Global Supply Chains and Regional Shocks*

Nadim Elayan†
Banco Central de Chile

José Ramón Morán‡
Banco de México

August 21, 2026
[Click here for the latest version]

Abstract

We explore how regional shocks affect the formation of global supply chains, recognizing the trade-off firms face when choosing the sourcing location of their inputs. On the one hand, sourcing from similar countries entails a lower risk of being exposed to regional shocks but on the other, it implies firms are less able to take advantage of the pattern of comparative advantage across the World. Using customs data on Mexican firms we document how firms weigh this trade-off, whether there is sectoral heterogeneity in this behavior, and study whether this heterogeneity is the result of sector-specific input complementarity. We build a model of global sourcing that accounts for comparative advantage being driven by geographical location and exposure to regional shocks. With our model, we explore the effects that an episode of increased regional risk, namely the COVID-19 global pandemic, had on firms' global supply chain formation.

JEL Codes: F12, F13, F15, F60
Keywords: global supply chains, nearshoring, supply shocks, global sourcing, correlated comparative advantage

*Special thanks to Alfonso Cebreros, Brian Fujiy, Andrei Levchenko, and Sebastián Sotelo for their invaluable comments and feedback. The data used in this paper are confidential and were made available through the Econlab at Banco de México. Inquiries regarding the terms and conditions for accessing this data should be directed to econlab@banxico.org.mx. The views and conclusions in this paper are solely those of the authors and do not necessarily reflect the opinions of Banco de México or its Board of Governors. All errors are our own.

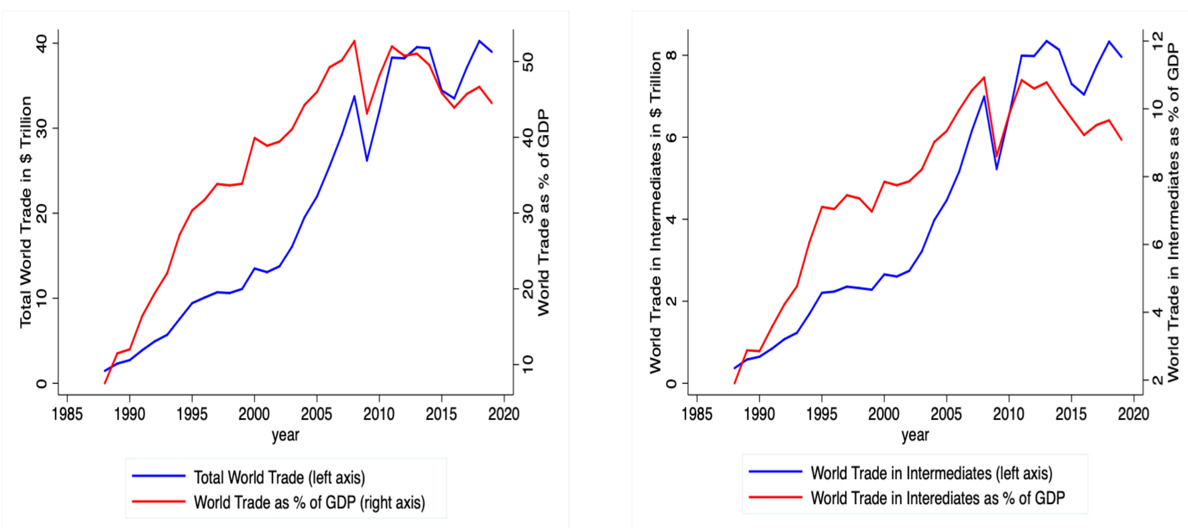
†Research Department, Agustinas 1180, 8340454 Santiago, Región Metropolitana, Chile. Contact at https://nadimelayan.github.io, or at nelayan@bcentral.cl.

‡Research Department, Avenida 5 de Mayo 18, Colonia Centro, CP 06000, Ciudad de México, México. Contact at https://sites.google.com/joseramonmoran, or at jmoran@banxico.org.mx.

1 Introduction

Globalization has witnessed a notable acceleration in recent decades, particularly in the integration of productive processes. Capitalizing on countries' comparative advantage patterns, production has progressively fragmented across the globe. For instance, steel production may occur in China, followed by transformation into an engine in the US, assembly into a car in Mexico, and eventual export back to the US.

In Figure 1, we present the trajectory of total world trade in trillion US dollars from 1988 to 2019 in the left panel, alongside the evolution of world trade in intermediates in the right panel. The surge is remarkable: over 30 years, global trade volume has surged from 1.46 trillion US dollars, accounting for 7.5% of world GDP, to 40 trillion US dollars, constituting 45% of world GDP. The growth in intermediate trade mirrors this trend, climbing from 0.37 trillion US dollars and 1.9% of world GDP in 1988 to 8 trillion US dollars and 9% of world GDP in 2019.



(a) World Trade (b) World Trade in Intermediates

Source: World Trade Organization

Figure 1: Evolution of World Trade and World Trade in Intermediates

In theory, the reason for why these production chains emerge stems from technological advancements and reductions in trade and transportation costs, enabling the exploitation of global comparative advantages. Building upon our previous illustration, the decision to assemble an engine in Mexico rather than the US may stem from cost advantages, prompting US firms to capitalize on this opportunity. In essence, the benefits of Global Value Chains (GVCs), or the integration of productive processes, lie in firms' ability to enhance production efficiency.

However, what about the associated costs? We contend that heightened fragmentation of the production process may elevate chain vulnerability, thereby potentially disrupting the entire production cycle. This vulnerability may materialize in scenarios where a chain relies on essential and interdependent inputs sourced from multiple locations, with the final product contingent upon the seamless provision of these inputs from each location. In this paper we focus on how these two opposing effects—increased efficiency versus increased risk of failure in the production process—shape firms' decisions on how dispersed their GVCs are.

Consider the following extreme scenarios in order to better illustrate our main tradeoff. Consider a scenario with 10 potential production locations, each with an idiosyncratic probability of failing to deliver their input of 10%. In this setup, if any location fails to deliver, the entire chain fails to produce the final good. The likelihood of a production failure hinges crucially on the correlation of these failure probabilities across locations. For instance, in the case of perfectly positive correlation—where shocks are synchronously aligned—the chain would fail only once every 10 periods. Conversely, if failure probabilities are completely independent, the chain would fail, on average, 6.5 periods out of 10. Finally, under a scenario of perfectly negative correlation, the chain would fail in every period.

We propose a mean-variance trade-off framework for supply chain formation, with geographical distance between countries serving as a key determinant. On one hand, firms are incentivized to choose supply chains with countries specializing in different segments of the production process to maximize comparative advantage gains. Our aggregate analysis shows that greater geographical distance between countries is associated with more divergent comparative advantage patterns, thus encouraging firms to source from countries specializing in disparate industries, likely situated further apart. On the other hand, firms also have incentives to form supply chains with countries experiencing more synchronous shocks, particularly in cases involving complementary or essential inputs. We find that countries situated closer to each other tend to have more synchronized business cycles and shocks. Consequently, firms seek to minimize failure rates by sourcing from geographically proximate countries.

We use customs data from Mexico at the exports and imports transaction level to show that firms with more geographically dispersed supply chains tend to exhibit higher export volumes. However, these firms also experience greater volatility and disruptions in their chains, on average. To quantify the firm-level dispersion of supply chains, we introduce a novel measure based on the geographical distance between all locations from which a Mexican firm sources its inputs.

We provide empirical support for our proposed trade-off at both the aggregate country level and the disaggregated firm level. At the country level, we observe that pairs of countries that are more distant from each other tend to specialize in more diverse industries, whereas pairs of countries that are closer geographically tend to exhibit more similar comparative advantage patterns. Additionally, we find that more distant country pairs experience less correlated (or more negatively correlated) economic shocks, whereas geographically closer country pairs tend to have more synchronized business cycles. At the firm level, we find that firms with more dispersed global value chains (GVCs) tend to achieve higher average export volumes. However, they also exhibit higher levels of volatility. Notably, we observe this trade-off primarily in more complex sectors, where production is presumed to rely more heavily on complementary inputs. This suggests that the trade-off is particularly prominent in environments where GVC locations supply complementary goods to a final goods producer, magnifying the impact of chain disruptions.

We develop a structural model of global input sourcing that incorporates our main mechanism. Building upon the framework of Antras et al. (2017), our model introduces two key elements. Firstly, we incorporate a multivariate Fréchet distribution with constant elasticity of substitution (CES) region nests of countries, which underlies the comparative advantage in input production. This feature encourages firms to source intermediates from countries that are geographically distant from each other, enhancing the exploitation of comparative advantage gains. Secondly, we consider that firms are exposed to regional shocks in the arrival of the inputs needed to produce their output. Given that shock synchronization is negatively correlated with geographical distance, firms are further motivated to form GVCs with less distant locations to minimize the risk of failures in delivery.

We show that in industries heavily reliant on complementary or essential inputs for production, a tradeoff arises in selecting sourcing locations. Highly dispersed GVCs tend to better exploit comparative advantages, resulting in higher average exports. However, these GVCs also tend to involve countries with less positively synchronized business cycles, leading to more frequent disruptions in the supply chain. These findings are particularly relevant in the aftermath of significant shocks like COVID-19, during which policymakers have intensely debated the merits of renationalizing production. We suggest that an alternative response to mitigate fears of supply chain disruptions could be the development of less global and more regional GVCs.

This paper contributes to the general literature regarding GVCs. We use Mexican customs data for our empirical strategy, as do Antràs (2020), Antràs and De Gortari (2020), De Gortari (2019), de Gortari (2020), Wang et al. (2017). The focus of these papers is on the length, upstreamness, value added, and the impact of trade agreements on global value chains; we instead focus on a novel mean-variance trade-off that firms face when deciding where to source their inputs from. We also document novel empirical facts about recent trends in GVC size, dispersion, and risk.

Our study contributes to two main branches of the literature. Firstly, we address a very relevant and prevailing topic regarding the status of GVCs in the aftermath of the COVID-19 crisis. By introducing a novel perspective and identifying an empirically unexplored trade-off, we offer an original assessment of the future trajectory of supply chain processes between nations. Secondly, we provide empirical evidence of a theoretical trade-off, as posited by Costinot (2009), between comparative advantage gains and the risk of failure arising from production fragmentation. Expanding upon this idea, we introduce the concept of shock synchronization between countries, thereby uncovering new implications for GVC formation.

This paper aims to answer a policy relevant question: how does the presence of regional shocks and uncertainty affect the formation of GVCs? Miroudot (2020) argues that during the COVID-19 pandemic industries in Thailand that suffered larger economic losses were not the ones with more complex or fragmented production processes and therefore the way we currently understand GVCs should not change. Evenett (2020) suggests that GVCs tend to be in danger during recessions as societies turn to more protectionist measures promising more robustness and less volatility. Bonadio et al. (2021) find that although the evident disruption of global supply chains magnified the last economic crisis re-nationalizing production would have not helped, in fact, they show that it would have made countries more dependent on domestic inputs that were also subject to lockdowns. Eppinger et al. (2020) and Hayakawa and Mukunoki (2021) also find that although the impact of GVC on the economy has been substantial during the global pandemic, re-nationalization would have brought even larger losses. We contribute to this discussion by empirically showing that firms can form more resilient GVCs without necessarily altering the size of them through increasing its shock synchronization. We also allow firms to choose endogenously the size and complexity of their GVCs so that the choice is no longer a binary between complete autarky or a fully globalized process.

