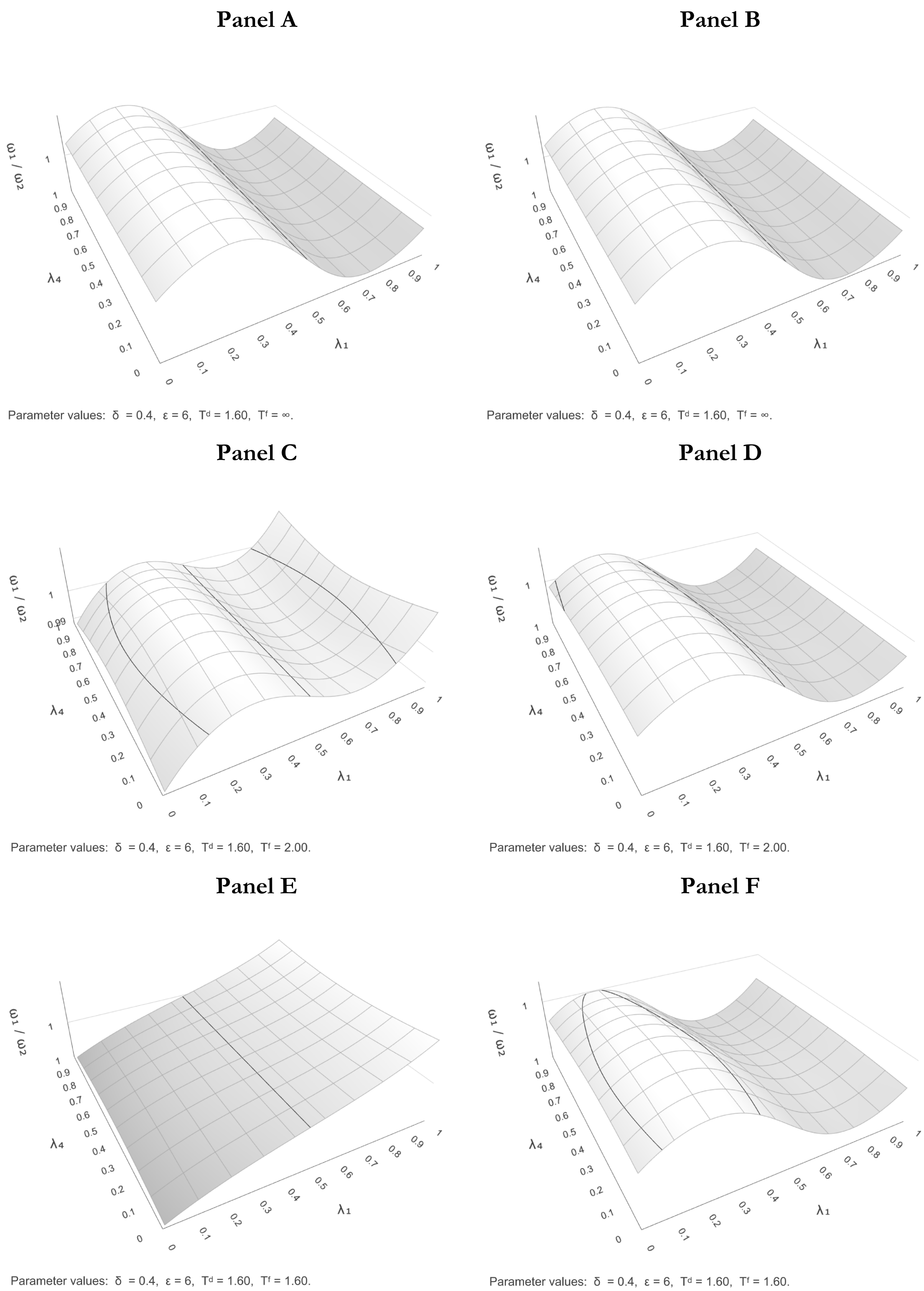
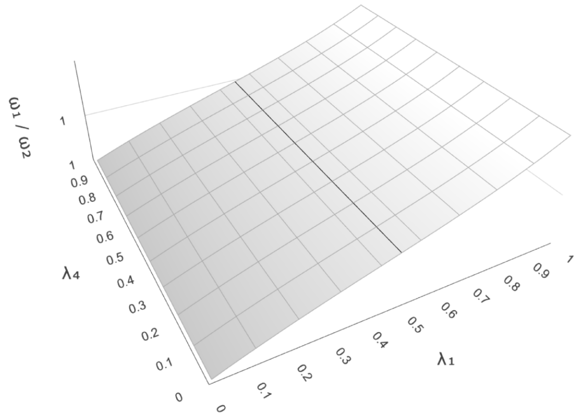


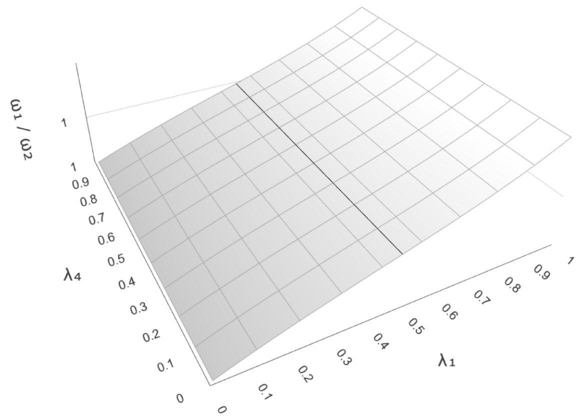
Figure A.2.1: Three-dimensional depiction of spatial equilibria ($\varepsilon = 6$)

Panel A



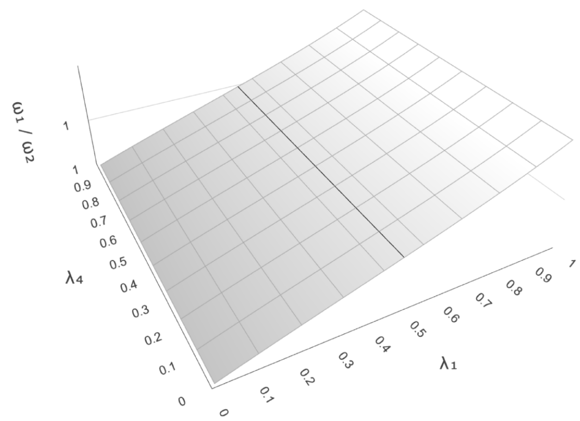
Parameter values: δ = 0.4, ε = 4, T^d = 1.60, T^f = ∞.

Panel B



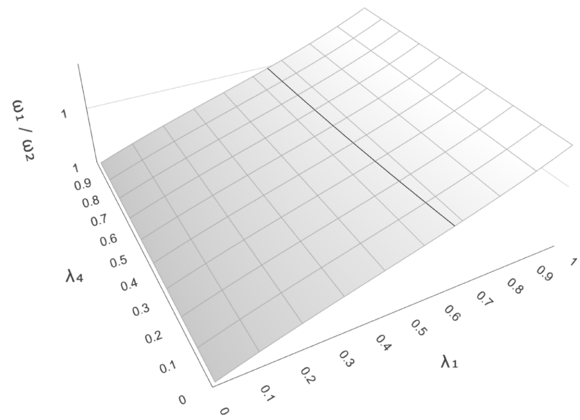
Parameter values: δ = 0.4, ε = 4, T^d = 1.60, T^f = ∞.

Panel C



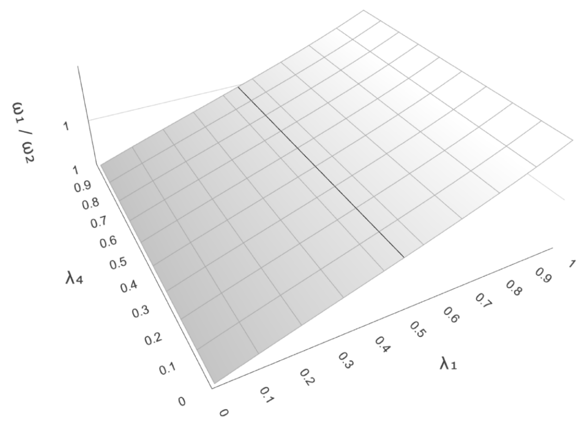
Parameter values: δ = 0.4, ε = 4, T^d = 1.60, T^f = 2.00.

Panel D



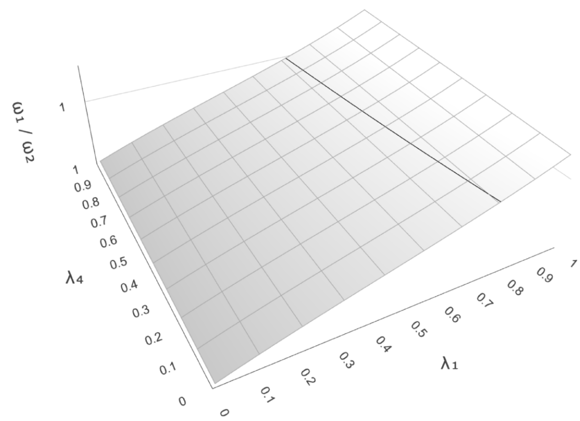
Parameter values: δ = 0.4, ε = 4, T^d = 1.60, T^f = 2.00.

Panel E



Parameter values: δ = 0.4, ε = 4, T^d = 1.60, T^f = 1.60.

Panel F



Parameter values: δ = 0.4, ε = 4, T^d = 1.60, T^f = 1.60.

Figure A.2.1: Three-dimensional depiction of spatial equilibria ($\varepsilon = 4$)

