TRANSPORT COSTS AND THE LOCATIONS OF UPSTREAM-DOWNSTREAM SECTORS IN CHINA AND INDONESIA

Jang Ping Thia and Yumin Hu*

Abstract

This paper presents evidence of different upstream-downstream location configurations for China and Indonesia, leveraging provincial input-output (IO) tables. While both upstream and downstream sectors in China exhibit a relatively even geographical distribution across provinces, Indonesia's sectors (even upstream raw material-processing ones) are notably concentrated in Jakarta and Java. To better account for these location differences, we develop a stylized spatial new economic geography model where an upstream competitive raw material processing and a downstream CES sector locate together or separately in different Nash equilibria. Both are located at the large market if transport costs are high, separately given moderate transport costs, and again together but towards the geographic center when transport costs are low. Given the pull of raw materials, low transport costs thus spread upstream and downstream sectors away from the large market. Conversely, low raw material access cost results in both locating in large market.

Keywords: Value chains; upstream, downstream; transport costs; spatial; economic geography; China; Indonesia
JEL classification: F12; F14; O18; O53; R12; R15

* Dr. Jang Ping Thia, Lead Economist and Manager of Economics Department, AIIB, email: jangping.thia@aiib.org; thia3003@gmail.com. Yumin Hu, Nankai University, email: yuminhu1999@outlook.com.

Authors' Note: We are grateful to Mert Kompil for advice and assistance on geospatial computations, Maricris Jan Tobias and Siddharth Poddar for editorial assistance, as well as Gilles Duranton and Stephen Redding for their guidance. We would also like to express our gratitude to the three anonymous reviewers who provided critical comments and suggestions. The data for this research is now available in a public depository for replication (https://data.mendeley.com/datasets/3yx53rgpfn/1).

1. Introduction

Where do firms along a value chain locate within an economy and what is the role of trade distance and transport? [1] We examine the location configurations of upstream and downstream sectors for China and Indonesia through provincial input-output (IO) tables. We provide evidence that both upstream and downstream sectors in China are more evenly distributed across provinces, while Indonesia's sectors (even upstream ones) are highly concentrated in Jakarta and Java. This motivates a stylized extension to a new economic geography (NEG) model incorporating a spatial dimension to account for these location patterns. The model provides additional perspectives on how trade cost shapes economic integration, resource use and regional development.

The model aligns with the reality that upstream sectors tend to be extractive or related to such processing. An upstream sector draws raw material endowments over an interval (representing area) to process these into an intermediate good that is then transported to the downstream industry featuring constant elasticity of scale (CES) differentiated final goods serving both a large and small market. This brings out the tension between locating toward the geographic center to minimize raw material access cost versus locating toward large markets or economic centers. There can be different Nash equilibria, and some of them point to the possibility that firms locate away from the larger market when transport costs fall (opposite of large market effects typical of NEG results).

To fully set the stage, we begin with the key idea that downstream industries favor more economically central locations (i.e., closer to large markets) due to transport costs. There is suggestive evidence that trade costs of final goods are more sensitive to distance and some cross-country evidence that economies that are more economically central have more downstream export structures.[2] There has not been more empirical work to establish this downstream-location hypothesis. We reason that the lack of more evidence is due to the many confounders in international trade. Industry locations are affected by tariffs as well as non-tariff barriers [Ossa (2011)]. Policies and governance add another layer of distortions to location choices [Melitz and Cuñat (2012)]. Furthermore, upstream sectors tend to be extractive or related to the processing of such raw materials.[3] Endowment-driven trade may be difficult to substitute and is hence less sensitive to distance.[4] Where upstream sectors locate can thus be driven by where endowments are found.

Within an economy, one would not expect tariffs or non-tariff barriers to trade (at the least, these should be significantly less formidable). There would be greater homogeneity of

---

[1] The earlier generation of research suggests that firms locate to where there is good infrastructure, linking this to comparative advantage and the patterns of trade [see Cook and Munnell (1990); Martins and Rogers (1995); Bougheas et al. (1999); Yeaple and Golub (2007)].
[2] Antràs and de Gortari (2020) underscore that welfare losses are high with trade tariffs as losses are amplified by the curtailment of multistage production across locations.
[3] There is a conjecture that upstream sectors in advanced economies would be more related to design, research and development etc. (i.e., non-extractive activities). However, note that the most upstream manufacturing industries in the United States are also related to raw materials [Antràs et al. (2012)].
[4] The IMF highlights key characteristics—inelastic demand and supply, large capital expenditure and lead times required for mining and extraction, concentrated number of suppliers, and few exporters meeting the needs of many importers, IMF World Economic Outlook October (2023). These characteristics are consistent with low trade elasticity with respect to distance.

regulatory, linguistic and cultural factors, and policy and governance uniformity (e.g., fiscal and monetary policies, labor standards, etc.). Within-economy analysis thus potentially offers a cleaner environment to test the downstream-location hypothesis and inform policy. Still, we know of no research in this direction and this paper attempts to fill this gap.[5]

Empirically, the paper focuses on two large emerging economies in Asia—China and Indonesia. Both cover large geographies, have large internal markets and are also engaged in trade. China's trade value is 38% of gross domestic product (GDP), while the corresponding figure for Indonesia is 45%, larger than that of the United States at 25%. Both China and Indonesia have large natural resource endowments that support value chain activities.[6] Their industries are plugged into global value chains (GVCs) [World Bank (2020); ADB and ISDB (2019); Kee and Tang (2016)].

However, China and Indonesia are different in terms of geography and transport infrastructure. China has a largely contiguous land mass (an east-west internal distance of around 5,000 kilometers) and a large non-coastal hinterland. Its exports are mainly channeled through the coastal cities. While Indonesia has roughly the same span, it has a tougher archipelago geography. Jakarta and the island of Java, where it sits, dominate economically and contribute more than half of the country's GDP. Extractive industries and agriculture are mainstay sectors in many of the outlying provinces. China also has better-developed internal transport infrastructure, ranked 24th compared to 55th for Indonesia [World Economic Forum (2019)].

Replicating Antràs and de Gortari (2020) by regressing the upstream measure of industrial composition (at provincial level) against measures of economic centrality, we find that economic centrality coefficients are right-signed for Indonesia but either zero or wrong-signed for China (Section 2). On the surface, Indonesia's provincial IO structure seems to support the downstream-location hypothesis but the corresponding provincial data for China's provinces do not. For China, even interior provinces have downstream intermediate shares and internal transport costs would likely be a key explanation.

Our model incorporates an upstream sector that processes extracted raw materials in line with IO stylized facts. The point-to-point transportation cost of a unit raw material to the upstream location is linear in distance but the need to cumulate raw materials over an area (represented by an interval distance) turns this cost function into a quadratic one. The economics thus favors a more geographically central location to minimize raw material costs, before shipping the intermediate goods downstream. However, the downstream industry prefers to be closer to the larger market to save on transport costs. These two forces create an interplay between geographic centrality, which reflects access to raw materials over an area versus economic centrality, which is based on proximity to markets. Multiple equilibria thus arise depending on the relative sizes of the markets, raw material access costs, and transportation costs of intermediates and final goods.

[5] Within-economy IO tables and location data would be needed for such analysis. The lack of such data could also have limited the number of such studies.
[6] See World Bank's Changing Wealth of Nations dataset [World Bank (2021)]. China is estimated to have non-renewable natural capital (oil, natural gas, coal, metals and minerals) worth USD2.5 trillion in 2018, and cropland, pastureland and timber wealth worth USD6 trillion. The corresponding figures are USD451 billion and USD874 billion for Indonesia.

In some equilibria, the upstream and downstream sectors co-locate. In other equilibria, they separate to different locations, with the downstream sector at the large market while upstream operating closer to the geographic center. It is even possible that both the upstream and downstream sectors co-locate at the larger market when transport costs are too high (such as in Indonesia). We show that the reduction of transport costs can shift equilibrium locations away from the large market, spreading development closer to the geographic middle and nearer to the remote location. The paper also shows that with a specific utility function, the decentralized location choice coincides with the utilitarian social planner. When transport costs are sufficiently low, the market solution maximizes equality between regions.

Our paper is related to several related strands in the literature. First, it pertains to economic geography. The model here does not feature any agglomeration or congestion to allow for more tractable Nash equilibrium results. Nevertheless, symmetric final good firms make the same location choice and equilibrium agglomeration occurs. Our baseline model uses two sectors and two locations to present key results more intuitively, featuring a menu of equilibria as opposed to solving for a single spatial general equilibrium [Caliendo and Parro (2014); Redding and Rossi-Hansberg (2017), etc.]. We provide an extension to the two-location baseline model to one where population (and hence consumption) is distributed along the entire spatial interval, with largely unchanged equilibrium properties.

Generally, the reduction of transportation costs in the final goods sector tends to accentuate large market effects in economic geography models [Krugman and Venables (1995); Venables (1996)]. Faber (2014) for example finds that highways in China have reduced activity in periphery regions. Our model nonetheless highlights a potential opposite effect where the lowering of transport costs brings sectors toward the geographic center given the pull of raw material access—a dispersion force that spreads development geographically.

Second, it relates to the literature on value chains [Baldwin and Venables (2013); Antràs and de Gortari (2020)]. An objective of this paper is to enrich the understanding of the downstream-location hypothesis but our motivation is related to the wider developmental context. Upstream activities are associated with resource exploitation with little economic spillovers. It has thus been argued that more is needed to integrate such activities with broader economic development [Sachs and Warner (2001); Singer and Donoso (2008); Amendolagine et al. (2019); Beverelli et al. (2019); McNerney et al. (2022), etc.]. This also relates to the concerns of “commodity-for-manufactures” literature [Costa et al. (2016)].

For Indonesia specifically, development banks have stressed the need to anchor more downstream value chain manufacturing activities [ADB and ISDB (2019); World Bank (2020)]. Policy makers are often keen to use raw materials to anchor downstream industries. Our model shows that a reduction in transport costs can distribute even downstream development toward more remote regions and points to the need for policy for regional development [Baldwin et al. (2005)].

Finally, it relates to the rich literature on transport infrastructure and trade costs [Fernald (1999); Michaels (2008); Donaldson and Hornbeck (2016); Donaldson (2018); Porteous (2019); Banerjee, Duflo and Qian (2020); Egger, Loumeau and Loumeau (2023); Thia and Ong Lopez (2023)]. In the context of value chains, trade costs also include non-explicit costs such as time delays, regulations, management, marketing etc., [Hummels and Schaur (2013);

Kalnins and Lafontaine (2013); Head and Mayer (2019)]. Consistent with the literature, the quality of transport infrastructure will determine where industries locate and these can give rise to regional inequalities [Redding and Venables (2004); Waugh (2010); Fan (2019); Porteous (2019)]. The wider literature shows that better infrastructure leads to more trade but poorer infrastructure can reduce gains from trade [Limao and Venables (2001); Celbis et al. (2013); Nordas and Piermartini (2004); Francois and Manchin (2013)].

Section 2 describes the data used and highlights the preliminary findings. Section 3 provides the model, characterizes the equilibrium and discusses the implications. Section 4 then provides the results of the regressions and discusses the model's implications. Finally, Section 5 concludes.[7]

2. Data and Preliminary Analysis

This section provides a description of the data and presents the various findings that motivate the model.

2.1 Input-output Tables and Sector Upstream Scores

For China, there is no official inter-province IO table. The inter-province 2015 IO table is drawn from Zheng et al. (2020), who used various provincial-level data (such as output and inter-province trade) to piece together the IO dataset. A total of 31 provinces and 42 sectors are covered. Of these, 27 sectors are included in our analysis paper, covering extraction and related processing in the primary sector, manufacturing and construction. Given the focus on industrial sectors and linkages, we exclude agriculture and services sectors from the analysis.

Using Antràs et al. (2012), the upstreamness measure for the sectors is computed and presented in Table 4 in Appendix A.1. Briefly, the upstreamness measure captures the number of times a sector's products churn through the IO table as intermediates, before they reach final consumption. As expected, the upstream sectors are related to resource extraction and processing, petroleum and gas processing products, coal mining and products and the like.[8]

The inter-regional IO table for Indonesia (2016) is available from the official statistical bureau, Bandan Pusat Statistik (BPS). It covers 52 sectors, with input-output relationships of 34 provinces. Likewise, we focus on extraction and related processing, manufacturing and construction sectors (totaling 24 sectors). Again, as can be seen in Table 5 of Appendix A.1, the most upstream sectors are also related to resource extraction or processing, specifically coal, gas, and metal mining, as well as related processing. Indonesia's upstreamness

[7] Appendix A.1 and A.2 provide the details for industry- and provincial-level upstreamness computations respectively. Appendix A.3 provides the solution to a special case of the model where there is no vertical separation between upstream and downstream sectors. Appendix A.4 provides the equilibrium characterization and numerical solutions for an extension to the model to allow for population locating along the entire interval (as opposed to population at two locations only in the main model).

[8] Services tend to be downstream and yet with high value-add, resulting in the positive correlation between upstreamness and downstreamness measures [Antràs & Chor (2018)]. As services sectors are excluded in our analysis, we do not address this positive correlation and instead rely on a single measure of upstreamness. In this paper, co-location refers to that for industrial sectors (rather than services sectors co-locating with industries).

measures tend to be lower, suggesting that the value chains in Indonesia are shorter as compared to in China.

2.2 Key Upstream and Downstream Location Differences Between China and Indonesia

We next investigate the provincial shares of upstream and downstream sales. Note that for this section, to focus sharply on intermediate goods in value chains, sales refer to the value of products sold to other provinces or as direct exports, but exclude product values sold within a province or directly sold for final consumption.

We define upstream sectors as those with upstreamness measures in the top half of the respective IO tables. Specifically, this refers to the top 13 sectors for China with upstreamness greater than 3.7 (Table 4 of Appendix A.1) and the top 12 sectors for Indonesia with upstreamness greater than 2.4 (Table 5 of Appendix A.1). The remaining sectors are defined as downstream. Formally,

$$\Omega^{UP,CHN} \equiv \{i \in \Omega^{CHN}: U_i > 3.7\}, \quad \Omega^{DOWN,CHN} = \Omega^{CHN} - \Omega^{UP,CHN}$$

$$\Omega^{UP,IDN} \equiv \{i \in \Omega^{IDN}: U_i > 2.4\}, \quad \Omega^{DOWN,IDN} = \Omega^{IDN} - \Omega^{UP,IDN}$$

where Ω^{CHN} and Ω^{IDN} are the sets of sectors in China and Indonesia, respectively and UP and $DOWN$ denote upstream and downstream sectors respectively.[9] For each province n, sales in upstream and downstream sectors are

$$X_n^{UP} = \sum_{i \in \Omega^{UP}} X_{ni}\,, \quad X_n^{DOWN} = \sum_{i \in \Omega^{DOWN}} X_{ni}$$

shares of sales in upstream and downstream

$$s_n^{UP} = \frac{X_n^{UP}}{\sum_n X_n^{UP}}, \quad s_n^{DOWN} = \frac{X_n^{DOWN}}{\sum_n X_n^{DOWN}}$$