Second, we provide empirical evidence of Costinot's (2009) trade-off between comparative advantage gains and risk of failure derived from the size of a production process, but in terms of GVC formation between countries. Sequential production processes with essential stages carried out in different locations à la Costinot et al. (2013) provide a better intuition for the main trade-off of our paper. These authors treat size and complexity of GVCs as exogenous whereas in our paper firms will decide endogenously according to comparative advantage and risk forces. Moreover, the risk of failure in a GVC will also come from a novel source, shock synchronization among GVC members.

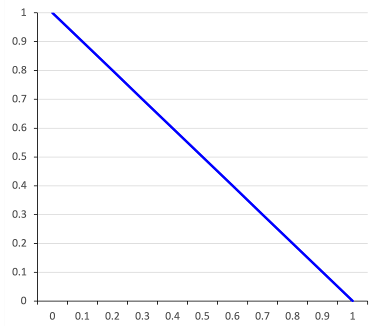
This paper is structured as follows. Section 2 introduces a simple model that underpins the core trade-off examined in this study. Section 3 outlines the empirical methodology and data sources employed to demonstrate this trade-off. Specifically, we use Mexican customs data, which provides detailed firm-level information on the sourcing of foreign inputs. Additionally, we draw on data from UN Comtrade and the World Input-Output Database (WIOD) to highlight aggregate trends regarding the geographical distance between country pairs and its impact on comparative advantage patterns and business cycle synchronization. Section 4 presents our primary model, built upon the framework proposed by Antras et al. (2017). Section 5 concludes.

2 Discussion of Our Mechanism

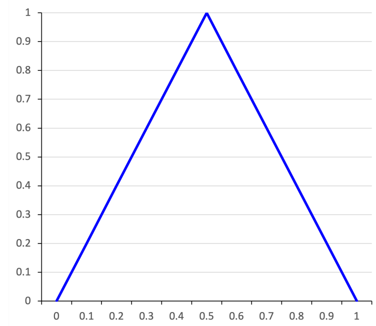
This section aims to clarify the factors influencing firms' endogenous selection of their global value chain (GVC) composition. We define a firm's GVC dispersion as the average geographical distance between all pairs of locations involved in the sourcing process. Thus, a GVC spanning considerable distances will exhibit high dispersion, while one involving closely located sites will display low dispersion. We posit that firms face a trade-off when deciding to expand their GVC dispersion. On one hand, increased dispersion allows firms to more effectively leverage each location's comparative advantage, thereby enhancing overall productivity. Conversely, higher dispersion raises the chain's vulnerability, as the delivery of the final product becomes contingent on the individual conditions of a greater number of disparate locations. The downside of a more dispersed GVC is particularly pronounced in scenarios where the production process relies heavily on complementary inputs or sequential steps.

We start with a simple example where a firm must get essential inputs from various locations to manufacture its final output. Consequently, the failure of any location to supply its input would impede the firm's ability to produce the final good. We consider a continuum of essential inputs, denoted by s, spanning the interval $[0, 1]$. When creating a global value chain (GVC), a firm must decide how to partition this interval, with each segment being supplied by a distinct location. For instance, firm i may opt to source its inputs s, or a grouping of inputs, from various locations l. Initially, we assume that price disparities between locations stem solely from varying marginal costs, which are contingent upon each location's unique comparative advantage in production.

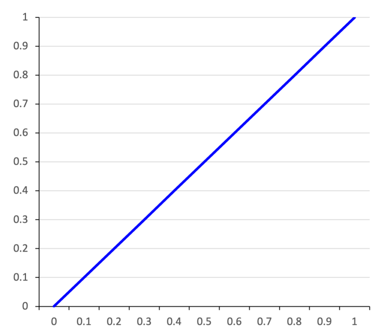
Consider three locations denoted by $l \in \{A, B, C\}$. Productivity, represented by $z_{l,s}$, is specific to each location (l) and input (s). Location productivity is defined as the average productivity across all inputs sourced from that location. Assuming a linear decline in productivity, location A attains its maximum productivity at $s = 0$. Location B peaks at



(a) Location A



(b) Location B



(c) Location C

Figure 2: Productivity (y-axis) when performing a given step (x-axis) of the productive process

$s = \frac{1}{2}$ and decreases linearly towards both extremes, while location C reaches its minimum at $s = 0$ and linearly increases to its maximum at $s = 1$. We represent this example in Figure 2.

This example allows us to compute the productivity levels of locations A, B, and C for a specific input value ($s = 0.2$). Location A achieves a productivity level of 0.8 at this input level, denoted as $z_{A,0.2} = 0.8$. In contrast, locations B and C achieve lower productivity levels of 0.4 and 0.2, respectively, represented as $z_{B,0.2} = 0.4$ and $z_{C,0.2} = 0.2$.

To compute productivity within input intervals, we calculate the average productivity over the interval, representing the area below the productivity curve divided by the length of the interval. For example, the average productivity within the interval $[0, \frac{1}{3}]$ for each location is: $z_{A,[0,\frac{1}{3}]} = \frac{5}{6}$, $z_{B,[0,\frac{1}{3}]} = \frac{1}{3}$, and $z_{C,[0,\frac{1}{3}]} = \frac{1}{3}$. Notably, the average productivity across the entire interval is $\frac{1}{2}$ for all three locations.

It is also worth noticing that as the dispersion of a value chain increases, the likelihood of output failure also rises. We assume that each location delivers its input with a probability of $\theta_l = 0.8$. Although this probability is likely location-specific, for simplicity, we initially assume it is uniform across all locations. Consequently, the effective productivity levels ($\tilde{z}$)—

productivity multiplied by the probability of delivery—are as follows within the interval $[0, \frac{1}{3}]$: $\tilde{z}_{A,[0,\frac{1}{3}]} = \frac{2}{3}$, $\tilde{z}_{B,[0,\frac{1}{3}]} = \frac{6}{25}$, and $\tilde{z}_{C,[0,\frac{1}{3}]} = \frac{6}{25}$.

We start examining the impact of increasing the number of locations in the GVC on the chain's average effective productivity. We assume that the probabilities of output delivery (θ_l) are independent across locations. Subsequently, we investigate the implications of a more dispersed chain, characterized by the inclusion of more or less similar locations. The similarity between two countries is contingent upon the similarity of their productive processes, such as having comparative advantages in inputs that are closer to each other within the unit interval. Additionally, countries are considered more similar if their probabilities of delivering output are more correlated, indicating similar responses to aggregate shocks.

2.1 Independent θ_l

We begin by assuming that the probability of output delivery in one location, such as A, is independent of that in any other location. Thus, in a chain comprising N countries, the probability of final output delivery is given by $\prod_{l=1}^{N} \theta_l$. For instance, if a firm sources inputs from 0 to $\frac{1}{2}$ in location A and from $\frac{1}{2}$ to 1 in location B, its effective productivity would be: $\tilde{z}_i = 0.8^2 * \frac{1}{2}(\frac{3}{4} + \frac{1}{2}) = 0.4$. Conversely, producing the entire product in either location would yield an average effective productivity level of 0.4.

The firm's decision problem involves choosing among the following options: sourcing all inputs from A, sourcing all inputs from B, sourcing all inputs from C, or sourcing from various locations in all possible combinations. Exclusive sourcing in any one location results in the same average productivity: $\tilde{z}_{A,[0,1]} = \tilde{z}_{B,[0,1]} = \tilde{z}_{C,[0,1]} = 0.8 * 0.5 = 0.4$. Another option is to split the sourcing between A and B. If the firm wants to only source from A and B, the optimal strategy would be to source inputs from 0 to $\frac{1}{3}$ in location A and the remainder in location B. This is because, upon examining the productivity schedules of both locations, it becomes evident that location A has a comparative advantage between inputs 0 and $\frac{1}{3}$—thus exhibiting higher productivity than B—while location B has a comparative advantage from inputs $\frac{1}{3}$ to 1. Therefore, $\tilde{z}_{A/B} = 0.8^2(\frac{1}{3}\frac{5}{6} + \frac{2}{3}\frac{7}{12}) = 0.427$.

In Table 1 we show all possible value chains combinations and show the average effective productivity of each case. Notice that we are already assuming that each interval of inputs s will be sourced at the location l with the highest productivity among the selected group.

The highest average effective productivity is achieved when the firm sources the first half of all inputs from location A and the second half from location C. However, in a scenario where delivery problems are absent ($\theta_l = 1 \; \forall l$), the highest average productivity is achieved by sourcing inputs from all locations by a significant margin. This example underscores

Chain Group	z_i	$\tilde{z}_i$
$A:[0,1],\ B:\emptyset,\ C:\emptyset$	0.5	0.4
$A:\emptyset,\ B:[0,1],\ C:\emptyset$	0.5	0.4
$A:\emptyset,\ B:\emptyset,\ C:[0,1]$	0.5	0.4
$A:[0,\frac{1}{3}],\ B:[\frac{1}{3},1],\ C:\emptyset$	0.66	0.42
$A:[0,\frac{1}{2}],\ B:\emptyset,\ C:[\frac{1}{2},1]$	0.75	**0.48**
$A:\emptyset,\ B:[0,\frac{2}{3}],\ C:[\frac{2}{3},1]$	0.66	0.42
$A:[0,\frac{1}{3}],\ B:[\frac{1}{3},\frac{2}{3}],\ C:[\frac{2}{3},1]$	**0.83**	0.43

Table 1: Optimal chain size when θ_j are independent.

two key insights. Firstly, the addition of an extra location to the chain tends to increase average productivity, but it also elevates the risk of non-delivery of the final good, rendering the overall impact on average effective productivity ambiguous in expectation. Secondly, a chain comprising the two most dissimilar locations, A and C, yields the greatest gains compared to other groupings. This result arises from the fact that location A possesses a stronger comparative advantage in inputs where location C performs poorly, and vice versa. Consequently, efficiency gains are maximized when locations' comparative advantage patterns exhibit greater disparity.

2.2 θ_j not independent

We now relax our previous assumption of independence among θ_j across locations. Instead, we consider two distinct states of nature, each occurring with equal probability, as depicted in Table 2. In one state, locations A and B exhibit a high probability of delivering their inputs, while location C delivers its inputs significantly less frequently. Conversely, in the second state, only location C delivers its input almost certainly. Thus, in this scenario, locations A and B share identical exposure to states of nature, whereas location C stands out as unique. It is worth noting that the average delivery probability remains consistent at 0.8 across all three locations, mirroring the previous assumption.

In Table 3 we compute again the average effective productivity of each possible chain group. We show that the optimal strategy remains to divide input demand between locations A and C. However, the average effective productivity has decreased due to the asymmetric shock exposure between these two locations. In contrast, locations A and B experienced a slight increase in effective productivity, from 0.42 to 0.43, attributable to greater

θ_l	θ_A	θ_B	θ_C
State 1	0.9	0.9	0.7
State 2	0.7	0.7	0.9

Table 2: θ_j in different states of nature.

synchronization in their shock exposure.

