Supply Chain Risk and the Pattern of Trade

Maximilian Eber and Hannes Malmberg*

November 25, 2015

Abstract

We analyze the interaction of supply chain risk and trade patterns. Our model yields a novel determinant of comparative advantage. In the model, countries with low supply chain risk specialize in risk-sensitive goods. We also show that risk-sensitivity is determined by the number of customized components used in production. Based on our theory, we construct a novel empirical measure of risk-sensitivity from input-output tables and measures of customization (Rauch, 1999). Using industry-level trade data and a variety of risk proxies, we show that countries with low supply chain risk indeed export risk-sensitive goods disproportionately. The model has policy implications: Countries that strive to attract a risk- sensitive industry such as car manufacturing can do so by improving supply chain reliability.

1 Introduction

This paper is motivated by a growing concern in the policy and business community over supply chain risk. The 2011 tsunami in Japan illustrates how important reliability is for modern production. For example, General Motors had to close a factory in Louisiana due to a lack of Japanese-made parts.[1] The Inter-American Development Bank notes that "firms fragmenting production internationally are likely to look for locations with adequate transport and logistics infrastructure to reduce disruptions in the supply chain" (Blyde, 2014). Similarly, the US Department of Commerce argues that "Expected gains from offshoring can often be erased by [...] unexpected delays."[2] In this paper, we study the trade pattern consequences of country differences in supply chain risk.

*The project was supported by the Private Enterprise Development for Low-Income Countries (PEDL) exploratory grant.

[1]Source: http://www.nytimes.com/2011/03/20/business/20supply.html (last accessed Nov 17th, 2015)

[2]http://www.esa.doc.gov/economic-briefings/assess-costs-everywhere-shipping (last accessed Nov 17th, 2015)

It is clear that countries vary in the degree of supply chain risk. Some countries offer high-quality infrastructure and predictable, quick bureaucratic services. In other countries, poor logistics systems and low government effectiveness mean that supply chain risk is high: components get stuck in the port; roads rain away; land rights are not transparent; import permits are delayed; and foreign currency availability is uncertain.

Variation in country-level supply chain risk induces comparative advantage as goods vary in their sensitivity to supply chain risk. Consider two industries, plain t-shirt production and cars: For t-shirts, there are very few separate intermediate inputs used in production. In this case, a risky supply chain is not too problematic. On the other hand, a modern car factory based on lean production principles requires a continuous flow of hundreds if not thousands of customized components. Supply chain reliability becomes key. If a country's infrastructure and institutions create severe supply chain risk, the country can be expected to have a comparative advantage in t-shirt production, and a comparative disadvantage in modern car production.

We formalize this intuition by constructing a model in which each sector produces a good using intermediate inputs. Intermediate input production is subject to *disruption risk,* which means that production (including delivery) fails with some probability. Proximate causes of production failures are delays in ports, infrastructure failures, strikes, political instability, and unpredictable bureaucratic procedures. A key feature of the model is that the effect of disruption risk depends on whether intermediate goods are *standardized* or *customized.* Standardized intermediates are homogenous and traded on centralized exchanges whereas customized intermediates are delivered directly from intermediate goods producers to final goods producers. For standardized goods, the centralized exchange insulates producers from disruption risk through a law of large numbers. There will be a steady supply of goods even if some suppliers fail for idiosyncratic reasons. This is not the case for customized components, which are often produced by only one or two suppliers. In this situation, idiosyncratic risk matters for production.

We derive a novel aggregation result that makes the model highly tractable. We show that aggregated supply and demand of a sector can be characterized by a representative firm with deterministic production, even though the underlying firms experience stochastic shocks. Supply chain risk enters the sectoral production function as productivity penalty. As all customized intermediate inputs are essential for production, this productivity penalty grows exponentially with the number of customized intermediate inputs used in production.[3] As a by-product of the theory, we derive a new measure of different industries' risk-

[3] For simplicity, we are assuming independent shocks to input suppliers.

sensitivity. Because each customized intermediate good represents an independent source of error, the risk-sensitivity of a product depends on the *number of customized components* used in production.

We embed this sectoral production structure in a simple trade model. In the model, we let goods vary in their number of customized inputs m, and countries vary in the disruption risk π. We show that productivity is log-submodular in π and m. Thus, we can use the insight from Costinot (2009a) to conclude that there will be negative sorting between π and m. In other words, risky countries (high π) will produce goods with few customized intermediate inputs (low m).

In the empirical part of this paper, we test this hypothesis in trade data using the methodology in Romalis (2004). In a first step, we construct a measure of how many customized intermediate inputs each industry uses from Input-Output tables and the definition of customized goods by Rauch (1999). To proxy for disruption risk, we use the World Governance Indicator (WGI) of government effectiveness and the World Bank's Logistics Performance Index (LPI). We then test whether countries with more effective government and logistics systems export relatively more goods with highly customized components. We show that this effect exists and is statistically and economically meaningful. The effect is present in a wide range of specifications, even when we (over-) control for country income levels.

The effects we find are of similar magnitude to other institutional determinants of trade patterns, for example contracting quality Nunn (2007). However, our theory builds on a different mechanism. Nunn empasizes that a bad contracting environment leads to a higher cost of customized intermediate inputs via lower levels of relation-specific investments. Therefore, he measures the *proportion in value terms* of inputs that comes from customized intermediate inputs. Our paper focuses on disruption risk and therefore considers the *number* of customized intermediate inputs.

Different perspectives on the sources of comparative advantage also imply a different policy levers. Many countries are actively trying to attract "sophisticated" industries such as advanced electronics or machinery equipment manufacturing. If relationship-specific investments are key—as implied by models such as Antràs (2003) and measured by Nunn (2007)—then the main task for governments is to improve the quality of the contracting environment and the rule of law. If supply chain risk also matters, as suggested by our results, then it is important for governments to improve the reliability of the business environment through, for example, more efficient bureaucracy, infrastructure, and promotion of third party business facilitators.

Our theory also has implications for the measurement of the quality of the business environment. Very few indicators focus on uncertainty and risk. For example, the World Bank's Doing Business Indicators have been used widely to illustrate the challenges for businesses in poor countries. It measures de jure time and cost to complete a wide range of functions such as time to export, import, receive electricity, and open a business. However, it only provides a single estimate per country of the time required for a given task, and it contains little information about the variability of its implementation. We stress the importance of risk and uncertainty in the business environment. Our findings suggest that characterizations of the business environment should include measures of risk. For example, surveys would benefit from reporting not only the average time to obtain a permit, but also the variance associated with the time to obtain such a permit and the chances of not obtaining a permit at all.

The paper proceeds as follows: We discuss related literature in section 2. Section 3 provides the model. We bring the model to the data in section 4. Finally, section 5 concludes.

2 Literature

The paper connects to a number of different literatures. As it analyzes the role of disruption risk, it connects to the literature on macroeconomics and uncertainty. The production structure in which all components are vital for production relates the paper to O-Ring theory. Furthermore, we analyze how institutional features interact with risk to shape countries' trade patterns. Therefore, the paper also adds to the literature on institutional sources of comparative advantage.

Risk and Uncertainty in economics First, the role of risk in shaping macroeconomic dynamics has been studied extensively, with an influential early contribution by Cooper and Haltiwanger (1993). This strand of the literature emphasizes lumpy investments with adjustment costs. The combination of fixed costs and adjustment frictions turn capital spending and hiring decisions into real options, whose values are affected by volatility.[4] Bloom (2009), for example, proposes a model in which time-varying volatility affects both capital investment and hiring decisions. Uncertainty shocks cause investment and hiring freezes with associated declines in economic activity. By contrast, we analyze a multi-good setting in which risk shapes the allocation of production across countries.

O-Ring theory and sequential production Our production process features a number of vital inputs that are necessary for production. Thus, the most closely related model is the O-Ring Theory proposed by

[4]See Bond and Van Reenen (2007) for a survey of the literature.

Kremer (1993). In contrast, however, we focus our attention on the on cross-sector implications as opposed to the labor market. Most importantly, we argue that sectors are distinguished by their use of customized and standardized components. This implies that some sectors are more insulated from uncertain input deliveries than others due to the market mechanism. The distinction is a the heart of our theory and it is important for the resulting pattern of specialization.

One paper that applies O-Ring like mechanisms to trade is Costinot (2009b). He proposes a model of comparative advantage where firms trade off the value of specialization against the risk of disruption when they select team sizes. Disruption comes from poorly enforced contracts, and as the gains from specialization comes from economizing on fixed training costs, his definition of sensitivity is the total training cost in an industry. In contrast, we have a different source of unreliability, and therefore focus on the number of specialized inputs as the source of sensitivity.

More generally, production processes in which all components are vital are related to sequential production, in which goods have to pass through a number of pre-defined steps. Economists have long noted the potential implications of sequential production, and also analyzed trade patterns in the context of sequential production models. Dixit and Grossman (1982) is an early attempt of analyzing the role of sequential production in shaping trade patterns. More recently, Costinot et al. (2012) and Antras and Chor (2013) have proposed novel models of sequential production and used them to interpret sorting along global supply chains.

Institutional sources of comparative advantage Thirdly, we analyze how variations in government effectiveness and logistics systems quality shape trade patterns. This connects the paper to the growing literature on institutional determinants of comparative advantage. Existing work on the institutional determinants of comparative advantage focuses on the role of technological differences (Eaton and Kortum, 2002), factor endowments (Romalis, 2004), contracting quality (Nunn, 2007; Antràs, 2003; Antras and Helpman, 2004), financial development (Manova, 2199), or labor market institutions (Cuñat and Melitz, 2012).

In particular, Levchenko (2007) treats institutional quality as a source of comparative advantage. In the model, he focuses on imperfect contract enforcement in the spirit of Grossman and Hart (1986). By contrast, in our model institutions (among other factors) determine the amount of risk that firms bear, which shapes comparative advantage. Blyde and Molina (2015) provide evidence that foreign direct investment is related to logistics infrastructure. This is in line with our thinking: production of complex goods is difficult when the environment is risky. We contribute to this topic by providing a highly tractable model and showing that logistics quality shapes comparative advantage in trade flows.

3 Model

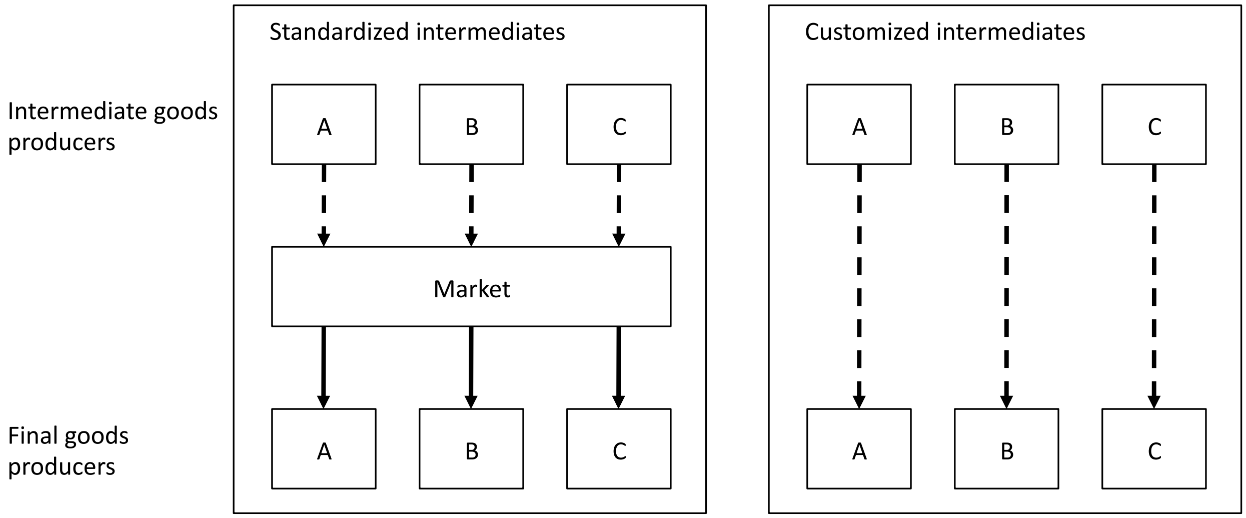
We construct a trade model where intermediate input production is risky. Countries vary in their degree of supply chain risk and goods vary in their risk-sensitivity. This generates specialization across countries according to comparative advantage. We first develop a parsimonius characterization of production with risky inputs. For each sector, we derive a sector-level aggregate production function, which summarizes how supply chain risk and industry characteristics interact to determine sector-level productivity. We then use these sectoral production functions in a trade model to predict how supply chains shape trade patterns.