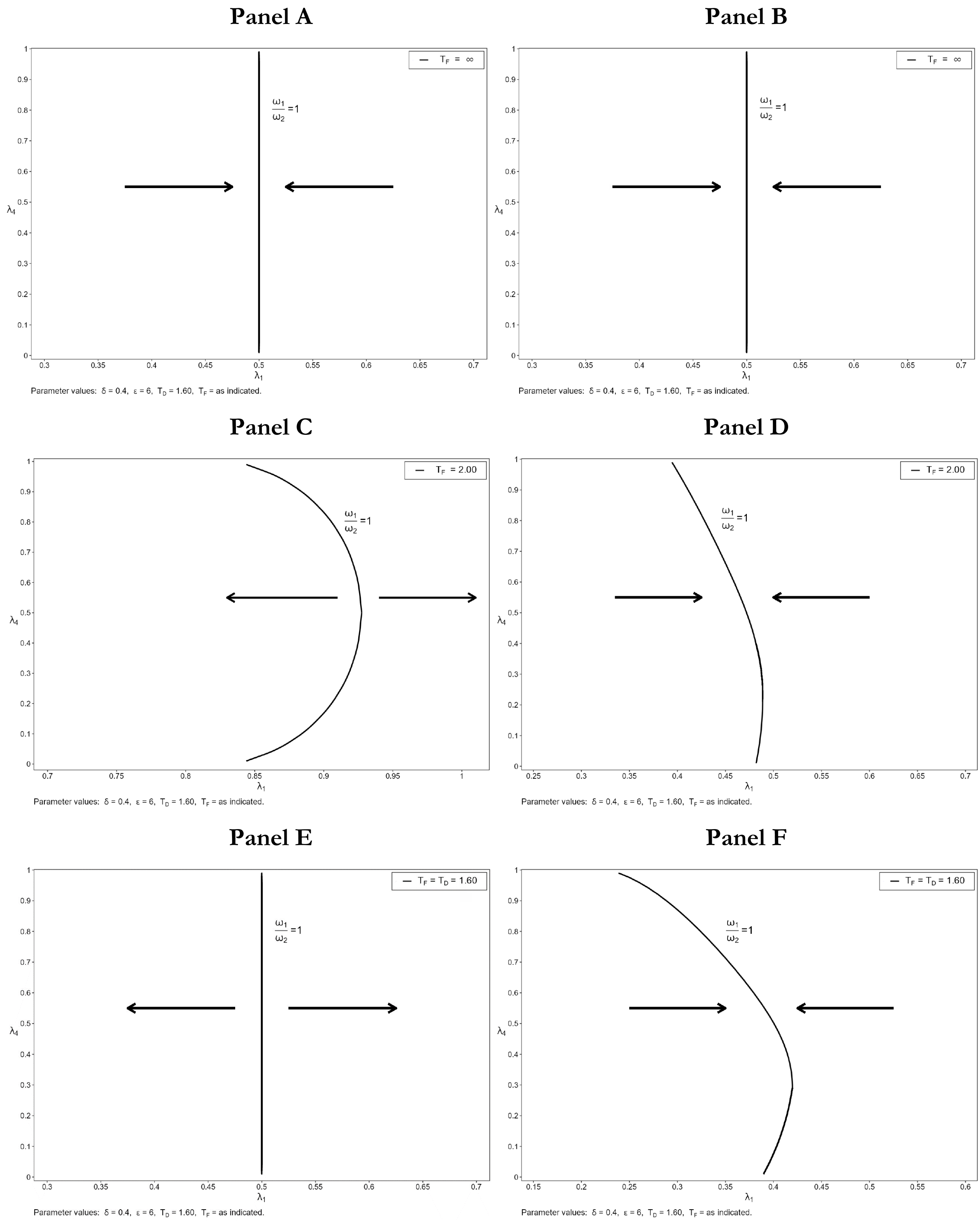
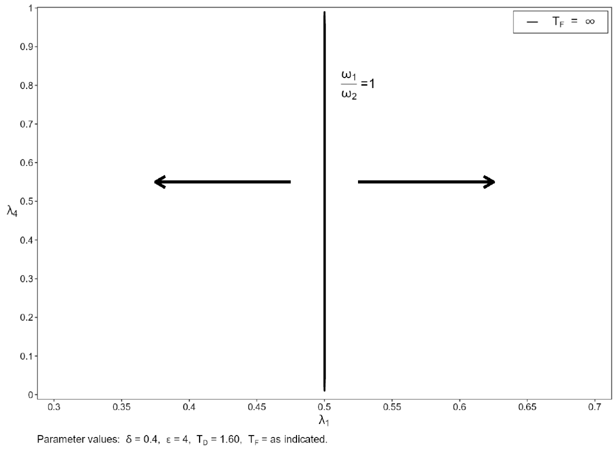
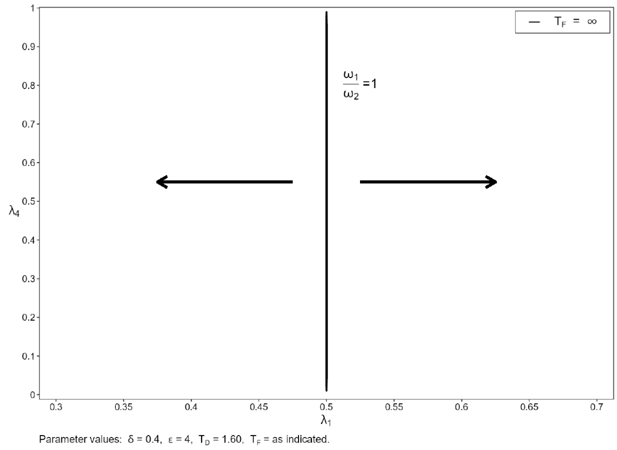


Figure A.2.1: Three-dimensional depiction of spatial equilibria ($\varepsilon = 6$)

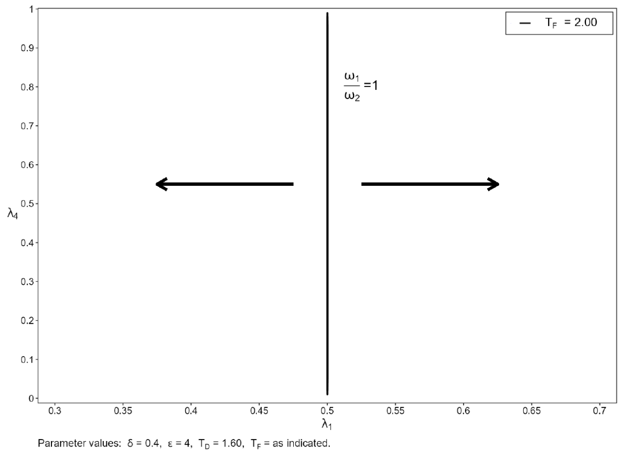
Panel A



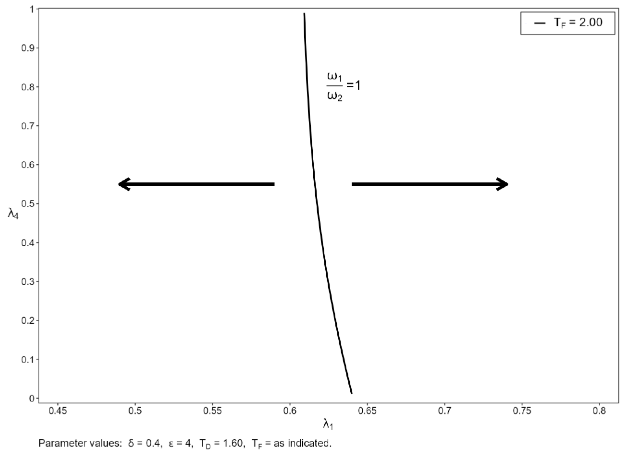
Panel B



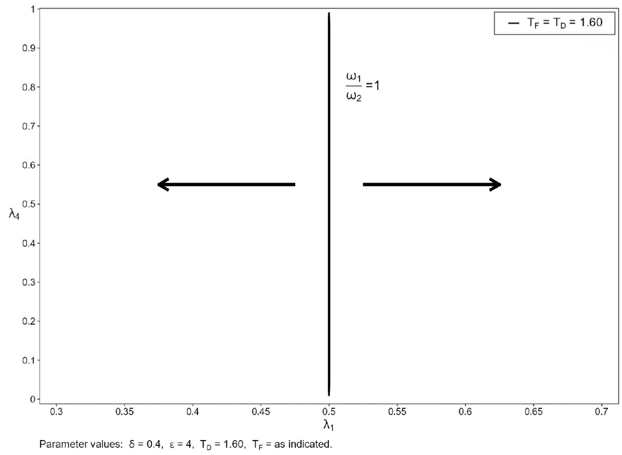
Panel C



Panel D



Panel E



Panel F

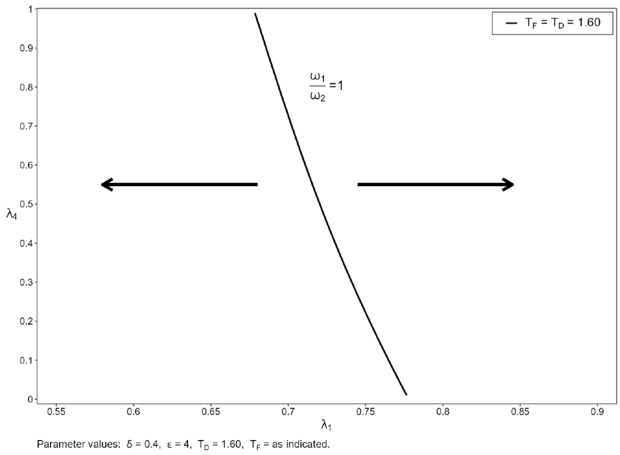


Figure A.2.1: Three-dimensional depiction of spatial equilibria ($\varepsilon = 4$)

Appendix B.1

Table B.1: Summary Statistics

Panel a)	*Mean*	*St. Dev*	*Min.*	*1st Quartile*	*Median*	*3rd Quartile*	*Max.*	*N*
			Distribution across Sample					
Basic Characteristics								
Afrobarometer (AFB)								
Age	36	14	17	26	33	43	101	39,417
Education (Level)	3.3	1.8	0.0	2.0	3.0	4.5	9.0	39,736
Female (0/1)	0.5	0.5	0.0	0.0	1.0	1.0	1.0	39,736
Demographic and Health Survey (DHS)								
Age	29	10	15	20	27	36	60	225,748
Education (Years)	5.6	4.1	0.0	2.0	6.0	8.0	26.0	225,657
Female (0/1)	0.8	0.4	0.0	1.0	1.0	1.0	1.0	225,748
Kagera Health and Development Survey (KHDS)								
Age	22	19	0	7	16	31	105	48,075
Education (Years)	5.3	3.3	0.0	3.0	6.0	7.0	22.0	30,154
Female (0/1)	0.5	0.5	0.0	0.0	1.0	1.0	1.0	48,119
Main Covariates								
AFB								
Urban (0/1)	0.27	0.44	0.00	0.00	0.00	1.00	1.00	39,740
Agglomeration (0/1)	0.28	0.45	0.00	0.00	0.00	1.00	1.00	38,644
Core Agglomeration (0/1)	0.13	0.33	0.00	0.00	0.00	0.00	1.00	39,740
Road Distance to nearest EAC Border Crossing (in km)	317	262	1	129	240	413	1362	39,740
Inverse rel. Distance to nearest EAC Border Crossing (0-1)	0.71	0.22	0.00	0.62	0.75	0.87	1.00	39,740
DHS								
Urban (0/1)	0.26	0.44	0.00	0.00	0.00	1.00	1.00	144,463
Agglomeration (0/1)	0.27	0.45	0.00	0.00	0.00	1.00	1.00	110,092
Core Agglomeration (0/1)	0.11	0.31	0.00	0.00	0.00	0.00	1.00	110,092
Road Distance to nearest EAC Border Crossing (in km)	301	254	0.1	121	222	396	1306	104,494
Inverse rel. Distance to nearest EAC Border Crossing (0-1)	0.72	0.22	0.00	0.63	0.77	0.88	1.00	104,494
KHDS								
Urban (0/1)	0.21	0.41	0.00	0.00	0.00	0.00	1.00	50,757
Core Agglomeration (0/1)	0.20	0.40	0.00	0.00	0.00	0.00	1.00	50,756
Road Distance to nearest EAC Border Crossing (in km)	144	100	0	80	100	170	1134	50,756
Inverse rel. Distance to nearest EAC Border Crossing (0-1)	0.87	0.09	0.00	0.85	0.91	0.93	1.00	50,756
Geographical Covariates								
AFB								
Distance to Harbor (in km)	683	337	1	431	722	948	1259	39,740
Distance to Navigable River (in km)	323	287	1	76	204	514	999	39,740
Distance to Major Lake (in km)	195	197	0	46	123	267	747	39,740
Absolute Latitude	2.6	2.7	0.0	0.6	1.4	3.7	11.5	39,740
Elevation (in m)	1180	536	0	1060	1196	1512	3914	39,740
Terrain Ruggedness (standardised)	-0.07	0.76	-0.64	-0.49	-0.30	0.03	6.51	39,740
Average Monthly Temperature (in Celsius)	24	3	8	22	24	25	32	39,740
Average Monthly Rainfall (in mm)	102	27	17	87	103	117	214	39,740
Growing Days (0-365)	292	70	14	247	305	358	365	39,740
Malaria Ecology	7.0	6.4	0.0	1.5	5.6	10.5	31.1	38,604
DHS								
Distance to Harbor (in km)	685	314	0	438	706	924	1259	110,092
Distance to Navigable River (in km)	308	260	1	91	212	478	961	110,092
Distance to Major Lake (in km)	183	183	0	45	120	253	706	110,092
Absolute Latitude	2.1	2.3	0.0	0.5	1.1	3.0	11.5	110,092
Elevation (in m)	1236	538	2	1071	1213	1562	3248	110,092
Terrain Ruggedness (standardized)	0.00	1.00	-0.75	-0.55	-0.29	0.12	11.43	110,092
Average Monthly Temperature (in Celsius)	22	3	11	20	22	23	30	110,092
Average Monthly Rainfall (in mm)	94	28	15	77	96	111	219	110,092
Growing Days (0-365)	291	76	10	243	310	365	365	110,092
Malaria Ecology	4.3	7.4	0.0	0.0	1.1	2.9	36.0	87,123
KHDS								
Distance to Harbor (in km)	909	135	0	913	919	956	1189	50,757
Distance to Navigable River (in km)	60	132	1	19	31	47	963	50,757
Distance to Major Lake (in km)	46	89	0	5	17	66	665	50,757
Absolute Latitude	1.8	1.0	0.1	1.3	1.5	1.9	10.1	50,757
Elevation (in m)	1289	226	3	1194	1262	1392	4249	50,754
Terrain Ruggedness (standardised)	0.00	1.00	-0.82	-0.75	-0.47	0.34	5.82	50,757
Average Monthly Temperature (in Celsius)	21	1	5	20	21	21	27	50,754
Average Monthly Rainfall (in mm)	113	31	48	83	112	141	168	50,754
Growing Days (0-365)	318	36	154	297	327	353	365	50,757
Malaria Ecology	4.0	2.8	0.0	2.6	3.9	4.3	23.0	50,678