Chain Group	z_i	$\tilde{z}_i$	
	Indep. θ_l	Indep. θ_l	Dep. θ_l
$A:[0,1]$, $B:\emptyset$, $C:\emptyset$	0.5	0.4	0.4
$A:\emptyset$, $B:[0,1]$, $C:\emptyset$	0.5	0.4	0.4
$A:\emptyset$, $B:\emptyset$, $C:[0,1]$	0.5	0.4	0.4
$A:[0,\frac{1}{3}]$, $B:[\frac{1}{3},1]$, $C:\emptyset$	0.66	0.42	0.43
$A:[0,\frac{1}{2}]$, $B:\emptyset$, $C:[\frac{1}{2},1]$	0.75	**0.48**	**0.47**
$A:\emptyset$, $B:[0,\frac{2}{3}]$, $C:[\frac{2}{3},1]$	0.66	0.42	0.42
$A:[0,\frac{1}{3}]$, $B:[\frac{1}{3},\frac{2}{3}]$, $C:[\frac{2}{3},1]$	**0.83**	0.43	0.42

Table 3: Optimal chain size when θ_j are not independent.

The rationale behind these findings is as follows: when the firm sources inputs solely from location A, the average probability of input delivery is 80%. Introducing location B to the chain reduces this probability by 15 percentage points on average. However, combining locations A and C—two locations with more disparate comparative advantage patterns—results in a larger average reduction of 17 percentage points in the probability of delivery.

In summary, locations with vastly different comparative advantage patterns yield the greatest gains in comparative advantage, as they specialize in distinct inputs crucial for the production process. However, they also face a higher risk of non-delivery due to their less synchronized shock exposure.

3 Empirical Evidence

This section of the paper is structured as follows. First, we describe the different data sources we use throughout the paper. Second, since our story rests on the assumption that both comparative advantage and risk are geographically correlated, we show country-level evidence supporting these relationships. Third, we use reduced-form estimations at the firm level in order to provide evidence for firms taking into account the trade-off between efficiency and risk when choosing their global supply chains. Furthermore, we explore the occurrence of the COVID-19 pandemic as a quasi-natural experiment and exploit variation in firms' exposure to desynchronized sourcing locations to provide evidence of its detrimental effect on firm performance.

3.1 Data Sources

Our goal is to construct a firm-level dataset on firms' sourcing choices, and in particular on how efficient and how risky a firm's supply chain is. In our context, a supply chain will be more efficient the less similar its members are in terms of comparative advantage. Likewise, a supply chain will be considered riskier the lower the correlation in the cyclical component of GDP. We also explore alternative proxies for measuring regional idiosyncratic risk such as the correlation in COVID-related lock-downs. The data sources we use for constructing our dataset are the following:

Customs Data: We use customs monthly data from Mexico from 2014 to 2020. It contains the universe of exports and imports at the transaction level. For every transaction, we observe the following: Firm ID, product at the HS 6-digit level, trade value, origin/destination of the transaction, and whether the transaction was classified as either for an intermediate or a final good. We focus on transactions corresponding to imports of intermediate goods, as these uncover firms' global sourcing strategies. We also use data on these firms' exports as we assign an industry to a firm depending on the products exported by it.

Trade Flow Data: This data comes from the BACI dataset from CEPII. It contains data on exports and imports at the country level, disaggregated at the HS 6-digit level. We use this data to compute country-specific measures of revealed comparative advantage using data on observed trade flows.

GDP Data: This data comes from the World Bank. We use GDP data in current USD. The purpose of the data is to compute the correlation between countries' cyclical component of GDP.

Gravity Data: This data comes from the Gravity dataset assembled by CEPII. For each country pair, it contains gravity variables such as geographical distance, whether countries share a common border, common language, common religion, common colonial ties, among others. We use data on pairwise geographical distance.

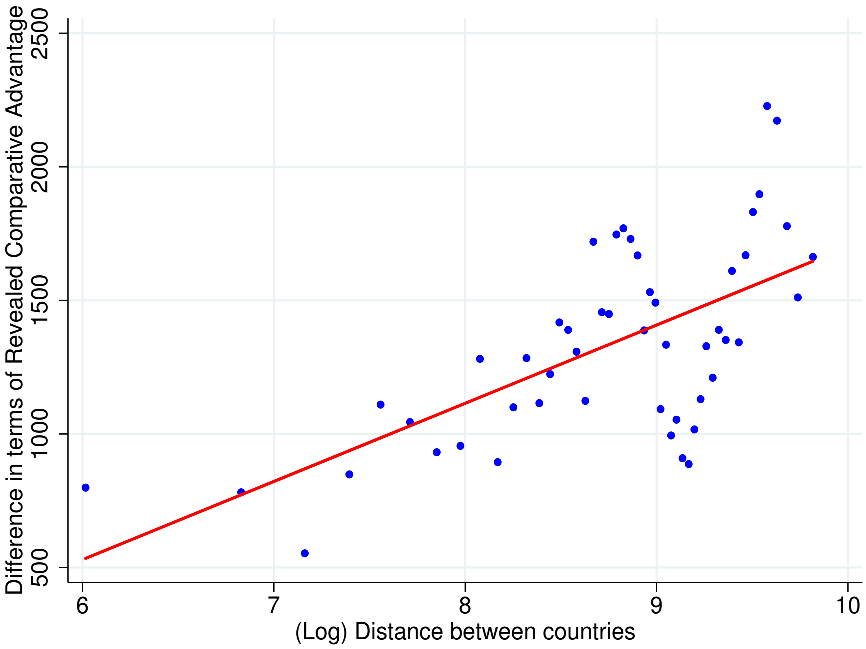
COVID Lockdown Data: Regarding data on COVID-related shutdowns we use Oxford's COVID-19 Government Response Tracker, in particular their Stringency Index which measures the severity of lockdowns at a country level. We further compute all pairwise correlations of this index across all countries in the world from the 1st of March 2020.

3.2 Country-Level Evidence of the Tradeoff

We present some country-level facts regarding our two main variables driving the tradeoff that a firm faces when choosing to form its GVC. One is the difference in revealed comparative advantage between a given pair of countries. This variable captures how different two countries are in terms of their industry comparative advantage. The higher the value the more different a pair of countries are with respect to each other. The other variable is the correlation in the cyclical component of countries' GDP. This variable captures how correlated a pair of countries' production is, or how likely it is that both countries experience a negative shock at the same time.

We first show that geographical distance is positively correlated with differences in Revealed Comparative Advantage for each pair of countries in the world. Figure 3 shows that, on average, countries that are further away from each other tend to specialize in different industries compared to countries that are closer to each other. We use a binscatter since we have 12,403 observations, which corresponds to 158 different countries ($\frac{158\times157}{2}$ = 12,403). For instance, the pair Senegal–Portugal has a log distance of 7.94 (2,803.7 km) and a difference in RCA of 1,937.5. The pair Uruguay–Argentina, by contrast, has a log distance of 5.37 (215.07 km) and is less different in terms of comparative advantage, with a difference in RCA of 327.28.

We also show that geographical distance is negatively correlated with the correlation in the cyclical component of GDP for each pair of countries in the world. Figure 4 shows that, on average, countries that are further away from each other tend to experience less synchronized business cycles. Following the previous example, the pair Senegal–Portugal, with a log distance of 7.94 (2,803.7 km), has a correlation in the cyclical component of GDP of 0.475. Argentina and Uruguay's business cycles, on the other hand, are relatively more correlated, with a pairwise correlation of 0.633. We show the same relationship but instead of using the cyclical correlation in GDP we use the stringency index on COVID-related shutdowns built by Oxford's COVID-19 Government Response Tracker. Figure B.1 shows the analogous relationship but using this new variable instead.



Note: Binscatter using 50 bins. Each bin represents a partition of the *x-axis* variable.

Figure 3: Geographical distance and differences in RCA

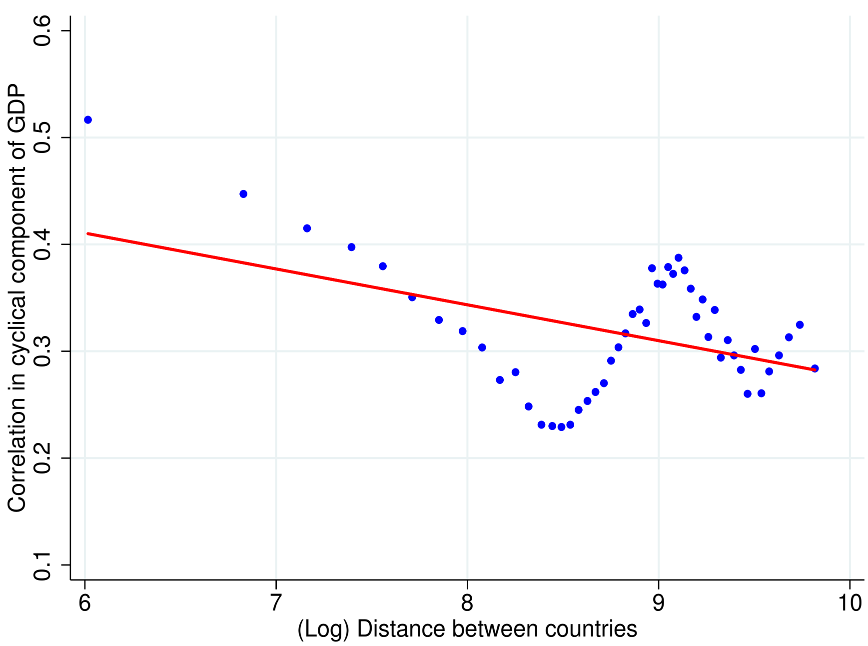


Figure 4: Geographical distance and cyclical GDP correlation

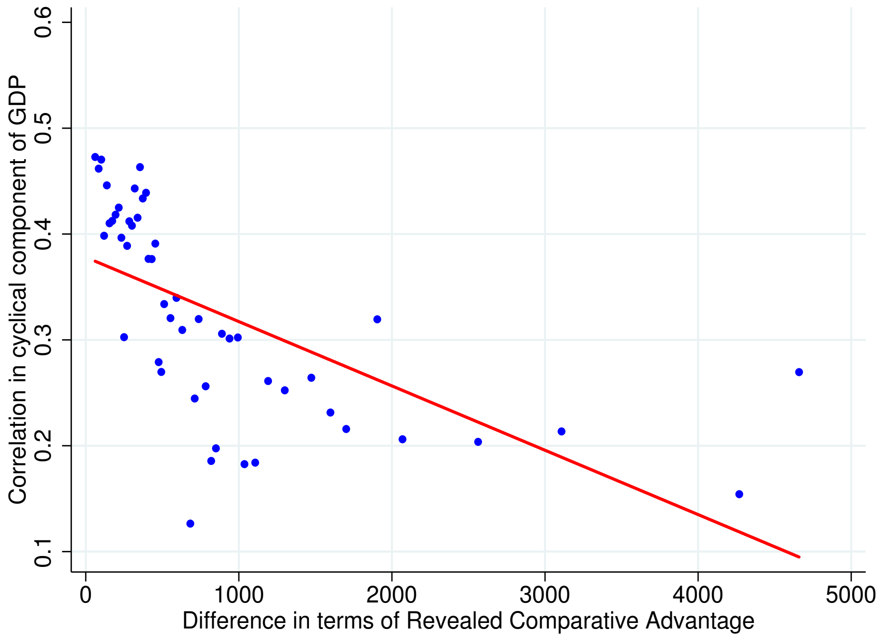


Figure 5: Comparative Advantage and cyclical GDP correlation

We hypothesize that geographical distance is the key variable generating our main trade-off. Countries that are closer to each other could potentially be good partners in a given GVC due to their highly correlated business cycles which in turn would result in delivering the final output more frequently for production processes that rely in complementary inputs. However, they will not bring high efficiency gains since they are very similar in comparative advantage terms. Conversely, countries that are very far from each other could also seem like good partners since they are very different in terms of comparative advantage; nevertheless, they will tend to experience negative shocks at different points in time, which will increase the probability of failing to deliver. It is important to notice that this trade-off should theoretically exist only for production processes that rely heavily on complementary steps. In Figure 5 we show the negative relationship between these two variables. We also show the same figure but using the COVID stringency index instead of the cyclical component of GDP in Figure B.2.