A sector s consists of a continuum of final goods producers which produce a good using labor and multiple intermediate inputs. The final goods producers combine intermediate inputs using a CES aggregator where inputs are gross complements. Therefore, every input is essential for production. Intermediate inputs are produced using labor, and the production process in the intermediate goods sector is risky. This means that for each intermediate good producer, there is a possibility that production—including delivery—will fail, and failures are independent across different suppliers.

For intermediate inputs, the model makes a distinction between standardized and customized intermediate inputs. Standardized intermediate inputs are traded on a centralized market, and all input producers ship to this market. Idiosyncratic delivery risks average out through the law of large numbers and there is a deterministic flow of products to the centralized market. This means that the final goods producers face no delivery risk for standardized intermediate inputs despite production and delivery risk on the producer side. The situation is different for customized intermediate inputs. Here, each final goods producer matches with a specific customized input producer and pre-committs to use this particular supplier. If there is a production disruption with this supplier, the final goods producer will not get anything of that particular input. Figure 1 illustrates the market structure.

With this production and market structure, final goods production succeeds only when all customized intermediate inputs are succesfully delivered. We define the failure probability π and assume that failures are independent. Then, production succeeds with probability $(1-\pi)^{m_s}$ where m_s is the number of customized intermediate inputs in sector s. Further, we assume that labor supply and other customized input supplies

Figure 1: Model Structure



are pre-committed before the resolution of production risk. Hence when production fails, those inputs are wasted. Even though the model features idiosyncratic risk, we show that aggregate output and aggregate labor demand of every sector can be summarized by an optimizing representative firm. This representative firm has a deterministic production function which is linear in labor. The supply chain risk re-appears as a productivity term proportional to $(1-\pi)^{m_s(1-\gamma)+\gamma}$. where γ is the share of standardized intermediate goods. The interpretation of the productivity penalty is that supply chain risk confers a $(1-\pi)^{m_s}$ penalty on the productivity of labor and customized intermediate inputs, as they are utilized only when production succeeds. There is a $1-\pi$ productivity penalty on standardized intermediate inputs due to production failure, but the centralized market means that this effect is not amplified. Combining these two penalties using the factor shares yields the aggregated productivity penalty.

Once we have characterized each sector using a representative firm, we can build a trade model that incorporates supply chain risk. We create a world economy in which sectors vary in their number of intermediate inputs m_s and countries vary in their degree of supply chain risk π_c. We represent the production technology of each country-sector pair using the previously derived representative firm. This gives us a trade model with country-industry-specific productivity penalties $(1-\pi_c)^{m_s(1-\gamma)+\gamma}$. We note that these productivity terms are *log-submodular* in π_c and m_s. It is well-known in trade theory that there is a a close connection between log-submodularity in productivity and negative sorting, and we prove that our model indeed features negative sorting between π_c and m_s. Countries with high supply chain risk will specialize in goods with a low number of customized inputs.

In section 3.1, we set up the production environment for a sector and derive a representative firm to characterize the sector's aggregate behavior. In section 3.2, we insert these sectors this into a trade model and derive the pattern of specialization.

3.1 Sector level supply function

A sector s features a unit interval of final goods producers $j \in [0, 1]$ (we will suppress this subscript when we talk about firm behavior). A final goods producers require labor, a composite of customized intermediate inputs X, and a composite of standardized intermediate inputs Z for production.

It will be important to distinguish between variables that are determined before and after the resolution of production and delivery uncertainty. In particular, the realized intermediate input supplies will be stochastic as they depend on the realization of a collection of production disruption shocks. We will use the convention to put a tilde ($\sim$) on top of variables to denote that they are determined after the resolution of uncertainty and therefore stochastic. The production function is given by

$$\tilde{y} = \kappa l^{\alpha} \tilde{X}^{\beta} \tilde{Z}^{\gamma}, \quad \alpha + \beta + \gamma = 1$$

We introduce the normalization $\kappa = \alpha^{-\alpha}\beta^{-\beta}\gamma^{-\gamma}(1 - \gamma\pi)m^{\frac{\beta}{1-\eta}} n^{\frac{\gamma}{1-\eta}}$ for notational convenience. The composite intermediate goods are given by

$$\begin{aligned} \tilde{X} &= \left(\sum_{i=1}^{m} \tilde{x}_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \\ \tilde{Z} &= \left(\sum_{i=1}^{n} \tilde{z}_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \end{aligned}$$

Firm decisions can take place before or after the resolution of uncertainty and, therefore, timing matters. In our model, firms decide on labor use and customized input orders before the resolution of uncertainty. They decide on standardized input purchases after the resolution of uncertainty. Our choice of timing is motivated by considering the possibility to reallocate inputs in case of input delivery failure. We think it is reasonable that labor is difficult to reallocate quickly, and customized goods of course involve pre-commitment as the producer needs to specialize a production batch to a particular supplier. However, for standardized inputs with deep markets, it is reasonable that inputs can be reallocated from firms with disruptions to those without disruptions relatively easily. Hence, the firm first decides on labor input l and customized input orders x_i^f. From the point of view of a firm, labor has a pre-determined wage w and the firm gives a take it or leave it offer to intermediate goods producers to pay $p_i^x x_i^f$ in case of successful delivery.[5] After the resolution of uncertainty, the final goods firm decides how much of the standardized intermediate inputs to buy. We denote this quantity $\tilde{z}_i^f$ to emphasize that it is a stochastic choice variable depending on the realization of production disruption shocks. The firm pays p_i^z per unit of standardized goods.[6] We assume that firms behave competitively in the standardized input market and that they can buy an arbitrary amount of goods at the prevailing price p_i^z. There is no delivery uncertainty, and in equilibrium p_i^z will adjust to clear the market. Taken together, the firm solves

$$\max_{l^f, x_i, \tilde{z}_i} \quad \mathbb{E}\left(P\tilde{y} - wl^f - \textstyle\sum_{i=1}^m p_i^x \tilde{x}_i - \sum_{i=1}^m p_i^z \tilde{z}_i\right) \qquad s.t. \qquad \tilde{y} = \kappa (l^f)^\alpha \tilde{X}^\beta \tilde{Z}^\gamma$$

We simplify this expression in steps to clarify the optimization problem. We first note that the randomness can be reduced to two cases: either all customized inputs arrive or at least one is missing. When a customized input is missing, production will fail ($\tilde{y} = 0$) regardless of the amount of standardized inputs. Clearly, the firm will then choose not to buy any standardized inputs. Thus, there is only one state of the world in which we buy standardized inputs: when all goods arrive. We write z_i without a tilde ($\sim$) to denote the purchased amount of standardized inputs in this case. As all failures of customized goods are independent and happen with probability π, the probability that all deliveries will suceed is $(1-\pi)^m$. We can rewrite the optimization problem as

$$\max_{l^f, x_i, z_i} (1-\pi)^m P y - wl^f - \sum_{i=1}^m p_i^x (1-\pi) x_i - \sum_{i=1}^n (1-\pi)^m p_i^z z_i \qquad s.t. \qquad y = \kappa (l^f)^\alpha X^\beta Z^\gamma$$

[5]The key assumptions are that we place all the bargaining strength on the buyer side and we do not introduce any contracting frictions. These assumptions can be relaxed to analyze the interaction between contracting frictions and production uncertainty. It is a non-consequential assumption that firms only pay when delivery is successful as firms are risk-neutral, but it is worth keeping in mind that if firms were risk-averse the pricing scheme would embody some form of risk-sharing. This notion could be useful to analyze the selection of payment terms in international trade. Lastly, the choice of writing total payment as $p_i(j) x_i^f(j)$ is only an inconsequential reparametrization of total payments $T_i(j)$.

[6]In a general version we would write $\tilde{p}_i^z$ to denote that the price of standardized inputs is determined after the realization of production shocks, but in this case there are no aggregate production shocks, and the price will be independent of the realized shocks with probability one.

Here, $(1-\pi)$ in the x_i-terms stems from our assumption that firms only pay customized goods suppliers upon succesful delivery, and X,Z are the deterministic versions of the input composites $\tilde{X}, \tilde{Z}$. With this formulation, we can derive the relative demand for different factors using standard methods.

$$\begin{aligned} \frac{x_i}{l^f} &= \frac{\beta}{\alpha} \frac{1}{m(1-\pi)} \frac{w}{p_i^x} \\ \frac{z_i}{l^f} &= \frac{\gamma}{\alpha} \frac{1}{n(1-\pi)^m} \frac{w}{p_i^z} \end{aligned}$$

The customized intermediate input sector has a linear production function in labor. This means that when they employ labor l_i^x, they produce output l_i^x with probability $1-\pi$ and zero output with probability π. The firms obtains an order x_i for which it is paid $p_i^x x_i$ upon delivery and 0 otherwise. Conditional on producing, it is always optimal for the firm to employ x_i units of labor to fill the order exactly. FIrms might also choose not to produce at all. Thus, they choose between accepting or not accepting an order. Mathematically, they solve:

$$\max_{x_i' \in \{0, x_i\}} x_i' \left((1-\pi) p_i^x - w \right).$$

Just as customized intermediate input producers, standardized intermediate input producers have linear production functions in labor and risky production. Thus, they employ l_i^z workers and produce l_i^z goods with probability $1-\pi$ and 0 goods with probability π. When successful, they sell their output to the centralized market at price p_i^z. Producers choose $l_i^z \geq 0$ to maximize their expected profit

$$\Pi_i^z = p(1-\pi) l_i^z - w l_i^z \tag{1}$$

We analyze a single sector which will be inserted into a trade model. Therefore, our primary interest is how the sector's aggregate labor demand and aggregate output vary with prices. That is, we are interested in:

$$\begin{aligned} Y &= \int_0^1 \tilde{y}(j) dj \\ L &= \int_0^1 l^f(j) dj + \sum_{i=1}^{m} \int_0^1 l_i^x(j) dj + \sum_{i=1}^{n} \int_0^1 l_i^z(j) dj \end{aligned}$$

and how they depend on final goods price P and wages w. Our main result is that the sector's aggregate behavior can be described by a representative firm where supply chain risk enters as a productivity term. We first define the aggregate net supply of the sector $\mathcal{S}_{sto}(P, w)$ as the set of sector outputs and labor demands that are consistent with profit maximization for some intermediate good prices. More formally, a pair of output and labor demand (Y, L) belongs to the net supply correspondence $\mathcal{S}_{sto}$, if we can find some intermediate input prices, order quantities, and labor demands such that:

