Technical Appendix for:

THE NETWORK STRUCTURE OF INTERNATIONAL TRADE

by Thomas CHANEY, Toulouse School of Economics

C Additional economic assumptions and discussions

C.1 A model of trade with informational barriers

In this section, I embed a simple Krugman (1980) model of trade into the model of network formation described in Section 2. The only assumption added to the Krugman model is that firms can only sell their output to a consumer they have met through the directed network described above. In the next section, I propose a simple model with informational asymmetries and moral hazard that would justify such a selective trading strategy.

Preferences: There is a continuum of consumers in each country, that share the same CES preferences, but differ in the set of goods they have access to. Consumer i has the following preferences over the set Ω_i of goods it has access to,

$$U_i = \left(\int_{\Omega_i} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}$$

where the elasticity of substitution, σ, is larger than 1.

Technology: There is a continuum of monopolistic firms that face the same increasing returns technology. The labor required to produce q units is

$$l(q) = \alpha + \beta q$$

Informational frictions: The only departure from the classical Dixit-Stiglitz-Krugman set-up is that firms can only sell their output to consumers they know (through the network described above), and consumers can only buy from firms they know (through the same network). More precisely, each consumer has access to a mass m, of goods ($m_i = \int_{\Omega_i} d\omega$ for consumer i), with m distributed within the population according to the distribution $F(m)$. In the same way, each firm has access to a mass m, of consumers, again with m distributed within the population according to the distribution $F(m)$. For simplicity, I assume that there are no additional barriers to trade, in the form of either a fixed or a variable trade cost.

Prices, quantities and utilities: Given that each consumer has access to a continuum of differentiated goods (only the measure of those goods varies across consumers), their demand for each good is iso-elastic.[33] Facing an iso-elastic demand function, each firm charges a constant mark-up over marginal cost, $p_i(\omega) = p = \frac{\sigma}{\sigma-1}\beta w$ for any (i,ω), where w is the wage rate, which I normalize to $1/\beta$ so that $p = 1$. Facing those prices, if consumer i has access to m_i goods, she will buy $q_i = 1/m_i$ units of each good $\omega \in \Omega_i$. The welfare of this consumer is then simply $U_i = U(m_i) = 1/m_i^{\frac{1}{1-\sigma}}$.

General equilibrium: Imposing free entry of firms in each location will pin down the number of firms. This is left as an exercise. Alternatively, I can assume that firms are born from the process described above, and that aggregate profits are redistributed lump sum to the consumers who collectively own all the firms in the economy. Alternatively, breaking the symmetry between all locations, one would have to impose trade balance between locations in order to solve for relative wages.

Aggregate welfare: Given the simple structure of the economy, I can perform a series of comparative statics experiments. First, I describe the average utility reached by consumers in this economy,

$$\mathbb{E}_F[U] = \int_{m\geq 0} m^{\frac{1}{\sigma-1}} dF(m)$$

Jackson and Rogers (2007) derive a series of properties for a distribution similar to $F(m)$, which allow me to describe the impact of changing the technological parameters of the network formation (μ, π) on aggregate welfare.

Proposition 4 *Average utility increases with the efficiency of the direct search process.*

Proof. For π fixed, if $\mu > \mu'$, then the distribution of contacts F associated with μ first order stochastically dominates the distribution F' associated with μ' (see Jackson and Rogers (2007), Theorem 7 page 905). Since the utility associated with having access to a mass m of goods, $m^{\frac{1}{\sigma-1}}$, is increasing in m, the average utility $\mathbb{E}_F[U]$ associated with μ is higher than he average utility $\mathbb{E}_{F'}[U]$ associated with μ'. ■

Proposition 5 *If goods are sufficiently substitutable, $\sigma > 2$, then average utility decreases with the relative efficiency of the remote and direct search, π. Otherwise, for $1 < \sigma < 2$, average utility increases with π.*

[33] In a model with a finite discrete number of goods, the price elasticity of demand would depend on the number of accessible goods, asymptoting to the constant elasticity case only for a large number of goods. Having a continuum of goods for each consumer assumes away this complication.