To better illustrate the tradeoff we can put it in terms of a specific example. Imagine that a firm splits its production process between 4 different locations: Argentina, Chile, Colombia and Uruguay. This group has an average pairwise difference of RCA of 1.25 and an average pairwise correlation of the cyclical component of their GDP of 0.61. If the firm decides to include Indonesia as the 5th location of the GVC, since it is a very different location, the average pairwise difference in RCA will increase to 1.27 but the average correlation of their business cycles will fall to 0.47. Therefore, including Indonesia will bring high comparative advantage gains but it will also add some risk of failing to deliver output since now the chain is exposed to different kind of shocks, or different timing of the shocks.

Conversely, adding Peru as the 5th location instead would decrease the average difference in RCA to 1.24, since it is a more similar country, but it would also increase the average correlation of their business cycles to 0.63. In this case adding Peru would not translate to huge gains in terms of comparative advantage but it would not increase the risk of not delivering the product by as much as when including Indonesia.

3.3 Efficiency, Risk, and the Formation of Supply Chains

We study how the trade-off between efficiency and risk shapes firms' choices regarding their supply chain. We first compute a country-specific measure of revealed comparative advantage following the work by Balassa (1965):

$$RCA_{ip} = \frac{X_{ip} / \sum_{p' \in P} X_{ip'}}{\sum_{i' \in I} X_{i'p} / \sum_{i' \in I} \sum_{p' \in P} X_{i'p'}} \tag{1}$$

where X_{ip} represents country i's exports of product p. The RCA index compares the share of product p in country i's total exports to the share of product p in world exports. If $RCA_{ip} > 1$, that is, if country i exports a product p relatively more intensively compared to the rest of the world, then country i is considered to have comparative advantage in product p.

We use the BACI data on observed trade flows and compute the revealed comparative advantage index for every country in our sample. We define products at the HS 4-digit level. We then compute how similar countries are in terms of their RCA. For every pair of countries in our sample, i and j, we compute:

$$\text{Sim}_{ij} = \left(\sum_{p \in P} |RCA_{ip} - RCA_{jp}|^2 \right)^{1/2} \tag{2}$$

We now compute our measure of supply-chain efficiency using our Mexican firm-level customs data. Let $S_t(f)$ represent the set of countries from which firm f sources intermediate inputs at time t. A firm's supply chain efficiency is computed following:

$$E_{ft} = \frac{2(|S_t(f)| - 2)!}{|S_t(f)|!} \sum_{\{i,j\} \subset S_t(f)} \text{Sim}_{ij} \tag{3}$$

For each firm f at each time period t we compute the average RCA dissimilarity between every pair of countries in the firm's supply chain. We use this as a measure of a firm's supply chain efficiency, as sourcing from countries with different patterns of comparative advantage implies taking advantage of the fact that while a country might have an advantage in supplying one product, another country might have it for a different product.

We also compute a proxy for how risky a firm's supply chain is, similarly to how we computed our measure of efficiency. This measure intends to capture the likelihood of a firm's supply chain experiencing disruptions. Using our World Bank data on countries' GDP, we first extract the cyclical component of it for each one of the countries from which Mexican firms are sourcing intermediate inputs. This is following the work in Hodrick and Prescott (1997).

Let HP_i stand for the cyclical component of country i's GDP. Our measure of a supply chain's level of risk is computed first by constructing an inverse measure of risk:

$$R_{ft}^{\text{Inv}} = \frac{2(|S_t(f)|-2)!}{|S_t(f)|!} \sum_{\{i,j\} \subset S_t(f)} \text{corr}(HP_i, HP_j)^{-1}, \tag{4}$$

and then normalizing it in order to deal with an issue of having negative and positive numbers altogether:

$$R_{ft}^{\text{Inv Norm}} = \frac{R_{ft}^{\text{Inv}} - R_{\text{Min}ft}^{\text{Inv}}}{R_{\text{Max}ft}^{\text{Inv}} - R_{\text{Min}ft}^{\text{Inv}}}. \tag{5}$$

Finally our measure of risk is:

$$R_{ft} = 1 - R_{ft}^{\text{Inv Norm}} \tag{6}$$

The correlation among countries' cyclical components of GDP serves as a proxy for the synchronization of shocks experienced by these countries. Our measure of supply chain riskiness assesses how, when the countries from which a firm sources intermediate inputs are less synchronized—indicating a lower correlation in their cyclical components—the firm faces a greater number of idiosyncratic shocks, leading to more disruptions in its production process. The variable $0 \leq R_{ft} \leq 1$ approaches 0 when the risk of a firm's f value chain in year t is very low, indicating that the firm sources from countries with highly synchronized business cycles. Conversely, it approaches 1 when the risk is high, signifying that the firm sources from countries with less synchronized business cycles.

Lastly, we also compute the degree of geographic dispersion of a firm's supply chain. As before, $S_t(f)$ represents the set of countries from which firm f sources inputs at time t. The level of geographic dispersion of a firm's supply chain is given by:

$$D_{ft} = \frac{2(|S_t(f)| - 2)!}{|S_t(f)|!} \sum_{\{i,j\} \subset S_t(f)} \text{distance}(i, j) \tag{7}$$

where distance(i, j) represents the distance between country i and country j, for which we use the gravity database from CEPII, measuring the physical distance between the capital cities of the two countries. As in Equations (3) and (4), our index computes the average across all pairs of countries in a firm's supply chain.

Table 4 provides a summary of our key variables. We also present the description when we limit the sample to only firms that source from more than one country, as this is the primary restriction in our regressions. On average, a firm sources from 13.21 different countries. However, the bottom 25% of firms only source from up to four different locations, and the median is only eight. This variation is largely driven by large firms, which source from many locations, with the maximum being 129. Additionally, we observe substantial variation in our other main variables, including dispersion, efficiency, and risk of the chain.

	Mean		Std. dev.	
	All	Size > 1	All	Size > 1
Size chain (# of countries)	13.21	14.06	13.28	13.33
Dispersion of the chain in kms (D_{ft})	7,004.48	7,491.45	2,591.65	1,880.26
Efficiency of the chain index RCA (E_{ft})	158.46	169.47	235.46	169.45
Risk of the chain (R_{ft}, index from 0 to 1)	0.29	0.31	0.11	0.08
Firm's Log of Exports ($)	14.03	14.19	3.58	3.57
Observations	34,640			
Firms	5,442			

Table 4: Description of our main variables

Figure 6 shows the yearly evolution of two key constructed variables: the average size and the average efficiency of the value chain. Both variables exhibit almost parallel trends, generally increasing over time, except for a significant decline during the COVID-19 pandemic in 2020. So, apart from the anomalous year of 2020, Mexican firms have progressively sourced inputs from a greater number of countries and from nations specialized in more diverse industries, thereby enhancing their comparative advantage. In Figure 7, we present the evolution of the other two principal variables: average dispersion and the average risk of the value chain. The trend in average dispersion is less clear, as it initially rises but then expe-

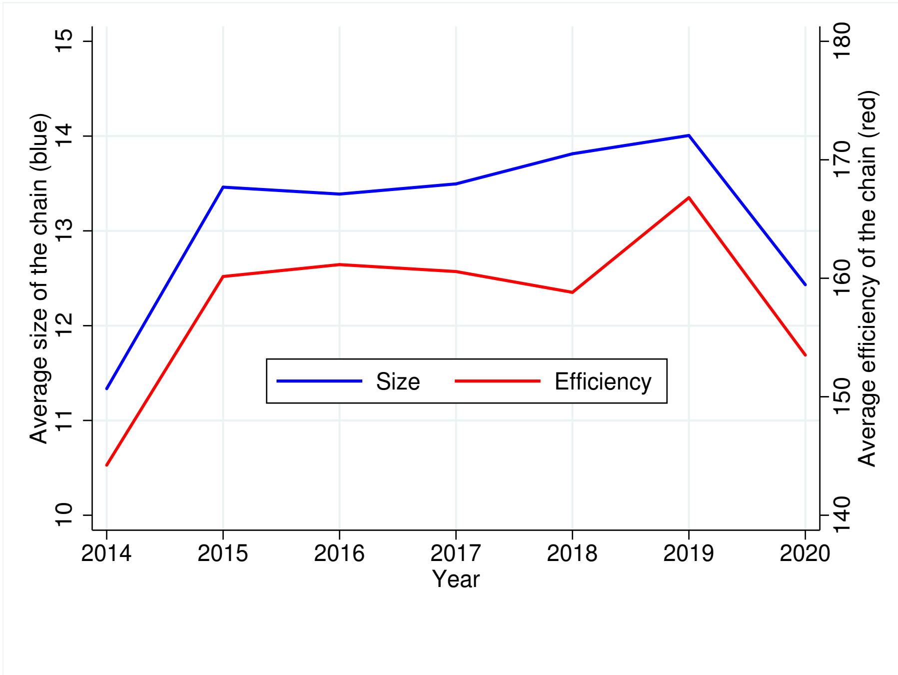


Figure 6: Yearly evolution of the average size and efficiency of a GVC

riences an almost identical decline in 2020, returning to levels observed in 2014. Conversely, the average risk of the chain has increased slightly since 2014, indicating that Mexican firms are now more inclined to source from locations with less positively synchronized business cycles. We also show these trends but only for firms sourcing from more than one location in Figures B.3 and B.4.

Our mechanism rests on the assumption that both the efficiency and risk of a supply chain are positively correlated with the degree of geographic dispersion. In Figure 8 we show that the interaction between the number of countries participating in a firm's GVC and the average geographic dispersion of these locations is positively correlated with the firm's volume of exports. This relationship, without claiming any causation, suggests that there might be benefits from forming GVCs with countries that are far away from each other. As we showed in Figure 3, countries that are far away tend to have more different comparative advantage patterns; forming GVCs with far away countries might therefore bring larger comparative advantage gains for the firm.