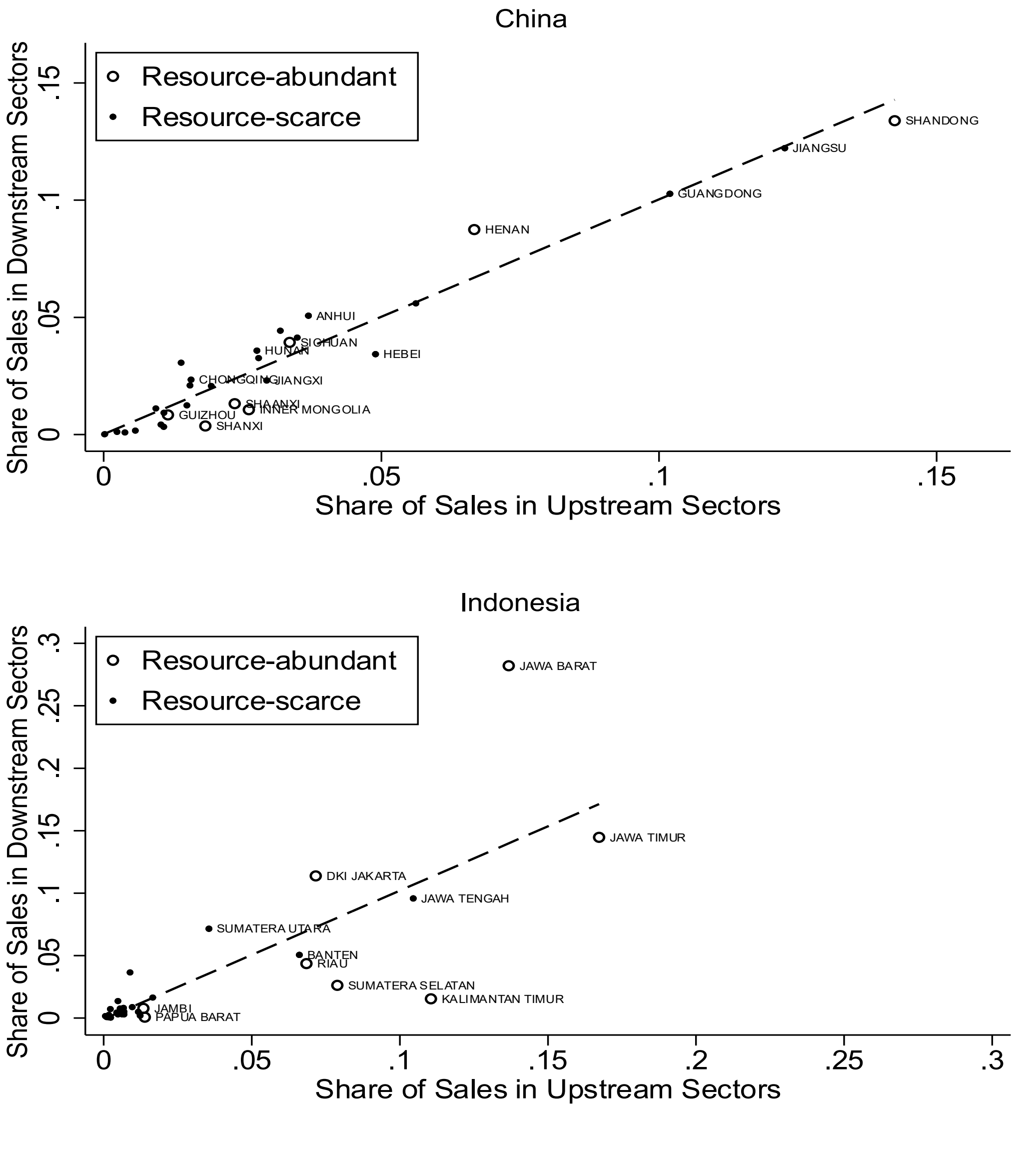
These shares provide the summary statistics on how much each province participates in cross-province value chains and exports. The correlations of upstream and downstream shares are then provided in Figure 1.

For China, there is a high degree of correlation between the provincial shares of upstream and downstream sectors (shares are close to the 45-degree line). Coastal provinces (Shandong, Jiangsu, Guangdong) see high shares in both upstream and downstream sales as to be expected given their geography. Interior provinces (such as Henan, Anhui, Hubei, Hunan, Jiangxi and Chongqing) also have moderate upstream and downstream shares.

[9] Sectors in China having longer value chains and hence higher numerical upstreamness measures (as described earlier). Hence, rather than imposing an arbitrary numerical upstreamness threshold to classify sectors into upstream or downstream, we broadly split the sectors in each economy into two halves based on sectoral upstreamness measures instead, with the top half being upstream and the bottom half being downstream.

In contrast, for Indonesia, most provinces have little participation in value chains (as indicated by the bunching up of scatter points close to 0). Kalimantan and Sumatra see a high share of upstream sales but little downstream sales. More remote regions (such as Aceh and Papua) and even the geographically central ones and resource-rich ones (like Sulawesi) see little value chain sales too. There is little evidence of upstream sectors anchoring in some resource-rich provinces, let alone downstream ones. Strikingly, provinces in Java (such as Java Barat, Java Timur, Jakarta, Jawa Tengah, Banten) see both high upstream and downstream sales.[10]

Figure 1: Province Shares of Upstream and Downstream Sales in China and Indonesia



Note: Provinces with abundant natural resources are defined as those where sales from natural resource sectors rank among the highest. Specifically, we focus on coal, oil and gas, which correspond to sectors 2 and 3 in China (as shown in Table 4) and sectors 8 and 9 in Indonesia (as shown in Table 5).

[10] One entity in Java with low upstream and downstream shares is the special administration region Yogyakarta.

2.3 Stronger Internal Market Access for China

We provide further evidence that China's sectors and their value chain participation, where it occurs, results in greater internal market access. We compute the average sales distance (ASD). Specifically, the bilateral provincial distances are weighted by the shares of sales, giving an indication of how far sales in that province travel on average. Formally, the market access sector i in province n is obtained as:

$$ASD_{ni} = \sum_{k} \frac{X_{nki}}{X_{ni}} Distance_{nk}$$

where X_{nki} refers to the province sales of sector i from origin n to destination k. Note that we exclude foreign sales in this computation so as to account for only internal sales distance. Thus, a province with higher sales of sector i to more distant provinces will be considered to have higher market accessibility, or sales distance, in that sector. Conversely, a province with most of its sales in sector i to nearby provinces will have a lower ASD.[11]

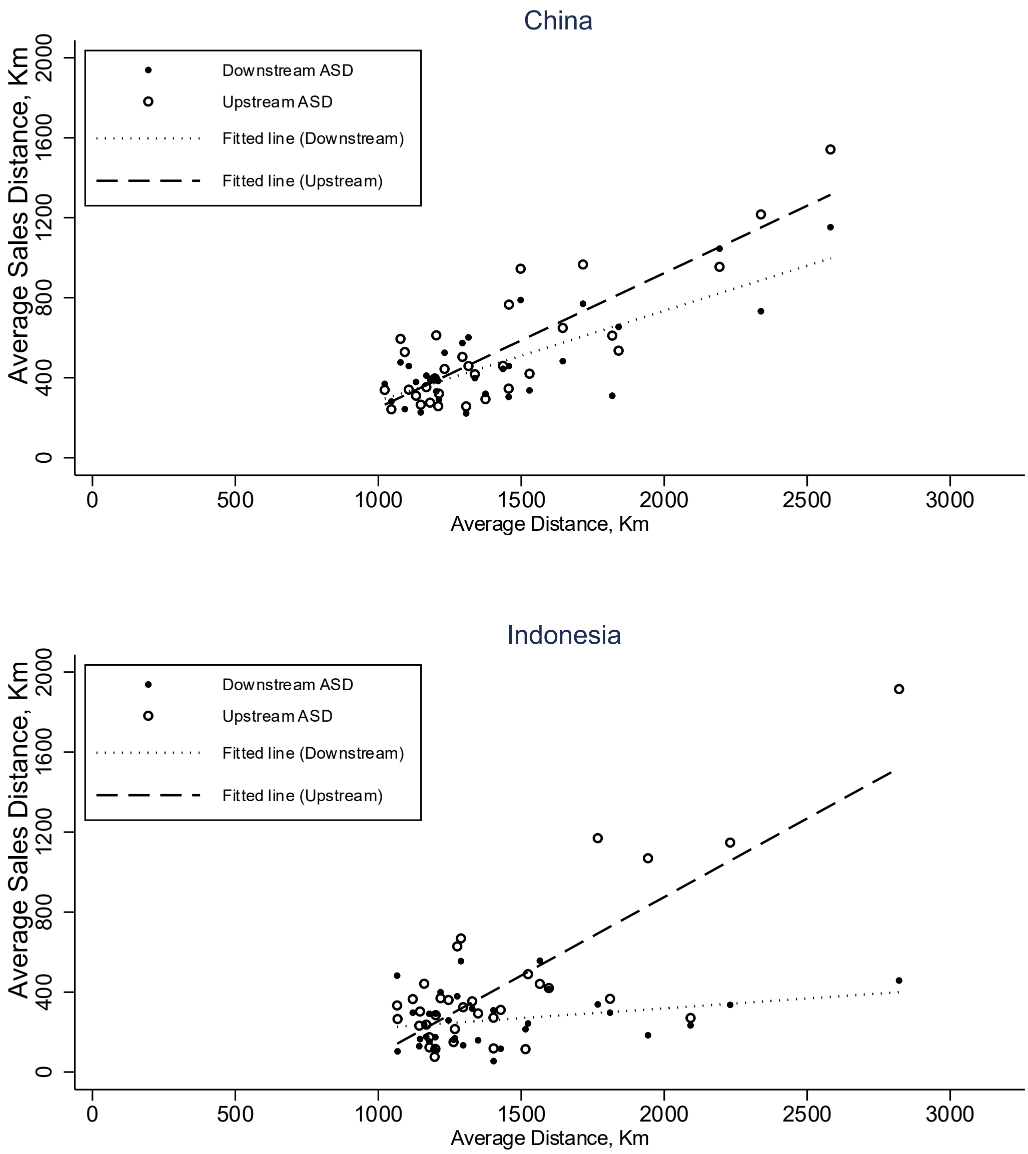
An equal weight of destination sales across province k would lead to a simple average sales distance being the same as average distance (note that when $\frac{X_{nki}}{X_{ni}} = \frac{1}{N}$ for all destination k, the market access will be simplified as $\frac{1}{N}\sum_{\mathrm{k}} Distance_{nk}$). On the other hand, with a relatively high weight (sales X_{nki}) assigned to the nearby provinces and relatively low weight (sales X_{nki}) assigned to the distant provinces, we will obtain a lower sales distance compared to average distance. The relationship between province level sector output weighted ASDs and province average distance (to all other provinces) is presented in Figure 2.

A few observations stand out. First, China's provinces have a higher average ASD (506 kilometers [km]), while Indonesia's provinces—including those on Java—have a lower average ASD (343 km). Conditioned on average provincial distances, sectors in China send their sales over longer distances while sectors in Indonesia produce and sell closer to markets. Second, in China, upstream sectors send sales over longer distances compared to downstream sectors, but the difference is not large, as evidenced by the fitted lines in Figure 2. Third, there is also less heterogeneity among upstream sectors in China.

For Indonesia, there are big differences between upstream and downstream location patterns as with their market access. Downstream sectors see considerably lower ASDs relative to average distance compared to China. Indonesia's upstream sectors have ASDs similar to China's but also exhibit greater heterogeneity in the ASD and average provincial distance relationship–some upstream sectors produce close to their downstream markets, while others send sales over larger distances. We will revisit these points in the discussion section.

[11] Note that direct exports are excluded in this computation. Within provincial sales (where $n = k$) are included. For the computation of ASD, internal provincial distance is measured as $0.67\sqrt{(area/\pi)}$ [Mayer and Zignago (2011)]. In some sense, ASD corresponds to the market access term of Redding and Venables (2004) where trade partners' and internal GDPs are weighted by distance to create a measure of market access. The key difference is that ASD measures actual realization of the distance of goods travelled, as opposed to just theoretical exporting potential. See also Ganapati and Wong (2023) for a review of distance of goods travelled internationally.

Figure 2: Average Sales Distances of Provinces in China and Indonesia



2.4 Province Average Upstreamness and Economic Centrality Measures

We further describe a province's average position in global production chains—that is, the sectors' upstreamness measures are weighted by their sales for each province to derive a weighted average upstreamness measure for the province. As per Antràs and de Gortari (2020), the provincial upstreamness is then regressed against economic centrality measures, and the results are provided in Table 1. Further details on various economic centrality measures and the regression specification are provided in Appendix A.2.

For Indonesia, the economically central provinces are more downstream structurally with the right negative coefficients (whether based on GDP or population-weighted centrality) across all specifications.[12] For China, on the other hand, coefficients are positive after controlling for GDP per capita and university rate and are “wrong-signed” (albeit not significant). The regressions for China’s provinces have practically no explanatory power as seen in the low R-squared statistics.

There is no evidence that more economically central provinces have a more downstream sales structure, based on the aggregated provincial-level sales structure. This finding could be due to several factors. It could be that economic centrality measures are not particularly informative for China, given that hinterlands (such as in Chongqing and Sichuan) also have large populations, though the explanation is not likely complete, given the data show that China’s coastal regions are indeed more economically central compared to those in the interior (see Figure 5). Second, sectors in China are indeed more spread out as we will show in later sections. Guided by these observations, we provide a stylized model that allows for different equilibria, some of which would see sectors that are more spread out from large markets.

Table 1: Provincial Upstreamness Regressions Against Economic Centrality Measures

Dep var: Provincial Upstreamness	Part (a): China				Part (b): Indonesia			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Centrality (GDP Weighted)	-0.010		0.091		-0.204**		-0.255***	
	(0.121)		(0.151)		(0.081)		(0.056)	
Centrality (Population Weighted)		-0.012		0.089		-0.219**		-0.258***
		(0.125)		(0.156)		(0.082)		(0.058)
GDP Per Capita			-0.140	-0.136			0.223***	0.216***
			(0.094)	(0.092)			(0.051)	(0.051)
University Rate			-0.000	-0.000			-0.002	-0.002
			(0.000)	(0.000)			(0.002)	(0.002)
Observations	31	31	31	31	34	34	34	34
R-squared	0.000	0.001	0.060	0.058	0.155	0.170	0.526	0.521

Notes: Robust standard errors in parentheses. $*** \ p < 0.01,\ ** \ p < 0.05,\ * \ p < 0.1$.

2.5 Distance and Travel Times

The evidence thus points to quite different upstream-downstream location configurations between China and Indonesia. China’s sectors are more spread out geographically and enjoy stronger internal market access. Indonesia, on the other hand, sees more concentration—even upstream sectors concentrate in Jakarta and Java rather than more outlying provinces where the resources are. We posit that the difference in transport costs—specifically between costs of raw material access versus transport costs to market—is key to understanding these different location configurations. We provide evidence to support this line of thinking before presenting a formal model.

For China, we consider Beijing, Shanghai and Guangzhou as the three large conurbations. Distance to a large city is then calculated as the direct distance from the province’s center to the nearest of the three large cities (i.e., minimum), computed using ArcGIS. The shortest

[12] Provincial upstreamness also correlates positively with higher provincial per capita GDP. This positive correlation is due to upstream provinces in Indonesia registering high per capita GDP arising from resource rents (see Figure 9 of Appendix A.2).

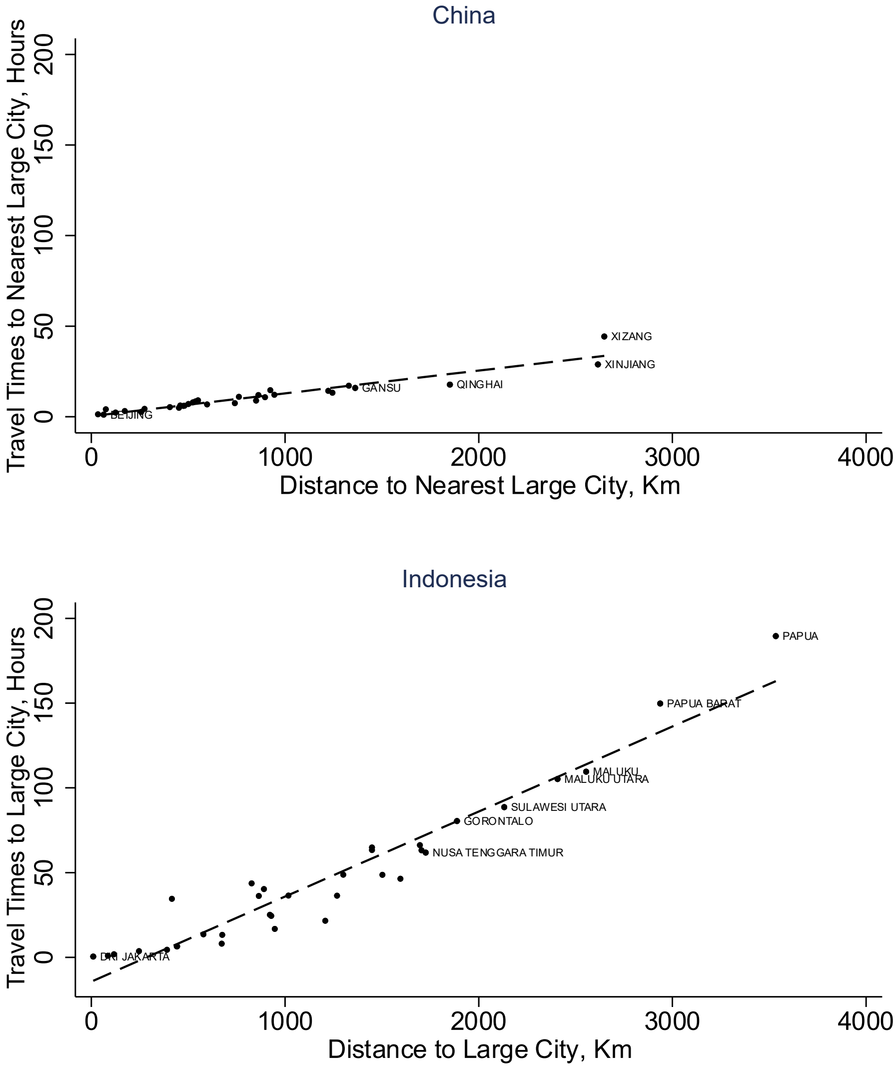
surface travel time to the nearest large city is computed using ArcGIS Network Analyst. For China, the surface travel times are more straightforward given the largely contiguous landmass, although the travel time from Hainan also includes sea travel time. A few entities, including Hong Kong, China and Macao, China, are not included in the analysis due to the practical constraints of not having input-output data with the rest of the provinces.