- The quantities and labor demands are optimal for both final goods producers and intermediate goods producers given intermediate intermediate input prices and aggregate prices P and w
- Total production of final goods is Y

$$Y = \int_0^1 \tilde{y}_j dj = (1-\pi)^m y \quad a.s.$$

- Total labor demand from final and intermediate good producers is L:

$$\int_0^1 \left(l^f(j) + \sum_{i=1}^m l_i^x(j) + \sum_{i=1}^n l_i^z(j) \right) dj = L \quad a.s.$$

- Standardized goods markets clear almost surely:

$$\int_0^1 \tilde{z}_i(j) dj = \int_0^1 l_i^z(j) \mathbb{I} \left(success_i(j) = 1 \right) dj \quad a.s. \quad i = 1, \dots, n$$

Here $success_i(j)$ is an indicator variable taking value 1 if there is no disruption for firm j in standardized intermediate input sector i. Exploiting the fact all firms behave symmetrically, we can write the labor demand equation and the market clearing equation for standardized inputs as

$$\begin{aligned} l^f + \sum_{i=1}^m l_i^x + \sum_{i=1}^n l_i^z &= L \quad a.s. \\ (1-\pi)^m z_i &= (1-\pi) l_i^z \quad a.s. \end{aligned} \tag{2}$$

The market clearing condition in the standardized input markets (2) is non-standard. The left-hand side reflects that only a fraction $(1-\pi)^m$ of firms demands standardized input goods, whereas the right-hand side reflects that a fraction $(1-\pi)$ of all standardized input producers are successful in their production. Furthermore, we use the formulation almost surely (a.s.) because there exist events in which, for example, all intermediate good production succeeds or fails. Due to a law of large numbers, all events that deviate from the mean have probability 0 and hence the formulation almost surely.

Now we want to show that this sector aggregate supply correspondence $\mathcal{S}_{sto}$ is identical to the aggregate supply correspondence of a representative firm with a linear deterministic production function

$$Y = (1-\pi)^{m(1-\gamma)+\gamma} L \tag{3}$$

The intuition behind this representative firm is that there is a $(1-\pi)^m$ production probability. For customized inputs and labor input, the productivity penalty is $(1-\pi)^m$ as they are pre-committed. For standardized intermediate inputs, the productivity penalty is just $(1-\pi)$ as firms do not pre-commit to use them. Given that the shares of labor, customized, and standardized intermediate inputs are α, β, γ, we obtain an aggregate productivity penalty

$$[(1-\pi)^m]^\alpha \left[(1-\pi)^m\right]^\beta [(1-\pi)]^\gamma = (1-\pi)^{m(1-\gamma)+\gamma}$$

using the fact that $\alpha+\beta+\gamma=1$. Given our proposed representative firm, the profit of the firm is given by

$$P(1-\pi)^{m(1-\gamma)+\gamma} L - wL$$

and we define the supply correspondence $\mathcal{S}_{rep}$ of the representative firm as all pairs Y and L that are consistent with profit maximization for the representative firm. More formally, (Y, L) belongs to $\mathcal{S}_{rep}$ if L maximizes profit and $Y = (1-\pi)^{m(1-\gamma)+\gamma} L$. We can now state our representative firm theorem:

Proposition 1. *(Representative Firm) The stochastic sector can be described by representative firm, i.e.*

$$S_{sto}(P, w) = S_{rep}(P, w) \quad \forall P, w > 0$$

Moreover, when $w/P = (1-\pi)^{m(1-\gamma)+\gamma}$, both the sector supply correspondence $S_{sto}(P, w)$ and the representative firm supply correspondence $S_{rep}(P, w)$ are given by

$$Y = (1-\pi)^{m(1-\gamma)+\gamma} L, \quad L \geq 0.$$

When $w/P < (1-\pi)^{m(1-\gamma)+\gamma}$, both correspondences are empty as there is no finite labor demand consistent with optimization. When $w/P > (1-\pi)^{m(1-\gamma)+\gamma}$, both correspondences are $\{(0,0)\}$ as zero production is the only firm choice consistent with optimization.

Proof. See Appendix. □

This result means that we can use the representative firm's production function to analyze the aggregate behavior of a sector. Provided we find a general equilibrium featuring prices P, w, and aggregate sectoral output and labor demand Y and L, we can find intermediate input prices and micro-level firm behavior which is optimal given P, w and produces the aggregate outcome (Y, L). Conversely, there is no micro-behavior that is consistent with optimization and produces other aggregate outcomes than $\mathcal{S}_{rep}$. Therefore, without loss of generality, we can assume that sectoral production is represented by (3) in the analysis of trade patterns.

3.2 Trade model

In this section, we use the representative firm from section 3.1 to derive trade patterns with risky supply chains. We posit a world economy in which industries differ in the number of customized intermediate inputs m and countries differ in terms of risk levels π. Under these conditions, we show that high-π countries will produce low-m goods.

3.2.1 Environment

There are k industries $m_1 < m_2 < \cdots < m_k$ indexed by the number of customized intermediate inputs. All goods have a common number n of substitutable intermediate inputs and common intermediate input shares for standardized inputs γ. There is a continuum of countries indexed by production risk $\pi \in [\underline{\pi}, \bar{\pi})$ with common labor supplies L. The production function for good m_j in country π is given by

$$Y_{\pi,j,} = (1-\pi)^{m_j+\gamma}\,\ell_{\pi,j} \tag{4}$$

and the representative firm in each sector maximizes profits

$$\Pi_{\pi,j} = p_j\,(1-\pi)^{m_j+\gamma}\,\ell_{\pi,j} - w_\pi \ell_{\pi,j}$$

Consumers in country π maximize

$$U(c_{\pi,1}, \ldots, c_{\pi,k}) \quad s.t. \quad \sum_{i=1}^{k} c_{\pi,j} p_j \leq w_\pi L$$

where U is strictly concave and satisfies the Inada conditions.

3.2.2 Equilibrium

An equilibrium in the economy consists of prices p_j, wages, w_π, labor allocation $l_{\pi,j}$, production $Y_{\pi,j}$, and consumption $c_{\pi,j}$ such that

- The labor allocation maximizes firm profits
- Production is given by representative firm: $Y_{\pi,j,} = (1-\pi)^{m_j+(1-\alpha)} \ell_{\pi,j}$
- Firms make zero profits

$$\begin{aligned} \Pi_{\pi,j} &\leq 0 \\ \Pi_{\pi,j} &= 0 \quad if\ l_{\pi,j} > 0 \end{aligned}$$

- Goods and labor markets clear

$$\begin{aligned} \int_{\underline{\pi}}^{\bar{\pi}} Y_{\pi,j} d\pi &= \int_{\underline{\pi}}^{\bar{\pi}} c_{\pi,j} d\pi \quad \forall j = 1, \dots k \\ \sum_{j=1}^{k} \ell_{\pi,j} &= L \qquad \forall \pi \in [\underline{\pi}, \bar{\pi}) \end{aligned}$$

- If good m_j is produced in country π, there exists δ such that m_j is produced in all countries $\pi' \in [\pi, \pi + \delta)$. This assumption is technical in nature and ensures that the function assigning countries to goods is right-continuous (see Costinot et al. (2012) for the use of a similar assumption),

3.2.3 Sorting proposition

We are interested in how countries sort according to comparative advantage. The following proposition describes the equilibrium allocation.

Proposition 2. *(Unique Equilibrium) There exists a unique equilibrium. It features k cutoff points*

$$\underline{\pi} = \pi_k < \pi_{k-1} < \dots < \pi_1 < \pi_0 = \bar{\pi}$$

such that

$$\begin{aligned} \ell_{\pi,j} &> 0 \quad if\ \pi \in [\pi_j, \pi_{j-1}) \\ l_{\pi,j} &= 0 \quad if\ \pi \notin [\pi_j, \pi_{j-1}) \end{aligned}$$

Proposition (2) states that unreliable (high-π) countries produce goods with few customized intermediate inputs (low-n). This is the prediction that we take to the data.

4 Empirical Evidence

In this section, we test our model of comparative advantage using country-industry export data. We follow the standard methodology in the empirical comparative advantage literature (Romalis, 2004; Nunn, 2007) and estimate the equation

$$log\left(y_{ig}\right) = \beta\left(r_i \times n_g\right) + \mu_i + \theta_g + \epsilon_{ig} \tag{5}$$

Here, $y_{i,g}$ denotes country i's exports in industry g, r_i is a measure of risk and n_g is the risk-sensitivity of industry g. We include country and industry fixed effects, μ_i and θ_g, respectively. Any country level variable that is common to all industries is subsumed in the country fixed effect. Importantly, this includes the total exports of the country. The industry fixed effects capture cross-industry effects that are common across countries. For example, easily shippable goods are generally exported in higher quantities than goods that are difficult to ship. Therefore, the coefficient β measures the *tilt* in countries' trade pattern towards certain industries depending on country-industry characteristics. The interpretation is the same as in Romalis (2004).

An example illustrates the logic of the specification: Suppose, for the sake of argument, that there are two industries, electronics and wheat production. The former is highly sensitive to disruptions while the latter is relatively robust. Assume further that there are two countries, a large safe and a small risky country. First, we might expect the large country to have higher exports in both industries. The country fixed effect takes this into account. Second, we might assume that electronics are generally more (or less) traded than wheat. The industry fixed effect takes this into account. The only effect that is left is the interaction of industry and country variables. The safe country is expected to export more electronics than wheat, since electronics are risk-sensitive. This is the effect that is captured by the coefficient β.

We adopt the convention that high values of r_i correspond to high reliability (low risk), and our theoretical prediction is $\beta > 0$: Countries with high scores on reliability measures specialize in industries that are sensitive to unreliability.

4.1 Data Sources and Concordances

To measure trade flows, we use the BACI dataset which is compiled by CEPII and based on the COMTRADE data (Gaulier and Zignago, 2010). We use total value of exports for each country in each HS 2007 six digit level industry. We use data for 2012. To categorize inputs as specialized vs customized we use Rauch's classification into goods which are traded on exchange, goods which are referenced in a trade journal, and goods which are neither (Rauch, 1999). For measurement of government properties, we use the World Banks' World Governance Indicators (WGI) (Kaufmann et al., 2011). The logistics quality is measured by the World Bank's Logistics Performance Index (Arvis et al., 2014). GDP and country factor endowments are obtained for 2011 data through Penn World Table 8.1 (Feenstra et al., 2015). We use expenditure-side real GDP at chained PPPs in million 2005 USD. We measure the capital stock per worker by dividign the total capital stock at current PPPs in millions of 2005 dolalrs with the number of engaged persons in millions. For human capital we use an index of human capital provided by the PWT constructed based on years of schooling (Barro and Lee, 2013) and returns to schooling and (Psacharopoulos, 1994). Our measures of number of inputs and their contract sensitivity is taken from the 2007 US Input-Output tables published by the BEA[7]. To measure capital and skill intensity across different industries we use the NBER CES database(Bartelsman and Gray, 1996). Capital intensity is defined as total value of capital divided by total payroll (dividing by payroll instead of number of workers give an approximation of human capital instead of physical labor input). The skill intensity of an industry is defined as the ratio of non-production payroll to total payroll.