Notes: *Table continued on next page.*

Panel b)	Mean	St. Dev	Min.	Distribution across Distance (within Quartiles) 1st Quartile	2nd Quartile	3rd Quartile	4th Quartile	N
Consumption								
AFB								
How often: Gone without […] (0-4)	1.12	0.95	0	1.06	1.02	1.15	1.23	37,466
How often: Gone without Food (0-4)	0.97	1.12	0	1.01	0.91	0.99	0.98	37,430
How often: Gone without Water (0-4)	1.12	1.31	0	0.93	1.00	1.18	1.36	37,447
How often: Gone without Medical Care (0-4)	1.26	1.22	0	1.24	1.16	1.27	1.35	37,392
DHS								
Wealth Quintile (1-5)	3.21	1.48	1	3.17	3.53	2.96	2.61	151,032
International Wealth Index (IWI)	23.46	2.80	18	23.29	23.68	23.32	23.08	225,748
Comparative Wealth Index (CWI)	-0.69	0.85	-2	-0.75	-0.27	-0.88	-0.93	71,846
KHDS								
Annual p.c. Consumption in 2010 TZS ('000)	481	456	36	472	502	401	529	33,434
Annual p.c. non-Food Consumption in 2010 TZS ('000)	175	291	6	168	179	134	209	33,606
Value of Occupied Dwelling in 2004 TZS ('000)	791	3623	0	691	1442	494	452	35,659
Value of Durable Assets in 2004 TZS ('000)	144	1319	0	112	196	55	203	35,672
Income & Work								
AFB								
Worked last Year (0/1)	0.55	0.50	0	0.52	0.56	0.56	0.55	26,880
Employed Work (0/1)	0.23	0.42	0	0.24	0.29	0.23	0.18	26,535
Occupation Level: Agr.-Worker-Prof. (1-3)	1.67	0.81	1	1.66	1.83	1.66	1.55	22,071
How often: Gone without Cash Income (0-4)	2.08	1.24	0	2.09	2.00	2.11	2.11	37,396
DHS								
Worked last Year (0/1)	0.75	0.43	0	0.75	0.75	0.79	0.78	163,933
Employed Work (0/1)	0.19	0.39	0	0.21	0.31	0.18	0.11	104,610
Occupation Level: Agr.-Worker-Prof. (1-3)	1.51	0.64	1	1.62	1.74	1.50	1.42	143,027
Paid in Cash (0/1)	0.53	0.50	0	0.53	0.61	0.49	0.45	132,363
KHDS								
Worked last Year (0/1)	0.26	0.44	0	0.26	0.24	0.19	0.33	20,259
Employed Work (0/1)	0.12	0.33	0	0.10	0.11	0.13	0.14	32,017
Occupation Level: Agr.-Worker-Prof. (1-3)	1.20	0.49	1	1.21	1.24	1.11	1.23	20,446
Salaried Work (0/1)	0.01	0.10	0	0.01	0.01	0.01	0.01	19,465
Monthly Salary in 2004 TZS ('000)	21.4	55.6	0.0	20.4	27.4	13.3	23.5	2,118
Agglomeration & Migration								
AFB								
Population Count	1.31	4.37	0	0.80	2.58	1.17	0.68	39,740
Population Density	1.53	5.09	0	0.94	3.00	1.38	0.80	39,740
DHS								
Population Count	1.22	3.57	0	1.01	2.39	0.49	0.92	110,076
Population Density	1.44	4.17	0	1.21	2.79	0.58	1.08	110,076
Has ever Migrated (0/1)	0.43	0.49	0	0.28	0.26	0.33	0.45	121,303
KHDS								
Population Count	0.98	1.58	0	0.84	0.82	1.08	1.20	50,757
Population Density	0.48	1.78	0	0.20	0.52	0.20	1.02	50,757
Has ever Migrated (0/1)	0.42	0.49	0	0.48	0.45	0.45	0.35	23,851
Main reason for Migration: *Economic* (0/1)	0.09	0.29	0	0.10	0.11	0.06	0.08	6,946
After moving to curr. Residence: *Paid Employment* (0/1)	0.09	0.28	0	0.09	0.08	0.03	0.12	2,388
Sentiments & Attitudes								
AFB								
Helps your Country: AU (0-3)	1.62	0.94	0	1.62	1.60	1.65	1.63	13,556
Helps your Country: REC (0-3)	1.66	0.92	0	1.67	1.63	1.67	1.65	12,003
Support for Regional Integration (1-5)	3.71	1.52	1	3.88	3.82	3.71	3.40	6,593
Would like as Neighbor: Immigrant (1-3)	3.38	1.30	1	3.43	3.31	3.29	3.46	11,708
Ease of Crossing Borders to live and Work (1-4)	2.21	0.93	1	2.35	2.20	2.15	2.13	4,784
Present vs. Past: Living Standards of People (1-4)	2.42	1.05	1	2.68	2.68	2.74	2.68	37,649
Present vs. Past: Economic Condition of Country (1-4)	2.43	1.17	1	2.41	2.45	2.44	2.41	36,810
KHDS								
Subjective HH. Wealth: *Today (2004)* (1-5)	2.62	0.59	1	2.67	2.67	2.54	2.62	3,313
Subjective HH. Wealth: *Ten Years ago (in 1994)* (1-5)	2.62	0.68	1	2.67	2.66	2.62	2.57	3,313
Subjective HH Life Satisfaction: Ladder (1-9)	3.79	1.49	1	3.85	3.94	3.61	3.79	3,313

Notes: The table depicts summary statistics corresponding to the main sample used in the estimations across the paper. The data encompasses geo-referenced individual- and household level responses from the three founding members of the East African Community (EAC), i.e. Kenya, Tanzania and Uganda. The first set of data stem from the Afrobarometer (AFB) survey sampled in 2000-2001 (only Tanzania and Uganda), 2002-03, 2005, 2008, 2011-12, 2014-15 and 2017 (i.e. Round 1 through Round 7). The second set of data stem from the Womens-, Men- and Household recodes of the Demographic and Health (DHS) Surveys sampled in Kenya in 1989, 1993, 1998, 2003, 2008-09, and 2014, in Tanzania in 1991-92, 1995, 1996, 1999, 2004-05, 2010, 2015-16, and 2017, and in Uganda in 1988-89, 1995-96, 1995, 2000-01, 2006, 2011, and 201. Note that DHS-GPS datasaets are available from 1999 and onwards, hence the overall reduced sample size in panel b) as well as for the respective geographical indicators in panel a). The last set of data stem from the Kagera Health and Development Survey (KHDS) initially sampled in yearly waves from 1991-1994 and repeated in 2004 and 2010, i.e. a total of four waves. Note that the KHDS summary statistics displayed includes repeat observations from tracked individuals (a maximum of 6 times). Remaining variations in the number of observations sizes stem from differences in response rates of variables as well as changes in questions asked across surveys. The geographic covariates displayed in panel a) come from an array of sources described in the data section of the paper. The distribution "across distance" in panel b) calculates mean values of the respective variable within quartiles of road distance to nearest EAC border crossings.

Table B.2: Robustness Checks (Afrobarometer)

	Afrobarometer					
	Dependent Variable					
Panel a)	*Freq. gone without:* [Wat./Food/ Med.] (0-4)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.- Work.-Prof.) (1-3)	*Freq. gone without:* [Cash Income] (0-4)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 2)**						
Core Agglom. (0/1) * EAC	-0.261***	-0.013	-0.204**	0.026	0.021	0.041
	(0.077)	(0.050)	(0.095)	(0.067)	(0.079)	(0.083)
Core Agglom. (0/1) * CU	-0.470***	-0.006	0.657***	0.002	0.057	-0.234**
	(0.097)	(0.045)	(0.247)	(0.055)	(0.090)	(0.105)
Core Agglom. (0/1) * CM	-0.338***	-0.076	0.978***	-0.258**	-0.009	0.043
	(0.065)	(0.069)	(0.269)	(0.110)	(0.087)	(0.079)
(a) Standard Erros clustered at EA-Level **(see Table A)**						
Core Agglom. (0/1) * EAC	-0.261**	-0.013	-0.204	0.026	0.021	0.041
	(0.126)	(0.041)	(0.132)	(0.044)	(0.084)	(0.159)
Core Agglom. (0/1) * CU	-0.470***	-0.006	0.657***	0.002	0.057	-0.234
	(0.125)	(0.044)	(0.198)	(0.046)	(0.080)	(0.155)
Core Agglom. (0/1) * CM	-0.338***	-0.076	0.978***	-0.258**	-0.009	0.043
	(0.119)	(0.036)	(0.192)	(0.041)	(0.078)	(0.150)
(b) No Controls **(see Table A)**						
Core Agglom. (0/1) * EAC	-0.392***	-0.011	-0.196***	0.018	0.029	-0.047
	(0.088)	(0.044)	(0.066)	(0.057)	(0.068)	(0.089)
Core Agglom. (0/1) * CU	-0.607***	0.005	0.661**	0.002	0.103	-0.318***
	(0.116)	(0.048)	(0.281)	(0.053)	(0.103)	(0.114)
Core Agglom. (0/1) * CM	-0.466***	-0.054	0.993***	-0.253***	0.060	-0.061
	(0.081)	(0.065)	(0.298)	(0.102)	(0.110)	(0.077)
(b) Extended Geographic Controls **(see Table A)**						
Core Agglom. (0/1) * EAC	-0.211***	-0.012	-0.274*	0.019	0.021	0.054
	(0.080)	(0.051)	(0.146)	(0.069)	(0.076)	(0.081)
Core Agglom. (0/1) * CU	-0.424***	-0.001	0.623***	0.000	0.069	-0.205**
	(0.094)	(0.047)	(0.198)	(0.060)	(0.084)	(0.088)
Core Agglom. (0/1) * CM	-0.304***	-0.076	0.920***	-0.262**	-0.018	0.046
	(0.073)	(0.070)	(0.233)	(0.109)	(0.082)	(0.083)
(d) Including Survey Weights **(see Table A)**						
Core Agglom. (0/1) * EAC	-0.245***	-0.016	-0.201**	0.016	-0.010	0.039
	(0.079)	(0.048)	(0.097)	(0.069)	(0.092)	(0.082)
Core Agglom. (0/1) * CU	-0.443***	0.001	0.784***	0.014	0.012	-0.187*
	(0.084)	(0.054)	(0.248)	(0.063)	(0.097)	(0.098)
Core Agglom. (0/1) * CM	-0.340***	-0.089	0.859***	-0.268**	-0.002	0.032
	(0.067)	(0.078)	(0.310)	(0.114)	(0.096)	(0.082)
(e) Excluding Low-Precision Localities **(see Table A)**						
Core Agglom. (0/1) * EAC	-0.008	0.036	-0.158**	0.074	-0.011	0.336***
	(0.092)	(0.036)	(0.072)	(0.058)	(0.092)	(0.123)
Core Agglom. (0/1) * CU	-0.238**	0.054	0.742**	0.029	-0.067	0.142
	(0.108)	(0.043)	(0.312)	(0.070)	(0.147)	(0.134)
Core Agglom. (0/1) * CM	-0.081	0.019	1.001***	-0.217**	-0.090	0.383***
	(0.078)	(0.050)	(0.160)	(0.089)	(0.107)	(0.110)