For Indonesia, Jakarta is defined as the large city. The distance and shortest surface travel time to Jakarta are computed. For provinces that are separated from Jakarta by sea (which is common for an archipelago), surface travel would typically include road travel time to a known seaport, sea transport at 20 kilometers per hour to another seaport in Java, and then road transport to Jakarta. We assume the connections are “smooth” and there are no delays or waiting times during transfers between roads and seaports. We also assume there is no international transit (such as through Singapore). In general, transport time data are computed based on assumed travel speed given the road types, as opposed to actual travel speed. In other words, actual congestion effects on the ground are not accounted for, which is an acknowledged limitation of this study.

We first provide the preliminary scatter plot highlighting the effect of travel times in Figure 3. The X-axis shows the distance to the large city, while the Y-axis provides the surface travel times as described. Indonesian provinces outside Java have a significantly higher travel time over equivalent distance, compared to China. This finding highlights Indonesia’s comparatively tougher internal geography, both due to physical geography as well as the effects of transport infrastructure.

For China, there is less dispersion of travel times, and most are close to the fitted distance-travel times line. Second, the increase in travel times per distance is much lower. For example, Xizang, Xinjiang and Qinghai have similar distances to large cities as Papua, Maluku and Sulawesi, but the corresponding travel times are half or less than that. Provinces on Java Island have much better connectivity to Jakarta as seen in the samples in the lower left side of the Indonesian chart.

Figure 3: Distance and Travel Times to Large City (China and Indonesia)



Note: The minimum distance to three cities (Beijing, Shanghai, Guangzhou) for China is employed to account for the economic effect of these markets (which are also key export gateways for China). While this narrows the domain of the X-axis for the chart on China (given it is the distance to the nearest), it does not otherwise affect the slope that shows the distance-to-travel times relationship, which is strictly due to transport.

3. The Baseline Model

We now present the model to reconcile stylized IO facts and locational differences with transport costs. An economy is characterized by two regions—Remote and City, subscripted with R and C, populated by identical consumers but with different population sizes with $L_C > L_R$. Consumers at both locations are not mobile and have unit income from labor and utility function

$$V = (1 - \mu) \ln A + \mu \ln X$$

where $X = \left[\int x_i^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}$ is the CES aggregated consumption of differentiated goods, with each firm manufacturing a variety i. The expenditures for each consumer are $1-\mu$ and μ for agriculture and the differentiated sector, respectively. A (agriculture) is the standard outside sector where trade is costless and acts as the numeraire ($P_A = 1$). We assume this is always active and hence equating wage rates in both regions.[13] Aggregated across, the demands from each region for each differentiated variety are given as

$$x_{iR} = \frac{p_{iR}^{-\sigma}}{P_R^{1-\sigma}} . \mu L_R \qquad x_{iC} = \frac{p_{iC}^{-\sigma}}{P_C^{1-\sigma}} . \mu L_C$$

Where P_R and P_C are the standard CES aggregated price.

3.1 Endowments, Production Location and Trade Costs

The Remote and City locations are characterized by a distance between them—D. The production location can be anywhere along this interval of distance, where d denotes the distance from Remote and $D-d$ the distance from City. Let $r, s = \frac{d}{D}$ denote the fractions of distance from Remote for the upstream and downstream sectors, respectively.

Along D, there is a continuum of different endowments or raw materials. Extraction is costless, fully competitive, but access to raw materials is costly. To bring a unit raw material from $r = 0$ to $r = 1$ (the full length of the interval), there will be a cost of τ_e proportionate to distance (bringing the unit endowment from $r = 0$ to $r = \frac{1}{2}$ will cost $\frac{\tau_e}{2}$). The need to cumulate raw material endowments across the continuum results in the following cost function:

$$e = \frac{\tau_e}{2} r^2 + \frac{\tau_e}{2}(1-r)^2 = \frac{\tau_e}{2}(2r^2 - 2r + 1)$$

The key point here is that the point access cost of unit raw material is linear in distance, but e is quadratic because of the need to cumulate resources over the interval.

With $\frac{\partial e}{\partial r} = \tau_e(2r-1)$, e is minimized at $r = \frac{1}{2}$, the mid-point location where the cost of the bundle of raw materials will be lowest. To help with conceptualization, these can be commodities' (hard or soft) extraction spread over a wide area, to be collected to the assembly location. It can even represent areawide energy harvesting (such as solar or wind farms over many different locations and transmitted to assembly). The continuum assumption is used for convenience. The same effect of a cumulative cost curve would also result if raw materials need to be brought in from various discrete positions along the interval—which is to say the cost of raw material access will also be lower near the geographic center.[14]

---

[13] With suitable scaling of labor productivity in the agriculture sector, one can think of a simplification where all labor is hired in this 'outside agriculture' sector at the same wage, which is then spent on agriculture and the CES sector. As will be elaborated later, this focuses the model on the CES demands and the interplay of trade costs between upstream and downstream sectors, without concern for potential migration induced agglomeration.

[14] Dutu (2015) provides the description and data on the distribution of natural resources across Indonesia. Li et al. (2013) provide a similar overview for China.

The raw materials transferred to location r are assembled competitively (upstream) to become an intermediate good, which must be shipped to location s where production of the final good takes place.[15] Inclusive of intermediate shipping cost, the input cost function for the final good (downstream) sector becomes

$$m = \frac{\tau_e}{2}(2r^2 - 2r + 1) + \tau_m|s - r|$$

The above is the cumulated access cost of bringing raw materials to location r, assembling the intermediate good, and then shipping it to location s—hence $\tau_m|s - r|$. The shipping cost of intermediates, point-to-point in nature, is linear. Finally, to ship a final good variety to Remote and City from this production location, the transport costs are

$$\tau_R = 1 + s.\tau_X \qquad \tau_C = 1 + (1 - s)\tau_X$$

where $\tau_X > 0$ is the linear cost of transportation of the final good from s. Note that τ_R and τ_C then take the iceberg form, as is standard in the literature. If production is at Remote ($s = 0$), $\tau_R = 1$, $\tau_C = 1 + \tau_X$. If production is at City ($s = 1$), $\tau_R = 1 + \tau_X$, $\tau_C = 1$. Where production happens thus affects the price levels and welfare of Remote and City. For the ease of exposition, where needed, we use "left" to denote a location that is relatively closer to Remote, and "right" to denote a location relatively closer to the City, along the interval distance. While we have denoted the transport cost of intermediates τ_m as separate from goods τ_X, these transport costs are likely to be similar. Combining τ_m and τ_X into a common term does not affect equilibrium conclusions.

We assume that no consumption takes place at the upstream or downstream production locations, as a simplification. Typically, with consumption at the production location, there will be a further agglomeration force as consumption demand reinforces the attractiveness of the production location. In Appendix A.4, we provide an extension to the model with a fixed population located along the interval, thereby softening the two-location assumption.

Production entry requires a fixed cost F per period, which pins down the number of varieties in equilibrium as is standard in the literature. Consistent with the model setup, we assume no migration or agglomeration externality, but firms will nonetheless "agglomerate" in the same location given they are symmetric in decisions. There is no congestion cost, again consistent with a sector A that is operative throughout.

3.2 Characterizing the equilibrium

With the assumptions above, we have added spatial features to a canonical two-location NEG model with home market effects. The standard pricing decisions for a firm are given as

$$p_{iR} = \frac{\sigma}{\sigma - 1}.m.[1 + s.\tau_X] = \frac{\sigma}{\sigma - 1}\left[\frac{\tau_e}{2}(2r^2 - 2r + 1) + \tau_m|s - r|\right].[1 + s.\tau_X]$$

$$p_{iC} = \frac{\sigma}{\sigma - 1}.m.[1 + (1 - s)\tau_X] = \frac{\sigma}{\sigma - 1}\left[\frac{\tau_e}{2}(2r^2 - 2r + 1) + \tau_m|s - r|\right].[1 + (1 - s)\tau_X]$$

[15] To borrow an analogy imperfectly from Baldwin and Venables (2013), one can think of upstream as the 'spider' part of the production where raw materials need to be collected over all locations, before the 'snake' part of sending the intermediate downstream. In our model, the 'spider' part of the cost is minimized at the geographic center.

which are the standard markups on input and iceberg transport costs. When $s \geq r$

$$\frac{\partial p_{iR}}{\partial s} = \frac{\sigma}{\sigma-1}.\left\{\tau_m.[1+s.\tau_X] + \left[\frac{\tau_e}{2}(2r^2-2r+1)+\tau_m|s-r|\right].\tau_X\right\}$$

$$\frac{\partial p_{iC}}{\partial s} = \frac{\sigma}{\sigma-1}.\left\{\tau_m.[1+(1-s)\tau_X] - \left[\frac{\tau_e}{2}(2r^2-2r+1)+\tau_m|s-r|\right].\tau_X\right\}$$

The profit function for each firm is given as

Equation 1

$$\pi_i = \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2r^2-2r+1)+\tau_m|s-r|\right].[1+s.\tau_X]\right]^{1-\sigma}\frac{\mu L_R}{P_R^{1-\sigma}} + \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2r^2-2r+1)+\tau_m|s-r|\right].[1+(1-s)\tau_X]\right]^{1-\sigma}\frac{\mu L_C}{P_C^{1-\sigma}} - F$$

Proposition 1: The decentralized choice of s coincides with the utilitarian social planner choice, given the logarithm-form utility function.

Taking the first order condition (FOC) of the profit function with respect to s gives

Equation 2

$$\frac{\partial \pi_i}{\partial s} = 0 \Leftrightarrow \frac{1}{P_R}.\frac{\partial p_{iR}}{\partial s}.L_R + \frac{1}{P_C}.\frac{\partial p_{iC}}{\partial s}.L_C = 0$$

This equation provides the implicit solution to the optimal choice of s, which pins down the decentralized equilibrium together with the zero-profit condition as below,

Equation 3

$$\pi_i = \frac{1}{\sigma}\frac{1}{N}\mu L_R + \frac{1}{\sigma}\frac{1}{N}\mu L_C - F = 0 \Rightarrow N = \frac{\mu(L_R+L_C)}{\sigma F}$$

and with the standard CES price indices as

$$P_R^{1-\sigma} = N\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2r^2-2r+1)+\tau_m|s-r|\right].[1+s.\tau_X]\right]^{1-\sigma}$$

and

$$P_C^{1-\sigma} = N\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2r^2-2r+1)+\tau_m|s-r|\right].[1+(1-s).\tau_X]\right]^{1-\sigma}$$

Consider the utilitarian planner's problem. For each individual, the utility of consuming X is the inverse of its CES perfect price index. Hence, summing across two population groups, the utilitarian social planner's optimal choice of s is formally defined as

$$\tilde{s} = \arg\max_s \; -L_R \, ln\, P_R - L_C \, ln\, P_C$$

subject to the zero-profit condition (Equation 3). The utilitarian planner's FOC is thus the same as the firm maximization problem ($\frac{\partial \bar{s}}{\partial s} = 0 \Leftrightarrow \frac{1}{P_R}.\frac{\partial p_{iR}}{\partial s}.L_R + \frac{1}{P_C}.\frac{\partial p_{iC}}{\partial s}.L_C = 0$), subject to the same zero-profit condition. (Q.E.D).

In some sense, this result is natural and unsurprising. As there are neither agglomeration nor congestion forces in this setup, there is no locational externality, and the market solution is the planner's one as well. Furthermore, this result is a narrow one in the sense that it applies only to the log utility specification, where the maximizing problem of the firms results in the perfect price indices at the denominator. The result is also a constrained maximum in the Dixit-Stiglitz sense.[16] Nevertheless, this finding provides comfort that the decentralized equilibrium maximizes utilitarian social welfare with log utility. Note that the Walrasian planner, on the other hand, would choose the mid-point to achieve $P_R = P_C$ to equalize the per consumer welfare between Remote and City.

Proposition 2: The optimal $s^*(r)$ for any given r should be selected from the subset $A(r) \equiv \{r, 1\}$. Downstream firms either locate with the upstream sector (r) or at the large market (i.e., City), but never closer to the smaller market (i.e., Remote) than the upstream sector.

Specifically, the optimization problem is

$$s^*(r) = \arg\max_s \pi_i(s, r)$$

Denote $\pi_s \equiv \frac{\partial \pi_i}{\partial s} = \frac{\partial \pi_i}{\partial p_{iR}}\frac{\partial p_{iR}}{\partial s} + \frac{\partial \pi_i}{\partial p_{iC}}\frac{\partial p_{iC}}{\partial s}$. If $s \geq r$, substituting $P_R^{1-\sigma}$, $P_R^{1-\sigma}$, N into the profit function and noting that $p_{iR}^{1-\sigma} p_R^{\sigma-1} = p_{iC}^{1-\sigma} p_{RC}^{\sigma-1} = N^{-1}$, we can express the marginal profit with respect to s as

Equation 4

$$\pi_s = (1-\sigma)\gamma N^{-1} \frac{(L_R + L_C)\tau_m}{\frac{\tau_e}{2}(2r^2 - 2r + 1) + \tau_m(s - r)} + (1-\sigma)\gamma N^{-1}\left[\frac{L_R \tau_X}{1 + s\tau_X} - \frac{L_C \tau_X}{1 + (1-s)\tau_X}\right]$$

where $\gamma \equiv \frac{\mu}{\sigma}$ is a constant. Suppose all downstream firms are at $s = r$, a downstream CES firm making a unilateral and small left location shift (toward Remote denoted with a minus sign) will see a change in profit as

$$\Delta\pi_i(s = r-) \propto -\tau_m (L_R + L_C) + \tau_X L_R - \tau_X L_C$$

The first term on the right-hand side is the reduction in profits caused by the increase in input cost due to the separation from upstream.[17] The second and third terms are the combined effects of market access, which is negative (because $L_C > L_R$). $\Delta\pi_i(s = r-)$ is thus strictly

[16] Dixit and Stiglitz (1977) show that the market equilibrium would have a lower number of varieties compared to the optimum. Nevertheless, the market number of varieties is a 'constrained optimum' in the absence of lump sum transfers to subsidize consumption. The utilitarian planner optimum in this paper is also a constrained one, in the absence of such lump sum transfers.

[17] Note that $\pi_s \approx \frac{\Delta\pi_i}{\Delta s}$ when Δs is sufficiently small. For the leftward (rightward) shift, Δs is negative (positive) in terms of market access between L_C and L_R. However, for input cost considerations, any shift away from r in either direction implies that Δs is positive.

negative—no downstream firm will move leftwards of r and locate closer to Remote than upstream firms.