Table 1: Data Sources and Industry Classifications

Dataset	Code
NBER CES	NAICS 1997 6-digit
IO-table	IO 2007 6-digit
Rauch	SITC rev.2 4-digit
BACI	HS 2007 6-digit

[7] http://www.bea.gov/industry/io_benchmark.htm (last acessed Nov 24th, 2015)

Table 1 provides a list of the industry level codes for the various datasets. The regressions are performed in NAICS 2012 6-digit and we use a set of concordances to map our industry level variables into NAICS 2012. We use a concordance between HS 2007 10-digit and and NAICS 2007 6-digit to convert the trade data to NAICS 2007 6-digit. We use a procedure where trade flows coded in HS 2007 6-digit are allocated equally to all 10-digit extensions, and these are then mapped to NAICS 2007 6-digit code. We create chains of concordances from NAICS 2007 to NAICS 2002 and NAICS 1997 to convert the capital and skill intensities to NAICS 2007, and use the trade flows coded in NAICS 2007 6-digit to create the weights used in these concordances. We create a concordance from NAICS 2007 to NAICS 2012 to convert all data into NAICS 2012.

For Rauch we use a concordance between SITC rev.2. 4-digit and HS 2007 6-digit to convert the measure into HS 2007 6-digit. Again we use the trade data now in HS 2007 6-digit to create the weighting scheme. We then map to IO 2007 via NAICS 2007 6-digit as we use the Rauch data in the IO-tabel to calcualte industry characteristics. The IO-data together with the Rauch variables are then mapped to NAICS 2012 via NAICS 2007.

In the Web Appendix we describe in detail which sources we use for the concordances, how concordances are weighted, and how the weights are used in the transformations. The code for creating the concordances and transforming variables across different coding systems is posted on our web pages.

4.2 Measuring Products' Sensitivity to Unreliability

Motivated by our theory, we propose a novel measure of industries' risk-sensitivity. In the model, we distinguish between standardized and customized comonents. Standardized components are traded in liquid markets. As a consequence, final goods producers are not materially affected by idiosyncratic supply failures. By contrast, customized components cannot be replaced easily and, therefore, the final goods producers is exposed to the risk that an component cannot be sourced. This could be due to outright failure of a supplier or, more realistically, failure of a port authority, bad infrastructure, and so forth. Furthermore, as all components are gross complements, a non-zero amount of each components is essential to production. Hence, the *number of customized components* that an industry uses determines its exposure to supply chain risk.

We classify components as customized using the methodology developed by Rauch (1999). For each industry, he records if a good is traded on an exchange or reference-priced in a trade journal. We define a component as customized if it belongs to an industry which Rauch records as neither traded on an exchange nor reference-priced. Using the US input-tables, we look up the list of other industries that a given industry buys from and count the number of those industries that are customized according to Rauch. Tables 5 and 6 list the top and bottom thirty industries sorted by the number of customized components. Our risk sensitivity measure leads to an intuitive classification of most industries. Motor vehicle components and semiconductor production, for example, are classified as sensitive to risk, whereas farming and cement manufacturing are not. Industry-level variables are visualized in figure 3b.

We infer the number of inputs used from industry level data. To the extent that firms are heterogenous, this introduces a problem of aggregation. Consider two firms in the same industry that use 50 inputs each. If the firms' business models are not exactly identical, only 30 out of 50 might be the same for the two firms. On aggregate, however, we would observe the industry using 70 inputs, despite each firm using only 50. To protect us against the extreme case when a very small fraction of firms in an industry uses a particular input, we re-estimate the main regression excluding input industries which contributes less than 0.1% and 0.01% of total intermediate input value. The results are robust against this modification. Moreover, as long as this shortcoming is similar across industries it will not affect our results, which is based on the ranking of industries.[8]

Our measure of sensitivity to unreliability can justifiably be called complexity as it denotes how many specialized components a product uses. Thus, we can contrast it with our proposed measures of complexity in the literature. Nunn (2007) develops one such measure. He also uses the Rauch (1999) measure of product differentiation. In contrast to our measure, he defines an industry's sensitivity to contract quality as the share of total input value that comes from customized inputs. The motivation behind his measure is that cost-saving investments in specialized goods production are relation specific, and will be provided less if contract protection is poor. In light of this, it is reasonable to use the proportion of component costs as lack of relation-specific investments can be expected to increase costs proportionally. While it is the theoretically motivated to weight the customized intermediate good content by value in the context of his study, our model suggests an independent role for the *number* of components that are customized. Indeed,

[8] Further problem that we have not addressed is firstly that an industry can use multiple components from a single industry. This would lead to an underestimation of the number of inputs. Lastly, there can be a problem in that the fineness of the IO-table classification is endogenous to the US production structure, which might mean that there is a bias in that US-concentrated industries appear to have more inputs because of how the IO-table is subdivided.

our mechanism operates through vital components not arriving, and as the absence of each vital component can disrupt production, the number of customized inputs is the relevant measure.

Another measure that has been used to capture complexity is one minus the Herfindahl-index of input suppliers. The Herfindahl index is a concentration index of an industry's input suppliers. It is high if an industry's intermediate good demand is skewed towards few industries. This measure of complexity is used in Blanchard and Kremer (1997) and Levchenko (2007). Levchenko (2007) explicitly discusses why they choose to use the Herfindahl-index instead of the number of intermediate inputs: "If intermediate input use is dominated by one or two inputs (high concentration), and all the other intermediates are used very little, then what really matters to the final good producer is the relationship it has with the largest one or two suppliers." A crucial point of our theory is that all suppliers of customized inputs matter, no matter how small they are. In fact, if all intermediates are vital (gross substitutes), then the reliability of small suppliers is just as important for productivity in final good production as the reliability of large suppliers. Former Apple executive Tony Fadell illustrated this point well when the Japanese tsunami threatened to disrupt global supply chains: "lacking some part, even if it costs just dimes or a few dollars, can mean shutting down a factory".[9] This is the notion of risk that our model proposes and, hence, the number of non-substitutable inputs determines risk-sensitivity.

In figure 2a, we compare our measure to the measure proposed in Nunn (2007). Generally, the correlation is very strong and positive. Both measures classify Automobile Manufacturing as risk-sensitive and contract-intensive. On the other end of the spectrum, Soybean Farming is classified as neither risk-sensitive and nor contract intensive. However, there are some differences as well. classification differs for the textile-related industries . Nunn (2007)'s measure classifies textile-related industries (NAICS 313, 314, 315, and 316; see bottom-right area in the graph) as complex whereas our measure categorizes textile-related industries as non-complex.

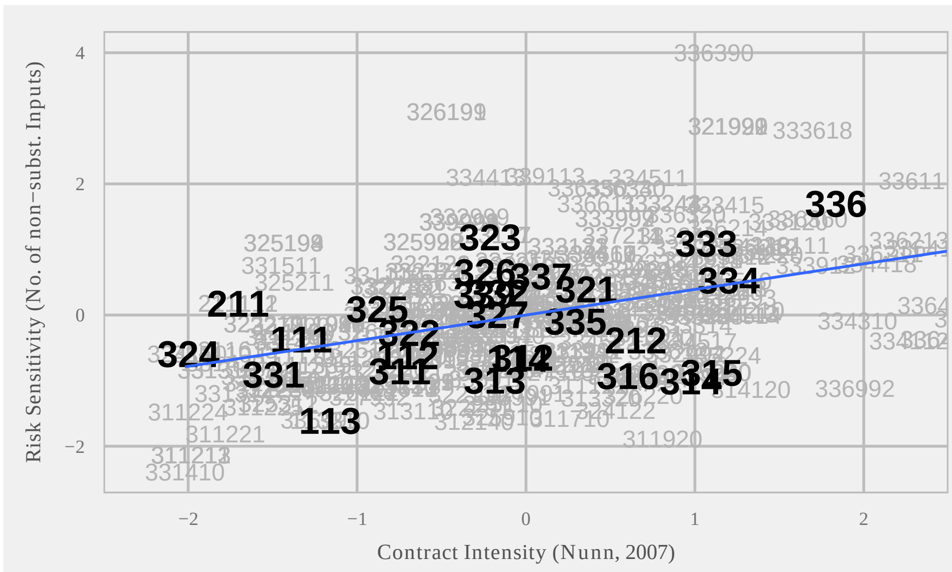
In figure 2b, we compare our measure to the Herfindahl measure used in Blanchard and Kremer (1997) and Levchenko (2007). The two measures are strongly correlated and tend to classify broad industries in similar ways. A notable exception is the transportation sector (NAICS 336), which our measure tends to classify as more risk-sensitive than than the Herfindahl Index.

[9] http://www.nytimes.com/2011/03/20/business/20supply.html?pagewanted=2&_r=0 (last accessed: November 19th, 2015)

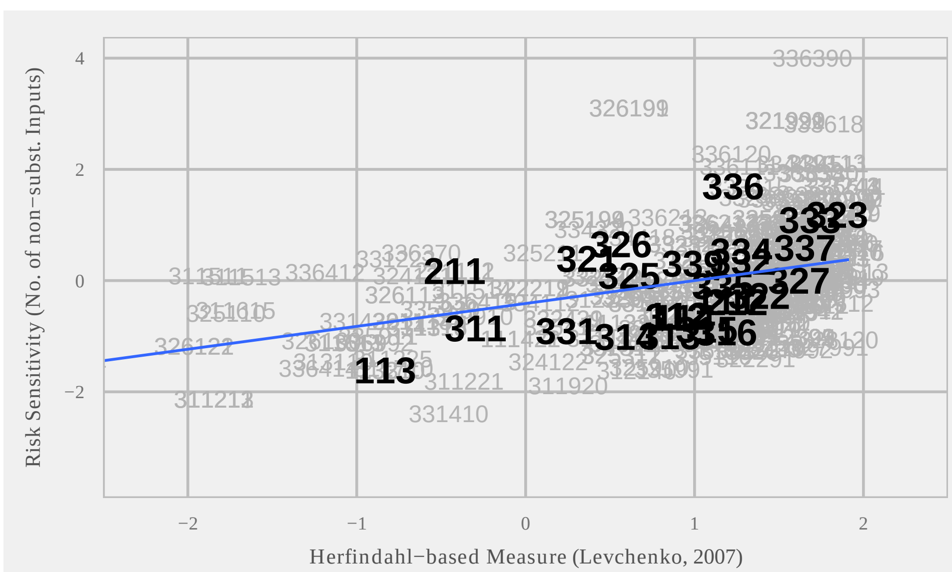
Figure 2: Comparing Complexity Measures

In this figure, we compare our measure of risk sensitivity—the number of non-substitutable inputs—to two measures that have been used in the literature. In the first panel, we compare our measure to contract intensity as defined by Nunn (2007). In the second panel, we compare our measure to the industry Herfindahl as defined by Blanchard and Kremer (1997) and Levchenko (2007). All measures are calculated at 6-digit level and standardized. We calculate trade-weighted averages at 3-digit level (printed in bold). We omit Petroleum and Coal Products Manufacturing (ID 324) to improve visibility but include it when calculating the line of best fit.