Notes: *Table continued on next page.*

Panel b)	Afrobarometer Dependent Variable *Freq. gone without:* [Wat./Food/ Med.] (0-4)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	*Freq. gone without:* [Cash Income] (0-4)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 2)**						
Core Agglom. (0/1) * EAC	-0.261*** (0.077)	-0.013 (0.050)	-0.204** (0.095)	0.026 (0.067)	0.021 (0.079)	0.041 (0.083)
Core Agglom. (0/1) * CU	-0.470*** (0.097)	-0.006 (0.045)	0.657*** (0.247)	0.002 (0.055)	0.057 (0.090)	-0.234** (0.105)
Core Agglom. (0/1) * CM	-0.338*** (0.065)	-0.076 (0.069)	0.978*** (0.269)	-0.258** (0.110)	-0.009 (0.087)	0.043 (0.079)
(f) Agglomerations vs. *Core* Agglomeration **(see Table A)**						
Agglomeration (0/1) * EAC	0.108 (0.078)	-0.056** (0.028)	0.012 (0.088)	-0.060* (0.034)	-0.046 (0.079)	-0.199* (0.115)
Agglomeration (0/1) * CU	-0.036 (0.092)	-0.006 (0.031)	0.064 (0.086)	-0.011 (0.035)	0.001 (0.075)	-0.350*** (0.115)
Agglomeration (0/1) * CM	0.024*** (0.080)	-0.011*** (0.027)	0.218*** (0.145)	-0.063*** (0.045)	0.042*** (0.071)	-0.334*** (0.117)
(g) Core Agglomeration ≤ 25km **(see Table A)**						
Core Agglom. ≤ 25km (0/1) * EAC	-0.301*** (0.076)	-0.037 (0.061)	-0.278** (0.130)	0.035 (0.076)	0.009 (0.079)	-0.021 (0.075)
Core Agglom. ≤ 25km (0/1) * CU	-0.520*** (0.098)	-0.051 (0.061)	0.591 (0.368)	-0.040 (0.071)	-0.049 (0.058)	-0.244** (0.105)
Core Agglom. ≤ 25km (0/1) * CM	-0.375*** (0.062)	-0.120*** (0.090)	1.023*** (0.345)	-0.342*** (0.126)	-0.101*** (0.093)	0.070*** (0.080)
(h) Core Agglomeration ≤ 10km **(see Table A)**						
Core Agglom. ≤ 10km (0/1) * EAC	-0.309*** (0.095)	-0.090 (0.060)	-0.146 (0.141)	0.007 (0.057)	-0.192*** (0.066)	-0.105 (0.076)
Core Agglom. ≤ 10km (0/1) * CU	-0.496*** (0.102)	-0.055 (0.044)	0.778 (0.476)	-0.055 (0.041)	-0.209*** (0.066)	-0.341*** (0.112)
Core Agglom. ≤ 10km (0/1) * CM	-0.368*** (0.068)	-0.126 (0.079)	1.369*** (0.432)	-0.351*** (0.097)	-0.313*** (0.114)	0.091 (0.082)
(i) Flexible Distance Specification **(see Table A)**						
Cont. Distance to Core Agglom. * EAC (in '00 km)	0.081 (0.086)	0.017 (0.015)	-0.087 (0.071)	0.021 (0.018)	0.048 (0.033)	0.024 (0.072)
Squared Cont. Distance to Core Agglom. * EAC (in '00 km)	-0.009 (0.008)	0.000 (0.002)	0.009 (0.006)	0.000 (0.002)	0.000 (0.003)	-0.003 (0.007)
(j) Post-CM Development						
Core Agglom. (0/1) * EAC	-0.261*** (0.077)	-0.013 (0.050)	-0.204** (0.096)	0.026 (0.067)	0.021 (0.078)	0.042 (0.082)
Core Agglom. (0/1) * CU	-0.470*** (0.097)	-0.006 (0.045)	0.657*** (0.247)	0.003 (0.055)	0.056 (0.091)	-0.233** (0.105)
Core Agglom. (0/1) * CM 1[2010-2014]	-0.344*** (0.066)	-0.044 (0.080)	0.970*** (0.342)	-0.256** (0.123)	-0.081 (0.088)	0.053 (0.094)
Core Agglom. (0/1) * 1[t ≥ 2015]	-0.332*** (0.066)	-0.097 (0.070)	0.990*** (0.382)	-0.259** (0.104)	0.037 (0.071)	0.027 (0.082)

Notes: This table offers an array of robustness tests on the main results of the paper for the differenc-in-difference effect for individuals living in core agglomerations. The full results (including border estimates) are given in the table referred below the description of each test. Data come from the Kenya, Uganda and Tanzania Afrobarometer surveys rounds 1 through 7 sampled between 2000 and 2017. The sample mean of the respective dependent variable is given in brackets above the estimates. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala), if not indicated otherwise. EAC switches to one for individuals sampled from the second half of 2001 to and including 2004, CU for individuals sampled from 2005 and including 2009, and CM for individuals sampled from 2010 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days, if not indicated otherwise, The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used, if not indicated otherwise. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table B.3: Robustness Checks (DHS)

	Demographic and Health Survey (DHS)					
	Dependent Variable					
Panel a)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 2)**						
Core Agglom. (0/1) * EAC	0.536***	0.073*	-0.140	0.096***	-0.046	0.000
	(0.118)	(0.040)	(0.320)	(0.027)	(0.062)	(0.027)
Core Agglom. (0/1) * CU	0.379***	0.037	0.876***	0.019	0.058	0.073*
	(0.104)	(0.039)	(0.340)	(0.059)	(0.111)	(0.042)
Core Agglom. (0/1) * CM	0.383***	0.006	1.109***	0.076**	0.007	0.029
	(0.114)	(0.041)	(0.172)	(0.038)	(0.088)	(0.035)
(a) Standard Erros clustered at EA-Level **(see Table A)**						
Core Agglom. (0/1) * EAC	0.536***	0.073*	0.346	0.096***	-0.046	0.000
	(0.127)	(0.040)	(0.218)	(0.024)	(0.058)	(0.038)
Core Agglom. (0/1) * CU	0.379***	0.037	1.061***	0.019	0.058	0.073**
	(0.101)	(0.032)	(0.243)	(0.022)	(0.052)	(0.035)
Core Agglom. (0/1) * CM	0.383***	0.006	1.100***	0.076***	0.007	0.029
	(0.092)	(0.029)	(0.191)	(0.018)	(0.047)	(0.032)
(b) No Controls **(see Table A)**						
Core Agglom. (0/1) * EAC	0.313	0.077	0.310	0.147***	-0.098	-0.065
	(0.209)	(0.050)	(0.368)	(0.031)	(0.094)	(0.046)
Core Agglom. (0/1) * CU	0.395**	0.052	1.025***	0.047	0.060	0.045
	(0.164)	(0.042)	(0.159)	(0.080)	(0.136)	(0.061)
Core Agglom. (0/1) * CM	0.357*	0.018	1.075***	0.099**	-0.024	0.009
	(0.197)	(0.047)	(0.123)	(0.046)	(0.116)	(0.043)
(c) Extended Geographic Controls **(see Table A)**						
Core Agglom. (0/1) * EAC	0.536***	0.073*	-0.140	0.096***	-0.046	0.000
	(0.118)	(0.040)	(0.320)	(0.027)	(0.062)	(0.027)
Core Agglom. (0/1) * CU	0.379***	0.037	0.876***	0.019	0.058	0.073*
	(0.104)	(0.039)	(0.340)	(0.059)	(0.111)	(0.042)
Core Agglom. (0/1) * CM	0.383***	0.006	1.109***	0.076**	0.007	0.029
	(0.114)	(0.041)	(0.172)	(0.038)	(0.088)	(0.035)
(d) Including Survey Weights **(see Table A)**						
Core Agglom. (0/1) * EAC	0.536***	0.073*	-0.140	0.096***	-0.046	0.000
	(0.118)	(0.040)	(0.320)	(0.027)	(0.062)	(0.027)
Core Agglom. (0/1) * CU	0.379***	0.037	0.876***	0.019	0.058	0.073*
	(0.104)	(0.039)	(0.340)	(0.059)	(0.111)	(0.042)
Core Agglom. (0/1) * CM	0.383***	0.006	1.109***	0.076**	0.007	0.029
	(0.114)	(0.041)	(0.172)	(0.038)	(0.088)	(0.035)
(e) Excluding Low-Precision Localities **(see Table A)**						
Core Agglom. (0/1) * EAC	0.635***	0.072*	0.340	0.088***	-0.039	0.000
	(0.121)	(0.040)	(0.313)	(0.031)	(0.069)	(0.031)
Core Agglom. (0/1) * CU	0.443***	0.038	1.128***	0.015	0.063	0.072
	(0.099)	(0.038)	(0.204)	(0.058)	(0.113)	(0.044)
Core Agglom. (0/1) * CM	0.321**	0.006	1.046***	0.048*	0.058	0.042
	(0.136)	(0.043)	(0.085)	(0.026)	(0.075)	(0.041)
(f) Agglomerations vs. *Core* Agglomeration **(see Table A6)**						
Agglomeration (0/1) * EAC	-0.227	-0.044	0.203	0.049	-0.147**	-0.003
	(0.243)	(0.043)	(0.184)	(0.038)	(0.060)	(0.049)
Agglomeration (0/1) * CU	-0.289**	-0.052*	0.095	0.054**	-0.117**	0.027
	(0.124)	(0.027)	(0.119)	(0.027)	(0.050)	(0.038)
Agglomeration (0/1) * CM	-0.281***	-0.041***	0.115***	0.014***	-0.070***	-0.004***
	(0.088)	(0.017)	(0.141)	(0.024)	(0.033)	(0.029)