Notice that only looking at this relationship with the dispersion without interacting it with the number of locations would not make economic sense. We understand that the comparison of geographic dispersion of GVCs is logical between firms with the same number of locations participating in their respective GVCs. Think about the following example. Following the construction of our dispersion variable in equation (7) the average dispersion

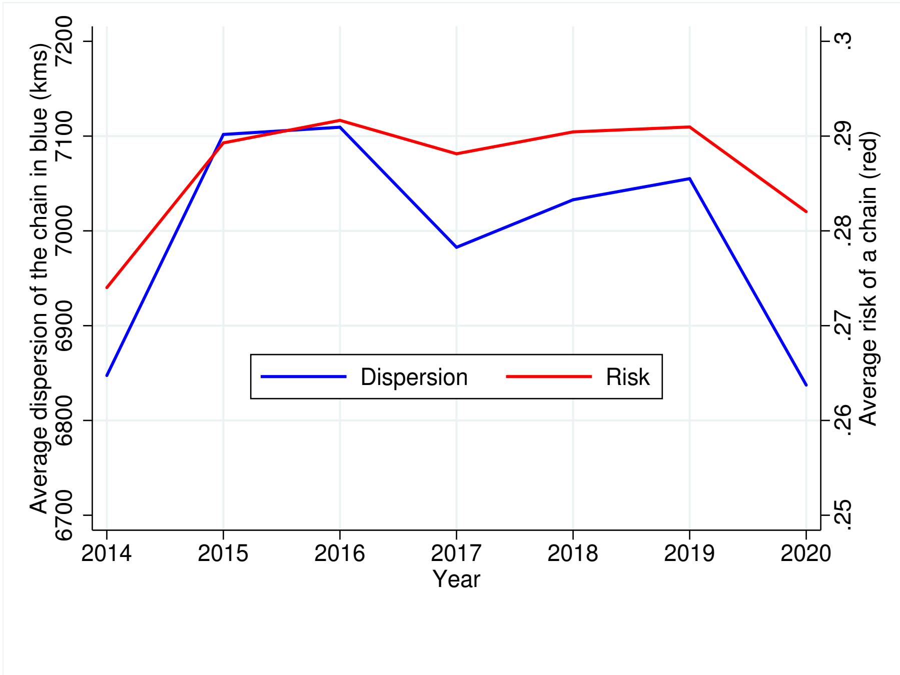
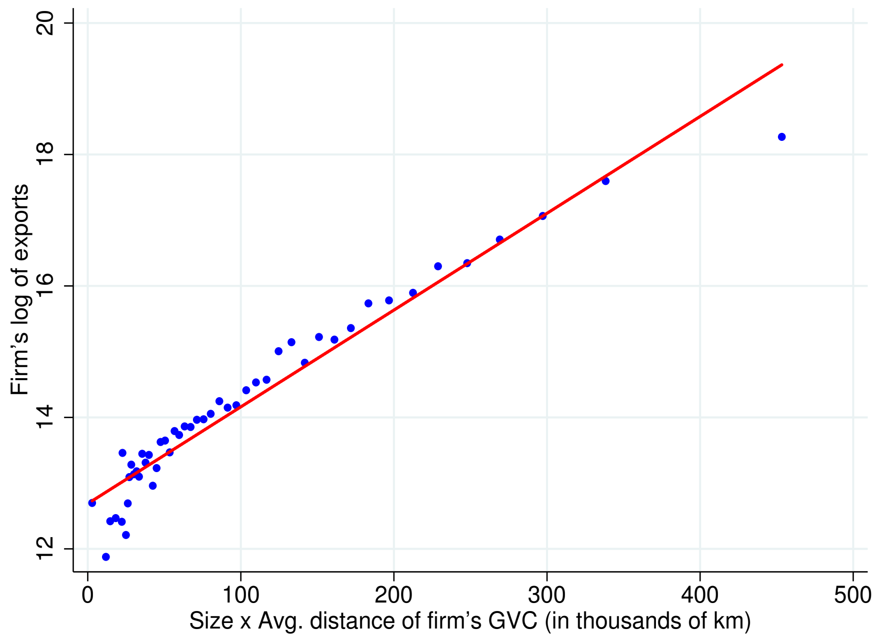


Figure 7: Yearly evolution of the average dispersion and risk of a GVC

measure of a firm sourcing only from Germany and Colombia would be of 9,505.21 kilometers since this is the distance between the two capitals. However, if the firm adds one more location to the chain, let's say Canada, now the average dispersion measure would be of 6,725.38 kilometers since it corresponds to the average of the distance between Germany and Colombia (9,505.21), Germany and Canada (6,134.98) and Canada and Colombia (4,539.95). However, we understand that now the dispersion of the GVC has increased rather than decreased as our variable alone would suggest. However, the interaction term would have increased from $2 \times 9{,}505.21 = 19{,}010.42$ to $3 \times 9{,}505.21 = 20{,}176.14$.

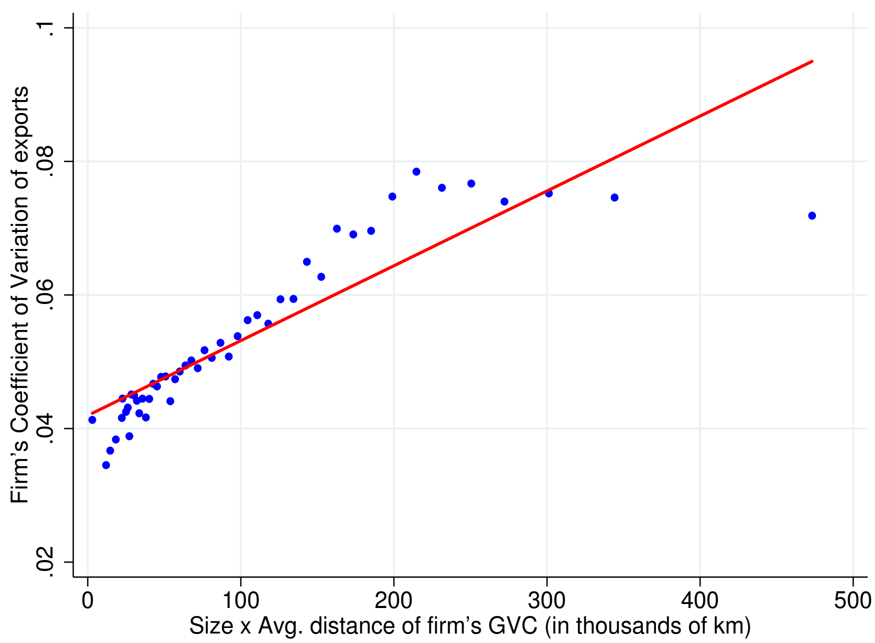
In Figure 9 we show that firms with more dispersed GVCs also tend to experience higher coefficients of variation in their volume of exports. This relationship suggests the presence of a downside from forming GVCs with countries that are far away from each other. As we also showed in Figure 4, countries that are far away tend to experience less synchronous shocks and therefore might impose higher volatility on firms' exports.

These figures suggest that firms are indeed facing a trade-off when choosing the degree of geographical dispersion in their supply chains. Firms with more geographically dispersed supply chains should have greater production efficiency since their suppliers exhibit more distinct patterns of comparative advantage. However, these firms should also face a higher degree of risk since their suppliers tend to have lower synchronization in terms of the cyclical component of GDP.



Note: Binscatter using 50 bins. Each bin represents a partition of the *x-axis* variable.

Figure 8: Dispersion of the GVC and firm performance



Note: Binscatter using 50 bins. Each bin represents a partition of the *x-axis* variable.

Figure 9: Dispersion of the GVC and firm volatility in exports

Nevertheless, as outlined in Section 2, GVC dispersion affects both efficiency and volatility, but this effect is contingent upon production relying on complementary inputs. In cases where production does not rely on such inputs, this trade-off should not be evident, or at least not through the mechanism we propose. To measure the complexity of output production, we construct a proxy using the first two digits of the Harmonized System (HS) code. Higher HS digits correspond to industries requiring a greater variety of complementary inputs, such as machinery and transportation, while lower HS digits correspond to industries producing animal and food products. For each firm, we select the HS code corresponding to the input of highest value in that year. We then conduct Ordinary Least Squares (OLS) regressions to analyze the following:

$$y_{ft} = \beta_0 + \beta_1 R_{ft} + \beta_2 \text{HS}_{ft} + \beta_3 R_{ft} \times \text{HS}_{ft} + \epsilon_{ft}, \tag{8}$$

where y_{ft} is our main firm performance variable of interest such as the logarithm of total exports or the coefficient of variation of exports and R_{ft} is our main risk variable that the firm is exposed to via its GVC dispersion as constructed in Equations (4), (5) and (6).

	Log of exports		Coef. of variation of exports	
	(1)	(2)	(3)	(4)
R_{ft}	-7.464***	-0.552	-75.039***	-21.772
	(0.754)	(0.590)	(9.148)	(12.341)
HS_{ft}	-0.091***	-0.011***	-0.524***	-0.328***
	(0.004)	(0.003)	(0.051)	(0.067)
$R_{ft} \times \text{HS}_{ft}$	0.278***	0.047***	2.197***	0.914***
	(0.013)	(0.010)	(0.164)	(0.214)
Clustered SE firm	No	Yes	No	Yes
Obs.	25,443	25,443	25,422	25,422
R-squared	0.048	0.035	0.019	0.007

Clustered standard errors (firm level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: GVC risk and firm performance

Table 5 reveals that firms with GVCs that include countries with less synchronized (or more negatively synchronized) business cycles exhibit higher export volumes and greater coefficients of variation, but only within industries of greater complexity (higher HS digits).

In contrast, this trade-off is absent in less complex industries (lower HS digits). Specifically, we find that $\hat{\beta}_3 > 0$, indicating that both interaction terms are positive. Columns (1) and (2) present results with the natural logarithm of exports as the dependent variable, while columns (3) and (4) use the coefficient of variation of exports. We employ the coefficient of variation instead of volatility to account for the tendency of firms exporting more to also have higher volatility levels. Columns (1) and (3) represent cross-sectional pool results, whereas columns (2) and (4) present panel regressions with standard errors clustered at the firm level.

The economic rationale behind these coefficients is that firms with GVCs linked to countries experiencing less positively synchronized business cycles (i.e., higher R_{ft}) tend to export more. However, they also exhibit higher coefficients of variation in their exports, particularly in industries with higher HS codes. This phenomenon represents a form of mean-variance trade-off specific to high HS industries. We attribute this to the heavier reliance on complementary inputs in the production processes of firms exporting products with higher HS digits compared to those with lower HS digits.

These findings demonstrate both statistical significance and economic importance. For instance, in a high HS code industry such as vehicle transportation, firms with a 1-standard-deviation higher risk in their global value chain (GVC)—indicating sourcing from less synchronized countries—export, on average, 161.3% more value than comparable firms in the tobacco industry. However, their coefficient of variation is also 11.07 points higher, representing 21% of its mean.

Analyzing the panel regression results, a firm that adjusts its sourcing strategy to increase its exposed risk by 1 standard deviation, and operates in the vehicle transportation industry, would, on average, increase its total exports by 24.3% more than if it operated in the tobacco industry. Additionally, its coefficient of variation of exports would rise by 4.6 points more, constituting a 9.6% increase from the within-firm mean.

From Table 5 we also notice that our proposed mean-variance trade-off of forming a GVC is only present in high HS industries, which we hypothesize are more complex and rely more heavily on complementary inputs. The estimation of coefficient $\hat{\beta}_1 < 0$ for the variable risk alone suggests that having countries within a firm's GVC with less synchronized business cycles should not pose this mean-variance tradeoff if the industry is sourcing mainly substitute inputs. In Table B1 we show that the main interaction term of $\hat{\beta}_3$ remains positive and statistically significant when we add the variable size of the chain and all the possible interaction terms.