Proof. For μ fixed, if $\pi > \pi'$, then the in-degree distribution F associated with π second order stochastically dominates the distribution F' associated with π' (see Jackson and Rogers (2007), Theorem 6 page 903). If $\sigma > 2$, the utility associated with having access to a mass m of goods, $m^{\frac{1}{\sigma-1}}$, is concave in m. The average utility $\mathbb{E}_F[U]$ associated with π is therefore lower than the average utility $\mathbb{E}_{F'}[U]$ associated with π'. If $\sigma < 2$, the utility is convex in m, and the opposite holds. ■

The intuition for those results is rather simple. First, increasing the efficiency of direct search, μ, increases the number of goods accessible to all consumers. Since consumers in this simple model have a love for variety, more links will unambiguously increase welfare for all consumers.

Second, decreasing the relative efficiency of remote search compared to direct search decreases the dispersion of the number of contacts across consumers. As explained in the main body of the text, remote search gives an advantage to agents who already have many contacts, which makes the access to new contacts more unequal. If goods are sufficiently substitutable, increasing the number of contacts brings about a smaller and smaller welfare gain. As a consequence, average utility is higher for a less "unequal" network. On the other hand, if goods are less substitutable, increasing the number of contacts brings about a larger and larger welfare gain. In that case, average utility will be higher for a more "unequal" network, where the welfare gain of the very connected agents dominates the welfare loss of the less connected agents.

Sales distribution: I can derive similar predictions for the distribution of sales across firms, as well as for aggregate production.

Firms differ in the mass of consumers they have access to, m. Moreover, each of their consumers themselves differ in the number of goods they have access to. Since by assumption all consumers have the same income, and since all goods have the same price, consumers with access to more goods will buy less of each good. The quantity of each good bought by a consumer who has access to $\tilde{m}$ goods is then simply $1/\tilde{m}$. The expected quantity sold to each consumer is therefore given by $\int_{\mathbb{R}} \frac{1}{\tilde{m}} dF(\tilde{m})$, and the total expected sales of a firm that has access to m consumers is,

$$pQ_F(m) = m \int_{\mathbb{R}} \frac{1}{\tilde{m}} dF(\tilde{m})$$

As for aggregate welfare, the characterization of the properties of the distribution $F(m)$ allows me to describe both aggregate sales and the sales distribution across firms.

First, the higher the efficiency of direct search, μ, the less a firm will sell to any single consumer: the more alternatives a consumer has, the fewer goods she will buy from any single supplier. Second, the lower the efficiency of remote search relative to direct search, π, the less a firm will sell to any single consumer: the lower π, the less dispersed the distribution $F(m)$ is; since a firm can increase its expected sales by shifting away from consumers who have many alternatives towards consumers who have few alternatives, the more dispersed the distribution $F(m)$ is, the more a firm can sell on average.

Note however that this simple model does not generate any interesting predictions on the intensive margin of sales, i.e. the average sales per consumer. More precisely, in expectation, the consumers reached by any firm have access to the same number of goods, irrespective of the number of consumers this firm reaches. Therefore, in expectation, all firms will sell the same quantities (and values) per consumer, irrespective of how many consumers a firm reaches. This result is obviously at odds with the fact that firms that sell to many markets tend to sell large quantities in each of these markets.

I leave an extension of this model that would incorporate a meaningful intensive margin of trade for future research.

C.2 Trading under the threat of moral hazard

In this section, I propose a simple model with informational asymmetries and moral hazard that rationalizes why a given firm would only trade with firms it has met through the network described in Section 2. This model is is purposefully simple and only meant as an illustration of a possible economic mechanism that would support the proposed dynamic network formation.

Set-up: There is a continuum of firms of mass 1. Each firm produces a differentiated good. Each firm can both buy differentiated inputs from other firms and sell its differentiated output to other firms. A good can be of either high quality (q_H) or low quality (q_L). Producing high quality goods is costly. The quality of a good is observable and can be contracted upon.

When a supplier meets a buyer, the match specific cost to the supplier of customizing its good for the client is c. The cost c is drawn over $\mathbb{R}^+$ from a known probability distribution G,

$$\Pr(\tilde{c} < c) = G(c)$$

For simplicity, I assume that the distribution of customization costs is independent across matches. This cost c is only observable to the supplier, and cannot be contracted upon. I normalize the cost of producing a low quality good to zero.

A high quality input has a value V for the client. The value of a low quality good for a client is normalized to zero. I assume for simplicity that those values are the same for all firms. Upon a successful match, one unit of output is traded. All firms are risk neutral. Upon meeting, the timing of the game played by a supplier and its client is as follows:

1. The client offers a price for a high quality good, and a price for a low quality good.

2. The supplier receives a customization cost draw, c, and decides whether to produce a high or low quality good.

3. After observing the good's quality, the client and supplier trade at the agreed prices.

I will look for sub-game perfect Nash equilibria of this game.