Notes: *Table continued on next page.*

Panel b)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.)	Paid in Cash (0/1)
	Demographic and Health Survey (DHS) – Dependent Variable					
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 2)**						
Core Agglom. (0/1) * EAC	0.536***	0.073*	-0.140	0.096***	-0.046	0.000
	(0.118)	(0.040)	(0.320)	(0.027)	(0.062)	(0.027)
Core Agglom. (0/1) * CU	0.379***	0.037	0.876***	0.019	0.058	0.073*
	(0.104)	(0.039)	(0.340)	(0.059)	(0.111)	(0.042)
Core Agglom. (0/1) * CM	0.383***	0.006	1.109***	0.076**	0.007	0.029
	(0.114)	(0.041)	(0.172)	(0.038)	(0.088)	(0.035)
(g) Core Agglomeration ≤ 25km **(see Table A)**						
Core Agglom. ≤ 25km (0/1) * EAC	0.556***	0.102**	0.145	0.092***	-0.016	0.022
	(0.174)	(0.047)	(0.353)	(0.027)	(0.076)	(0.035)
Core Agglom. ≤ 25km (0/1) * CU	0.257**	0.060	1.017***	0.009	0.040	0.083*
	(0.128)	(0.049)	(0.188)	(0.063)	(0.132)	(0.044)
Core Agglom. ≤ 25km (0/1) * CU	0.319*	-0.004	1.099***	0.075**	-0.023	0.013
	(0.147)	(0.049)	(0.187)	(0.036)	(0.084)	(0.040)
(h) Core Agglomeration ≤ 10km **(see Table A)**						
Core Agglom. ≤ 10km (0/1) * EAC	0.553**	0.031	-0.057	0.080***	-0.083	-0.018
	(0.231)	(0.043)	(0.403)	(0.027)	(0.058)	(0.063)
Core Agglom. ≤ 10km (0/1) * CU	0.160	0.030	1.056***	-0.017	-0.008	0.058
	(0.158)	(0.029)	(0.336)	(0.057)	(0.125)	(0.071)
Core Agglom. ≤ 10km (0/1) * CM	0.259*	-0.024	1.494***	0.058**	-0.087	-0.022
	(0.141)	(0.027)	(0.578)	(0.030)	(0.054)	(0.075)
(i) Flexible Distance Specification **(see Table A)**						
Cont. Distance to Core Agglom.* EAC (in '00 km)	-0.419**	-0.129***	-0.168**	-0.109***	0.003	-0.026
	(0.195)	(0.033)	(0.083)	(0.027)	(0.061)	(0.042)
Squared Cont. Distance to Core Agglom.* EAC (in '00 km)	0.117**	0.033***	0.015*	0.009	0.018	0.022*
	(0.046)	(0.007)	(0.008)	(0.007)	(0.012)	(0.012)
(j) Varying Intervention Year **(see Table A)**						
Core Agglom. (0/1) * EAC	0.508***	0.066*	-0.139	0.095***	-0.065	-0.002
	(0.099)	(0.036)	(0.327)	(0.028)	(0.049)	(0.024)
Core Agglom. (0/1) * CU	0.371***	0.035	0.911**	0.019	0.047	0.072*
	(0.103)	(0.038)	(0.370)	(0.062)	(0.109)	(0.040)
Core Agglom. (0/1) * CM 1[2010-2014]	0.433***	0.009	1.559***	0.134***	-0.061	0.062
	(0.123)	(0.037)	(0.351)	(0.032)	(0.072)	(0.054)
Core Agglom. (0/1) * post-CM 1[t ≥ 2015]	0.343**	0.001	0.782***	0.040	0.034	0.006
	(0.146)	(0.042)	(0.071)	(0.027)	(0.081)	(0.036)
(k) Excluding post-EAC Migrants **(see Table A)**						
Core Agglom. (0/1) * EAC	0.605***	0.066	-0.292	0.110**	-0.037	0.027
	(0.145)	(0.045)	(0.490)	(0.044)	(0.064)	(0.046)
Core Agglom. (0/1) * CU	0.375***	0.046	0.876*	0.037	0.034	0.137***
	(0.122)	(0.038)	(0.464)	(0.091)	(0.099)	(0.047)
Core Agglom. (0/1) * CM	0.256**	-0.021	1.162***	0.062	0.023	0.116**
	(0.114)	(0.060)	(0.031)	(0.041)	(0.070)	(0.053)
(l) Full DHS Sample (incl. AIS, KAP, and MIS) **(see Table A)**						
Core Agglom. (0/1) * EAC	0.433***	0.073*	0.346	0.067*	-0.046	-0.002
	(0.141)	(0.040)	(0.312)	(0.038)	(0.062)	(0.026)
Core Agglom. (0/1) * CU	0.314***	0.037	1.061***	0.014	0.058	0.053
	(0.105)	(0.039)	(0.181)	(0.043)	(0.111)	(0.033)
Core Agglom. (0/1) * CM	0.361***	0.006***	1.100***	0.061***	0.007***	0.029***
	(0.142)	(0.041)	(0.134)	(0.031)	(0.088)	(0.034)

Notes: This table offers an array of robustness tests on the main results of the paper for the differenc-in-difference effect for individuals living in core agglomerations. Data come from the Kenya, Uganda and Tanzania Demographic and Health surveys (DHS) sampled between 1999 and 2020. The sample mean of the respective dependent variable is given in brackets above the estimates. The sample mean of the respective dependent variable is given in brackets above the estimates. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala), if not indicated otherwise. EAC switches to one for individuals sampled from the second half of 2001 to and including 2004, CU for individuals sampled from 2005 and including 2009, and CM for individuals sampled from 2010 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days, if not indicated otherwise, The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used, if not indicated otherwise. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table B.4: Robustness Checks (KHDS)

Panel a)	Annual p.c. Consumption (in 2010 TZS '000)	dur. Assets (in 2004 TZS '000)	Employed Work (0/1)	Salaried Work (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Population Density (sdz.)
	Kagera Health and Development Survey (KHDS)					
	Dependent Variable					
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 3)**						
Core Agglom. (0/1) * EAC 1[2004]	152.450***	754.636***	-0.014	0.026***	0.135***	0.420***
	(52.776)	(12.682)	(0.023)	(0.008)	(0.021)	(0.050)
Core Agglom. (0/1) * CU 1[2010]	264.587***		-0.039		0.069	0.474***
	(70.040)		(0.025)		(0.067)	(0.050)
(a) Standard Erros clustered at EA-Level **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	152.450*	754.636*	-0.014	0.026	0.135***	0.420**
	(91.280)	(447.442)	(0.033)	(0.018)	(0.021)	(0.212)
Core Agglom. (0/1) * CU 1[2010]	264.587**		-0.039		0.069	0.474**
	(108.275)		(0.040)		(0.067)	(0.227)
(b) No Controls **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	122.774***	559.384***	-0.012	0.024***	0.109***	0.411***
	(44.895)	(40.702)	(0.027)	(0.008)	(0.021)	(0.035)
Core Agglom. (0/1) * CU 1[2010]	212.121***		-0.021***		0.035***	0.528***
	(71.097)		(0.031)		(0.079)	(0.080)
(c) Extended Geographic Controls **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	172.061***	713.341***	-0.012	0.026***	0.143***	0.424***
	(39.808)	(69.683)	(0.024)	(0.009)	(0.025)	(0.044)
Core Agglom. (0/1) * CU 1[2010]	269.142***		-0.036		0.080	0.459***
	(54.303)		(0.028)		(0.072)	(0.039)
(f) Urbanities vs. *Core* Agglomeration **(see Table A)**						
Urban (0/1) * EAC 1[2004]	164.031***	712.533***	-0.014	0.020**	0.128***	0.392***
	(62.480)	(36.124)	(0.026)	(0.009)	(0.024)	(0.039)
Urban (0/1) * CU 1[2010]	259.193***		-0.056*		0.080	0.411***
	(81.286)		(0.029)		(0.070)	(0.036)
(g) Core Agglomeration ≤ 25km **(see Table A)**						
Core Agglom. ≤ 25km * EAC 1[2004]	152.845***	754.636***	-0.013	0.026***	0.133***	0.424***
	(53.086)	(12.682)	(0.023)	(0.008)	(0.021)	(0.054)
Core Agglom. ≤ 25km (0/1) * CU 1[2010]	265.826***		-0.037		0.074	0.488***
	(70.593)		(0.026)		(0.067)	(0.059)
(h) Core Agglomeration ≤ 10km **(see Table A)**						
Core Agglom. ≤ 10km (0/1) * EAC 1[2004]	142.565***	737.227***	-0.010	0.027***	0.110***	0.426***
	(53.914)	(17.145)	(0.023)	(0.009)	(0.015)	(0.141)
Core Agglom. ≤ 10km (0/1) * CU 1[2010]	211.579***		-0.042*		0.009	0.495***
	(73.643)		(0.025)		(0.080)	(0.158)
(i) Flexible Distance Specification **(see Table A)**						
Cont. Distance to Core Agglom. * EAC (in '00 km)	-15.980	68.004	0.016	0.003	-0.005	0.049
	(112.097)	(133.455)	(0.017)	(0.004)	(0.019)	(0.051)
Squared Cont. Distance to Core Agglom. * EAC (in '00 km)	0.007		0.000		0.000	0.000
	(0.053)		(0.000)		(0.000)	(0.000)
(k) Excluding post-EAC Migrants **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	124.366***	772.930***	0.009	0.017*	0.143***	0.421**
	(43.012)	(91.216)	(0.022)	(0.010)	(0.026)	(0.169)
Core Agglom. (0/1) * CU 1[2010]	225.536***		-0.049*		0.141*	0.455**
	(47.400)		(0.026)		(0.074)	(0.224)
(m) Logged Dependent Variables **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	0.216*	0.318	-	-	-	0.501***
	(0.114)	(0.349)	-	-	-	(0.051)
Core Agglom. (0/1) * CU 1[2010]	0.306***		-		-	0.522***
	(0.113)		-		-	(0.054)

Notes: *Table continued on next page.*

	Kagera Health and Development Survey (KHDS)					
	Dependent Variable					
Panel b)	Annual p.c. Food Consumption (in 2010 TZS '000)	Annual p.c. non-Food Consumption (in 2010 TZS '000)	Value of Dwelling (in 2004 TZS '000)	Worked in last Year (0/1)	Monthly Salary (in 2004 TZS '000)	Years lived in curr. Residence (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Coefficients **(see Table 3)**						
Core Agglom. (0/1) * EAC 1[2004]	112.914***	27.642	3630.431***	-0.057	7.548	-0.456
	(41.725)	(37.101)	(101.287)	(0.048)	(26.717)	(2.988)
Core Agglom. (0/1) * CU 1[2010]	125.590***	130.221***		-0.008		-0.171
	(40.742)	(43.891)		(0.045)		(2.939)
(a) Standard Erros clustered at EA-Level **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	112.914**	27.642	3630.431	-0.057	7.548	-1.079
	(46.716)	(54.240)	(2880.013)	(0.075)	(28.603)	(1.648)
Core Agglom. (0/1) * CU 1[2010]	125.590**	130.221*		-0.008		-0.404
	(48.701)	(73.028)		(0.051)		(1.636)
(b) No Controls **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	95.129***	20.115	3358.533***	-0.004	29.024	-0.435
	(33.011)	(28.904)	(573.316)	(0.026)	(22.260)	(1.073)
Core Agglom. (0/1) * CU 1[2010]	109.181***	96.844***		0.065***		-0.479***
	(36.436)	(42.779)		(0.039)		(0.952)
(c) Extended Geographic Controls **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	120.084***	37.350	3971.477***	-0.064	10.089	-1.209
	(35.210)	(36.610)	(577.551)	(0.041)	(26.403)	(1.489)
Core Agglom. (0/1) * CU 1[2010]	127.440***	130.428***		-0.011		-0.689
	(37.672)	(35.080)		(0.047)		(1.486)
(d) Urbanities vs. Core Agglomeration **(see Table A)**						
Urban (0/1) * EAC 1[2004]	112.070***	39.684	3429.282***	-0.077*	8.746	-0.760
	(39.545)	(40.223)	(330.496)	(0.041)	(26.955)	(1.370)
Urban (0/1) * CU 1[2010]	138.237***	111.714**		-0.012		-0.299
	(46.960)	(45.793)		(0.045)		(1.513)
(e) Core Agglomeration ≤ 25km **(see Table A)**						
Core Agglom. ≤ 25km * EAC 1[2004]	112.922***	28.019	3630.431***	-0.057	7.548	-0.995
	(41.646)	(37.051)	(101.287)	(0.048)	(26.717)	(1.439)
Core Agglom. ≤ 25km (0/1) * CU 1[2010]	125.537***	131.488***		-0.007		-0.351
	(40.600)	(44.264)		(0.044)		(1.459)
(f) Core Agglomeration ≤ 10km **(see Table A)**						
Core Agglom. ≤ 10km (0/1) * EAC 1[2004]	105.338***	25.895	3228.771***	-0.058	7.148	-0.380
	(38.931)	(45.664)	(821.149)	(0.042)	(21.082)	(1.427)
Core Agglom. ≤ 10km (0/1) * CU 1[2010]	107.641***	95.504**		-0.001		-0.116
	(39.978)	(47.210)		(0.037)		(1.467)
(g) Flexible Distance Specification **(see Table A)**						
Cont. Distance to Core Agglom. * EAC (in '00 km)	-15.980	68.004	0.016	0.003	0.085***	0.003
	(112.097)	(133.455)	(0.017)	(0.004)	(0.027)	(0.004)
Squared Cont. Distance to Core Agglom. * EAC (in '00 km)	0.007		0.000		0.000***	
	(0.053)		(0.000)		(0.000)	
(k) Excluding post-EAC Migrants **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	107.813***	3.365	4599.535***	-0.077	-2.626	-0.456
	(36.869)	(17.126)	(1202.895)	(0.071)	(27.644)	(2.988)
Core Agglom. (0/1) * CU 1[2010]	139.025***	75.908**		0.009		-0.171
	(35.868)	(35.180)		(0.047)		(2.939)
(h) Logged Values **(see Table A)**						
Core Agglom. (0/1) * EAC 1[2004]	0.271*	0.025	0.547	-0.057	0.192	-0.456
	(0.140)	(0.139)	(0.566)	(0.048)	(0.435)	(2.988)
Core Agglom. (0/1) * CU 1[2010]	0.300**	0.154		-0.008		-0.171
	(0.126)	(0.180)		(0.045)		(2.939)