A downstream firm making a unilateral small location shift rightwards (toward City, denoted with a plus sign) will see a change in profit as

$$\Delta\pi_i(s = r+) \propto -\tau_m(L_R + L_C) - \tau_X L_R + \tau_X L_C$$

This change can be positive or negative, depending on whether the net gain in market access ($-\tau_X L_R + \tau_X L_C$) outweighs the increase in input costs caused by separation from upstream.

Suppose all firms are at $s = 1$, a firm making a unilateral slight leftward location shift (towards Remote) will see a change in profit as

$$\Delta\pi_i(s = 1-) \propto \tau_m\,(L_R + L_C) + \tau_X L_R - \tau_X L_C$$

Because downstream firms are at the corner, any leftward move must bring them closer to upstream and the first term on the right-hand side represents the savings in input cost. In total, the right-hand side terms can be positive or negative, depending on whether the reduction in input cost outweighs reduced market access.

However, note that if $\Delta\pi_i(s = 1-) > 0$, then $\Delta\pi_i(s = r+) < 0$; these are complementary conditions. If there is incentive to move left of City, then there is incentive to stick to r exactly. The converse is also true, if $\Delta\pi_i(s = 1-) < 0$, then $\Delta\pi_i(s = r+) > 0$. If there is incentive to move right of r, there is incentive to stick to City (corner solution). Downstream firms will either locate at City ($s = 1$) or with the upstream sector at r. (Q.E.D).

Proposition 3: The best response of $s^*(r)$ is increasing in τ_X.

Note that $\Delta\pi_i(s = r+)$ is increasing in τ_X, a higher τ_X strengthens incentives to locate towards City. Note that as $\Delta\pi_i(s = 1-)$ is decreasing in τ_X, a higher τ_X reduces the incentives to move out of City. The increase in transportation cost for the final good always creates stronger incentives for downstream firms to locate towards City.[18] (Q.E.D).

Proposition 4: The optimal $r^*(s)$ for any given s should be selected from the subset B(s) $\equiv$ {r_L, r_H, s}. The upstream sector either locates with the downstream firms at s, or at the left boundary $r_L \equiv \frac{1}{2} - \frac{1}{2}\frac{\tau_m}{\tau_e}$, or at the right boundary $r_H = \frac{1}{2} + \frac{1}{2}\frac{\tau_m}{\tau_e}$.

To recap, the cost function of the upstream sector is

$$m = \frac{\tau_e}{2}(2r^2 - 2r + 1) + \tau_m|s - r|$$

The FOC conditioning on $r \leq s$ is

$$2r\tau_e - \tau_e - \tau_m = 0$$

which gives $r = r_H$. Similarly, the FOC conditioning on $r \geq s$ is

[18] In Appendix A.3, we show that this proposition holds also when there is no vertical separation (locationally speaking) between upstream and downstream sectors.

$$2r\tau_e - \tau_e + \tau_m = 0$$

which gives $r = r_L$.

Therefore, we separately discuss the optimal location choice $r^*(s)$ of the upstream sector when downstream firms locate at: (i) $s \leq r_L$; (ii) $r_L < s < r_H$; and (iii) $s \geq r_H$. Specifically:

(a) If $s \leq r_L$, the cost function m decreases when $r \in [0, s]$, decreases when $r \in [s, r_L]$, and increases when $r \in [r_L, 1]$. Hence, the optimal location that minimizes the cost m is obtained at $r = r_L$ (which is an unstable equilibrium as will be seen shortly).
(b) If $r_L < s < r_H$, the cost function m decreases when $r \in [0, s]$, and increases when $r \in [s, 1]$. Hence, the optimal location that minimizes the cost m is obtained at $r = s$.
(c) If $s \geq r_H$, the cost function m decreases when $r \in [0, r_H]$, increases when $r \in [r_H, s]$, and increases when $r \in [s, 1]$. Hence, the optimal location that minimizes the cost m is obtained at $r = r_H$. (Q.E.D).

Proposition 5: The optimal (s^*, r^*) is a Nash equilibrium between upstream and downstream such that (a) if $s^* > r^*$, the optimal (s^*, r^*) should be $(s^* = 1, r^* = r_H)$ or (b) if $s^* = r^*$, the optimal solution should be selected from the range of $[0.5, r_H]$, and any solution smaller than 0.5 is unstable.

The first part (a) is based on Proposition 3; the best location response of downstream firms is to operate either with upstream firms or at the large market. Upstream firms cannot exceed r_H because input costs will increase too much otherwise. Hence, if $s^* > r_H$, the only solution is $(s^* = 1, r^* = r_H)$. Note that it is possible that $r_H > 1$ given some parameters, in which case both upstream and downstream locate together in City $(s^* = 1, r^* = 1)$.

For any $s^* = r^* < 0.5$ (including the left boundary r_L), the equilibrium is dominated by another equilibrium position $(1 - s^*) = (1 - r^*) > 0.5$. Consider that $s^* = r^* = 0.4$, this equilibrium location is dominated by $s^* = r^* = 0.6$, because while they both share the same input costs (being same distance from the center) the latter is closer to City. Hence, any (and all) downstream CES firms have an incentive to shift to the latter location. The upstream sector, being entirely competitive and having constant returns to scale can operate at either location indifferently. (Q.E.D).

Any equilibrium $s^* = r^* < 0.5$ (and hence including r_L) is unstable, since all downstream firms have an incentive to move closer to City. As a corollary, this result also shows that it would be difficult to have market-determined location of value chain activities (both the upstream and/or downstream sectors) closer to remote regions within an economy. The geographic center is the closest point that sectors would locate towards Remote.

However, note that this finding is not necessarily a pessimistic one—if downstream firms locate closer towards the geographic center, the welfare for Remote would also improve (and the economy moves closer to Walrasian welfare maximization as discussed).

Proposition 6: For the Nash equilibrium $s^* = r^*$, the optimal solution is given by the solution to $s^* = \arg\max_s \pi_i(s, s)$ over the relevant range $[0.5, r_H]$.

Given that both upstream and downstream sectors co-locate in this equilibrium, the profit function is reduced to

Equation 5

$$\pi_i = \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2s^2-2s+1)\right].[1+s.\tau_X]\right]^{1-\sigma}\frac{\mu L_R}{P_R^{1-\sigma}} + \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2s^2-2s+1)\right].[1+(1-s)\tau_X]\right]^{1-\sigma}\frac{\mu L_C}{P_C^{1-\sigma}} - F$$

This equation is relatively more straightforward to solve, with the FOC with respect to s producing an implicit solution in $s \in (0.5, r_H)$.[19] Note that s^* is increasing in τ_X. Again, the reduction in transport costs leads to a location away from the large market and more toward the geographic center. (Q.E.D).

3.3 Model Implications

The Nash equilibrium reflects two tensions—the upstream sector balancing raw material access cost versus intermediate transport cost, and the downstream firms balancing access to two different markets and input cost.
At high transport cost τ_X, the downstream sector locates at City. The upstream sector also locates at City at high τ_m relative to τ_e as it becomes costly to transport intermediates (r_H can be 1 in such a scenario). This is a co-location equilibrium. In this co-location at the City scenario, it is still cheaper to transport raw materials to City (despite the high cost).

If raw material access cost τ_e is high, the configuration would be one of separation where the downstream sector continues to locate at City while the upstream sector moves closer to r_H along the interval and hence closer to the geographic center. We argue that these equilibria likely characterize the location patterns for Indonesia, where upstream and downstream sectors locate near Jakarta, or where upstream sectors locate in some outlying provinces (such as Kalimantan) while the downstream sectors are near Jakarta.

At lower transport cost τ_X, the downstream sector co-locates with upstream at $0.5 < r_H < 1$. We argue that this characterizes China's equilibrium locations, where interior provinces see moderate shares of both upstream and downstream sectors. Finally, as it should be clear from Propositions 2 and 3, the larger the relative size of L_C, the more efficient transport of industrial goods (i.e., lower transport costs) must be to incentivize firms to move away from City.

The Nash equilibrium ($s^* = r^* = \frac{1}{2}$) is feasible only when transport costs in the final sector τ_X approach zero. Note that this solution also maximizes both social welfare functions—utilitarian and Walrasian, another message of this paper.

While this paper focuses exclusively on locations within the economy, there are also potential interactions with global trade and value chain locations. In our model, City is attractive because of its market size. Yet it is also likely that City is the gateway for exports (for example, coastal cities in China and Jakarta for Indonesia). This strengthens the downstream sectors'

[19] We provide the numerical and fuller analytical solution to this case in Appendix A.3. Over the relevant range, this solution also coincides with the scenario where there is full vertical integration between upstream and downstream. As we will explain in Appendix A.3, raw material access costs τ_e will no longer affect location choices under the scenario of where both upstream and downstream sectors must operate at the same location.

incentives to locate toward City. Conversely, should export potential weaken, City will become less attractive.

Our analysis also highlights a darker implication, that the easy access to raw materials (low τ_e) could in fact result in both upstream and downstream production locating at the large market. One can think of asset-specific infrastructure that directly improves raw material access but otherwise does not improve overall transportation (mine mouth or mineral rail line, gas pipelines, commodity ports, etc.). By providing easy access to raw materials, certain infrastructure can shift development toward the large market (City), resulting in an inequality of outcomes. On an international spatial scale, this also relates to "commodity for manufactures" concerns—which refers to extracting resources for exports while increasing dependency on manufactured imports.

Having described the equilibrium properties and implications of the baseline model, Appendix A.4 provides the model extension where the population is assumed to be distributed along the entire interval. Qualitatively, the equilibrium implications hold, except that equilibrium conditions in the extended model depend on the distribution of population (as opposed to just the population at City). Appendix A.5 provides visualizations of the model setup and the various Nash equilibria.

4. Empirical Evidence

We complete the analysis by revisiting the regression on economic centrality and locations of sectors. Here, we make a departure from the earlier province-level aggregated regression (which we showed was wrong-signed for China). Having seen how China's and Indonesia's provinces differ in their shares of upstream and downstream sectors, we instead provide a regression at a more granular scale, at the province-sector level.

Equation 6

$$\ln X_{ni} = \beta_0 + \beta_1 \ln DistCity_n + \beta_2 (\ln DistCity_n \times U_i) + \gamma_1 \ln POP_n + \gamma_2 \ln AvgDist_n + \xi_i + \epsilon_{ni}$$

where X_{ni} is the provincial sales of sector i in province n, $DistCity_n$ is the direct distance from province n to the large city, and U_i is the upstreamness measure of sector i (sales include inter-province, within-province intermediate sales and exports, but not within-province final consumption sales—in order to focus on value chain sales).

For China, $DistCity_n$ is the minimum distance between the province to any of the large conurbations (Beijing, Shanghai or Guangzhou). For Indonesia, correspondingly, $DistCity_n$ is the distance to Jakarta. In certain specifications, $DistCity_n$ is replaced by travel times to the city instead. In a specification, distance is used as the instrumental variable for travel times.[20]

Our regression specification incorporates several province-level control variables such as population (POP_n) and the average bilateral distance to all other provinces ($AvgDist_n$). Population accounts for scale effects. At a minimum, this controls for the effects of market size

[20] In our model, the various transport costs and raw material access costs are taken as exogenous, resulting in the various equilibria. Transport costs can potentially be endogenous to the agglomeration of economic activities. Nevertheless, the instrumental variable regressions show that the coefficients for travel times are similar to OLS ones.

and hence normalizes sales across provinces. As will be explained later, this addition is also important in the context of home market effects common in the NEG literature. Similarly, controlling for the average provincial distance to other markets, which measures the geographic location of each province relative to others, helps distinguish geographic from economic centrality. Finally, all regressions also include sector fixed effects (ξ_i) and heteroskedasticity robust standard errors. The descriptive statistics for all the variables used are provided in Table 2 with the regression results in Table 3.

Table 2: Variables and Summary Statistics

Variable	Description	Unit	Obs	Mean	Std. Dev.	Min	Max
Part (a): China, 2015							
X_{ni}	Province sales by sectors[(1)]	Billion USD	837	16.08	32.95	0.00	414.62
POP_n	Population	Million	31	44.55	29.15	3.30	116.78
GDP_n	GDP	Billion USD	31	356.22	291.87	16.60	1189.74
$DistCity_n$	Distance to large city[(2)]	Kilometer	31	781.86	654.51	34.05	2648.23
$TimesCity_n$	Travel times to large city[(2)]	Hour	31	10.19	8.65	1.17	44.33
$AvgDist_n$	Average distance to all other provinces	Kilometer	31	1424.70	385.18	1022.28	2581.61
$AvgTimes_n$	Average travel times to all other provinces	Hour	31	18.02	5.93	12.78	40.97
U_i	Sectoral upstreamness	-	27	3.77	1.04	1.12	5.41
Part (b): Indonesia, 2016							
X_{ni}	Province sales by sectors[(1)]	Billion USD	816	0.43	1.23	0.00	12.47
POP_n	Population	Million	34	5.14	7.47	0.44	32.25
GDP_n	GDP	Billion USD	34	21.01	29.85	1.62	115.79
$DistCity_n$	Distance to large city[(2)]	Kilometer	34	1200.74	834.15	9.75	3534.20
$TimesCity_n$	Travel times to large city[(2)]	Hour	34	45.93	43.55	0.51	189.58
$AvgDist_n$	Average distance to all other provinces	Kilometer	34	1425.22	379.89	1065.62	2821.51
$AvgTimes_n$	Average travel times to all other provinces	Hour	34	65.88	25.05	44.76	166.29
U_i	Sectoral upstreamness	-	24	2.34	0.86	1.10	4.20

Notes: (1) Province sales by sectors are defined as sales of intermediates within a province and to other provinces plus exports. The average exchange rate for 2015 was CNY1 = USD0.1592. The average exchange rate for 2016 was IDR1 = USD0.00007519. (2) For China, $DistCity_n$ ($TimesCity_n$) is the minimum distance (travel time) between the province to one of the large conurbations (Beijing, Shanghai or Guangzhou). For Indonesia, $DistCity_n$ ($TimesCity_n$) is the distance (travel times) to Jakarta.

Table 3: Regressions of Provincial Sales

Dep var: ln X_{ni}	Part (a): China						Part (b): Indonesia					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ln $DistCity_n$	-0.155	-1.289***			0.668		-0.277	-0.734**			1.479	
	(0.189)	(0.379)			(1.015)		(0.194)	(0.286)			(1.017)	
ln $DistCity_n \times U_i$		0.301***			-0.042			0.195			-0.662	
		(0.105)			(0.306)			(0.135)			(0.576)	
ln $TimesCity_n$			-0.311	-1.865***	-2.683*	-1.750***			-0.279	-0.835**	-1.954**	-0.686**
			(0.255)	(0.373)	(1.330)	(0.478)			(0.176)	(0.310)	(0.883)	(0.253)
ln $TimesCity_n \times U_i$				0.412***	0.464	0.406***				0.238	0.742	0.169
				(0.122)	(0.393)	(0.121)				(0.147)	(0.484)	(0.125)
ln POP_n	1.803***	1.803***	1.783***	1.783***	1.761***	1.793***	1.577***	1.577***	1.484***	1.484***	1.500***	1.474***
	(0.260)	(0.260)	(0.253)	(0.253)	(0.241)	(0.127)	(0.228)	(0.229)	(0.266)	(0.266)	(0.308)	(0.261)
ln $AvgDist_n$	-2.262***	-2.262***	-2.081***	-2.081***	-2.071***	-2.218***	-0.641	-0.641	-0.575	-0.575	-0.573	-0.551
	(0.726)	(0.726)	(0.716)	(0.717)	(0.713)	(0.503)	(1.082)	(1.082)	(1.035)	(1.035)	(1.044)	(1.072)
Sector Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Estimation	OLS	OLS	OLS	OLS	OLS	IV	OLS	OLS	OLS	OLS	OLS	IV
Observations	837	837	837	837	837	837	816	816	816	816	816	816
R-squared	0.488	0.495	0.490	0.498	0.500	0.351	0.486	0.487	0.486	0.489	0.491	0.209

Notes: Robust standard errors in parentheses are clustered at the province level. $*** \ p < 0.01$, $** \ p < 0.05$, $* \ p < 0.1$.