(a) Number of Non-Substitutable Inputs vs. Contract Intensity



(b) Number of non-substitutable inputs vs. Herfindahl



4.3 Measuring Countries' Reliability

We are interested in measuring disruption risk in different countries. In this context, we need to take a stand on likely causes of production and delivery disruption. For this, we focus on two country characteristics: logistics systems quality and overall government effectiveness. The motivation for including the quality of logistics system is clear: disruption is more likely if third-party logistics providers have low quality, goods clear customs slowly, and transportation infrastructure is subject to frequent failures.[10] We also include government effectiveness which we define as the quality of bureaucratic procedures and government provided services. We include this firstly as red tape is another possible cause of production disruptions. Disruption risks in this area include delays in permits for starting production, or delays in permits for bringing in inputs and foreign worker. It also captures poorly functioning bureaucracy in customs, as well as uncertain land rights. The quality of government provided services is important as failure in electricity, water supply and infrastructure also are sources of potential disruption.

When it comes to measurement, we proxy logistics systems quality with the World Bank's Domestic Logistics Performance Index (Arvis et al., 2012). The index is based on surveys with global freight forwarders and express carriers, and combines it with quantitative measures of some components of supply chains. As of 2014 it encompasses 160 countries. For bureaucratic quality, we use the Government Effectiveness-measure from the Worldwide Governance Indicators (Kaufmann et al., 2009). It is an aggregated measure derived from a large number of measures including the quality of bureaucracy, extent of red tape, infrastructure quality, and the quality of various government provided services. Figure 3a visualizes the distributions of the main country-level variables in our data.

4.4 Results

In table 2, we present the main results for the baseline specification (equation 5). We are interested in the interaction of industries' risk-sensitivity measured by the number of customized components (num_cust) and countries' reliability. Our two preferred measures of country reliability are government effectiveness (effectiveness) and logistics performance (lpi). We report interactions with four additional World Governance Indicators: regulatory quality (regquality), political stability and absence of violence (stability), voice

[10] In the current model, intermediate input suppliers are all domestic, which means that the final goods supplier does not get through customs to obtain intermediate inputs. However, even in cases where you only source domestically, we believe it is plausible that customs problems affect reliability through its effect on your intermediate input suppliers. Explicit modeling of this channel would involve intermediate good trade and bilateral delivery risks which are not in the current model. Formally showing how customs risk interacts with intermediate goods trade is an interesting area of further research.

and accountability (voice), and control of corruption (corruption). All these indicators proxy for an environment that is amenable to the production of risk-sensitive products.

The results are consistent with the hypothesis that risk-sensitive industries are disproportionately produced by reliable countries. Consider an industry that is one standard deviation above the mean in terms of risk-sensitivity (num_cust). Increasing a country's government effectiveness (effectiveness) by one standard deviation is associated with 10.4% (column 1) more exports in this industry, compared to a country with an average Logistics Performance Index. Increasing a country's Logistics Performance Index (lpi) by one standard deviation is associated with 10.1% (column 6) more exports in this industry, compared to a country with an average Logistics Performance Index. The coefficients are of very similar magnitude for the other institutional variables that we use to proxy for reliability. The main coefficient is statistically significant at the 1% level for all measures considered.

4.5 Relationship to other results in the literature

As previously discussed, Nunn (2007) tests whether contracting quality affects the pattern of trade. Might we just be capturing the effect that stable countries also tend to have good contracting environments? In table 2, we replicate Nunn (2007)'s main result (column 3). Countries with high scores on the rule of law index (ruleoflaw) tend to export contract-intensive goods (nunn). Given that both our country-measures (rule of law vs. government effectiveness) and our industry measures (contract intensity vs. risk sensitivity) are correlated, our main result in column 1 might be spurious. However, as we show in column 6, the two estimates remain quantitatively similar and significant when analyzed jointly (column 6). This result suggests that our mechanism is distinct from the role of contracting.

Of course, the two explanations are not mutually exclusive. In fact, when component producers fail to deliver a component on time they typically also violate a contract. However, in Nunn, poor contracting is analyzed in terms of relationship-specific investments and we consider our theory as an additional explanation for observed trade patterns. The distinction matters since the policy implications differ: Nunn (2007) implies that countries can attract sophisticated industries by improving contract enforcement. Our story, by contrast, suggest that a crucial policy lever is the reduction of supply chain risk.

Table 2: Baseline Regression

This table presents estimates of the main specification (equation 5). The unit of observation is a country-industry. The outcome variable is the natural logarithm of total trade volume. The variable num_cust refers to the industry's number of customized inputs, which is the measure of risk-sensitivity implied by our model. Risk-sensitivity is interacted with country-characteristics. Our two preferred measures of country reliability are government effectiveness (effectiveness) and logistics performance (lpi). We report interactions with four additional World Governance Indicators for robustness: regulatory quality (regquality), political stability and absence of violence (stability), voice and accountability (voice), and control of corruption (corruption). Standard errors are double-clustered at the country and industry level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent Variable: Log Exports					
	(1)	(2)	(3)	(4)	(5)	(6)
num_cust_x_effectiveness	0.108***					
	(0.029)					
num_cust_x_regquality		0.110***				
		(0.029)				
num_cust_x_stability			0.098***			
			(0.028)			
num_cust_x_voice				0.093***		
				(0.026)		
num_cust_x_corruption					0.087***	
					(0.025)	
num_cust_x_lpi						0.105***
						(0.029)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,458	35,458	35,458	35,458	35,458	35,458
R^2	0.969	0.969	0.969	0.969	0.969	0.969
Adjusted R^2	0.969	0.969	0.969	0.969	0.969	0.969
Residual Std. Error	1.653	1.653	1.654	1.654	1.654	1.653

Table 3: Comparison with Nunn (2007)

This table compares our estimates to Nunn (2007), who posits that countries with good legal systems (ruleoflaw) specialize in contract-intensive goods (nunn). Contract intensity is defined as the share of inputs that are neither traded on exchanges nor reference priced. Our main result is provided in column 1. We replicate the main result of Nunn (2007) in column 4. In column 6, we estimate the effect of both $nunn * ruleoflaw$ and $num_{cust} * effectiveness$ at the same time. Standard errors are double-clustered at the country and industry level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent Variable: Log Exports					
	(1)	(2)	(3)	(4)	(5)	(6)
num_cust_x_effectiveness	0.108^{***}					0.085^{***}
	(0.029)					(0.028)
num_cust_x_ruleoflaw		0.100^{***}				
		(0.026)				
nunn_x_effectiveness			0.103^{***}		0.159	
			(0.030)		(0.115)	
nunn_x_ruleoflaw				0.094^{***}	-0.055	0.061^{**}
				(0.028)	(0.107)	(0.028)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,458	35,458	35,458	35,458	35,458	35,458
R^2	0.969	0.969	0.969	0.969	0.969	0.969
Adjusted R^2	0.969	0.969	0.969	0.969	0.969	0.969
Residual Std. Error	1.653	1.653	1.653	1.654	1.653	1.652

Table 4: Interaction with GDP

This table conducts additional robustness checks. Column (1) reports our benchmark results. Column (2) we estimate specialization with Heckscher-Ohlin variables similarly to Romalis (2004): skill abundance (ln_hl) interacted with industries' skill intensity (int_sk) as well as capital abundance (ln_kl) interacted with capital intensity (int_cap). In column (3), we interact our measure of complexity (num_cust) with the logarithm of countries' per capita income (ln_y). In column (4), we re-estimate our benchmark specification while also controlling for the income interaction. In column (5), we also add Heckscher-Ohlin controls. Standard errors are double-clustered at the country and industry level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent Variable: Log Exports					
	(1)	(2)	(3)	(4)	(5)	(6)
num_cust_x_effectiveness	0.108***		0.100***		0.088***	0.086***
	(0.029)		(0.028)		(0.032)	(0.032)
ln_hl_x_int_sk		0.517***	0.473***			0.471***
		(0.163)	(0.162)			(0.161)
ln_kl_x_int_cap		−0.004	0.006			0.007
		(0.024)	(0.022)			(0.022)
num_cust_x_ln_y				0.082***	0.020	0.015
				(0.025)	(0.028)	(0.028)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,458	35,458	35,458	35,458	35,458	35,458
R^2	0.969	0.969	0.969	0.969	0.969	0.969
Adjusted R^2	0.969	0.969	0.969	0.969	0.969	0.969
Residual Std. Error	1.653	1.653	1.650	1.653	1.653	1.650

4.6 Robustness Checks

The main result is biased if our regressors are correlated with the error term, and this can happen in many different ways. First, our measures of reliability can be correlated with other country characteristics which give a comparative advantage in goods with many customized components. Second, the number of customized components can be correlated with other industry features, and high reliability can give a comparative advantage due to these industry features as well. There can also be some mixture of these two effects, for example that high government reliability is correlated with high financial development, and that a large number of customized components is correlated with having high external financing needs. We assess the robustness of our results by including other sources of comparative advantage in our regression specification.

First, we test whether Heckscher-Ohlin effects can explain the results by controlling for the interaction between factor endowments and factor intensity of different industries similar to Romalis (2004). It could be the case that reliable countries are simply countries with a large endowment of skilled labor and risk-sensitive industries tend to be skill-intensive. In column (2) of table 4, we replicate the result that skill-abundant countries specialize in skill-intensive industries (coefficient on ln_hl_x_int_sk). Importantly, our main estimate (num_cust_x_effectiveness) barely changes when we control for factor endowments. In unreported results, we confirm that the same is true for other measures of reliability.

Second, we use the logarithm of income as a catch-all term for variables that might proxy for being a rich country. It should be noted that this is over-controlling: We argue that one reason for why countries are rich is that they are stable, which lets them specialize in complex goods. Despite that, as shown in table 4, our main result remains statistically significant (column 5). Quantitatively, the estimate becomes only marginally weaker when we (over-) control for log income. This allows us to exclude any alternative explanation that is strongly correlated with GDP.

5 Conclusion

This paper provides a tractable model of the effect of supply chain risk on trade patterns. We show how the behavior of a sector with idiosyncratic delivery risk can be described by a representative firm. Supply chain risk enters as a productivity penalty, which grows exponentially with the number of specialized inputs. Therefore, the appropriate measure of supply chain sensitivity on a sectoral level the number of specialized inputs. In an international setting, the theory implies that low risk countries specialize in risk-sensitive industries, and this prediction is borne out in the data.

Our paper has a number of policy implications: Most importantly, it suggests that reducing risk attracts industries that produce risk-sensitive goods. The paper also implies that measures of the business of the environment would be more informative if they described the variability in outcomes. World Bank's Doing Business Survey, for instance, contains the time to start a business. However, it does not contain the risk of severe delays during the process, which might be equally important.

Looking ahead, there are many natural extensions to the paper. Some production and delivery risks are only relevant for cross-border trade. For example, customs procedures might be slow and frictions to international contracting can make deliveries uncertain. Therefore, it would be interesting to consider a model with trade in inputs and different disruption probabilities when goods are traded or used domestically. We conjecture that such an extension could have interesting implications for how variations in supply chain risk are connected to the spatial organization of production and the structure of intermediate input trade.

Given the endogeneity problems in our empirical work, we are also interested in extensions to improve identification. One such extension would be to use the panel dimension of trade data. The World Governance Indicators goes back to 1996 and the BACI trade data goes back to 1995. This would allow us to test whether countries that improve on institutional measures also see a concomitant rise in trade of risk sensitive goods.

References

Antràs, Pol (2003) "Firms, Contracts, And Trade Structure," *The Quarterly Journal of Economics*, Vol. 118, pp. 1375–1418.

Antras, Pol and Davin Chor (2013) "Organizing the Global Value Chain," *Econometrica*, Vol. 81, pp. 2127–2204.