Notes: This table offers an array of robustness tests on the main results of the paper for the differenc-in-difference effect for individuals living in core agglomerations. Data come from the Kagera Health and Development Surveys (KHDS) collected in four waves across 1991-1994, as well as one wave in 2004 and 2010, respectively. In columns (1) through (3) outcome variables represent aggregate household information provided by the head of the household, in columns (4) through (6) they are administered on an individual level. Certain indicators were not sampled in the survey wave of 2010, which is why there is no estimate given for these columns. The sample mean of the respective dependent variable is given in brackets above the estimates. EAC Border (0-1) is the inverse, relative within country distance to the nearest EAC border crossing. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala), if not indicated otherwise. For the initial KHDS survey waves 'Bukoba' - the capital of Kagera representes the core agglomeration. EAC 1[2004] switches on for individuals (re-)sampled in 2004. CU 1[2010], switches on for individuals (re-)sampled in 2010, the second re-interview period of the KHDS. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days, if not indicated otherwise. The regressions also include an indicator whether the household is living in proximity to (former) refugee camps. The regressions testing household-level outcomes, columns (1) through (3), include household fixed effects, the regressions testing individual-level outcomes, columns (4) through (6), include individual fixed effects. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used, if not indicated otherwise. ***, **, * represents significance at the 1, 5

Table B5: Labor Market Effects (AFB and DHS)

	Dependent Variable					
	Afrobarometer			DHS		
	Income					
	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	*Freq. gone without:* [Cash Income] (0-4)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.55]	[1.66]	[2.09]	[0.75]	[1.51]	[0.53]
EAC Border (0-1) * EAC 1[2001-2004]	0.072	0.017	-0.350	0.413***	-0.184	-0.163
	(0.077)	(0.134)	(0.294)	(0.113)	(0.188)	(0.136)
EAC Border (0-1) * CU 1[2005-2009]	0.091	0.040	-0.234	0.161***	-0.291***	-0.269***
	(0.073)	(0.119)	(0.312)	(0.054)	(0.106)	(0.097)
EAC Border (0-1) * CM 1[t ≥ 2010]	-0.001	0.289***	-0.456	0.098**	-0.149**	-0.141
	(0.076)	(0.087)	(0.291)	(0.049)	(0.075)	(0.092)
Core Agglom. (0/1) * EAC 1[2001-2004]	0.026	0.021	0.041	0.096***	-0.046	0.000
	(0.067)	(0.079)	(0.083)	(0.027)	(0.062)	(0.027)
Core Agglom. (0/1) * CU 1[2005-2009]	0.002	0.057	-0.234**	0.019	0.058	0.073*
	(0.055)	(0.090)	(0.105)	(0.059)	(0.111)	(0.042)
Core Agglom. (0/1) * CM 1[t ≥ 2010]	-0.258**	-0.009	0.043	0.076**	0.007	0.029
	(0.110)	(0.087)	(0.079)	(0.038)	(0.088)	(0.035)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	26,563	21,232	35,975	125,539	95,717	96,277
R-Squared	0.23	0.25	0.09	0.22	0.26	0.17
R-Squared -Within	0.23	0.25	0.09	0.22	0.26	0.17

Notes: The results in each column are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.

Table B6: Further Results (KHDS)

	Dependent Variable					
	Kagera Health and Development Survey (KHDS)					
	Consumption			Income		Agglomeration
	Annual p.c. Food Consumption (in 2010 TZS '000)	Annual p.c. non-Food Consumption (in 2010 TZS '000)	Value of Dwelling (in 2004 TZS '000)	Worked in last Year (0/1)	Monthly Salary (in 2004 TZS '000)	Years lived in curr. Residence (Years)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[349.89]	[204.38]	[649.97]	[0.26]	[21.38]	[7.81]
EAC Border (0-1) * EAC 1[2004]	-16.202	-518.546	5064.620	-0.203	-60.533	-0.544
	(229.293)	(1067.545)	(4465.440)	(0.378)	(149.436)	(18.589)
EAC Border (0-1) * CU 1[2010]	-202.509	-737.668		-0.231		1.058
	(241.112)	(1132.954)		(0.343)		(19.004)
Agglomeration (0/1) * EAC 1[2004]	112.914***	27.642	3630.431***	-0.057	7.548	-0.456
	(41.725)	(37.101)	(101.287)	(0.048)	(26.717)	(2.988)
Agglomeration (0/1) * CU 1[2010]	125.590***	130.221***		-0.008		-0.171
	(40.742)	(43.891)		(0.045)		(2.939)
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Individual Fixed Effects	NO	NO	NO	YES	YES	YES
Household Fixed Effects	YES	YES	YES	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	5,494	5,524	4,500	16,330	1,782	5,411
Observations - Fixed Effects	3,817	3,830	2,449	11,599	1,190	3,763
R-Squared	0.87	0.83	0.55	0.86	0.95	0.89
R-Squared -Within	0.05	0.11	0.06	0.27	0.52	0.09

Notes: The results in each column are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 3.

Table B7: Urbanities in Border Regions – Triple Difference Estimates

	Dependent Variable					
	Afrobarometer			DHS		
	Consumption	Income	Agglomeration	Consumption	Income	Agglomeration
Panel a)	*Freq. gone without:* [Wat./Food/Med.]	Employed Work (0/1)	Population Density (sdz.)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.12]	[0.23]	[0.00]	[3.21]	[0.19]	[0.00]
EAC Border (0-1) * EAC	0.032	0.093**	-0.012	-0.193	-0.030	-0.026
	(0.300)	(0.043)	(0.065)	(0.247)	(0.032)	(0.070)
Triple Interaction						
EAC Border (0-1) * EAC * Urban (0/1)	0.498	-0.147	0.383	-0.465	-0.004	0.791*
	(0.368)	(0.092)	(0.417)	(0.355)	(0.071)	(0.405)
Isolated Effect of the EAC on Urbanities in Border Regions						
Combined Effect:						
EAC Border (0-1) * EAC + *Triple Interaction*	0.530	-0.054	0.371	-0.658**	-0.035	0.765*
	[0.15]	[0.55]	[0.37]	[0.02]	[0.65]	[0.05]
Observations	36,042	25,465	38,234	104,483	71,738	104,467
R-Squared	0.13	0.15	0.27	0.48	0.18	0.33
R-Squared -Within	0.11	0.11	0.25	0.47	0.13	0.32

	Afrobarometer			DHS		
	Income					
Panel b)	Worked in last Year (0/1)	Occupation Type (Agr.-Worker-Prof.)	*Freq. gone without:* [Cash Income]	Worked in last Year (0/1)	Occupation Type (Agr.-Worker-Prof.)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.55]	[1.66]	[1.12]	[0.75]	[1.51]	[0.53]
EAC Border (0-1) * EAC	0.054	0.207**	-0.378	0.091*	-0.045	-0.071
	(0.064)	(0.099)	(0.321)	(0.052)	(0.047)	(0.088)
Triple Interaction						
EAC Border (0-1) * EAC * Urban (0/1)	-0.146	-0.383	0.110	0.074	-0.273*	-0.240**
	(0.147)	(0.284)	(0.349)	(0.094)	(0.141)	(0.099)
Isolated Effect of the EAC on Urbanities in Border Regions						
Combined Effect:						
EAC Border (0-1) * EAC + *Triple Interaction*	-0.092	-0.177	-0.268	0.165*	-0.318**	-0.311***
	[0.53]	[0.51]	[0.39]	[0.05]	[0.02]	[0.00]
Observations	26,563	21,232	35,975	125,539	95,717	96,277
R-Squared	0.23	0.29	0.10	0.23	0.32	0.20
R-Squared -Within	0.08	0.26	0.07	0.20	0.24	0.13
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table analyzes the differential effect of EAC Border (0-1)for individuals living in urban regions. Row one shows the uninteracted effect of EAC Border (0-1), i.e. effect of the EAC for individuals at rural border regions, row two shows the differential effect for being in an urban area. Row three depicts the combined effect of the two constituent terms, i.e. the effect of EAC Border (0-1) for individuals in urban regions. The results in each column and panel are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 2.

Table B8: Placebo Tests – Disaggregated DiD in contiguous Countries

	Dependent Variable					
	Afrobarometer			DHS		
	Consumption	Income	Agglomeration	Consumption	Income	Agglomeration
Panel a)	*Freq. gone without:* [Wat./Food/Med.] (0-4)	Employed Work (0/1)	Population Density (sdz.)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[1.25]	[0.24]	[0.00]	[3.20]	[0.15]	[0.00]
EAC Border (0-1) * EAC 1[2001-2004]	0.274*	-0.030	-0.356	0.855***	-0.049	-0.357**
	(0.161)	(0.084)	(0.262)	(0.160)	(0.073)	(0.163)
EAC Border (0-1) * CU 1[2005-2009]	0.036	0.165**	-0.249	0.693***	0.080	0.078
	(0.133)	(0.082)	(0.248)	(0.187)	(0.058)	(0.235)
EAC Border (0-1) * CM 1[t ≥ 2010]	0.179	0.056	-0.139	0.676***	0.000	-0.137
	(0.143)	(0.076)	(0.259)	(0.139)	(0.037)	(0.126)
Core Agglom. (0/1) * EAC 1[2001-2004]	-0.114	-0.033	0.875***	-0.760**	-0.245*	-1.477**
	(0.145)	(0.042)	(0.338)	(0.326)	(0.135)	(0.682)
Core Agglom. (0/1) * CU 1[2005-2009]	-0.025	-0.008	0.529**	0.135	-0.076	-0.352
	(0.069)	(0.036)	(0.267)	(0.194)	(0.116)	(0.447)
Core Agglom. (0/1) * CM 1[t ≥ 2010]	-0.159	-0.056	0.770**	0.166	-0.053	-0.146
	(0.109)	(0.048)	(0.318)	(0.170)	(0.073)	(0.247)
Observations	28,541	18,994	28,573	234,346	126,587	233,841
R-Squared	0.09	0.19	0.23	0.30	0.14	0.33
R-Squared -Within	0.06	0.15	0.20	0.29	0.14	0.32
Panel b)	Worked in last Year (0/1)	Type (Agr.-Work.-Prof.) (1-3)	*Freq. gone without:* [Cash Income] (0-4)	Worked in last Year (0/1)	(Agr.-Work.-Prof.) (1-3)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[0.49]	[1.62]	[2.12]	[0.60]	[1.45]	[0.41]
EAC Border (0-1) * EAC 1[2001-2004]	-0.117	0.001	-0.575***	-0.055	-0.058	-0.007
	(0.090)	(0.170)	(0.220)	(0.129)	(0.210)	(0.249)
EAC Border (0-1) * CU 1[2005-2009]	0.107	-0.012	-0.839***	-0.051	0.076	-0.034
	(0.081)	(0.156)	(0.149)	(0.079)	(0.108)	(0.126)
EAC Border (0-1) * CM 1[t ≥ 2010]	0.037	0.058	-0.462***	-0.027	-0.081	-0.122
	(0.081)	(0.133)	(0.155)	(0.084)	(0.107)	(0.120)
Core Agglom. (0/1) * EAC 1[2001-2004]	0.034	0.028	-0.365**	0.130**	-0.268**	-0.246
	(0.058)	(0.059)	(0.169)	(0.055)	(0.129)	(0.159)
Core Agglom. (0/1) * CU 1[2005-2009]	0.093**	0.001	-0.448***	0.041	0.013	0.016
	(0.041)	(0.076)	(0.044)	(0.063)	(0.073)	(0.095)
Core Agglom. (0/1) * CM 1[t ≥ 2010]	-0.106	-0.040	-0.320***	0.100**	0.075	0.059
	(0.070)	(0.085)	(0.082)	(0.047)	(0.070)	(0.094)
Observations	19,191	12,715	28,274	310,644	187,498	196,153
R-Squared	0.18	0.31	0.11	0.17	0.24	0.18
R-Squared -Within	0.18	0.31	0.11	0.17	0.24	0.18
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table conducts a 'placebo' analysis by testing for a spatially differentiated effect across contiguous, non-EAC countries within the time frame of the EAC's establishment and expansion. As such, in columns (1) through (3), data come from the Malawi, Mozambique and Zambia Afrobarometer surveys rounds 1 through 7 sampled between 1999 and 2018. In columns (4) through (6), data come from the Ethiopia, Malawi, Mozambique, Rwanda, and Zambia Demographic and Health surveys (DHS) sampled between 2000 and 2019. The sample mean of the respective dependent variable is given in brackets above the estimates. EAC Border (0-1) is the inverse, relative within country distance to the nearest border crossing of a contiguous EAC country. Core Agglom. (0/1) is a dummy indicating individuals living in the core agglomeration of their respective country (i.e. Addis Abeba, Kigali, Lilongwe, Lusaka, Maputo). EAC 1[2001-2004] switches to one for individuals sampled from the second half of 2001 to and including 2004, CU 1[2005-2009] for individuals sampled from 2005 and including 2009, and CM 1[t ≥ 2010] for individuals sampled from 2010 onwards. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The results in each column and panel are produced by a separate regression. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table B9: Region-Based Estimates (Placebo Countries)