We adopt an alternative approach to explore the increased vulnerability to supply chain disruptions for firms with more risky GVCs. In addition to the regression model in Equation (8), we introduce the variable COVID, a binary variable that equals 1 for the 2020 year and 0 otherwise. We focus on the interaction term represented by α_5, which captures the performance of firms with more vulnerable supply chains—those formed with countries characterized by less synchronized business cycles. The regression model is as follows:

$$y_{ft} = \alpha_0 + \alpha_1 R_{ft} + \alpha_2 \mathrm{HS}_{ft} + \alpha_3 R_{ft} \times \mathrm{HS}_{ft} + \alpha_4 \mathrm{COVID}_t + \alpha_5 \mathrm{COVID}_t \times R_{ft} + \epsilon_{ft}. \quad (9)$$

Table 6 illustrates that our estimation for α_5 aligns with the rationale proposed in this study. The negative value of $\hat{\alpha}_5$ indicates that during the COVID-19 year, firms with riskier GVCs—those sourcing from more desynchronized locations—experienced the greatest decline in export value. Firms with a 1-standard-deviation higher risk in their global value chain (GVC) exported, on average, 62.1% less value during 2020. From the panel regression results in column (2), a firm that adjusted its sourcing strategy to increase its risk exposure by 1 standard deviation decreased its total exports by 18.9% on average during 2020.

	Log of exports	
	(1)	(2)
R_{ft}	-7.124***	-0.331
	(0.768)	(0.587)
HS_{ft}	-0.091***	-0.011***
	(0.004)	(0.003)
$R_{ft} \times \mathrm{HS}_{ft}$	0.277***	0.045***
	(0.013)	(0.010)
COVID_t	0.520**	-0.218*
	(0.254)	(0.128)
$\mathrm{COVID}_t \times R_{ft}$	-2.173***	-1.212***
	(0.821)	(0.397)
Clustered SE firm	No	Yes
Obs.	25,443	25,443
R-squared	0.049	0.020

Clustered standard errors (firm level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: GVC risk, firm performance, and COVID-19

4 A Model of Global Sourcing

In our model, firms choose the set of countries from which they can source their intermediate inputs by weighing the trade-off between sourcing efficiency and production risk. We adopt the workhorse model of global sourcing in Antras et al. (2017) and introduce a geographical pattern of comparative advantage by following Lind and Ramondo (2023) in assuming that unit labor costs are correlated within geographical regions. To introduce production risk, we assume that input deliveries are subject to idiosyncratic shocks, which are also correlated with geographic dispersion across countries. We are interested in the formation of supply chains, so for simplicity, we assume firms only sell in the home country while possibly importing their intermediate inputs from foreign countries.

4.1 Household Preferences

Households derive utility from consuming differentiated varieties of manufactured goods according to a CES utility function, which functions as a composite of all the varieties $\omega \in \Omega$ available in the home country.

$$U = \left(\int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \tag{10}$$

where $\sigma > 1$ represents the elasticity of substitution between varieties. This functional form results in the following demand function for variety ω:

$$q(\omega) = EP^{\sigma-1} p(\omega)^{-\sigma} \tag{11}$$

where E refers to the share of expenditure in manufactured varieties, P is the Dixit-Stiglitz price index for the manufacturing sector, and $p(\omega)$ is the price of variety ω.

4.2 Sourcing of Intermediate Inputs

We assume firms in the home country produce a single differentiated variety according to monopolistic competition. Firms have an idiosyncratic productivity φ drawn from a Pareto distribution as in Melitz (2003), which impacts their cost of producing their variety. Production requires a continuum of intermediate inputs, and for simplicity we assume its measure to be equal to one. Intermediate inputs are imperfect substitutes, with a constant and symmetric elasticity of substitution equal to ρ. We denote by $a_j(\nu)$ the amount of labor required for the production of one unit of intermediate input ν at sourcing location j.

As in Eaton and Kortum (2002), the price a firm pays for a given intermediate input is the cheapest price available across the sourcing locations from which a firm can source intermediate inputs. We impose the assumption that suppliers of intermediates are competitive, and thus price at marginal cost. Let $\mathcal{J}_i(\varphi)$ represent the set of countries from which the firm can source intermediate inputs, then the price a firm pays for input ν is:

$$z_i(\nu, \varphi; \mathcal{J}_i(\varphi)) = \min_{j \in \mathcal{J}_i(\varphi)} \{\tau_{ij} a_j(\nu, \varphi) w_j\} \tag{12}$$

where τ_{ij} is the iceberg-type trade cost between the home country and sourcing location j, and w_j is the wage per unit of labor at j. Firms aggregate their purchases of intermediate inputs according to a CES production function and thus, the marginal cost function of the firm is given by:

$$c_i(\varphi) = \frac{1}{\varphi}\Big(\int_0^1 z_i(\nu, \varphi; \mathcal{J}_i(\varphi))^{1-\rho} d\nu\Big)^{\frac{1}{1-\rho}} \tag{13}$$

which depends negatively on firm productivity φ.

4.3 Comparative Advantage

We follow Lind and Ramondo (2023) in assuming unit-labor costs are correlated across sourcing locations according to their geography. Sourcing locations $1, ..., J$ draw their unit-labor cost of producing input ν following a multivariate Fréchet distribution:

$$\Pr(a_1(\nu) \geq a_1, ..., a_J(\nu) \geq a_J) = e^{-G(a_1,...,a_J)} \tag{14}$$

where

$$G(a_1, ..., a_J) = \sum_{k=1}^{K} \Big[\sum_{i=1}^{J} \sum_{j=1}^{J} (\omega_{kij} T_j a_j^{\theta})^{\frac{1}{1-\rho_k}} \Big]^{1-\rho_k} \tag{15}$$

This specification allows us to introduce draws for unit-labor costs correlated within geographical locations. There are K geographical regions and J locations, and ω_{kij} is an indicator function that gives a value of 1 if the pair of locations i, j belongs to the same geographical region k and 0 otherwise. Finally, parameter ρ_k governs the correlation in unit-labor costs within geographical region k. A high ρ_k implies that unit-labor costs are highly correlated within a region. On the contrary, as ρ_k approaches 0, unit-labor costs become close to independent within region k.

This distributional assumption for unit-labor costs captures the benefits of having a geographically dispersed supply chain through the comparative advantage channel. Supply chains in which sourcing locations have less correlated unit-labor costs will result in a lower marginal cost of production, intuitively, as production will be cheaper due to a wider range of inputs being cheaper than if the firm were to source them from one region exclusively. We show a more detailed explanation using the main equations and derivations of the model in Appendix A.1.

4.4 Production Risk

We introduce production risk by assuming that deliveries of intermediate inputs are subject to idiosyncratic shocks, and these shocks are correlated according to geography. Specifically, the probability of an intermediate input being delivered on time depends on the country of origin and its geographical and economic similarity to other sourcing countries.

As in the standard model by Melitz (2003), we assume that firms take as given the demand function in Equation (11). For each unit of its variety that the firm wants to produce, it must source one unit of each intermediate input ν. Therefore, the demand faced by the firm for its final product also determines how many units of each intermediate input the firm seeks to purchase from a given sourcing location. Under this assumption, purchases of input ν are given by:

$$q^S(\nu) = q(\omega)\big(1 - \xi(\nu)\big), \tag{16}$$

where $\xi(\nu)$ is the delivery shock experienced by the cheapest supplier of input ν. The presence of these shocks implies that the firm will always under-serve the local market, with the degree of excess demand being given by the size of these shocks to deliveries of intermediates. Production of the final good is given by:

$$\begin{aligned} q^S &= \Big(\int_0^1 [EP^{\sigma-1}p^{-\sigma}\big(1 - \xi(\nu)\big)]^{1-\rho} d\nu\Big)^{\frac{1}{1-\rho}} \\ &= EP^{\sigma-1}p^{-\sigma}\Big(\int_0^1 \big(1 - \xi(\nu)\big)^{1-\rho} d\nu\Big)^{\frac{1}{1-\rho}} \\ &= EP^{\sigma-1}p^{-\sigma}\Phi(\mathcal{J}_i(\varphi)) \end{aligned} \tag{17}$$

where $\Phi(\mathcal{J}_i(\varphi)) = \left(\int_0^1 \left(1 - \xi(\nu)\right)^{1-\rho} d\nu \right)^{\frac{1}{1-\rho}}$ is a function of a firm's sourcing strategy. It can be thought of as the average shock to deliveries, which itself depends on the set of countries from which a firm sources intermediate inputs and the degree of similarity across them in terms of these shocks.

We model the shocks $\xi(\nu)$ as following a multivariate normal distribution where the shocks are correlated across locations. Specifically, let $\xi(\nu) \sim \mathcal{N}(\mu, \Sigma)$, where μ is the vector of means and Σ is the covariance matrix. The elements of Σ are determined by the geographical and economic distances between countries. More formally, we can define:

$$\Sigma_{ij} = \sigma_i \sigma_j \rho_{ij}, \tag{18}$$

where σ_i and σ_j are the standard deviations of the shocks in countries i and j, and ρ_{ij} is the correlation coefficient between the shocks in these countries. The correlation coefficient ρ_{ij} can be modeled as decreasing with the geographical distance d_{ij} between countries i and j:

$$\rho_{ij} = \exp(-\kappa d_{ij}), \tag{19}$$

where κ is a parameter that captures the rate at which the correlation decreases with distance. This implies that more geographically dispersed locations experience less synchronized shocks.

For example, if a firm sources inputs from two countries, A and B, the shocks ξ_A and ξ_B might have the following joint distribution:

$$\begin{pmatrix} \xi_A \\ \xi_B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \sigma_A \sigma_B \exp(-\kappa d_{AB}) \\ \sigma_A \sigma_B \exp(-\kappa d_{AB}) & \sigma_B^2 \end{pmatrix} \right). \tag{20}$$

In this setup, the average shock to deliveries $\Phi(\mathcal{J}_i(\varphi))$ is a function of the set of countries from which the firm sources intermediate inputs and the correlation structure of the shocks. The firm's sourcing strategy will thus affect the average level of production risk, with more geographically dispersed sourcing strategies leading to less synchronized shocks and higher production risk. This captures the tradeoff between exploiting comparative advantage and managing production risk in the formation of global value chains. We provide a more detailed explanation in Appendix A.2.

4.5 Firm Profits and Choice of a Supply Chain

Firms choose the set of countries from which they can source their intermediate inputs, i.e., their global supply chain, in order to maximize their expected profits as we assume firms are unable to ex-ante observe the realization of shocks to input deliveries across the different sourcing locations. A final-good producer in the home country can acquire the capability to supply inputs from country j only after incurring a fixed cost equal to f_j units of labor.