Antras, Pol and Elhanan Helpman (2004) "Global Sourcing," *Journal of Political Economy*, Vol. 112, pp. 552–580.

Arvis, Jean-François, Monica Alina Mustra, Lauri Ojala, Ben Shepherd, and Daniel Saslavsky (2012) "Connecting to Compete 2012 : Trade Logistics in the Global Economy,"Technical report.

Arvis, Jean-François, Daniel Saslavsky, Lauri Ojala, Ben Shepherd, Christina Busch, and Anasuya Raj (2014) "Connecting to Compete 2014 : Trade Logistics in the Global Economy–The Logistics Performance Index and Its Indicators."

Barro, Robert J. and Jong Wha Lee (2013) "A new data set of educational attainment in the world, 1950-2010," *Journal of Development Economics*, Vol. 104, pp. 184–198.

Bartelsman, Eric J and Wayne Gray (1996) "The NBER Manufacturing Productivity Database."

Blanchard, O. and M. Kremer (1997) "Disorganization," *The Quarterly Journal of Economics*, Vol. 112, pp. 1091–1126.

Bloom, Nick (2009) "The Impact of Uncertainty Shocks," *Econometrica*, Vol. 77, pp. 623–685.

Blyde, Juan and Danielken Molina (2015) "Logistic infrastructure and the international location of fragmented production," *Journal of International Economics*, Vol. 95, pp. 319–332.

Blyde, Juan S. (2014) "The Drivers of Global Value Chain Participation: Cross-Country Analyses," in *Synchronized Factories*, Cham: Springer International Publishing, pp. 29–73.

Bond, Stephen and John Van Reenen (2007) "Microeconometric Models of Investment and Employment," in *Handbook of Econometrics*, Vol. 6, pp. 4417–4498.

Cooper, Russell and John Haltiwanger (1993) "The Aggregate Implications of Machine Replacement : Theory and Evidence," *American Economic Review*, Vol. 83, pp. 360–382.

Costinot, a., J. Vogel, and S. Wang (2012) "An Elementary Theory of Global Supply Chains," *The Review of Economic Studies*, Vol. 80, pp. 109–144.

Costinot, Arnaud (2009a) "An Elementary Theory of Comparative Advantage," *Econometrica*, Vol. 77, pp. 1165–1192.

Costinot, Arnaud (2009b) "On the origins of comparative advantage," *Journal of International Economics*, Vol. 77, pp. 255–264.

Cuñat, Alejandro and Marc J. Melitz (2012) "Volatility, Labor Market Flexibility, and the Pattern of Comparative Advantage," *Journal of the European Economic Association*, Vol. 10, pp. 225–254.

Dixit, Avinash K and Gene M Grossman (1982) "Trade and Protection with Multistage Production," *Review of Economic Studies*, Vol. 49, pp. 583–594.

Eaton, Jonathan and Samuel Kortum (2002) "Technology, Geography, and Trade," *Econometrica*, Vol. 70, pp. 1741–1779.

Feenstra, Robert C, Robert Inklaar, and Marcel P. Timmer (2015) "The Next Generation of the Penn World Table," *American Economic Review*, Vol. 105, pp. 3150–3182.

Gaulier, Guillaume and Soledad Zignago (2010) "BACI: International Trade Database at the Product-Level (the 1994-2007 Version)."

Grossman, Sanford J. and Oliver D. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, Vol. 94, pp. 691–719.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi (2009) "Governance matters VIII : aggregate and individual governance indicators 1996-2008," Policy Research Working Paper Series 4978, The World Bank.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi (2011) "The Worldwide Governance Indicators: Methodology and Analytical Issues," *Hague Journal on the Rule of Law*, Vol. 3, pp. 220–246.

Kremer, M (1993) "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, Vol. 108, pp. 551–575.

Levchenko, Andrei A. (2007) "Institutional Quality and International Trade," *Review of Economic Studies*, Vol. 74, pp. 791–819.

Manova, Kalina, "Credit Constraints, Heterogeneous Firms, and International Trade," *NBER Working Papers*.

Nunn, Nathan (2007) "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade," *The Quarterly Journal of Economics*, Vol. 122, pp. 569–600.

Psacharopoulos, George (1994) "Returns to investment in education: A global update," *World Development*, Vol. 22, pp. 1325–1343.

Rauch, James E. (1999) "Networks versus markets in international trade," *Journal of International Economics*, Vol. 48, pp. 7–35.

Romalis, John (2004) "Factor Proportions and the Structure of Commodity Trade," *American Economic Review*, Vol. 94, pp. 67–97.

Additional Results

Table 5: Top 30 Industries by Risk Sensitivity

Code	Name	Sensitivity
336390	Other Motor Vehicle Parts Manufacturing	4.0
326191	Plastics Plumbing Fixture Manufacturing	3.1
326199	All Other Plastics Product Manufacturing	3.1
321920	Wood Container and Pallet Manufacturing	2.9
321991	Manufactured Home (Mobile Home) Manufacturing	2.9
321992	Prefabricated Wood Building Manufacturing	2.9
321999	All Other Miscellaneous Wood Product Manufacturing	2.9
333618	Other Engine Equipment Manufacturing	2.8
336120	Heavy Duty Truck Manufacturing	2.3
334413	Semiconductor and Related Device Manufacturing	2.1
334511	Search, Detection, Navigation, Guidance, Aeronautical, and Nautical Sy	2.1
336111	Automobile Manufacturing	2.1
339113	Surgical Appliance and Supplies Manufacturing	2.1
336330	Motor Vehicle Steering and Suspension Components (except Spring) Manuf	1.9
336340	Motor Vehicle Brake System Manufacturing	1.9
336350	Motor Vehicle Transmission and Power Train Parts Manufacturing	1.9
333241	Food Product Machinery Manufacturing	1.7
333243	Sawmill, Woodworking, and Paper Machinery Manufacturing	1.7
333244	Printing Machinery and Equipment Manufacturing	1.7
333415	Air-Conditioning and Warm Air Heating Equipment and Commercial and Ind	1.7
336611	Ship Building and Repairing	1.7
332999	All Other Miscellaneous Fabricated Metal Product Manufacturing	1.5
333992	Welding and Soldering Equipment Manufacturing	1.5
333997	Scale and Balance Manufacturing	1.5
333999	All Other Miscellaneous General Purpose Machinery Manufacturing	1.5
336320	Motor Vehicle Electrical and Electronic Equipment Manufacturing	1.5
336360	Motor Vehicle Seating and Interior Trim Manufacturing	1.5
333120	Construction Machinery Manufacturing	1.4
339991	Gasket, Packing, and Sealing Device Manufacturing	1.4
339992	Musical Instrument Manufacturing	1.4

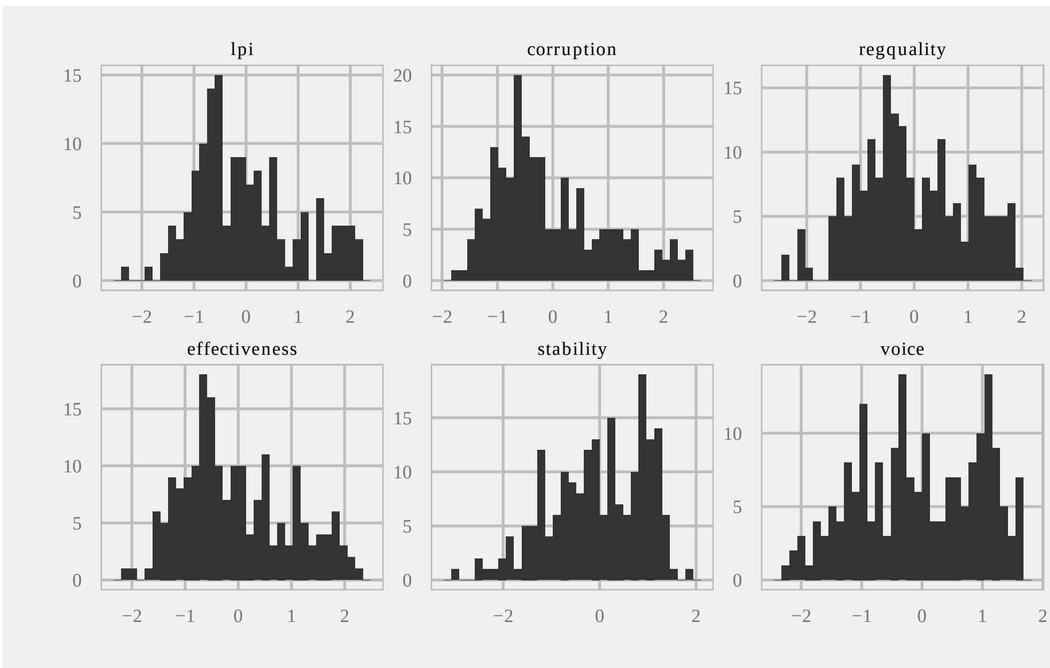
Table 6: Bottom 30 Industries by Risk Sensitivity

Code	Name	Sensitivity
336991	Motorcycle, Bicycle, and Parts Manufacturing	-1.2
331314	Secondary Smelting and Alloying of Aluminum	-1.2
327992	Ground or Treated Mineral and Earth Manufacturing	-1.2
327991	Cut Stone and Stone Product Manufacturing	-1.2
326220	Rubber and Plastics Hoses and Belting Manufacturing	-1.2
326122	Plastics Pipe and Pipe Fitting Manufacturing	-1.2
326121	Unlaminated Plastics Profile Shape Manufacturing	-1.2
322110	Pulp Mills	-1.2
325312	Phosphatic Fertilizer Manufacturing	-1.3
325311	Nitrogenous Fertilizer Manufacturing	-1.3
313320	Fabric Coating Mills	-1.3
335110	Electric Lamp Bulb and Part Manufacturing	-1.4
322291	Sanitary Paper Product Manufacturing	-1.4
311225	Fats and Oils Refining and Blending	-1.4
324122	Asphalt Shingle and Coating Materials Manufacturing	-1.5
313110	Fiber, Yarn, and Thread Mills	-1.5
311224	Soybean and Other Oilseed Processing	-1.5
336414	Guided Missile and Space Vehicle Manufacturing	-1.6
335991	Carbon and Graphite Product Manufacturing	-1.6
325910	Printing Ink Manufacturing	-1.6
312140	Distilleries	-1.6
311710	Seafood Product Preparation and Packaging	-1.6
113310	Logging	-1.6
113210	Forest Nurseries and Gathering of Forest Products	-1.6
311221	Wet Corn Milling	-1.8
311920	Coffee and Tea Manufacturing	-1.9
311213	Malt Manufacturing	-2.1
311212	Rice Milling	-2.1
311211	Flour Milling	-2.1
331410	Nonferrous Metal (except Aluminum) Smelting and Refining	-2.4

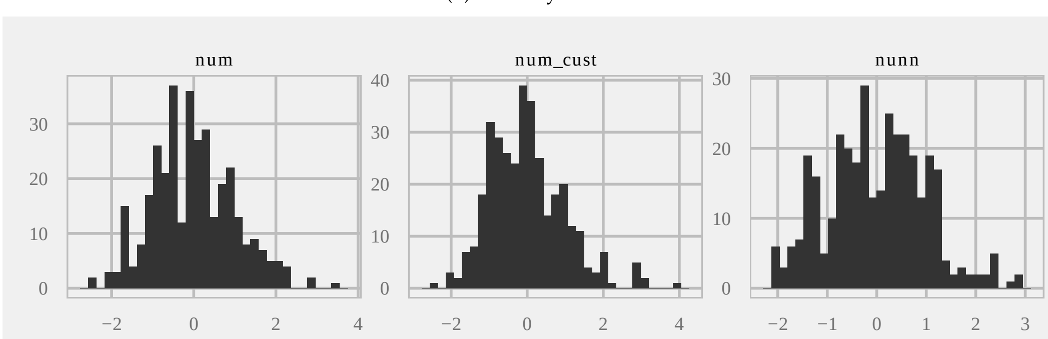
Figure 3: Country and Industry Variables

This figures presents histograms of country and industry characteristics. All variables are standardized for ease of comparison. Data sources for country-level variables: From World Governance Indicators (Kaufmann et al., 2009), we collect Government Effectiveness (effectiveness), Regulatory Quality (regquality), Political Stability and Absence of Violence/Terrorism(stability), Voice and Accountability(voice), Control of Corruption(corruption), and Rule of Law (ruleoflaw). We add the World Bank's Logistics Performance Index (lpi; see Arvis et al., 2012). In robustness checks, we also use Penn World Tables (Feenstra et al., 2015) to account for skilled labor (ln_hl), capital(ln_kl), and the logarithm of per capita GDP (ln_y). Data sources for industry-level variables: We define the number of inputs (num) using the US input-output tables. The number of customized inputs is calculated by counting the number of inputs that are neither reference-priced nor traded on an exchange according to Rauch (1999). Contract intensity (nunn) is calculated as in Nunn (2007). Skill and capital intensity (sk_int, cap_int; unreported) are sourced from Romalis (2004).