	Dependent Variable					
	DHS (Region-based)					
	Consumption	Income	Agglomeration	Income		
	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Paid in Cash (0/1)
Panel a)	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.22]	[0.18]	[0.00]	[0.66]	[1.51]	[0.46]
Border Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1991-2000]	0.230*	-0.004	0.028	-0.022	-0.008	-0.087*
	(0.111)	(0.013)	(0.145)	(0.016)	(0.047)	(0.048)
Core Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1991-2000]	-0.159	-0.056	0.964	0.099**	-0.086	-0.072
	(0.171)	(0.034)	(0.641)	(0.039)	(0.066)	(0.059)
Observations	282,480	167,590	289,536	348,016	225,336	223,908
R-Squared	0.27	0.15	0.44	0.15	0.27	0.20
R-Squared -Within	0.25	0.14	0.41	0.13	0.23	0.13
Panel b)						
Border Region (0/1) * pre-EAC 1[1996-2000] pre-EAC 0[1991-1995]	-0.225	-0.004	0.087	-	0.177	0.014
	(0.152)	(0.020)	(0.156)	-	(0.130)	(0.062)
Border Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1991-1995]	0.041	-0.009	0.099	-0.022	0.139	-0.077
	(0.217)	(0.017)	(0.215)	(0.016)	(0.116)	(0.045)
Core Region (0/1) * pre-EAC 1[1996-2000] pre-EAC 0[1991-1995]	0.346	0.088	0.884	-0.107**	0.211**	0.188**
	(0.308)	(0.072)	(0.659)	(0.038)	(0.096)	(0.088)
Core Region (0/1) * EAC 1[t ≥ 2001] EAC 0[1991-1995]	0.107	0.011	1.597*	-0.008	0.072	0.078
	(0.260)	(0.050)	(0.784)	(0.025)	(0.049)	(0.085)
Observations	282,480	167,590	289,536	348,016	225,336	223,908
R-Squared	0.27	0.15	0.45	0.15	0.27	0.20
R-Squared -Within	0.26	0.14	0.41	0.13	0.23	0.13
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table conducts a 'placebo' analysis by testing for a spatially differentiated effect across contiguous, non-EAC countries within the time frame of the EAC's establishment and expansion. Specifically, the analysis makes use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999 and additionally conducts 'pre-tests' towards the difference-in-differences approach. The data thereby come from the full sample of Ethiopia, Malawi, Mozambique, Rwanda and Zambia DHS surveys sampled between 1992 and 2019, making use of AIS, KAP and MIS rounds as well. The sample mean of the respective dependent variable is given in brackets above the estimates. Border Region (0/1) switches to one for individuals living in a region with a median road distance to the nearest border crossing of a contiguous EAC country below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region (0/1) is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Addis Abeba, Kigali, Lilongwe, Lusaka, Maputo). EAC 1[t ≥ 2001] switches to one for individuals sampled from the second half of 2001. Pre-EAC 1[1996-2000] switches to one for individuals sampled in survey years between 1996 and including 2000. As such, in panel a), the reference group of the estimates are comprised of individuals sampled in the full pre-EAC period, i.e. from 1991 to 2000, while in panel b), the reference group is formed by individuals sampled between 1991 and including 1995. Hence, the DiD estimate on 'pre-EAC' in panel b) represents the pre-test. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for clustering at the 'region' level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table B10: Disaggregated Region-Based Estimates

	Dependent Variable					
	DHS (Region-based)					
	Consumption	Income	Agglomeration	Income		
Panel a)	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Paid in Cash (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.12]	[0.19]	[0.00]	[0.76]	[1.52]	[0.53]
Border Region (0/1) * EAC 1[2001-2004]	0.113	-0.030	-0.010	-0.091	-0.039	-0.106
EAC 0[1991-2000]	(0.199)	(0.023)	(0.117)	(0.092)	(0.055)	(0.064)
Border Region (0/1) * CU 1[2005-2009]	0.207	-0.004	-0.164	-0.120**	0.121	0.010
CU 0[1991-2000]	(0.218)	(0.027)	(0.098)	(0.054)	(0.072)	(0.070)
Border Region (0/1) * CM 1[t ≥ 2010]	-0.044	-0.028	-0.030	-0.068	-0.051	-0.136**
CM 0[1991-2000]	(0.178)	(0.030)	(0.113)	(0.058)	(0.059)	(0.065)
Core Region (0/1) * EAC 1[2001-2004]	0.633***	0.142***	3.444***	-0.064	0.149***	0.059
EAC 0[1991-2000]	(0.116)	(0.025)	(0.441)	(0.065)	(0.037)	(0.072)
Core Region (0/1) * CU 1[2005-2009]	0.399***	0.127***	1.878	-0.094	0.210**	0.124***
CU 0[1991-2000]	(0.131)	(0.023)	(1.410)	(0.063)	(0.086)	(0.032)
Core Region (0/1) * CM 1[t ≥ 2010]	0.316**	0.105***	2.064	-0.006	0.111***	0.040
CM 0[1991-2000]	(0.139)	(0.022)	(1.568)	(0.047)	(0.038)	(0.035)
Observations	258,820	104,440	282,866	236,646	142,478	136,163
R-Squared	0.28	0.16	0.52	0.21	0.25	0.17
R-Squared -Within	0.27	0.10	0.51	0.18	0.16	0.08
Panel b)						
Border Region (0/1) * pre-EAC 1[1996-2000]	0.098	0.040	0.126	-0.296**	0.037	-0.013
pre-EAC 0[1991-1995]	(0.125)	(0.029)	(0.105)	(0.121)	(0.057)	(0.078)
Border Region (0/1) * EAC 1[2001-2004]	0.181	-0.006	0.042	-0.314**	-0.018	-0.112
EAC 0[1991-1995]	(0.220)	(0.028)	(0.106)	(0.150)	(0.070)	(0.086)
Border Region (0/1) * CU 1[2005-2009]	0.230	0.015	-0.120	-0.307**	0.032	-0.122
CU 0[1991-1995]	(0.249)	(0.037)	(0.098)	(0.129)	(0.101)	(0.096)
Border Region (0/1) * CM 1[t ≥ 2010]	-0.003	-0.011	0.028	-0.290**	-0.036	-0.139
CM 0[1991-1995]	(0.213)	(0.033)	(0.093)	(0.133)	(0.067)	(0.086)
Core Region (0/1) * pre-EAC 1[1996-2000]	0.440*	0.066***	0.275	-0.245*	0.167*	-0.049
pre-EAC 0[1991-1995]	(0.223)	(0.022)	(0.259)	(0.121)	(0.088)	(0.064)
Core Region (0/1) * EAC 1[2001-2004]	0.899***	0.171***	3.540***	-0.238	0.226***	0.039
EAC 0[1991-1995]	(0.214)	(0.022)	(0.519)	(0.143)	(0.075)	(0.083)
Core Region (0/1) * CU 1[2005-2009]	0.674***	0.154***	2.439	-0.250*	0.300***	0.120**
CU 0[1991-1995]	(0.206)	(0.020)	(1.608)	(0.146)	(0.091)	(0.053)
Core Region (0/1) * CM 1[t ≥ 2010]	0.570**	0.131***	1.988	-0.180	0.166***	-0.005
CM 0[1991-1995]	(0.241)	(0.020)	(1.583)	(0.129)	(0.052)	(0.054)
Observations	255,844	102,005	279,890	232,864	139,407	133,083
R-Squared	0.29	0.16	0.52	0.21	0.24	0.17
R-Squared -Within	0.27	0.10	0.51	0.18	0.16	0.07
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table makes use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999 and additionally conducts 'pre-tests' towards the difference-in-differences approach. The data thereby come from the full sample of Kenya, Tanzania, and Uganda DHS surveys sampled between 1989 and 2020, making use of AIS, KAP and MIS rounds as well. The sample mean of the respective dependent variable is given in brackets above the estimates. Border Region (0/1) switches to one for individuals living in a region with a median road distance to EAC border cossings below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region (0/1) is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Nairobi, Dar es Salaam and Kampala). Pre-EAC 1[1996-2000] switches to one for individuals sampled in survey years between 1996 and including 2000. EAC 1[2001-2004] switches to one for individuals sampled from the second half of 2001 to and including 2004, CU 1[2005-2009] for individuals sampled from 2005 and including 2009, and CM 1[t ≥ 2010] for individuals sampled from 2010 onwards. As such, in panel a), the reference group of the estimates are comprised of individuals sampled in the full pre-EAC period, i.e. from 1991 to 2000, while in panel b), the reference group is formed by individuals sampled between 1991 and including 1995. Hence, the DiD estimate on 'pre-EAC' in panel b) represents the pre-test. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for clustering at the 'region' level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table B11: Disaggregated Region-Based Estimates (Placebo Countries)