Across all specifications, and for both China and Indonesia, population coefficients are larger than one, showing that larger provinces have more than proportionate sales, and thus hinting at increasing economies of scale or home market effects typical in the NEG literature. In other words, larger provinces host proportionately more value chain activities. This effect is also larger in China as compared to Indonesia.

Compared to regressions of aggregated upstreamness measures against centrality (whether at province or country level), we argue that the regressions here provide a more discriminating test of the downstream-location hypothesis. We pay specific focus on the interaction term $(ln\,DistCity_n \times U_i)$.

One can first see the results in column (1), $\ln DistCity_n$ for China is negative but insignificant. Without the upstream measure, the distance from City variable has little explanatory power. When the interaction term $(\ln DistCity_n \times U_i)$ is included in column (2), the distance from city variable becomes negative and significant and the interaction term (positive coefficient) highlights the fact that upstream sectors are less affected.

We see a different effect for Indonesia. While the inclusion of the upstreamness measure in column (8) results in a more negative coefficient compared to column (7) as with China, the key difference is that the interaction term $(ln\,DistCity_n \times U_i)$ is insignificant for Indonesia. In other words, upstream sectors can be as affected by distance to the large market as downstream ones.

Taken together, the location pattern in China is consistent with model predictions under moderate trade costs. While distances to large markets affect sectors, and more so the downstream ones, the distribution of upstream and downstream sectors is nonetheless more geographically even across provinces. Even inland provinces have some upstream and downstream sector shares (Figure 1). Both upstream and downstream sectors distribute their sales to various locations, resulting in average sales distances being more similar to average geographic distances (Figure 2).

The location pattern in Indonesia is consistent with high trade costs. Resource-abundant provinces either host high shares of upstream sectors (e.g., Kalimantan) or low (e.g., Papua, Jambi). Our model suggests that the nature of the equilibria will depend on raw material access cost. Hence, the average sales distances of upstream sectors are more heterogenous depending on equilibria. On the other hand, Java provinces have high shares of upstream sectors, and downstream sectors clearly concentrate in Java. As a result, the average sales distance is low for downstream sectors since production mostly happens close to market.

At this stage, it is useful to provide some reconciliation with the provincial aggregate regressions in Table 1. Downstream sectors have larger sales, accounting for around 55% to 60% of all sales in both China and Indonesia. Hence, economically central locations may host significant shares of upstream activities and still appear to have a downstream structure on average (because downstream sales are larger). This is a generalization of Indonesia's location configuration and is again consistent with the model outcome of high transport costs. For China, less economically central locations may host some downstream activities, do not appear particularly upstream on average, but still have large shares of upstream activities consistent with the model outcome of more moderate trade costs. The more granular province-sector regressions in this paper bring out these important nuances.

When distance from City is replaced by travel times, the results are largely the same for both China and Indonesia but become sharper with improved R-square statistics. For China, both the coefficients for travel times and upstreamness interaction term are larger compared to the specification using distance. For Indonesia, the relevant coefficients are also larger, but the interaction term notably continues to be insignificant.

When both distance and travel times are included—column 5 for China and column 11 for Indonesia—travel times coefficients are negative and significant, notwithstanding the high collinearity between distance and travel times (distance coefficients become insignificant). In other words, travel times perform better than distance as an explanatory variable for industrial locations. This underscores the importance of infrastructure, as opposed to distance, for industrial locations.
In columns (6) and (12) for China and Indonesia, respectively, distance is used as an instrumental variable for travel times to address potential concerns that travel times can be endogenous to economic activities (because policy makers may improve transport systems for provinces with more economic potential). Comparing columns (4) and (6) for China, and columns (10) and (12) for Indonesia, travel time coefficients are slightly higher under OLS estimation compared to the IV. This finding suggests some evidence of endogeneity, thus overstating the effects of transport infrastructure. Otherwise, the IV results continue to be largely in line with OLS ones, and the interaction term is significant for China but not Indonesia.

5. Conclusion

We provide evidence on how provinces in China and Indonesia differ in their shares of upstream and downstream sectors, pointing to differences in upstream-downstream location configurations. We provide a stylized model that incorporates the IO characteristics of an upstream extraction processing and downstream consumer differentiated sector in a value chain together with a spatial representation of locations. The model highlights the tension between economic centrality (which the downstream sector prefers) and geographic centrality (which has the benefit of lower-cost input for the upstream sector). While the two-sector two-location model is stylized in nature, it brings out the potential for multiple equilibria depending on various types of trade costs. The upstream and downstream sectors co-locate or operate at separate locations, depending on trade costs.

With more granular sectoral data and regressions, our conclusions are more nuanced. The more granular province-sector regressions affirm the downstream-location hypothesis for China. The regressions with granular data do not reject the downstream-location hypothesis for Indonesia per se, but also do not cleanly establish this either because even upstream sectors can be fairly attached to downstream locations (no separation). This is consistent with the model's equilibrium with high trade costs.

While an understanding of international economic geography and upstream-downstream export structure is undoubtedly important, we argue that it is equally important (if not more) to understand where value chains locate within economies. For large developing economies, where value chains locate given transport costs is critically important for the development of regions. To the best of our knowledge, this research is also the first to provide a within-economy study of the downstream-location hypothesis.

We show that the downstream-location hypothesis is not absolute or automatic. If within-economy transport costs are too high, upstream sectors will locate at economically central locations together with downstream ones. There will be no separation, and development will become highly uneven.

Physical distances are immutable, but transport times can be shortened with improved infrastructure. We emphasize how improved transport times can help shift more activities, first upstream, then downstream, away from large economic centers and more towards the regions. Travel times have higher explanatory power compared to distance and with coefficients in line with expectations. IV estimations for travel times result in slightly smaller coefficients but continue to affirm the direction of results otherwise.

China’s provinces see greater participation in value chain output. Interior provinces host moderate shares of both upstream and downstream activities. Indonesia undoubtedly has a tougher internal geography, and higher transport costs have resulted in a different set of value chain locations. Downstream and even upstream value chains are largely concentrated in Java for Indonesia. With some exceptions, most resource-endowed outlying provinces in Indonesia also see relatively little upstream value chain sales. The research underscores the need to reduce transport costs to facilitate value chain development away from large economic centers. There will also be a need for public policy support for the development of regions outside Java to reduce regional inequality.

Appendices

A.1 China and Indonesia Industries Upstreamness Measures

Following Antràs et al. (2012), the upstreamness measure of each industry i is the weighted average of the number of stages from final demand at which i inputs would enter final use. Specifically,

$$U_i = 1 \cdot \frac{F_i}{Y_i} + 2 \cdot \frac{\sum_{j=1}^{N} d_{ij}\, F_j}{Y_i} + 3 \cdot \frac{\sum_{j=1}^{N} \sum_{k=1}^{N} d_{ik}\, d_{kj} F_j}{Y_i} + \cdots$$

where F_i is the value of that output that goes directly to final use, and $d_{ij}Y_j$ is sector i's total output that is purchased by industry j. U_i can be equivalently written as

$$U_i = 1 + \sum_{j=1}^{N} \frac{d_{ij}Y_j}{Y_i} U_j$$

Let U be the column vector with U_i as its i-th entry, $\mathbf{1}$ be the column vector of ones, and let Δ be the matrix with $d_{ij}Y_j/Y_i$ in entry (i,j)

$$U = \mathbf{1} + \Delta U$$

Solving for U leads to

$$\mathrm{U} = [\mathrm{I} - \Delta]^{-1}\mathbf{1}$$

In the open economy, Δ is adjusted where the (i,j) entry of Δ is precisely the value of commodity i used in j's production $(\mathrm{d_{ij}Y_j})$. The denominator of each entry is replaced by $\mathrm{Y_i} - \mathrm{X_i} + \mathrm{M_i}$, where $\mathrm{X_i}$ and $\mathrm{M_i}$ denote exports and imports of sector i output. We find that the range of upstreamness spans from a minimum of 1.097 to a maximum of 4.202 for Indonesian industries.

Similarly, within the context of China, this measure varies from a minimum of 1.123 to a maximum of 5.409. The manufacturing industries are ranked from the most upstream to the least upstream for China and Indonesia respectively. Notably, tobacco, construction, furniture and transport are among the most downstream industries, with almost all their output going directly to final demand. Conversely, the most upstream industries, including fossil fuel extraction and metal ore mining, play pivotal roles in the raw material processing phase of production.

Note that the upstreamness measure for each sector is computed at a national (rather than provincial) level for China and Indonesia. As mentioned, provinces within a national economy face a more similar regulatory environment (compared to between economies). Firms and human capital also relocate more readily within the economy; and firms do not face tariff barriers when selling across provinces. Therefore, the upstreamness measure of each sector should be similar across provinces.

For provinces, we compute upstreamness measures as the weighted average of the shares in each sector and their respective nationally determined upstreamness (see Appendix A.2). Importantly, having a standard national measure of upstreamness for each sector also allows for comparison across provinces.

Table 4: Upstream Ranking of Sectors in China

Ranking	Sector ID	China IO2015 Sector	Upstream
1	3	Petroleum and Natural Gas Extraction Products	5.409
2	2	Coal Mining and Selection Products	5.250
3	23	Waste and Scrap	5.098
4	4	Metallic Ore Selection Products	5.046
5	24	Metal Products, Machinery, and Equipment Repair Services	4.992
6	25	Production and Supply of Electric Power and Heat	4.707
7	7	Textile Products	4.590
8	12	Chemical Products	4.576
9	11	Petroleum, Coking Products, and Nuclear Fuel Processing Products	4.549
10	10	Paper, Printing, and Educational and Sports Products	4.326
11	20	Communication Equipment, Computers, and Other Electronic Equipment	4.251
12	14	Smelting and Rolling Products of Metals	3.946
13	26	Production and Supply of Gas	3.754
14	5	Non-Metallic Mineral and Other Ore Selection Products	3.642
15	22	Other Manufacturing Products	3.609
16	27	Production and Supply of Water	3.592
17	15	Metal Products	3.368
18	21	Instruments and Meters	3.323
19	9	Wood Processing and Furniture	3.150
20	19	Electrical Machinery and Equipment	3.137
21	6	Food and Tobacco	2.994
22	16	General Equipment	2.976
23	13	Non-Metallic Mineral Products	2.862
24	8	Textile, Clothing, Shoes, Hats, Leather, Down, and Their Products	2.857
25	17	Special Equipment	2.370
26	18	Transportation Equipment	2.326
27	28	Construction	1.123

Table 5: Upstream Ranking of Sectors in Indonesia

Ranking	Sector ID	Indonesian IO2016 Sector	Upstream
1	9	Coal and Lignite Mining	4.202
2	8	Oil and Gas Extraction and Geothermal	3.785
3	10	Metal Ore Mining	3.680
4	29	Gas Supply and Ice Production	3.543
5	28	Electricity	3.265
6	18	Paper and Paper Products Industry, Printing and Recorded Media Reproduction	2.704
7	20	Rubber, Rubber Goods and Plastic Industry	2.574
8	22	Basic Metal Industry	2.529
9	12	Coal and Crude Oil Processing	2.513
10	17	Wood and Wood Products Industry, Bamboo, Rattan, and Similar Woven Product Industry	2.493
11	11	Other Mining and Quarrying	2.480
12	19	Chemical, Pharmaceutical, and Traditional Medicine Industry	2.474
13	21	Non-Metallic Mineral Product Industry	2.291
14	24	Machinery and Equipment Industry Except Electrical	2.062
15	23	Metal Goods, Computer, Electronic and Optical Goods, Electrical Equipment Industry	2.010
16	15	Textile and Garment Industry	1.998
17	16	Leather, Leather Goods and Footwear Industry	1.685
18	13	Food and Beverage Industry	1.658
19	25	Transport Equipment Industry	1.643
20	27	Other Processing Industry, Machinery and Equipment Repair and Installation Services	1.585
21	30	Water Supply, Waste Management, Waste, and Recycling Services	1.508
22	26	Furniture Industry	1.205
23	31	Construction	1.139
24	14	Tobacco Processing Industry	1.097

A.2 China and Indonesia Provincial Upstreamness Measures and Economic Centrality

Formally, the provincial upstreamness measure is obtained as

$$U_n = \sum_i \frac{X_{ni}}{X_n} U_i$$

where X_{ni} is the province sales (sales to other provinces and export) of sector i by province n. Thus, provinces with higher cross-province sales and exports in upstream sectors will be measured as more upstream. In line with the literature, we compute two measures of economic centrality—GDP-weighted and population-weighted—respectively as

$$Centrality_n^{GDP} = \sum_k \frac{GDP_k}{TotalGDP} Distance_{nk}^{-1},$$

$$Centrality_n^{Population} = \sum_k \frac{Population_k}{TotalPopulation} Distance_{nk}^{-1}$$

which captures a province's proximity to other provinces with either a large GDP or large population (or both). Figure 4 and Figure 5 provide a geographic view of provincial upstreamness and economic- or GDP-weighted centrality in China, respectively, while Figure 6 and Figure 7 are for Indonesia. This is followed by the correlations of upstreamness and GDP-weighted centrality (Figure 8).