Expected variable profits given a certain supply chain are given by:

$$\begin{aligned}\mathbf{E}[\pi(\varphi)] &= (p(\varphi) - c(\varphi))q^S \\ &= \left[\frac{\sigma}{\sigma-1}c(\varphi) - c(\varphi)\right] EP^{\sigma-1}p^{-\sigma}\mathbf{E}[\Phi(\mathcal{J}(\varphi))] \\ &= \frac{1}{\sigma-1}c(\varphi)EP^{\sigma-1}p^{-\sigma}\mathbf{E}[\Phi(\mathcal{J}(\varphi))] \end{aligned} \tag{21}$$

Equation (21) captures the two forces present in our main trade-off. On the one hand, choosing a more geographically dispersed supply chain will decrease the marginal cost of producing $c_i(\varphi)$ as firms can source a larger share of intermediate inputs from possibly cheaper suppliers. On the other, under the appropriate distributional assumption on delivery shocks, object $\mathbf{E}[\Phi(\mathcal{J}_i(\varphi))]$ is decreasing in the level of geographical dispersion of the supply chain due to the concavity inherited by the CES production function assumption.

Let $2^{\mathcal{J}}$ represent the power set of the set of countries $\mathcal{J} = \{1, \ldots, J\}$, i.e., the set of all possible supply chains. A firm's supply chain $\mathcal{J}^*(\varphi)$ is the solution to the following maximization problem:

$$\mathcal{J}^*(\varphi) = \underset{\mathcal{J}\in 2^{\mathcal{J}}}{\arg\max}\left[\frac{1}{\sigma-1}c(\varphi)EP^{\sigma-1}p^{-\sigma}\mathbf{E}[\Phi(\mathcal{J}(\varphi))] - w\sum_{j\in\mathcal{J}} f_j\right] \tag{22}$$

That is, a firm chooses its supply chain, i.e., the set of countries from which it can source intermediate inputs, to maximize its expected profits, taking into account the trade-off between a cheaper marginal cost of production and production risk.

Extending the model to deliver policy-relevant counterfactuals requires a numerical solution together with an estimation of the key parameters governing the trade-off described above. We leave this for future work.

5 Conclusions

This paper makes several novel contributions to the literature on global value chains (GVCs) by proposing a mean-variance trade-off in GVC formation using geographical dispersion as the key variable. We argue that firms face a trade-off between exploiting comparative advantage by sourcing inputs from distant locations and minimizing the risk of supply chain disruptions due to asynchronous business cycles. Our empirical analysis, using both country-level data and detailed firm-level customs data from Mexico, provides evidence supporting this trade-off.

We also present new empirical facts regarding recent trends in GVCs, highlighting how, before COVID-19, firms were increasingly sourcing inputs from a larger number of locations and from locations that were more geographically dispersed, thereby enhancing their comparative advantage but also exposing themselves to higher risks of disruption. These trends reversed sharply in 2020 in response to the trade collapse driven by global lockdowns and restrictions.

For future research, a more refined proxy for sectoral heterogeneity and complementarity is essential. The current use of the HS code, while useful, may not fully capture the degree of complementarity in input requirements and production complexities.

We also propose numerically solving the model to explore the effects of varying key parameters that drive the tradeoffs in GVC formation. By employing the method of moments, we can estimate these parameters more accurately and analyze their impact on the mean-variance tradeoff. This approach would allow us to simulate different scenarios and better understand the conditions under which firms optimize their GVC strategies, providing deeper insights into the policy implications and potential strategies to mitigate supply chain risks.

References

P. Antràs. Conceptual aspects of global value chains. *The World Bank Economic Review*, 34(3):551–574, 2020.

P. Antràs and A. De Gortari. On the geography of global value chains. *Econometrica*, 88 (4):1553–1598, 2020.

P. Antras, T. C. Fort, and F. Tintelnot. The margins of global sourcing: Theory and evidence from us firms. *American Economic Review*, 107(9):2514–64, 2017.

B. Balassa. Trade liberalisation and "revealed" comparative advantage 1. *The manchester school*, 33(2):99–123, 1965.

B. Bonadio, Z. Huo, A. A. Levchenko, and N. Pandalai-Nayar. Global supply chains in the pandemic. *Journal of international economics*, 133:103534, 2021.

A. Costinot. On the origins of comparative advantage. *Journal of International Economics*, 77(2):255–264, 2009.

A. Costinot, J. Vogel, and S. Wang. An elementary theory of global supply chains. *Review of Economic studies*, 80(1):109–144, 2013.

A. De Gortari. Disentangling global value chains. Technical report, National Bureau of Economic Research, 2019.

A. de Gortari. Global value chains and increasing returns. Technical report, Working Paper, Dartmouth College, 2020.

J. Eaton and S. Kortum. Technology, geography, and trade. *Econometrica*, 70(5):1741–1779, 2002.

P. Eppinger, G. J. Felbermayr, O. Krebs, and B. Kukharskyy. Covid-19 shocking global value chains. 2020.

S. J. Evenett. 13 what's next for protectionism? watch out for state largesse, especially export incentives. *COVID-19 and trade policy: Why turning inward won't work*, page 179, 2020.

K. Hayakawa and H. Mukunoki. Impacts of covid-19 on global value chains. *The Developing Economies*, 59(2):154–177, 2021.

R. J. Hodrick and E. C. Prescott. Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16, 1997.

N. Lind and N. Ramondo. Trade with correlation. *American Economic Review*, 113(2): 317–353, 2023.

M. J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, 71(6):1695–1725, 2003.

S. Miroudot. Resilience versus robustness in global value chains: Some policy implications. *COVID-19 and trade policy: Why turning inward won't work*, pages 117–130, 2020.

Z. Wang, S.-J. Wei, X. Yu, and K. Zhu. Characterizing global value chains: production length and upstreamness. Technical report, National Bureau of Economic Research, 2017.

A Mathematical Appendix

A.1 Comparative advantage and ρ_k

The expected value of the minimum cost can be derived using the properties of the Fréchet distribution. For simplicity, consider the case where the Fréchet scale parameters T_j are the same across locations. Then, the expected value of the minimum cost can be approximated as follows:

$$\mathbf{E}[\min_{j \in \mathcal{J}_i(\varphi)} \{\tau_{ij} a_j(v, \varphi) w_j\}] \approx \left(\sum_{j \in \mathcal{J}_i(\varphi)} (\tau_{ij} w_j)^{-\theta} \right)^{-\frac{1}{\theta}} \tag{23}$$

Equation (23) indicates that the expected minimum cost decreases as the number of sourcing locations $|\mathcal{J}_i(\varphi)|$ increases.

To simplify notation, let:

$$Z(\mathcal{J}_i(\varphi)) \equiv \left(\sum_{j \in \mathcal{J}_i(\varphi)} (\tau_{ij} w_j)^{-\theta} \right)^{-1/\theta}$$

The marginal cost function can then be written as:

$$c_i(\varphi) = \frac{1}{\varphi} \left(\int_0^1 Z(\mathcal{J}_i(\varphi))^{1-\rho} d\nu \right)^{\frac{1}{1-\rho}}$$

Assuming $Z(\mathcal{J}_i(\varphi))$ is independent of ν:

$$c_i(\varphi) = \frac{1}{\varphi} Z(\mathcal{J}_i(\varphi)) \tag{24}$$

Taking the derivative of $c_i(\varphi)$ with respect to the number of sourcing locations $|\mathcal{J}_i(\varphi)|$:

$$\frac{\partial c_i(\varphi)}{\partial |\mathcal{J}_i(\varphi)|} = \frac{1}{\varphi} \frac{\partial Z(\mathcal{J}_i(\varphi))}{\partial |\mathcal{J}_i(\varphi)|}$$

Given $Z(\mathcal{J}_i(\varphi)) = \left(\sum_{j \in \mathcal{J}_i(\varphi)} (\tau_{ij} w_j)^{-\theta} \right)^{-1/\theta}$, the derivative with respect to $|\mathcal{J}_i(\varphi)|$ is:

$$\frac{\partial Z(\mathcal{J}_i(\varphi))}{\partial |\mathcal{J}_i(\varphi)|} = \frac{\partial}{\partial |\mathcal{J}_i(\varphi)|} \left(\sum_{j \in \mathcal{J}_i(\varphi)} (\tau_{ij} w_j)^{-\theta} \right)^{-1/\theta}$$

Applying the chain rule, we get:

$$\frac{\partial Z(\mathcal{J}_i(\varphi))}{\partial|\mathcal{J}_i(\varphi)|} = -\frac{1}{\theta}\left(\sum_{j\in\mathcal{J}_i(\varphi)}(\tau_{ij}w_j)^{-\theta}\right)^{-(1+\theta)/\theta} \cdot \sum_{j\in\mathcal{J}_i(\varphi)}\frac{\partial(\tau_{ij}w_j)^{-\theta}}{\partial|\mathcal{J}_i(\varphi)|} \tag{25}$$

Since adding an additional country to the sourcing set reduces the overall cost:

$$\sum_{j\in\mathcal{J}_i(\varphi)}\frac{\partial(\tau_{ij}w_j)^{-\theta}}{\partial|\mathcal{J}_i(\varphi)|} > 0$$

Therefore, the derivative of the marginal cost with respect to the number of sourcing locations is negative:

$$\frac{\partial c_i(\varphi)}{\partial|\mathcal{J}_i(\varphi)|} < 0$$

The parameter ρ_k affects the correlation of unit labor costs within a region. Lower ρ_k values imply less correlation, enhancing the comparative advantage gains. Thus, the derivative's magnitude depends on ρ_k:

$$\frac{\partial c_i(\varphi)}{\partial|\mathcal{J}_i(\varphi)|} \propto -\left(\sum_{k=1}^{K}\rho_k\left(\sum_{j\in\mathcal{J}_i(\varphi)}(\tau_{ij}w_j)^{-\theta}\right)^{\frac{1}{1-\rho_k}}\right)^{-1} \tag{26}$$

This indicates that lower ρ_k values (less correlation) lead to a larger decrease in marginal cost as the number of sourcing locations increases, highlighting the comparative advantage gains of constructing a geographically dispersed GVC.

A.2 Production risk and ρ_k

The correlation structure affects the synchronization of shocks. More geographically dispersed sourcing strategies lead to less synchronized shocks. We start by looking at the covariance structure and how it affects $\Phi(\mathcal{J}_i(\varphi))$.

A higher correlation ρ_{ij} means that shocks are more synchronized, which raises the variance of the aggregate delivery shock faced by the chain. Conversely, lower correlation (which occurs with greater geographical dispersion) leads to less synchronized shocks across sourcing locations.

Let $\mathbf{E}[\Phi(\mathcal{J}_i(\varphi))]$ denote the expected value of the average shock to deliveries. The correlation ρ_{ij} affects this expectation:

$$\mathbf{E}[\Phi(\mathcal{J}_i(\varphi))] \propto \left(\int_0^1 \mathbf{E}[(1-\xi(\nu))^{1-\rho}]d\nu\right)^{\frac{1}{1-\rho}}$$

Using the properties of the normal distribution and its moments, we can approximate the expectation of $(1-\xi(\nu))^{1-\rho}$:

$$\mathbf{E}[(1-\xi(\nu))^{1-\rho}] = \exp\left((1-\rho)\mu - \frac{(1-\rho)^2}{2}\sigma^2\right)$$

Here, μ is the mean of the shocks, and σ^2 is the variance of the shocks, which depends on the correlation structure.