	Dependent Variable					
	DHS (Region-based)					
	Consumption	Income	Agglomeration	Income		
	Wealth Index (1-5)	Employed Work (0/1)	Population Density (sdz.)	Worked in last Year (0/1)	Occupation Type (Agr.-Work.-Prof.) (1-3)	Paid in Cash (0/1)
Panel a)	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean of Dep. Var.	[3.22]	[0.18]	[0.00]	[0.66]	[1.51]	[0.46]
Border Region (0/1) * EAC 1[2001-2004]	0.306**	0.005	0.065	-0.045	-0.022	-0.057
EAC 0[1991-2000]	(0.107)	(0.018)	(0.195)	(0.062)	(0.051)	(0.068)
Border Region (0/1) * CU 1[2005-2009]	0.252	0.010	0.063	-0.023	0.007	-0.051
CU 0[1991-2000]	(0.176)	(0.030)	(0.107)	(0.027)	(0.042)	(0.047)
Border Region (0/1) * CM 1[t ≥ 2010]	0.218	-0.007	0.016	-0.019	-0.004	-0.102*
CM 0[1991-2000]	(0.150)	(0.012)	(0.158)	(0.026)	(0.063)	(0.056)
Core Region (0/1) * EAC 1[2001-2004]	-0.051	-0.049	0.486	0.004	-0.021	0.002
EAC 0[1991-2000]	(0.233)	(0.053)	(0.840)	(0.036)	(0.099)	(0.112)
Core Region (0/1) * CU 1[2005-2009]	0.317	0.087*	1.668***	0.058	0.112	0.110
CU 0[1991-2000]	(0.235)	(0.043)	(0.464)	(0.055)	(0.109)	(0.073)
Core Region (0/1) * CM 1[t ≥ 2010]	-0.295	-0.091**	0.883	0.129***	-0.138*	-0.119**
CM 0[1991-2000]	(0.200)	(0.039)	(0.760)	(0.041)	(0.070)	(0.053)
Observations	282,480	167,590	289,536	348,016	225,336	223,908
R-Squared	0.27	0.16	0.45	0.16	0.27	0.20
R-Squared -Within	0.26	0.15	0.42	0.14	0.23	0.13
Panel b)						
Border Region (0/1) * pre-EAC 1[1996-2000]	-0.223	-0.005	0.083	#N/A	0.177	0.014
pre-EAC 0[1991-1995]	(0.152)	(0.019)	(0.157)	#N/A	(0.130)	(0.062)
Border Region (0/1) * EAC 1[2001-2004]	0.118	-0.002	0.134	-0.045	0.124	-0.048
EAC 0[1991-1995]	(0.205)	(0.016)	(0.225)	(0.062)	(0.088)	(0.062)
Border Region (0/1) * CU 1[2005-2009]	-0.049	-0.043	-0.010	-0.056*	0.076	-0.101***
CU 0[1991-1995]	(0.319)	(0.026)	(0.190)	(0.031)	(0.075)	(0.027)
Border Region (0/1) * CM 1[t ≥ 2010]	-0.005	-0.006	0.109	-0.014	0.149	-0.089
CM 0[1991-1995]	(0.241)	(0.026)	(0.217)	(0.024)	(0.128)	(0.069)
Core Region (0/1) * pre-EAC 1[1996-2000]	0.345	0.088	0.888	-0.098**	0.211**	0.189*
pre-EAC 0[1991-1995]	(0.306)	(0.071)	(0.664)	(0.036)	(0.096)	(0.088)
Core Region (0/1) * EAC 1[2001-2004]	0.215	0.018	1.121	-0.095**	0.137	0.152
EAC 0[1991-1995]	(0.379)	(0.054)	(0.790)	(0.042)	(0.120)	(0.155)
Core Region (0/1) * CU 1[2005-2009]	-0.175	-0.059**	0.638	-0.021	-0.053	-0.014
CU 0[1991-1995]	(0.281)	(0.027)	(0.582)	(0.028)	(0.040)	(0.067)
Core Region (0/1) * CM 1[t ≥ 2010]	0.123	0.034	2.044**	0.033	0.072	0.068
CM 0[1991-1995]	(0.280)	(0.062)	(0.921)	(0.028)	(0.058)	(0.079)
Observations	254,644	150,220	261,700	309,353	202,333	200,813
R-Squared	0.27	0.16	0.47	0.16	0.26	0.20
R-Squared -Within	0.26	0.15	0.44	0.14	0.23	0.12
Individual Controls	YES	YES	YES	YES	YES	YES
Geographic Controls	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: This table makes use of the non-GPS survey rounds of the Demographic and Health Surveys (DHS) sampled before 1999 and additionally conducts 'pre-tests' towards the difference-in-differences approach. The data thereby come from the full sample of Ethiopia, Malawi, Mozambique, Rwanda and Zambia DHS surveys sampled between 1992 and 2019, making use of AIS, KAP and MIS rounds as well. The sample mean of the respective dependent variable is given in brackets above the estimates. Border Region (0/1) switches to one for individuals living in a region with a median road distance to the nearest border crossing of a contiguous EAC country below the 10th percentile of all (within-country) GPS-border distances in the sample. Core Region (0/1) is a dummy indicating individuals living in the region which hosts the core agglomeration of their respective country (i.e. Addis Abeba, Kigali, Lilongwe, Lusaka, Maputo). Pre-EAC 1[1996-2000] switches to one for individuals sampled in survey years between 1996 and including 2000. EAC 1[2001-2004] switches to one for individuals sampled from the second half of 2001 to and including 2004, CU 1[2005-2009] for individuals sampled from 2005 and including 2009, and CM 1[t ≥ 2010] for individuals sampled from 2010 onwards. As such, in panel a), the reference group of the estimates are comprised of individuals sampled in the full pre-EAC period, i.e. from 1991 to 2000, while in panel b), the reference group is formed by individuals sampled between 1991 and including 1995. Hence, the DiD estimate on 'pre-EAC' in panel b) represents the pre-test. The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for clustering at the 'region' level. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively.

Table B12: Border Heterogeneity (AFB and DHS)

	Dependent Variable											
	Afrobarometer						DHS					
	Consumption		Income		Agglomeration		Consumption		Income		Agglomeration	
	Freq. gone without: [Wat./Food/Med.] (0-4)		Employed Work (0/1)		Population Density (sdz.)		Wealth Index (1-5)		Employed Work (0/1)		Population Density (sdz.)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample Mean of Dep. Var.	[1.12]	[1.12]	[0.23]	[0.23]	[0.00]	[0.00]	[3.44]	[3.44]	[0.19]	[0.19]	[0.00]	[0.00]
TZA-UGA ≤ 100km (0/1) * EAC	-0.023	0.213	0.062	0.022	-0.020	-0.023	-0.486*	-0.412	-	-	-0.455***	-0.437***
	(0.119)	(0.136)	(0.039)	(0.039)	(0.088)	(0.102)	(0.259)	(0.266)	-	-	(0.111)	(0.096)
TZA-KEN ≤ 100km (0/1)* EAC	-0.061	-0.075	0.064**	0.053	-0.167	-0.167	-0.489*	-0.447	-0.217***	-0.195***	0.039	0.105
	(0.111)	(0.118)	(0.029)	(0.033)	(0.257)	(0.260)	(0.295)	(0.316)	(0.053)	(0.057)	(0.198)	(0.255)
KEN-UGA ≤ 100km (0/1) * EAC	-	-	-0.012	-0.008	-0.118*	-0.117*	-0.460**	-0.378*	-0.036	-0.014	-0.178	-0.103
	-	-	(0.010)	(0.008)	(0.063)	(0.065)	(0.197)	(0.222)	(0.030)	(0.033)	(0.142)	(0.166)
Core Agglom. (0/1) * EAC	-0.360***	-0.555***	0.000	0.016	0.285**	0.287***	0.344**	0.306*	0.055	0.043	0.429	0.437
	(0.074)	(0.093)	(0.041)	(0.021)	(0.113)	(0.103)	(0.154)	(0.169)	(0.042)	(0.040)	(0.331)	(0.328)
Individual Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Refugee Camp Controls	NO	YES	NO	YES	NO	YES	NO	NO	NO	NO	YES	YES
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	36,042	36,042	14,813	14,813	14,930	14,930	29,956	29,956	14,691	14,691	29,956	29,956
R-Squared	0.12	0.12	0.11	0.12	0.26	0.26	0.34	0.34	0.19	0.19	0.36	0.36
R-Squared -Within	0.11	0.11	0.10	0.11	0.24	0.24	0.33	0.33	0.15	0.15	0.35	0.35

Notes: This table analyzes the heterogeneity across the three EAC border regions.TZA-UGA (0/1), TZA-KEN (0/1) and KEN-UGA (0/1) thereby switch to one for individuals living within 100km road distance to the respective border crossing (within either country).The results in each column and panel are produced by a separate regression. The sample mean of the respective dependent variable is given in brackets above the estimates. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes

Table B13: Opinion Polling (AFB and KHDS)

	Dependent Variable						
	Afrobarometer						
	DiD		Contemporary Opinion				
	Present vs. Past: Life Satsfaction (1-4)	*Present vs. Past:* People's Living Standards (1-4)	*Support for:* Regional Integration (1-5)	*Ease of:* Crossing Borders (1-4)	*Helps your Country:* REC / EAC (0-3)	*Helps your Country:* African Union (0-3)	*Would like as Neighbor:* Immigrants/Foreign Worker (1-5)
Panel a)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample Mean of Dep. Var.	[2.70]	[2.71]	[3.71]	[2.21]	[1.66]	[1.62]	[3.38]
EAC Border (0-1)			0.087	0.368***	-0.109	-0.115	-0.245
			(0.249)	(0.083)	(0.095)	(0.092)	(0.160)
Core Agglom. (0/1)			0.000	-0.039	-0.033	-0.004	-0.198***
			(0.059)	(0.039)	(0.038)	(0.027)	(0.057)
EAC Border (0-1) * EAC 1[t ≥ 2001]	-0.356**	0.745***					
	(0.146)	(0.255)					
Core Agglom. (0/1) * EAC 1[t ≥ 2001]	-0.050	-0.209*					
	(0.130)	(0.112)					
Observations	36,213	23,370	6,362	4,766	11,687	13,069	11,647
R-Squared	0.08	0.08	0.10	0.06	0.03	0.03	0.02
R-Squared -Within	0.01	0.02	0.01	0.02	0.00	0.00	0.01

	Kagera Health and Development Survey (KHDS)						
	DiD	Contemporary Opinion					
	Main reason for Migration: Economic (0/1)	*After moving to curr. Residence:* Paid Employment (0/1)	*After moving to curr. Residence:* Paid form. Employment (0/1)	*Activity in curr. residence:* Working (0/1)	*HH. Wealth compared to 10 years ago (1994)* (1-5)	*HH. Wealth today (2004)* (1-5)	*HH. Life Satisfaction (2004)* (1-9)
Panel b)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample Mean of Dep. Var.	[0.20]	[0.14]	[0.07]	[0.83]	[2.62]	[2.62]	[3.79]
EAC Border (0-1)		-0.564**	-0.408**	0.547*	0.274	0.237	0.783
		(0.241)	(0.208)	(0.327)	(0.278)	(0.265)	(0.819)
Core Agglom. (0/1)		0.061	-0.069	-0.051	0.160	0.292**	0.444
		(0.073)	(0.072)	(0.103)	(0.188)	(0.143)	(0.415)
EAC Border (0-1) * EAC 1[2004]	0.855						
	(3.719)						
EAC Border (0-1) * CU 1[2010]	1.143						
	(3.944)						
Core Agglom. (0/1) * EAC 1[2004]	-0.519						
	(0.901)						
Core Agglom. (0/1) * CU 1[2010]	-0.392						
	(0.600)						
Individual Controls	YES	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES	YES
Household Fixed Effects	YES	NO	NO	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	1,347	900	900	774	2,833	2,833	2,833
Observations - Fixed Effects	1,202	2	2	2	2	2	2
R-Squared	0.89	0.16	0.22	0.03	0.02	0.08	0.10
R-Squared -Within	0.10	0.15	0.22	0.02	0.02	0.08	0.10

Notes: This table analyzes opinions and sentimnts of survey respondents. As these data are not available for all tested survey rounds, a difference-in-differences (DiD) estimate can only be conducted for columns (1) and (2) in panel a) as well as column (1) in panel b). The results in each column and panel are produced by a separate regression. All regressions include individual-level controls for respondents' age, gender, as well as education, and also include the geographic controls average monthly temperature, average monthly rainfall, elevation, ruggedness, and the number of growing days. The regressions also include country-year fixed effects. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See full notes below Table 1 and Table 3, respectively.

Table B14: Border Price Pass-Through (KHDS)

	Dependent Variable			
	KHDS			
	Unit Price of Item in Tanzanian Shilling			
	Maize (TZS)	*Millet* (TZS)	*Tea & Coffee* (TZS)	*Sugar* (TZS)
	(1)	(2)	(3)	(4)
Sample Mean of Dep. Var.	[207.25]	[122.83]	[24.60]	[52.50]
EAC Border (0-1) * EAC 1[2004]	-0.019	-1.066*	0.308	-0.048
	(0.090)	(0.563)	(0.928)	(0.094)
Core Agglom. (0/1) * EAC 1[2004]	-0.036***	-	-0.080*	-0.021***
	(0.006)	-	(0.047)	(0.004)
Individual Controls	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES
Household Fixed Effects	NO	NO	NO	NO
Country-Year Fixed Effects	YES	YES	YES	YES
Observations	3,267	193	948	1,722
Observations - Fixed Effects	1,990	127	554	899
R-Squared	0.99	0.99	0.99	0.99
R-Squared -Within	0.01	0.05	0.05	0.00

Notes: This table tests for a differential price-pass through of traded, homogenous, staple goods, following the establishment of the EAC. The results in each column are produced by a separate regression. Data come from the Kagera Health and Development Surveys (KHDS) consumption components which askes respondents for seasonal prices of goods. These data are collected in four waves across 1991-1994, as well as in 2004. All regressions include country-year fixed effects.The results in each column and panel are produced by a separate regression. Binary dependent variables are estimated through a Linear Probability Model (LPM). The standard errors reported allow for spatial correlation, i.e. Conley standard errors are used. ***, **, * represents significance at the 1, 5 and 10 percent level, respectively. See Table 3 for full table notes.