Figure 4: Upstreamness by Province, China

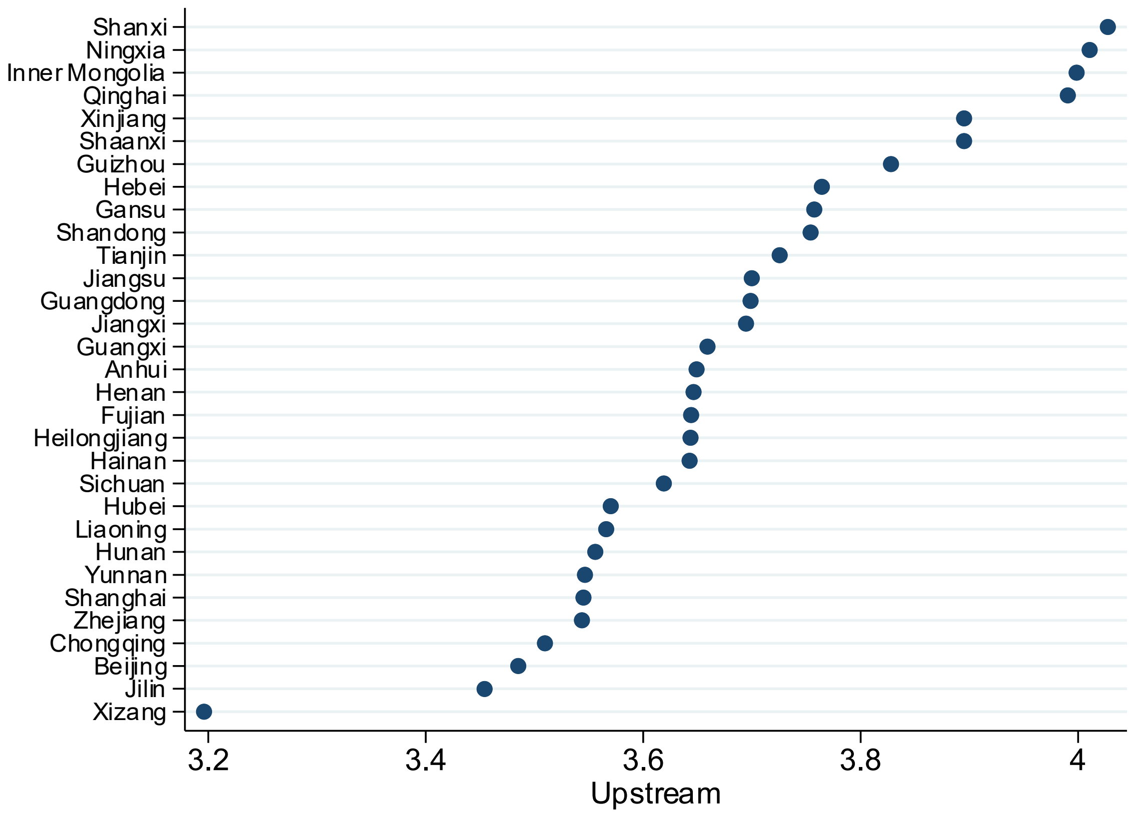


Figure 5: GDP-weighted Centrality by Province, China

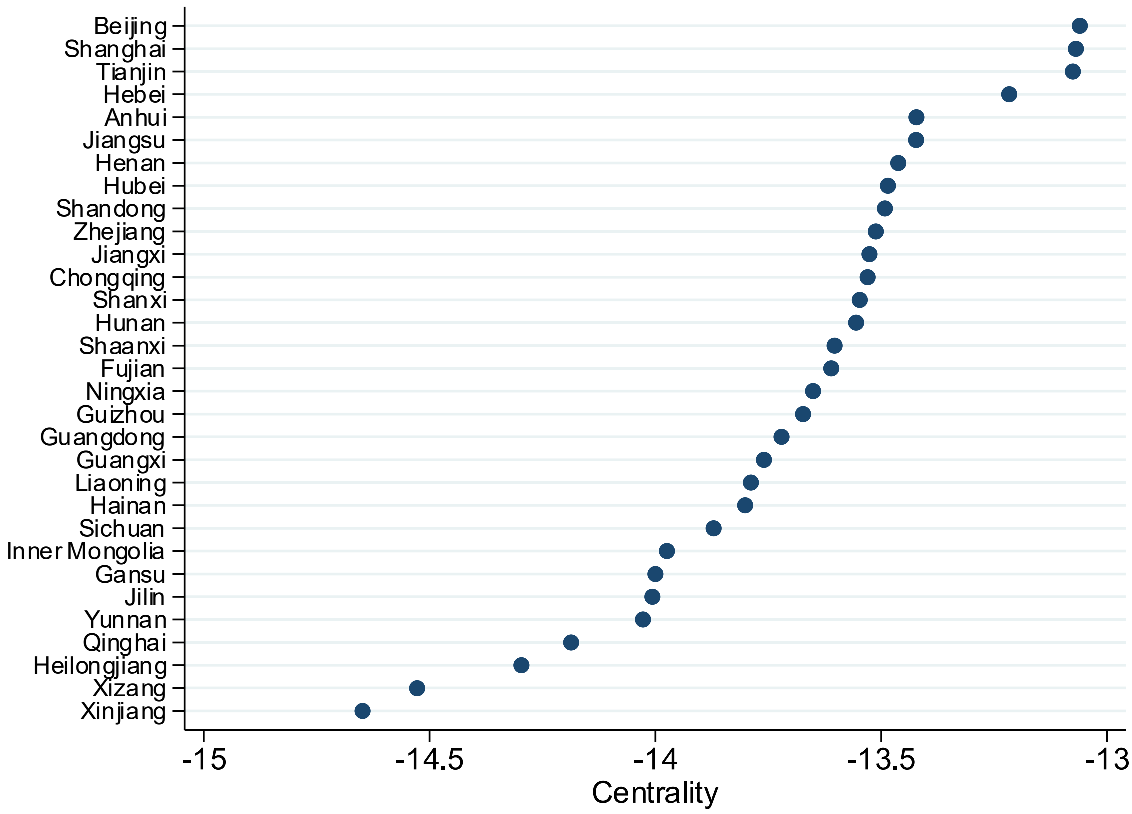


Figure 6: Upstreamness by Province, Indonesia

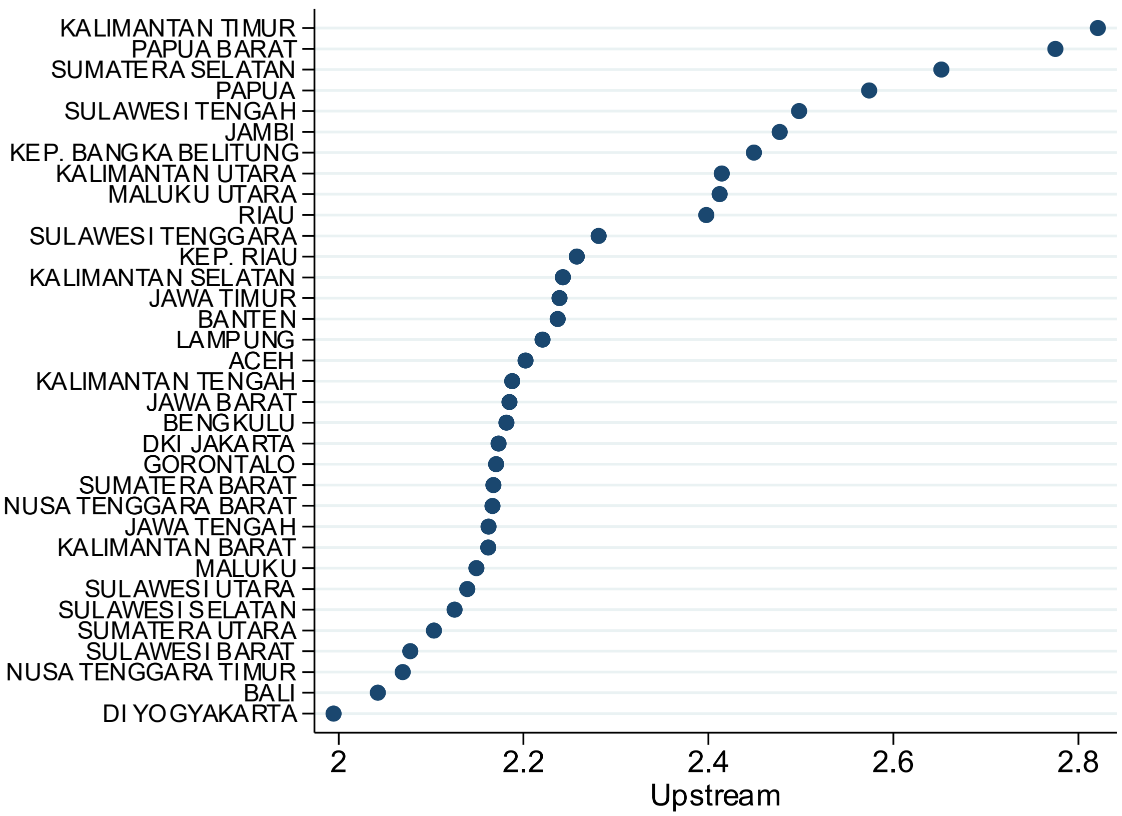


Figure 7: GDP-weighted Centrality by Province, Indonesia

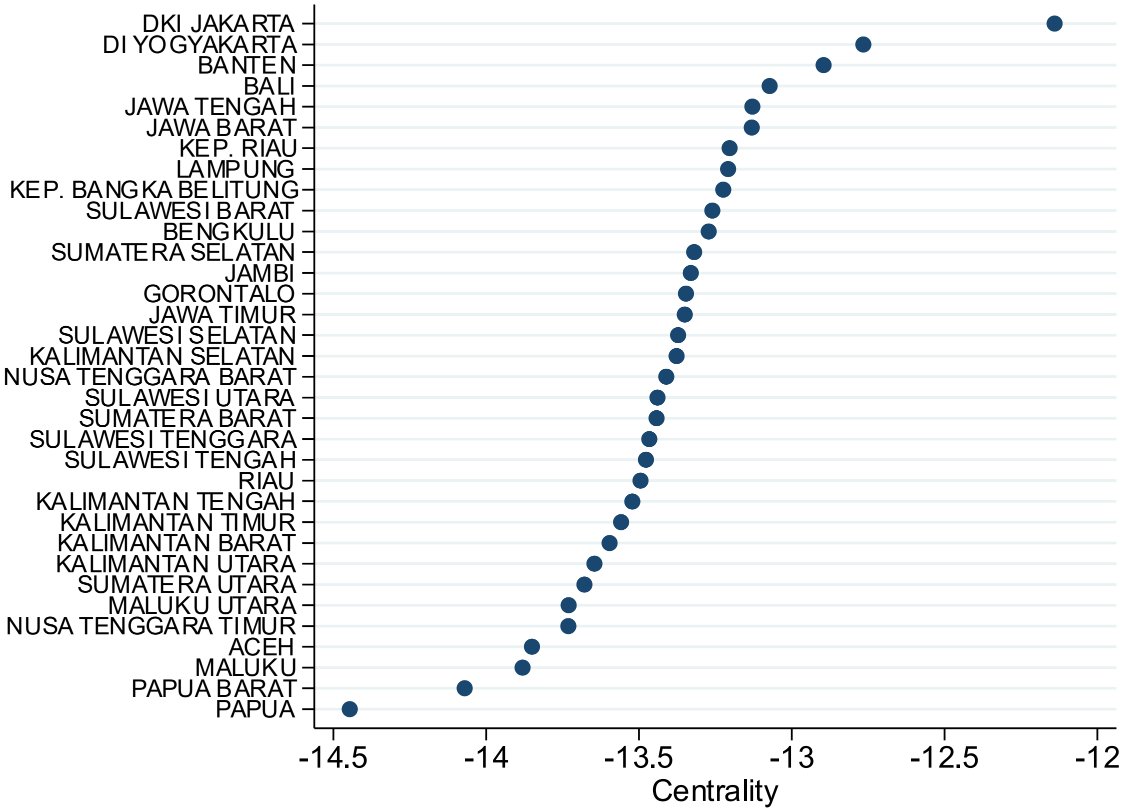
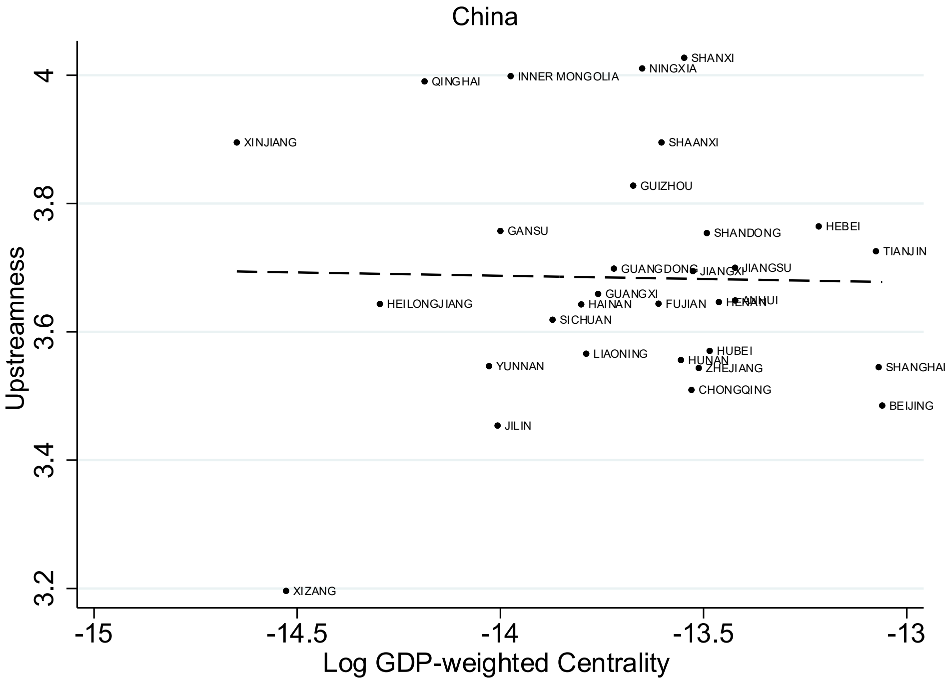
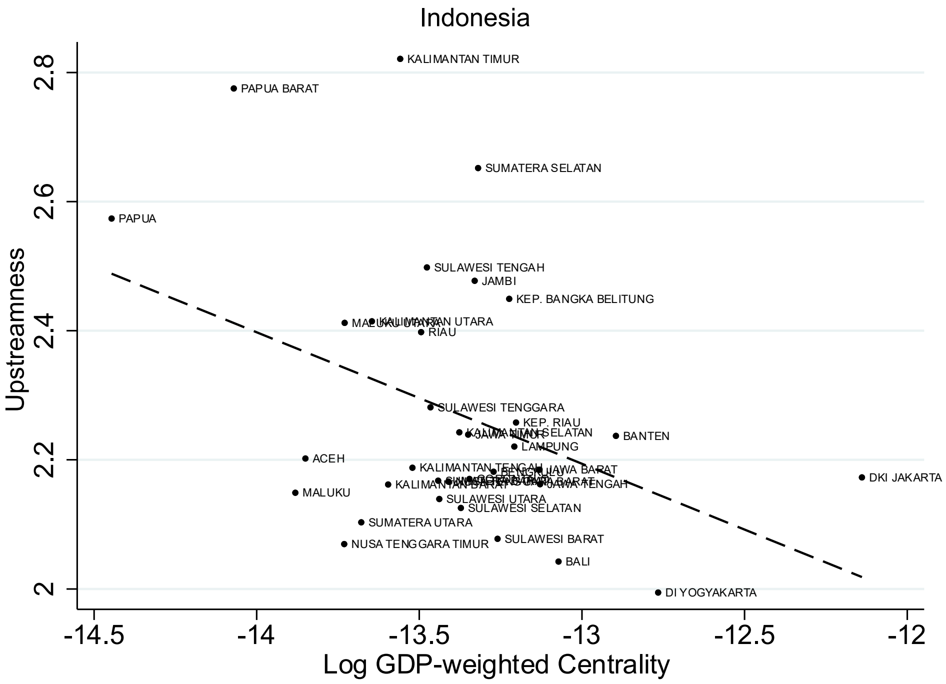


Figure 8: Correlations Between Provincial Upstreamness and GDP-weighted Centrality





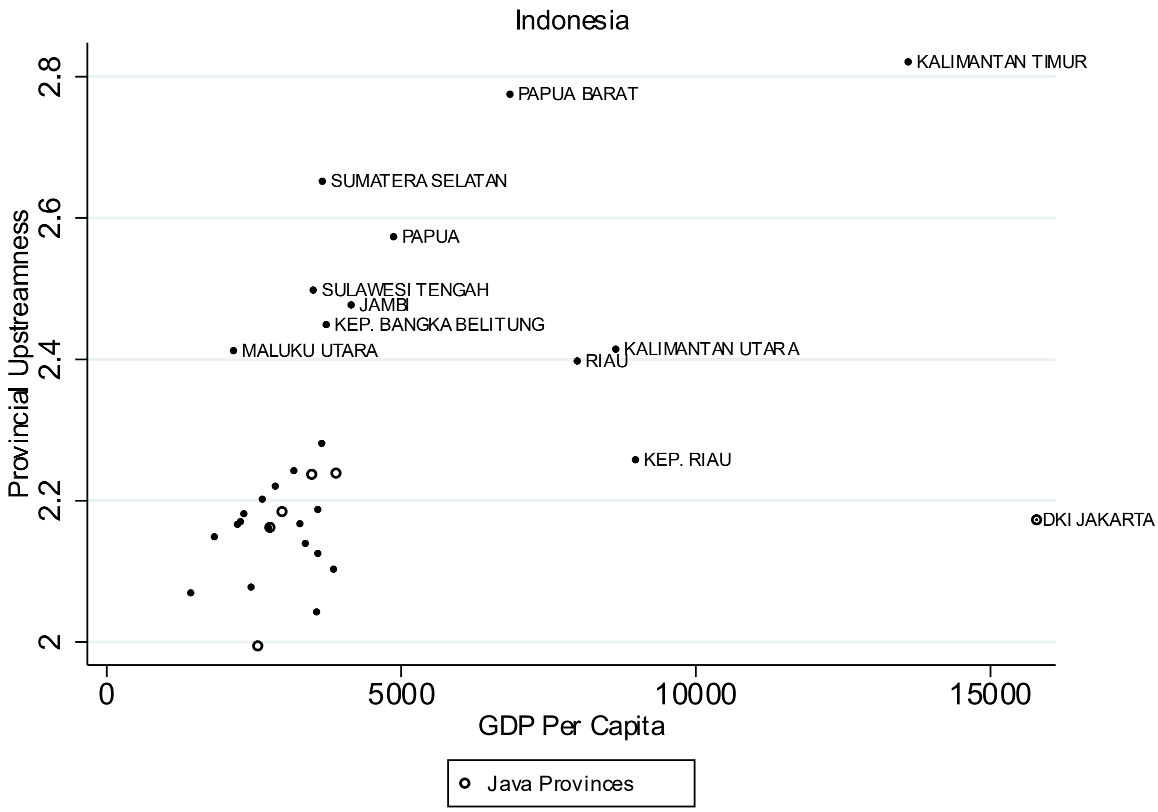
With these measures, we test whether more economically central provinces have a comparative advantage in relatively more downstream sectors by examining the correlation across provinces between provincial economic centrality and upstreamness.