(a) Country-level



(b) Industry-level



Theory Appendix

Proof of Proposition 1

We seek to show that for all $\Gamma = (P, w, \pi)$,

$$\Omega_{det}(\Gamma) = \Omega_{sto}(\Gamma) = \begin{cases} \emptyset & \text{if } \frac{w}{P} < (1-\pi)^{m+(1-\alpha)} \\ \{(L, F(L;\gamma)) : L \geq 0\} & \text{if } \frac{w}{P} = (1-\pi)^{m+(1-\alpha)} \\ \{(0,0)\} & \text{if } \frac{w}{P} > (1-\pi)^{m+(1-\alpha)} \end{cases}.$$

We first note that it is obvious that

$$\Omega(\Gamma) = \begin{cases} \emptyset & \text{if } \frac{w}{P} < (1-\pi)^{m+(1-\alpha)} \\ \{(L, F(L;\gamma)) : L \geq 0\} & \text{if } \frac{w}{P} = (1-\pi)^{m+(1-\alpha)} \\ \{(0,0)\} & \text{if } \frac{w}{P} > (1-\pi)^{m+(1-\alpha)} \end{cases}.$$

Indeed, if real wage is below unit cost, no finite L solves the firm's problem. If real wage is above unit cost, 0 is the only profit maximizing production level. If real wage equals unit cost, firms are indifferent about production size.

Thus, the interesting thing is to show $\Omega_{det}(\gamma) = \Omega_{sto}(\gamma)$. We go through the three cases of $\frac{w}{P}$ and show that $\Omega_{det}(\Gamma) \subseteq \Omega_{sto}(\Gamma)$ and $\Omega_{sto}(\Gamma) \subseteq \Omega_{det}(\Gamma)$ for each case.

Case 1: $\frac{w}{P} < (1-\pi)^{m+\gamma}$

It is trivial that $\emptyset \subseteq \Omega_{sto}(\Gamma)$. To prove that $\Omega_{sto}(\Gamma) = \emptyset$, we note that if $(Y, L) \in \Omega_{sto}(\Gamma)$ we need $p_i^x \leq w/(1-\pi)$ for $i = 1, \ldots, m$ and $p_i^z(j) \leq \frac{w}{1-\pi}$ for $i = 1, \ldots, n$ as there would otherwise be infinite labor demand in the intermediate goods sector. But with this assumption, unit cost in the final goods sector becomes

$$\begin{aligned} \frac{w^\alpha \left(\sum_{i=1}^m \left((1-\pi)p_i^x(j)\right)^{1-\eta}\right)^{\frac{\beta}{1-\eta}} \left(\sum_{i=1}^n (p_i^z)^{1-\eta}\right)^{\frac{\gamma}{1-\eta}}}{m^{\frac{\beta}{\eta-1}} n^{\frac{\gamma}{\eta-1}} (1-\pi)^m} &\leq \frac{w}{(1-\pi)^{m+\gamma}} \\ &< P \end{aligned}$$

which means that labor demand is unbounded in the final goods sector. Thus, no finite L is consistent with optimization.

Case 2: $\frac{w}{P} = (1-\pi)^{m+\gamma}$

First, we want to show that $\Omega_{det}(\Gamma) \subseteq \Omega_{sto}(\Gamma)$, that is we want to show that $(L, F(L;\Gamma)) \in \Omega_{sto}(\Gamma)$ for any L. To do this, consider prices $p_i^x(j) = \frac{w}{1-\pi}$ for $i = 1, \ldots, m$ (more precisely that the offered payment is $x_i^F(j)p_i^x(j)$) and$p_i^z = \frac{w}{1-\pi}$ for $i = 1, \ldots, n$, and allocations

$$\begin{aligned}
l^F(j) &= \alpha L && (6)\\
x_i^F(j) &= \frac{\beta L}{m} \quad i = 1, \ldots, m && (7)\\
l_i^x(j) &= \frac{\beta L}{m} && (8)\\
z_i^F(j) &= \frac{\gamma L(1-\pi)}{n} \quad i = 1, \ldots, n && (9)\\
l_i^z(j) &= \frac{\gamma L}{n} \quad i = 1, \ldots, n && (10)
\end{aligned}$$

It is clear that labor demand sums to L. Intermediate goods producers are indifferent between different production levels, so they optimize.The final goods producer's problem is equivalent to solving a deterministic problem with price $P(1-\pi)^m$ and where the price of customized components is modified to $p_i^x(j)(1-\pi)$ to reflect that the final goods producer only pays in case of delivery. Given the symmetry within the classes of standardized and intermediate components, it is clear that the firm chooses the same amount x, z of all of them. So the firm solves the problem

$$\max_{l^f, X, Z} P(1-\pi)^m \kappa(\alpha, \beta, \gamma, m, n) l^\alpha x^\beta m^{\frac{\beta\eta}{\eta-1}} z^\gamma m^{\frac{\gamma\eta}{\eta-1}} - lw - mx(1-\pi)p_i^x(j) - nzp_i^z$$

Standard optimization gives that $\frac{l^f}{x} = \frac{\alpha}{(\beta/m)}$ and $\frac{l^f}{z} = \frac{\alpha}{(\gamma/n)}$, and we can check that profits are zero for all l^f when these two conditions are satisfied. Thus, the proposed allocation solves the final goods producer's problem.

Total production is given by

$$\int_0^1 \tilde{y}(j)dj \quad = \quad \kappa(\alpha,\beta,\gamma,\eta)(1-\pi)^m(\alpha L)^\alpha \left(m\left(\frac{\beta L}{m}\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\beta\eta}{\eta-1}} \left(n\left(\frac{\gamma L(1-\pi)}{n}\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\gamma\eta}{\eta-1}}$$

$$= \Omega_{dewt} \quad (1-\pi)^{m+\gamma}L$$

$$= \quad F(L;\Gamma).$$

Hence, $(L, F(L;\Gamma)) \in \Omega_{sto}(\Gamma)$ and $\Omega_{det}(\Gamma) \subseteq \Omega_{sto}(\Gamma)$.

Second, we want to show that $\Omega_{sto}(\Gamma) \subseteq \Omega_{det}(\Gamma)$. So consider an arbitrary (L, Y). If $L = 0$, then $Y = 0$ trivially and we are done, as $(0, 0) \in \Omega_{det}(\Gamma)$. So let us assume that $Y, L > 0$. As $L > 0$, we need that $l^f(j) > 0$ for some j. Let $\mathcal{S}$ be set of j for which this is true and assume without loss of generality that $\mathcal{S}$ =$[0, 1]$ (size is indeterminate, but if $\mathcal{S} \neq [0, 1]$ we can just divide everything with the measure of $\mathcal{S}$).

If final goods producers optimally choose positive labor component, optimality implies that they also choose positive amounts of all intermediate components. Thus, market clearing implies that for all i, there exists some j, such that$l^z(j) > 0$ which means that $p_i^z = \frac{w}{1-\pi}$ for all i. Similarly, $l_i^x(j) > 0$ for all i, j which mean that offers are given with $p_i^x(j) = \frac{w}{1-\pi}$. The necessary condition for optimality for final goods producers derived above gives us the relative demand for labor and different intermediate goods components. Using the market clearing condition for intermediate goods products and labor, we get that the labor allocations and intermediate good demands are given by equations (6)-(10). This means that total production is $Y = F(L;\Gamma)$, and $(L, Y) \in \Omega_{det}(\Gamma)$.

Hence, $\Omega_{sto}(\Gamma) = \Omega_{det}(\Gamma)$.

Case 2: $\frac{w}{P} > (1-\pi)^{m+\gamma}$

We want to show that $\Omega_{sto}(\Gamma) = \{(0, 0)\}$. We first show that $(0, 0) \in \Omega_{sto}(\Gamma)$. Now, suppose that $p_i^z(j) = p_i^z(j) = \frac{w}{1-\pi}$ for all i, j. Then no production lies in the optimal set for all intermediate good producers. Furthermore, we can check that no production is also optimal for the final good producers by noting that their unit cost exceeds their price. Thus, $(0, 0) \in \Omega_{sto}(\Gamma)$.

Next, we want to show that $\Omega_{sto}(\Gamma) \subseteq \{(0, 0)\}$. We prove this by contradiction. Suppose that $(L, Y) \in \Omega_{sto}(\Gamma)$ with $L > 0$. This means that $l^f(j) > 0$ for some j. This also means that for each $i \in \{1, \ldots, n\}$, there exists a $j' \in [0, 1]$ such that$l_i^z(j') > 0$. Hence, $p_i^z = \frac{w}{1-\pi}$ for all i. Furthermore, optimality together with the restriction that the customized goods producers accept their offers, requires that $l_i^x(j) = \frac{w}{1-\pi}$ for all $i \in \{1, \dots, m\}$. But now we can check that these prices make it optimal for final goods producers to choose zero production, thus contradicting our assumption that $L > 0$.

Hence, we again have $\Omega_{sto}(\Gamma) = \Omega_{det}(\Gamma)$.

Proof of Proposition [Sort trade]

We proceed in steps. First we prove that each country has positive production and that each good is produced in equilibrium. Then we characterize the sorting behavior and show that the equilibrium is unique.

Lemma 3. *For each m_j, there exists an π with $\ell(\pi, m_j) > 0$, and for each π, there exists an m_j with $\ell(\pi, m_j) > 0$.*

The Inada condition means that every good is produced in equilibrium, which proves the first part of the proposition. The second part of the proposition follows directly from the labor clearing condition.

Second, we prove a lemma that captures the sorting of m_j and π. It states that if there are a high risk and a low risk country, as well as a complex and a simple good, then if the low risk country produce the simple good in equilibrium, the high risk country will not produce the complex good. This excludes reversals of comparative advantage and is used to prove sorting.