To capture the effect of sourcing potential on $\mathbf{E}[\Phi(\mathcal{J}_i(\varphi))]$, we take the derivative with respect to the number of sourcing locations $|\mathcal{J}_i(\varphi)|$:

$$\frac{\partial \mathbf{E}[\Phi(\mathcal{J}_i(\varphi))]}{\partial|\mathcal{J}_i(\varphi)|}$$

Given that more dispersed locations lead to less synchronized shocks, the correlation ρ_{ij} decreases as $|\mathcal{J}_i(\varphi)|$ increases. This effect can be captured as follows:

$$\frac{\partial \mathbf{E}[\Phi(\mathcal{J}_i(\varphi))]}{\partial|\mathcal{J}_i(\varphi)|} \propto -\left(\int_0^1 \mathbf{E}[(1-\xi(\nu))^{1-\rho}]d\nu\right)^{\frac{1}{1-\rho}} \cdot \frac{\partial \rho_{ij}}{\partial|\mathcal{J}_i(\varphi)|}$$

Since $\rho_{ij} = \exp(-\kappa d_{ij})$ and increases in geographical dispersion d_{ij} lead to decreases in ρ_{ij}:

$$\frac{\partial \rho_{ij}}{\partial|\mathcal{J}_i(\varphi)|} < 0$$

Thus, the derivative of the expected average shock with respect to the number of sourcing locations is positive, indicating higher average shocks with more dispersion:

$$\frac{\partial \mathbf{E}[\Phi(\mathcal{J}_i(\varphi))]}{\partial |\mathcal{J}_i(\varphi)|} > 0$$

Combining these results, we see that the expected average shock to deliveries $\Phi(\mathcal{J}_i(\varphi))$ increases as the firm's sourcing potential becomes more geographically dispersed. This reflects the negative effect of less synchronized shocks due to increased geographical dispersion.

In summary, while geographical dispersion can reduce production costs through comparative advantage (positive effect), it can simultaneously increase production risk due to less synchronized shocks (negative effect). These trade-offs are crucial in determining the optimal sourcing strategy for firms.

B Additional Figures and Tables

Figures

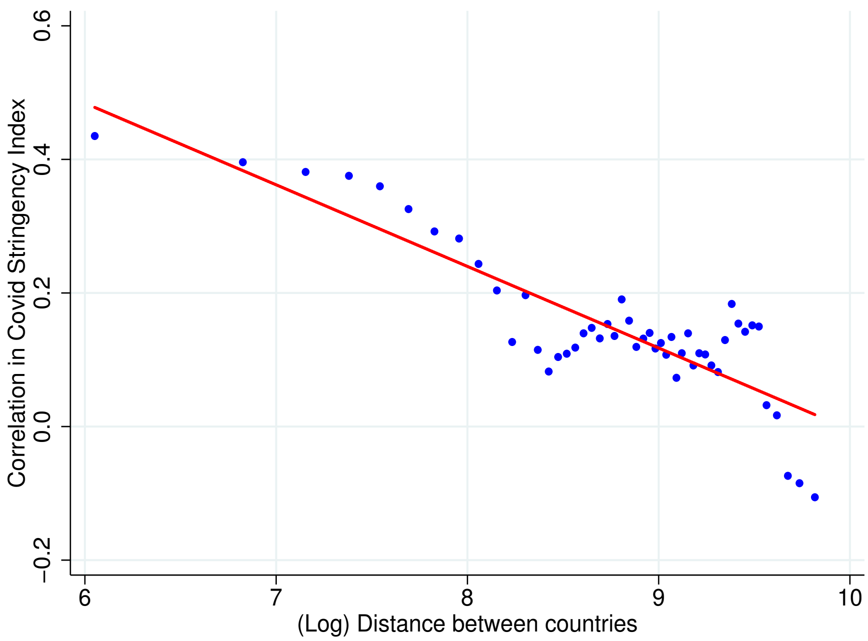


Figure B.1: Geographical distance and COVID Stringency Index correlation

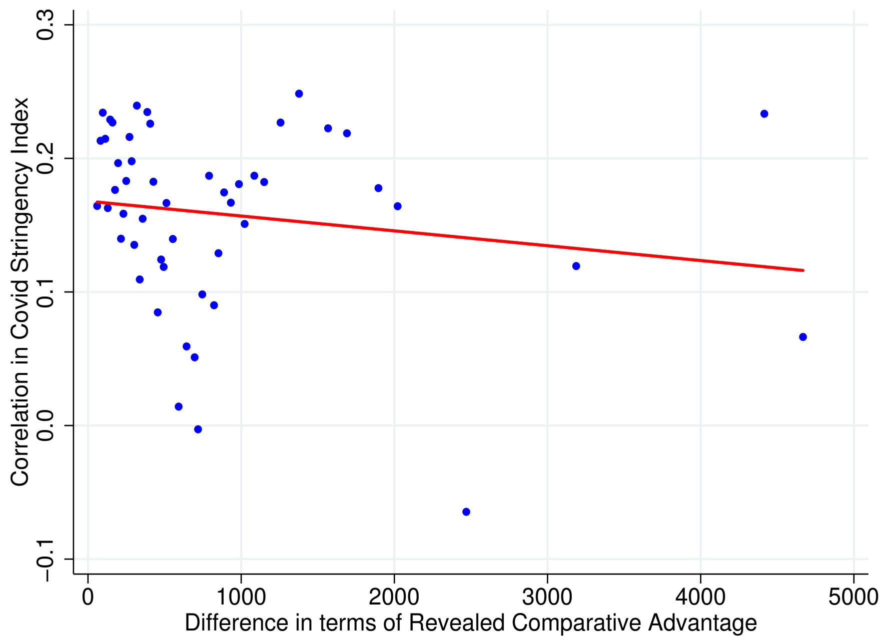


Figure B.2: Comparative Advantage and COVID Stringency Index correlation

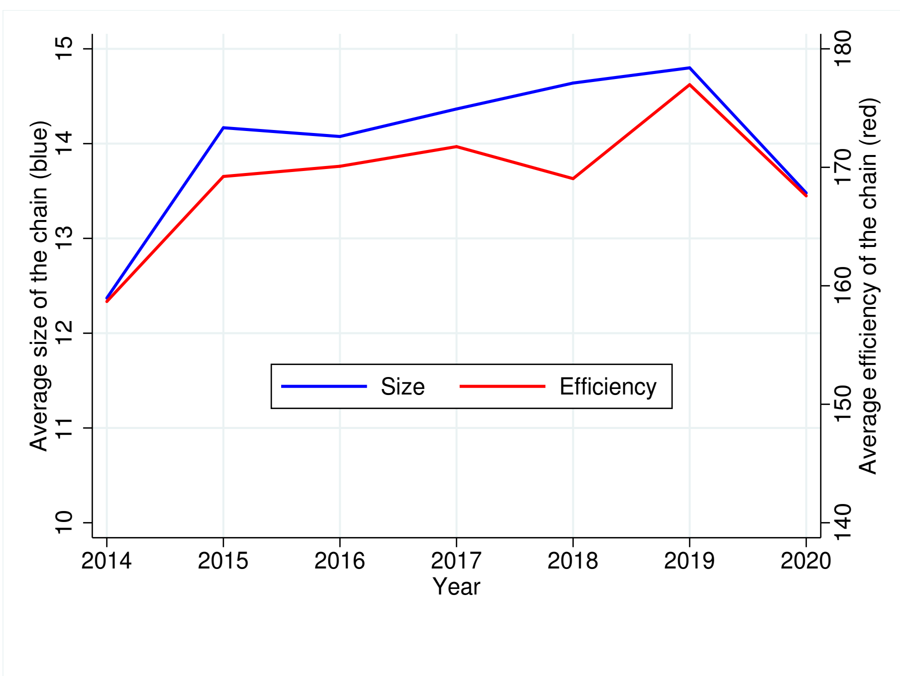


Figure B.3: Yearly evolution of the average size and efficiency of a GVC for more than 1 location

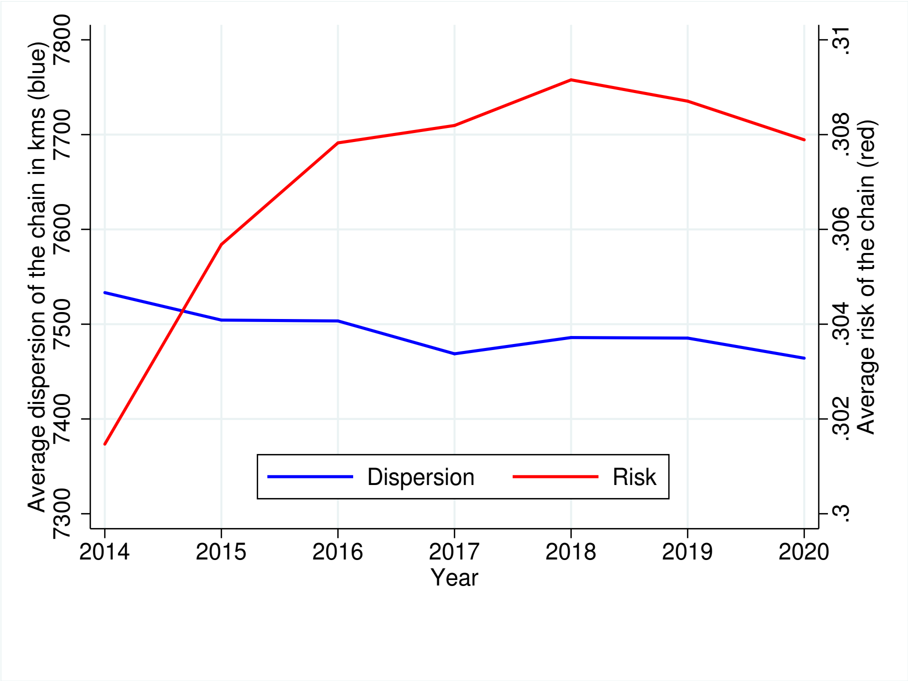


Figure B.4: Yearly evolution of the average dispersion and risk of a GVC for more than 1 location

Tables

	Log of exports		Coef. of variation of exports	
	(1)	(2)	(3)	(4)
R_{ft}	-3.197***	0.191	-30.630***	-4.965
	(0.890)	(0.677)	(11.010)	(15.384)
HS_{ft}	-0.056***	-0.006*	-0.441***	-0.236***
	(0.005)	(0.003)	(0.062)	(0.082)
$R_{ft} \times \mathbf{HS}_{ft}$	**0.103*** **	**0.027** **	**1.067*** **	**0.464***
	(0.015)	**(0.011)**	**(0.191)**	**(0.257)**
Size_{ft}	0.235***	0.223***	-0.937*	2.998***
	(0.039)	(0.028)	(0.521)	(0.725)
$\text{Size}_{ft} \times \text{HS}_{ft}$	-0.001**	0.001	0.052***	0.020**
	(0.001)	(0.000)	(0.007)	(0.010)
$\text{Size}_{ft} \times \text{R}_{ft}$	-0.171	-0.197***	2.312	-2.854
	(0.112)	(0.075)	(1.473)	(1.967)
$\text{Size}_{ft} \times \text{R}_{ft} \times \text{HS}_{ft}$	0.001	-0.002*	-0.110***	-0.036
	(0.001)	(0.001)	(0.021)	(0.027)
Clustered SE firm	No	Yes	No	Yes
Obs.	25,443	25,443	25,422	25,422
R-squared	0.204	0.150	0.072	0.063

Clustered standard errors (firm level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B1: GVC risk and firm performance: Robustness