Equation 7

$$\ln U_n = \beta_0 + \beta_1 \ln Centrality_n + \Gamma Z_n + \epsilon_n$$

The results of the regression in Equation 7 is provided in Table 1. As discussed in the main text, upstream provinces in Indonesia register higher per capita GDP (Figure 9).

Figure 9: Upstreamness of Indonesia Provinces and Per Capita GDP



A.3 Special Case – No Vertical Separation

We consider a special case in which there is no vertical separation between the upstream and downstream sectors—i.e., the final good sector has to bring raw materials to its production location. This special case is interesting because first, some sectors will have incentives to organize themselves in such an integrated manner and this special case provides the solution to the equilibrium location in such an instance. Second, this special case also provides the solution to the equilibrium of $(s^* = r^*)$ as described in Proposition 5(b) and Proposition 6, over the relevant solution range $[0.5, r_H]$.

With full vertical integration of the upstream and downstream sectors, the profit function fully incorporates the resource collection costs and is reduced to

$$\pi_i = \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2s^2-2s+1)\right].\left[1+s.\tau_X\right]\right]^{1-\sigma}\frac{\mu L_R}{P_R^{1-\sigma}} + \frac{1}{\sigma}\left[\frac{\sigma}{\sigma-1}\left[\frac{\tau_e}{2}(2s^2-2s+1)\right].\left[1+(1-s)\tau_X\right]\right]^{1-\sigma}\frac{\mu L_C}{P_C^{1-\sigma}} - F$$

as per Equation 5 in the main text. Note that this is the case if $(s = r)$. The first order condition (FOC) is given as

$$\frac{\partial \pi_i}{\partial s} = (1-\sigma)\gamma N^{-1}\frac{(L_R+L_C)(4s-2)}{2s^2-2s+1} + (1-\sigma)\gamma N^{-1}\left[\frac{L_R\tau_x}{1+s\tau_x} - \frac{L_C\tau_x}{1+(1-s)\tau_x}\right]$$

Which can be further rewritten as

$$\frac{\partial \pi_i}{\partial s} = (1-\sigma)\gamma N^{-1}W_R\frac{As^3+Bs^2+Cs+D}{(2s^2-2s+1)(1+s\tau_X)(1+(1-s)\tau_x)}$$

Here, the coefficients A, B, C, D are specified as

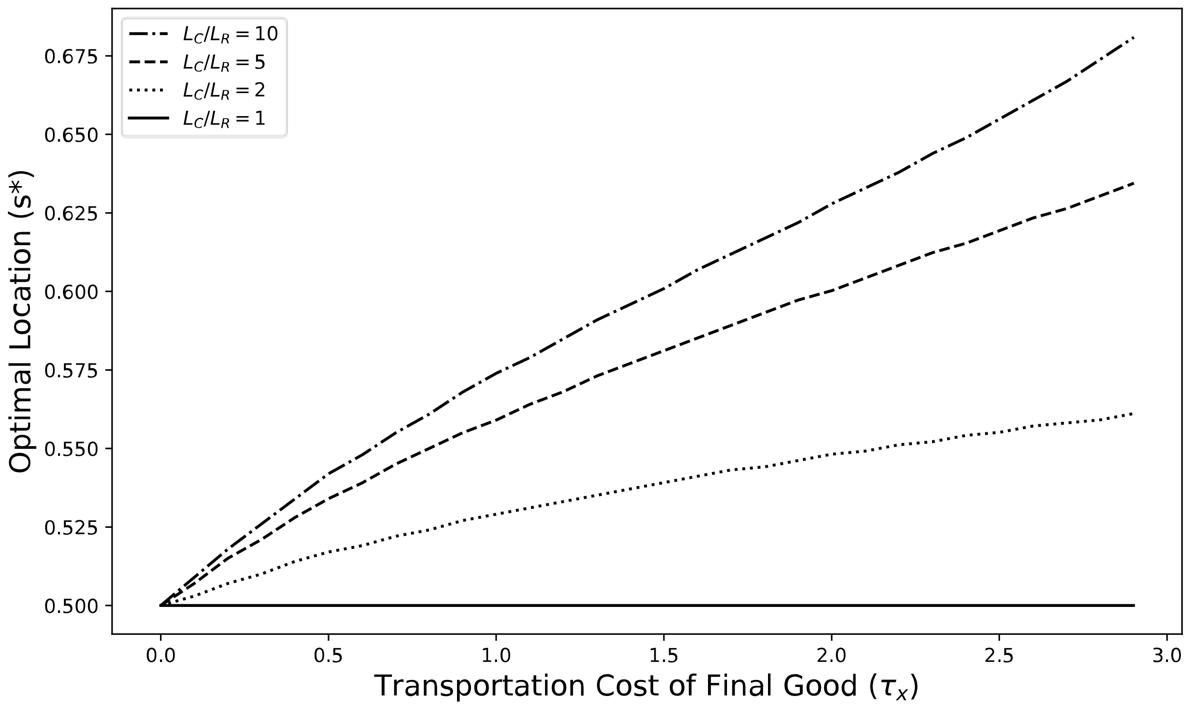
$$A = -\left(6\frac{L_C}{L_R}+6\right)\tau_x^2;\ B = \left(8\frac{L_C}{L_R}+10\right)\tau_x^2 - \left(2\frac{L_C}{L_R}-2\right)\tau_x;$$

$$C = -\left(3\frac{L_C}{L_R}+5\right)\tau_x^2 + \left(6\frac{L_C}{L_R}+2\right)\tau_x + \left(4\frac{L_C}{L_R}+4\right);\ D = \tau_x^2 - \left(3\frac{L_C}{L_R}+1\right)\tau_x - \left(2\frac{L_C}{L_R}+2\right)$$

Hence, the optimal location s is then determined as a function of (i) the ratio of City to Remote size (L_C/L_R), and (ii) the transportation cost of final good (τ_X). As can be seen in Figure 10, there is a "smooth" location solution depending on transport costs of final goods (τ_X), as opposed to the more "discrete" location solutions of the Nash equilibrium presented in the main text. This is not surprising because the downstream cannot be locationally separated and now faces a quadratic input cost function directly, and as a result, it no longer has the binary option of locating with upstream or at City.

Figure 10 provides the numerical solution to the above problem with various ratios of market sizes. As vertical separation is not possible as in this special case, the relevant range for co-location will be $[0.5, 1]$ as per Figure 10.

Figure 10: Location Equilibria with Different Trade Costs and Market Size Ratios



Notes: This figure presents the numerical solutions to s under the assumption of co-location between upstream and downstream sectors (or no vertical separation), with varying values for τ_X and City to Remote market size ratios L_C/L_R, and with other parameters held constant at $\sigma = 5, \tau_e = 1, \tau_m = 0.1$. Note that $r_H = 0.55$ given the parameters. If sectors are allowed to be at separate locations, upstream will locate at $r^* = 0.55$ while downstream will locate entirely at City ($s^* = 1$).

We reconnect this special case to the main model where vertical separation is possible. To be clear, the upstream sector would not locate at values larger than the boundary r_H because the quadratic access cost would outweigh the transport cost of intermediates past this point. If the solution to co-location is larger than r_H, the sectors will locate separately with upstream at r_H and downstream at City as per Proposition 5. It would be better to just incur τ_m to send the assembled intermediate to the downstream location.

Hence, any co-location in the context where vertical separation is permissible only occurs for the relevant range of $[0.5, r_H]$ as per Proposition 5. Over this relevant range, sectors can locate separately but will choose to co-locate nonetheless. Note however that r_H can take the value of 1 depending on the parameters, resulting in co-location at City. Note also that raw material access cost τ_e is not a factor to determine location in this special case because the profit function becomes homogenous of degree one in τ_e. Raw material access cost changes locational choices only when there is the possibility of locationally separating the upstream and downstream sectors.

A.4 Extending the Model to Population Distribution Along the Interval

We extend the baseline model to allow for the presence of fixed populations (and hence consumption) along the interval, instead of only at Remote or City. The assumption of population distributed along the interval also softens the no consumption at production location assumption in the main model. The model will become less restrictive as a result, though there will be a need to rely more on numerical solutions.
Along the interval, there is a fixed population at each point denoted by $L(t)$. We assume that $L(t)$ is monotonically increasing from Remote to City to avoid definition issues (e.g., non-City points having large populations) and to avoid "local" equilibria during optimization. We use a hat to denote the equilibrium variables or labels of this extended model. The optimization problem for the model with population distributed along the interval is given as

$$\hat{\pi}_i^*(r) = \arg\max_s \pi_i(s, r)$$

Where the profit function is given as

Equation 8

$$\hat{\pi}_i = \int_0^1 \hat{\pi}_i(t)dt - F = \int_0^s \hat{\pi}_i(t)dt + \int_s^1 \hat{\pi}_i(t)dt - F$$

Each firm will charge a t location specific $\hat{p}_i(t)$, which is a standard mark up on input cost $m = \frac{\tau_e}{2}(2r^2 - 2r + 1) + \tau_m|s - r|$ and the iceberg cost between production and consumption locations $\hat{\tau}_X(t) = 1 + \tau_X|s - t|$. As a result, there will also be a location-specific CES price $\hat{P}(t)$—the further away the consumption location is away from production, the higher the CES price. From each location t, the profit earned by the firm is

$$\hat{\pi}_i(t) = \frac{1}{\sigma} . \hat{p}_i(t)^{1-\sigma} . \hat{P}(t)^{\sigma-1} . \mu . L(t)$$

Note that $\frac{\partial \hat{\pi}_i(t)}{\partial \hat{p}_i(t)} = \frac{1-\sigma}{\sigma}.\hat{p}_i(t)^{-\sigma}.\hat{P}(t)^{\sigma-1}.\mu.L(x) = \frac{1-\sigma}{\sigma}.N^{-1}.\hat{p}_i(t)^{-1}.\mu.L(t)$. Furthermore, $\pi_s(t) \equiv \frac{\partial \hat{\pi}_i}{\partial s} = \frac{\partial \hat{\pi}_i(t)}{\partial \hat{p}_i(t)}\frac{\partial \hat{p}_i(x)}{\partial s}$ and combined with Leibniz rule gives

Equation 9

$$\pi_s \propto -\mu.\int_0^1 L(t).\frac{\tau_m}{m}dt - \mu.\int_0^s L(t).\left[\frac{\tau_X}{1+\tau_X(s-t)}\right]dt + \mu.\int_s^1 L(t).\left[\frac{\tau_X}{1+\tau_X(t-s)}\right]dt$$

If constrained only to two locations at Remote and City, the above equation collapses exactly to Equation 4. In the extended model, the terms are also similar. The first term on the RHS is the effects on profits of any departure from where the upstream sector operates. Any move away from the upstream location results in an increase in intermediate costs and hence a reduction in profit for final sales across the entire interval.

The second term is the reduction of profits caused by the increase in shipping costs for the interval left of s, with any rightward s move. The third term is the increase in profits caused by the reduction in shipping costs for the interval right of s, with any rightward s move. In Proposition 2, it is clear that the combined effects of the second and third terms are positive with an increase in s, because $L_C > L_R$. In this extended case, it is less clear-cut and will depend on the distribution of population $L(t)$.

Proposition $\widehat{2}$: The optimal solution $\widehat{s^*}(r)$ for any given r should be selected from the subset $\{r, s^*(r)\}$ where $s^*(r)$ corresponds to the location such that $\pi_s(s^*(r)) = 0$ and that $0.5 \leq s^*(r) < 1$.

First, because population is rising monotonically, the optimal solution $s^*(r)$ will be weakly larger than 0.5 as downstream firms always choose to be closer to the thicker side of the population distribution.

Second, note that from Equation 9, $\pi_s(s=1) \propto -\mu.\int_0^1 L(t).\left(\frac{\tau_m}{m} + \frac{\tau_X}{1+\tau_X(s-t)}\right)dt < 0$. This shows that it is never optimal for downstream to be at City even as they locate at the thicker side of the distribution. This is because the population is now distributed across the interval (and not just concentrated at City in the earlier two-location model). Hence, $0.5 \leq s^*(r) < 1$.

If $\pi_s(s = r_H) > 0$, which occurs when $\int_r^1 L(t)dt$ is large enough via inspection of Equation 9, the downstream sector has the incentive to locate right of r_H. There thus exists an optimal $s^*(r) \in (r_H, 1)$ through the solution to the FOC such that $\pi_s(s^*(r)) = 0$ and the maximum profit is achieved. In this case, the upstream sector will locate at r_H while the downstream sector locates at $r_H < s^*(r) < 1$.

If $\pi_s(s = r_H) < 0$ and $\pi_s < 0$ for any $s \in (r_H, 1]$, there is incentive for the downstream sector to seek locations left of r_H. In this case, given that location lies left of r_H, it would be possible for upstream and downstream to locate together to minimize input cost, which is optimal. The optimal solution will be co-location of upstream and downstream such that $0.5 < s^* = r^* < r_H$. (Q.E.D.).

Proposition $\hat{3}$: The optimal solution $\widehat{s^*}(r)$ is increasing in τ_X when the population distribution is skewed towards City.

With the implicit function theorem, $\partial s^*/\partial\tau_X$ is given as

$$\frac{\partial s^*}{\partial\tau_X} = -\frac{\partial\pi_s/\partial s}{\partial\pi_s/\partial\tau_X}$$

The numerator term $\frac{\partial\pi_s}{\partial s} < 0$ and the denominator term $\partial\pi_s/\partial\tau_X$ depend on the relative importance of $\int_0^s L(t).dt$ and $\int_s^1 L(t).dt$. Inspecting s, if $\int_s^1 L(t).dt$ is sufficiently large to dominate, then $\frac{\partial\pi_s}{\partial\tau_X} > 0$, indicating that $\frac{\partial s^*}{\partial\tau_X} > 0$. (Q.E.D.)

Characterization of Equilibrium for Extended Model

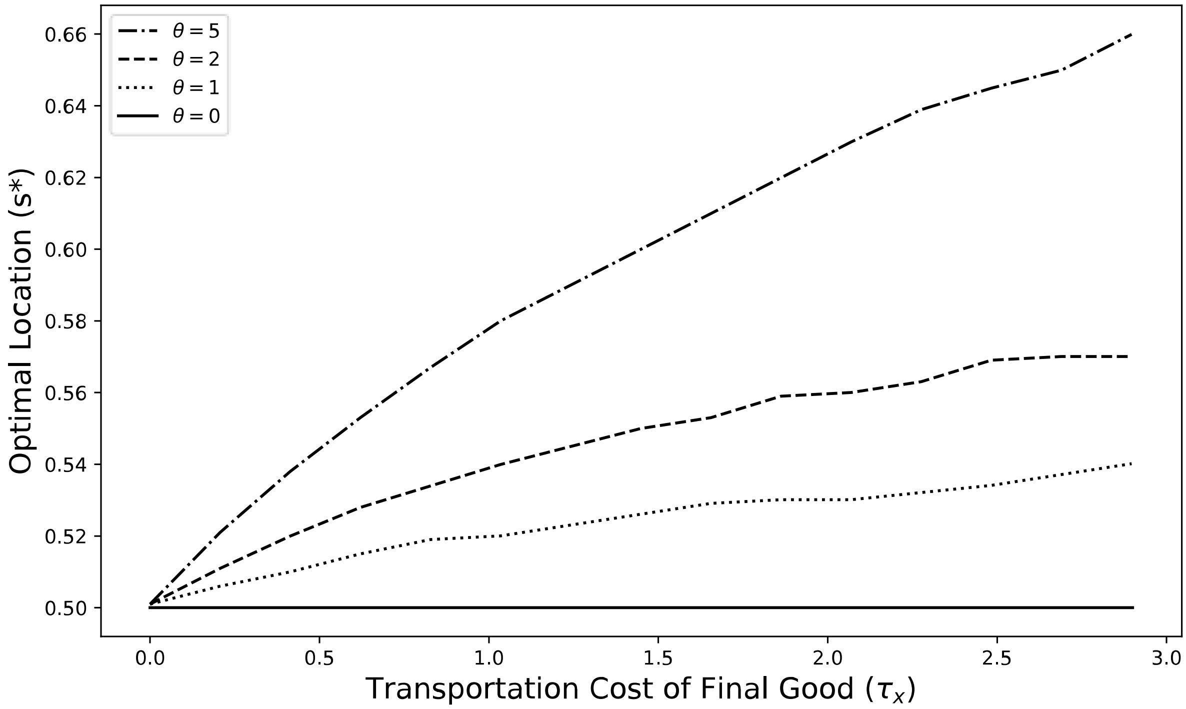
Proposition $\hat{2}$ is a modification of Proposition 2, with the key difference that Proposition $\hat{2}$ rules out locating at City. Proposition $\hat{3}$ is a modification of Proposition 3—it provides the same general result but Proposition $\hat{3}$ relies on the population distribution along the interval to be skewed towards City (as opposed to just dependent on the large population at City itself).
These two modified propositions can be intuitively explained by the fact that, unlike the two-location case, where only the large population at City matters, the downstream sector's location choice in the extended model will depend on the distribution of population along the interval. Proposition 4 is unchanged with the model extension, as the upstream sector will continue to minimize cost.