Lemma 4. *Suppose that $\pi' < \pi$ and $m_j < m_{j'}$. Then $\ell(\pi', m_j) > 0$ implies $\ell(\pi, m_{j'}) = 0$.*

Proof. We proceed by contradiction. Suppose $\ell(\pi', m_j), \ell(\pi, m_{j'}) > 0$. Then the no profit conditions give us

$$\begin{aligned}
p_{m_{j'}}(1-\pi)^{m_{j'}} &= w_\pi \\
p_{m_j}(1-\pi)^{m_j} &\leq w_\pi \\
p_{m_{j'}}(1-\pi')^{m_{j'}} &\leq w_{\pi'} \\
p_{m_j}(1-\pi')^{m_j} &= w_{\pi'}
\end{aligned}$$

From which we derive the contradiction

$$\frac{p_{m_{j'}}}{p_{m_j}} \geq (1-\pi)^{m_j - m_{j'}}$$
$$\frac{p_{m_{j'}}}{p_{m_j}} \leq (1-\pi')^{m_j - m_{j'}}$$

This is a contradiction as $\pi' < \pi$ and $m_j < m_{j'}$ implies that $(1-\pi')^{m_j - m_{j'}} < (1-\pi)^{m_j - m_{j'}}$ so no price ratios satisfy the two inequalities simultaneously. □

Corollary 5. *$\ell(\pi, m_j) > 0$ implies $\ell(\pi, m_{j'}) = 0$ for all $j' \neq j$. I.e. each country only produces one good.*

Proof. Suppose that $m_j < m_{j'}$ are both produced in country π, i.e. $\ell(\pi, m_j), \ell(\pi, m_{j'}) > 0$. Our assumption of continuity means that we can find δ such that $\ell(\pi', m_j) > 0$ for all $\pi' \in (\pi, \pi + \delta)$. But then $\ell(\pi, m_{j'}), \ell(\pi', m_j) > 0$ which contradicts the lemma.

With aid of this corollary, we can obtain a full characterization of the sorting behavior. I.e. that there exist

$$\underline{\pi} = \pi_k < \pi_{k-1} < \cdots < \pi_1 < \pi_0 = \bar{\pi}$$

such that

$$\ell(\pi; m_j) > 0 \quad if\ \pi \in [\pi_j, \pi_{j-1})$$
$$l(\pi; m_j) = 0 \quad if\ \pi \notin [\pi_j, \pi_{j-1})$$

Define the correspondence

$$\Psi(\pi) = \left\{ j \in \{1, \ldots, k\} : \ell(\pi, m_j) > 0 \right\}$$

From previous results, this correspondence is always non-empty, single valued, and weakly decreasing in π. Define

$$\pi_1 = \inf \{\pi : \Psi(\pi) = 1\}$$

The set $\{\pi : \Psi(\pi) = 1\}$ is non-empty, and the assumption that $\ell(\pi, m_j) > 0$ implies $\ell(\pi', m_j) > 0$ for $\pi' \in [\pi, \pi+\delta)$ implies that Ψ is right-continuous, so $\Psi(\pi_1) = 1$. By weak monotonicity $\Psi(\pi) = 1$ for $\pi \in [\pi_1, \pi_0)$. Define π_2 analogously and continue in the same way. The Proposition is thus proved. $\square$

Web Appendix - Concordance construction

To generate concordances and map data across coding system, we create a general mathematical framework to treat the problem. In this Web Appendix, we describe how the general system works, and then we show how we use it to convert our particular data.

The basic building block of our concordance system is a many-to-many concordance between coding systems A and B where we have weights on both A and B. We call such concordances *two-weighted concordances*. An example of such a concordance is provided below:

A	B	A_w	B_w
1	a	10	70
2	b	20	50
2	c	20	100
3	c	15	40
4	d	5	70
5	d	25	70
6	e	30	90

Note that each code in system A can be converted to multiple B codes (in this example, code "2" in System A maps to both code "b" and "c" in System B). The converse is also true: both code "4" and "5" map to code "e". The weights code how important the respective industries are. This could for example be total value of shipments, total trade value, etc. Notice the weights are both on A and B, and that they are constand whenever they stand for the same industry.

We can define this mathematically as there being two sets A, B with measures w_A, w_B giving the mass on each code, and a concordance being a correspondence

$$\phi : A \rightrightarrows B.$$

We will write results in terms of this mathematical definition, but also in terms of examples to show the working of the system.

We will go through three operations relating to two-weighted concordances:

1. How to transform quantity variables such as total industry sales using a two-weighted concordance
2. How to transform property variables such as capital intensity using a two-weighted concordance
3. How to create a two-weighted concordance using a unweighted concordance and a weighting scheme for one of the variables (e.g. when we want to create a two-weighted concordance between HS and SITC and only have total trade in HS codes).

5.0.1 Transform quantity variables using two-weighted concordances

Starting with quantity variables, suppose that we have total trade flows in industry code A. We then want to allocate it across different codes in coordinate system B. In this case, for each element in A we look at all elements in B that it maps to. It then allocates the quantity in A across the elements in B in proportion to their weights. The quantity attributed to element B is then the sum of the contributions over all elements in A.

For example, suppose we have the following measures of total value of shipments in coordinate system A

A	vship
1	1000
2	3000
3	6000
4	2000
5	3000
6	4000

and we want to convert it to B using the previous correspondence. We will explain what value of shipments we will attribute to industry c in system B. The pre-image of "c" is "2" and "3" in system A, so we can look how much of the shipments of these two A-industries that will be attributed to "c". Industry "2" ships 3000 in value, and it corresponds to both industry "b" and "c" in System B. As the relative weights of "b" and "c" are 50 and 100 respectively, 1000 will be attributed to "b" and 2000 will be attributed to "c". However, in the concordance, we see that 3 only maps to "c", so all 6000 shipments from 3 will be attributed to c. Hence, total attribution to "c" is $2000 + 6000 = 8000$.

We can write this in terms of the mathematical representation Φ as well, together with the weights μ_A and μ_B. If

$$f_A : A \to \mathbb{R}$$

is an arbitrary quantity measure on A we convert it to B by

$$f_B(y) = \sum_{x \in \Phi^{-1}(y)} f_A(x) \times \frac{\mu_B(y)}{\sum_{y' \in \Phi(y)} \mu_B(y')}.$$

The equation is quite difficult to parse, but it says that we take all the values from the pre-image to y. The value of each of those pre-images x attributed to y is equal to the relative weight of $\mu_B(y)$ compared to the total weights of those codes in B that x maps to.

5.0.2 Transform property variables using two-weighted concordances

The situation is different when we have so-called property variables, for example capital intensity, skill intensity or other industry level properties. We can see how these differs by means of an example. Suppose that we have a concordance between HS 2007 six-digit and HS 2007 ten-digit data. If we want have data on trade flows on six-digit level and want to convert these to ten-digit level. Then, the reasonable thing is to split it up across the ten digits according to some weighting scheme.

However, if we instead have measured capital intensity on the six-digit level, the natural thing is to give this capital intensity as a prediction for the capital intensity in all ten-digit descendant categories (if we have no additional information on capital intensity on ten-digit level). Similarly, if we wanted to convert from ten-digit to six-digit, trade flows ought to be summed, whereas for properties it is appropriate to take a weighted average of industry-level properties on the ten digit level.

Thus, we see that property variables translate across coding systems in a fundamentally different way from quantity variables. We define the transformation scheme for property variables by saying that for each code $y \in B$ in the target system, we define its property as a weighted average of the properties that its pre-images

$x \in A$, where we use the weights on A as a weighting scheme. For example, in our example concordance, we would attribute c a property which is the weighted average of $2, 3$ in System A, using the measures $\mu_A(\{2\}) = 20$ and $\mu_A(\{3\}) = 15$ as weights.

More formally, if we have a property measure

$$g_A : A \to \mathbb{R}$$

defined on A, then we translate it to B using ϕ by the equation

$$g_B(y) = \frac{\sum_{x \in \phi^{-1}(y)} g_A(x)\mu_A(x)}{\sum_{x \in \phi^{-1}(y)} \mu_A(x)}.$$

5.0.3 Construct a two-side weighted concordance from a one-sided weighted concordance

Above we defined how you translate between different coordinate systems if you have a two-sided weighted concordance. However, sometimes we only have a one-sided concordance. For example, if we have total trade data in HS 2007 six-digtit and want to create a concordance between HS 2007 6-digit and NAICS 2007 it might be that we do not have data to create a natural weighting scheme for NAICS 2007 data.

For this case, we have a procedure to create a two-sided weighted concordance from a one-sided weighted concordance. It is quite similar to the quantity transformation above. Suppose that we have a concordance ϕ and a measure μ_A on A and want to create a measure μ_B on B. The question is how much weight we should attribute to each element $y \in B$. In this case, we go through each element $x \in A$ and take its weight $\mu_A(x)$ and portion it out equally on all elements y' in B that x maps to. This gives us how much weight element x gives to element y, that is $\frac{\mu_A(x)}{|\phi(x)|}$ where $|\phi(x)|$ gives how many codes x maps to. By summing over all x we get the total contribution to y. In mathematical terms

$$\mu_B(y) = \sum_{x \in \phi^{-1}(y)} \frac{\mu_A(x)}{|\phi^{-1}(x)|}.$$

5.0.4 Practical implementation

The process above allows us to define three primitive operations: creating a two-sided concordance, using it to convert between property variables, and use it to convert between quantity variables. We can use these three operations to create arbitrary chains of concordances between data. Below we list the actual concordances we create, which weights are used, and how we use these concordances to translate everything into NAICS 2012 six-digit data.

Created concordance sequence:

1. Create concordance between HS 2007 six-digit and HS 2007 ten-digit from one sided concordance with total world trade as weight on HS 2007 six digit.
2. Create concordance from HS 2007 10-digit to NAICS 2007 six digit from a one sided concordance using [...] as a basic concordance and the HS 2007 10-digit weights obtained from previous exercise
3. Create concordance from NAICS 2007 six digits to NAICS 2002 six digits using a one sided concordance with [...] as basic concordance and the NAICS 2007 six digits weights obtained from previous step
4. Create concordance from NAICS 2002 six digit to NAICS 1997 six digit analogously to previous step
5. Create concordance from NAICS 2007 six digit to NAICS 2012 six digit analogously to previous step
6. Create concordance between IO 2007 six-digit and NAICS 2007 six digit directly using [....] as basic concordance, total production as weight on IO-codes and previously constructed weights from step 2 for NAICS 2007 six digit
7. Create concordance between HS 2007 six digit and SITC rev.2 four digits using a one-sided concordance with [...] as basic concordance and total world trade as weight on HS 2007 six digit.

Once we created these concordances, we can translate all variables to NAICS 2012 six-digit code. We use the following transitions.

Source data set	Code	Path
NBER CES	NAICS 1997 6 digits	NAICS 1997 6 digits
		NAICS 2002 6 digits
		NAICS 2007 6 digits
		NAICS 2012 6 digits
IO-table	IO 2007 6 digits	IO 2007 6 digits
		NAICS 2007 6 digits
		NAICS 2012 6 digits
BACI Trade data	HS 2007 6 digits	HS 2007 6 digits
		HS 2007 10 digits
		NAICS 2007 6 digits
		NACIS 2012 6 digits
Rauch	SITC rev 2 4 digits	SITC rev 2 4d
		HS 2007 6 digits
		HS 2007 10 digits
		NAICS 2007 6 digits
		NAICS 2012 6 digits