As a result, the equilibrium results of the extended model will look like the two-location one, with the exception that location at City is ruled out, and that downstream equilibrium location will depend on population distribution. An increase in trade cost τ_X will continue to incentivize downstream towards City if population distribution is sufficiently skewed towards City.

Numerical Example

We provide a numerical solution by assuming that the population along the interval takes on the following function $L(t) = L_R e^{\theta t}$, where θ determines how fast the population is increasing as one moves toward City. The optimal solution $s^*(r)$ is increasing in both τ_X and θ. Similarly, we can show that the direction of $\frac{\partial\pi_s}{\partial\theta}$ is dominated by the first term of π_s, $-\mu.\int_0^1 L(t).\frac{\tau_m}{m}dt$. Therefore, $\frac{\partial\pi_s}{\partial\theta} < 0$, and $s^*(r)$ is increasing in θ. The numerical solutions to various population distribution parameters and final goods trade cost are provided in Figure 11.

Figure 11: Location Equilibria with Different Trade Costs and Population Distribution Parameters



Notes: This figure presents the numerical solutions to s for the extended model where population is distributed along the entire interval, with varying values for τ_X and population parameter θ, and with other parameters held constant at $\sigma = 5, \tau_e = 1, \tau_m = 0.1$. Note that $r_H = 0.55$ given the parameters, this is the rightmost point of upstream industry. Hence, any solutions of $s > r_H = 0.55$ will imply separate locations between upstream and downstream.

A.5 Additional Materials–equilibrium Sketches

We provide a sketch of the model (Figure 12) and a few equilibria sketches (Figure 13) to support readers. The various equilibria sketches are not meant to highlight precise solutions (which are provided in the paper), but to facilitate readers' understanding.

Figure 12: Sketch Model setup

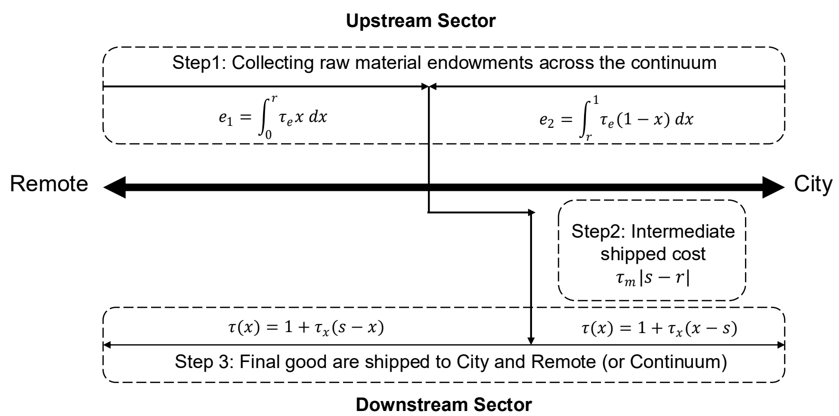
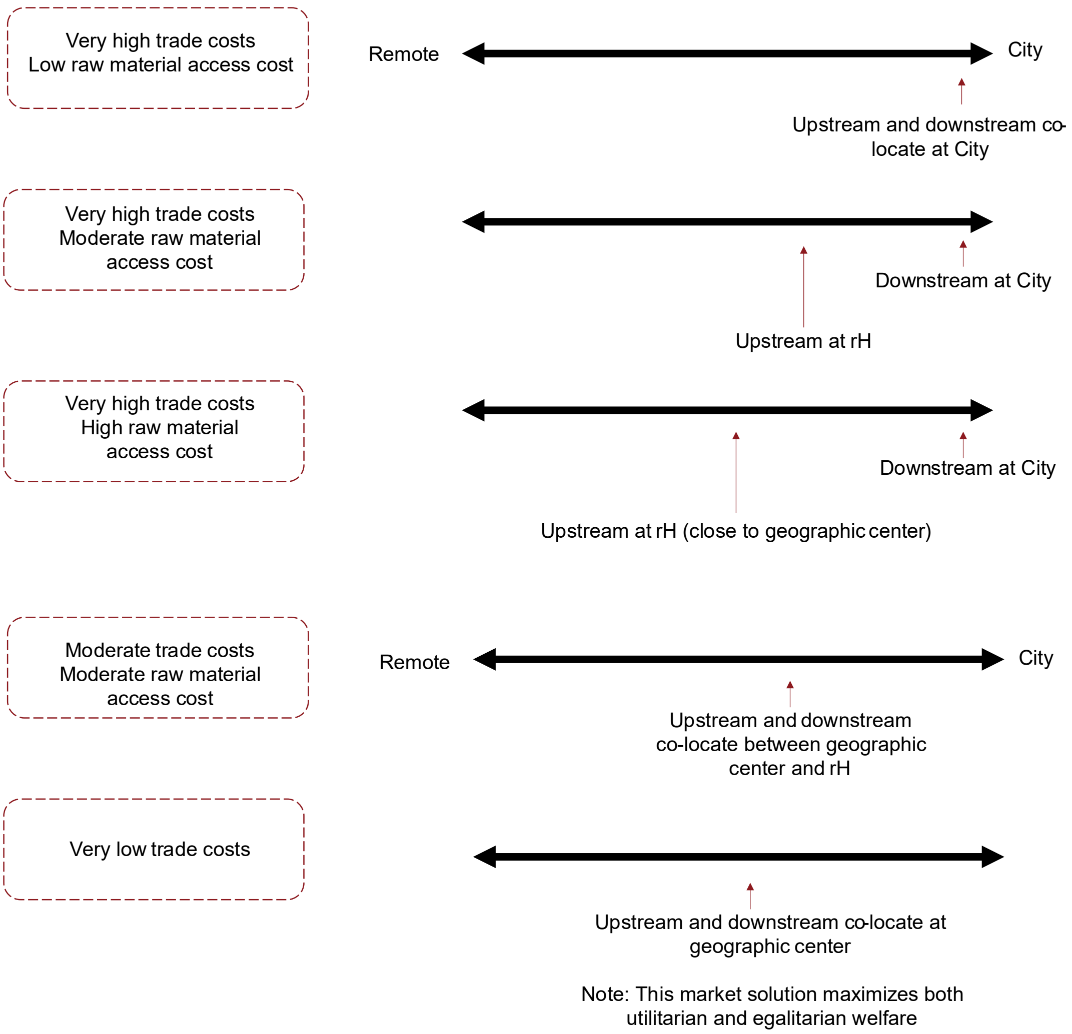


Figure 12 shows the tension between upstream and downstream locations. We discuss the sketches in sequence:

(a) When trade costs are very high (coupled with low raw material access costs), both upstream and downstream sectors co-locate at the City location.
(b) With higher raw material access costs, the upstream sector separates from City and locates closer to the geographic center (at r_H) to reduce input costs. The r_H location balances the transport cost of intermediates to City location against the raw material access cost.
(c) With prohibitively high raw material access costs, the upstream sector is always located at the geographic center, regardless of other trade costs.
(d) With lower trade costs (or if vertical separation is not feasible), both upstream and downstream co-locate at an optimal location that is between the geographic center the right boundary r_H.
(e) With negligible trade costs, both upstream and downstream co-locate at the geographic center. Input cost is minimized at this location. Welfare reaches a maximum for both utilitarian and egalitarian welfare measures.

Figure 13: Sketches of Various Nash Equilibria



References

Amendolagine, V., Presbitero, A., Rabellotti, R., & Sanfilippo, M. (2019). Local Sourcing in Developing Countries: The Role of Foreign Direct Investments and Global Value Chains. *World Development, 113*, 73 - 88.

Antràs, P., & Chor, D. (2018). On the Measurement of Upstreamness and Downstreamness in Global Value Chains. *World Trade Evolution: Growth, Productivity and Employment, 5*, 126 - 194.

Antràs, P., & de Gortari, A. (2020). On the Geography of Global Value Chains. *Econometrica, 84 (4)*, 1553 - 1598.

Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the Upstreamness of Production and Trade Flows. *American Economic Reviews: Papers and Proceedings 2012*, 412 - 416.

Asian Development Bank, Islamic Development Bank. (2019). *The Evolution of Indonesia Participation in Global Value Chains.* Asian Development Bank and Islamic Development Bank.

Baldwin, R., & Venables, A. (2013). Spiders and Snakes: Offshoring and Agglomeration in the Global Economy. *Journal of International Economics, 90 (2)*, 245 - 254.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., & Frederic, R.-N. (2005). *Economic Geography and Public Policy.* Princeton University Press.

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics, 145*.

Beverelli, C., Stolzenburg, V., Koopman, R., & Neumueller, S. (2019). Domestic Value Chains as Stepping Stones to Global Value Chain Integration. *The World Economy, 42 (5)*, 1467 1494.

Bougheas, S., Demetrides, P., & Morgenroth, E. (1999). Infrastructure, Transport Costs and Trade. *Journal of International Economics, 47*, 169 - 189.

Caliendo, L., & Parro, F. (2014). Estimates of the Trade and Welfare Effects of NAFTA. *Review of Economic Studies, 82*, 1 - 44.

Celbis, M., Nijkamp, P., & Poot, J. (2013). How Big Is the Impact of Infrastructure on Trade? Evidence from Meta-Analysis. *Merit Working Papers No.32*.

Cook, L., & Munnell, A. (1990). How Does Public Infrastructure Affect Regional Economic Performance. *New England Economic Review, Federal Reserve Bank of Boston*, 11 - 33.

Costa, F., Garred, J., & Pessoa, J. (2016). Winners and Losers from a Commodity-for-Manufactures Trade Boom. *Journal of International Economics, 102*, 50 - 69.

Dixit, A., & Stiglitz, J. (1977). Monopolistic Competition and Optimum Product Variety. *American Economic Review, 67 (3)*, 297 - 308.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Effects of Transport Infrastructure. *American Economic Review, 108*, 899-934.

Donaldson, D., & Hornbeck, R. (2016). Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal Of Economics, 131 (2)*, 799-858.

Dutu, R. (2015). Making the Most of Natural Resources in Indonesia. *OECD Economics Department Working Papers No. 1236*.

Egger, P., Loumeau, G., & Loumeau, N. (2023). China's Dazzling Transport-Infrastructure Growth: Measure and Effects. *Journal of International Economics, 142*.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies, 81 (3)*, 1046 - 1070.

Fan, J. (2019). Internal Geography, Labor Mobility and the Distributional Impact of Trade. *American Economic Journal: Macroeconomics, 11 (3)*, 252 - 288.

Fernald, J. (1999). Roads to Prosperity? Assessing the Link Between Public Capital and Productivity. *American Economic Review, 89 (3)*, 619-638.

Francois, J., & Manchin, M. (2013). Institutions, Infrastructure, and Trade. *World Development, 46*, 165-175.

Ganapati, S., & Wong, W. (2023). How Far Goods Travel: Global Transport and Supply Chains from 1965 - 2020. *Journal of Economic Perspectives, 37 (3)*, 3 - 30.

Head, K., & Mayer, T. (2019). Brands in Motion: How Frictions Shape Multinational Production. *American Economic Review, 109 (9)*, 3073 - 3124.

Hummels, D., & Schaur, G. (2013). Time as a Trade Barrier. *American Economic Review, 103 (7)*, 2935-2959.

International Monetary Fund. (2023). *World Economic Outlook, October.* International Monetary Fund.

Kalnins, A., & Lafontaine, F. (2013). Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance. *American Economic Journal: Microeconomics, 5 (3)*, 157 - 179.

Kee, H., & Tang, H. (2016). Domestic Value Added in Exports: Theory and Firm Evidence from China. *American Economic Review, 106 (6)*, 1402 - 1436.

Krugman, P., & Venables, A. (1995). Globalisation and The Inequality of Nations. *The Quarterly Journal of Economics, 110 (4)*, 857-880.

Li, C., Wang, A., Chen, X., Chen, Q., Zhang, Y., & Li , Y. (2013). Regional Distribution and Sustainable Development Strategy of Mineral Resources in China. *China Geographical Science, 23*, 470 - 481.

Martin, P., & Rogers, C. (1995). Industrial Location and Public Infrastructure. *Journal of International Economics, 39*, 335 - 351.

Mayer, T., & Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist database. *CEPII Working Paper No. 25*, 25.

McNerney, J., Savoie, C., Caravelli, F., Carvalho, V., & Farmer, J. (2022). How Production Networks Amplify Economic Growth. *Proceedings of the National Academy of Sciences. 119*.

Melitz, M., & Cuñat, A. (2012). Volatility, Labor Market Flexibility, and the Pattern of Comparative Advantage. *Journal of the European Economic Association, 10 (2)*, 225 - 254.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics, 90 (4)*, 683-701.

Nordas, H., & Piermartini, R. (2004). *Infrastructure and Trade.* Geneva: World Trade Organization.

Nuno, L., & Venables, A. (2001). Infrastructure, Geographical Disadvantages, Transport Costs, and Trade. *The World Bank Economic Review, 15 (3)*, 451-479.

Ossa, R. (2011). A "New Trade" Theory of GATT / WTO Negotiation. *Journal of Political Economy, 119 (1)*, 122 - 152.

Porteous, O. (2019). High Trade Costs and Their Consequences: An Estimated Dynamic Model of African Storage and Trade. *American Economic Journal: Applied Economics, 12 (3)*, 324-366.

Redding, S., & Esteban, R.-H. (2017). Quantitative Spatial Economics. *Annual Review of Economics, 9.*, 21 - 58.

Redding, S., & Venables, A. J. (2004). Economic Geography and International Inequality. *Journal of International Economics, 62*, 53-82.

Sachs, J., & Warner, A. (2001). The Curse of Natural Resources. *European Economic Review, 45*, 827 - 838.

Singer, M., & Donoso, P. (2008). Upstream or Downstream in the Value Chain. *Journal of Business Research, 61 (6)*, 669 - 677.

Thia, J., & Ong Lopez, A. (2023). Infrastructure and Trade Liberalization. *Structural Change and Economic Dynamics, 66*, 134 - 150.

Venables, T. (1996). Equilibrium Locations of Vertically Linked Industries. *International Economic Review, 37 (2)*, 341 - 359.

Waugh, M. (2010). International Trade and Income Differences. *American Economic Review, 100*, 2093-2124.

World Bank. (2020). *World Development Report 2020: Trading for Development in the Age of Global Value Chains.* Washington, DC.

World Bank. (2021). *The Changing Wealth of Nations 2021: Managing Assets of the Future.* Washington, DC.

World Economic Forum. (2019). *The Global Competitiveness Report 2019.* World Economic Forum.

Yeaple, S., & Golub, S. (2007). International Productivity Differences, Infrastructure and Comparative Advantage. *Review of International Economics, 15 (2)*, 223 - 242.

Zheng, H., Zhang, Z., Wei , W., Song, M., Dietzenbacher, E., Wang, X., . . . Guan, G. (2020). Regional Determinants of China's Consumption-based Emissions in the Economic Transition. *Environment Research Letters, 15 (7)*.