Solution to the match specific game: The supplier will produce a high quality good for any price above its cost draw. This happens with probability $G(p_H)$. Conditional on receiving a high quality good, the surplus of the client is $(V - p_H)$. The client therefore chooses p_H^* so as to maximize her expected profits,[34]

$$p_H^* = \arg\max_{p_H} (V - p_H) G(p_H)$$

Since a low quality good has no value, the client sets $p_L^* = 0$, and no low quality good is ever produced or traded. Facing those prices, the expected surplus of a match for a supplier, S, is given by,

$$S = \int_0^{p_H^*} (p_H^* - c)\, dG(c)$$

Direct search frictions: Each period, suppliers engage in a costly search for potential clients. The marginal cost of finding m new clients is $s(m)$, with $s(0) = 0$, $s \geq 0$, $s' > 0$ and $\lim_{m \to \infty} s(m) = +\infty$. Given the expected profit from finding a successful match S, a supplier will sample a mass M_d of firms at random, defined by $s(M_d) = S$. Given that a successful match is formed with a probability $G(p_H^*)$ with each client met, a supplier forms a mass $\gamma\mu$ of successful matches,

$$\gamma\mu \equiv \frac{M_d}{G(p_H^*)}$$

[34] Note that an interior solution exists under some regularity conditions on V and $G(\cdot)$. $V = 4$ and $G(c) = 1 - 1/c$ for instance admits the simple solution $p_H^* = 2$.

Note that I do not consider the role of geography in this simple example. It would be trivial to add a geographic dimension, where the M_d matches, and the subset of $\gamma\mu$ successful matches, are distributed over space according to the probability distribution $g(\cdot)$.

Remote search: Given the direct search frictions, clients who themselves are suppliers to other clients are in a privileged position to leverage the information about their own network of clients. The game played by the upstream supplier, her client, and the downstream clients of her client is similar to the game above. The initial client makes a take-it-or-leave-it offer to each of her suppliers to reveal information about each of her own clients for a fee $S = \int_0^{p_H^*} (p_H^* - c)\, dG(c)$. After being connected, the supplier and her new client bargain as above, and the new client sets a price p_H^* for a high quality good, which the supplier accepts only if her match specific cost draw is below p_H^*. S is exactly the surplus that a supplier can expect from meeting with a random client. Since the initial client does not observe the cost c that the supplier would have to incur to customize her good for a new client, no additional information permeates. Moreover, the initial client extracts all the informational rent from the supplier.

Each period, a supplier will therefore meet a fraction $\gamma\mu\pi \equiv G(p_H^*)$ of the clients of each of her own clients. Given the implicit assumption about constant returns to scale, there is no strategic consideration for initial clients to reveal (at a cost) their own client base to their suppliers. They simply extract a fee for each contact they reveal, without the fear of losing market shares to their own contacts.

Discussion: The proposed model would explain why each period, firms search for a fixed number of contacts directly, and in addition, get connected to a subset of the contacts of their contacts via remote search. This model mimics the one proposed in Section 2, and offers a simple micro-foundation for the parameters μ and π. Note however that the geographic dynamics that arises are somewhat more complex: remote search evolves over time as the contacts of each firm acquire a growing network of contacts themselves. I explore the properties of this more complex network in Chaney (2013, The Gravity Equation in International Trade: An Explanation," TSE *mimeo*).

C.3 Comparison with existing trade models

In this section, I derive formally the predictions of the two most prominent existing firm level trade models, Melitz (2003) and Bernard, Eaton, Jensen and Kortum (2003), regarding the cross-sectional distribution of the number of foreign markets reached by French exporters. I show that neither of those models makes robust prediction regarding this distribution without imposing specific parametric assumptions. I argue that there is no a priori reason to make any such assumption. For the purpose of this discussion, I assume for simplicity a firm's number of contacts, m, is equal to the number of countries it exports to, M.

Comparison with Melitz (2003) and Chaney (2008)

Strictly speaking, Melitz (2003) predicts a degenerate distribution $f(M)$ where all exporters export to each and every country in the world. This is due obviously to the simplifying assumption that all trade barriers, country sizes and labor productivities are perfectly symmetric. Chaney (2008) develops a simple extension of Melitz (2003) with asymmetric countries and trade barriers. I will describe the prediction of this model regarding the p.d.f. $f(M)$.

Set-up: As a reminder, the set-up in Chaney (2008) is as follows. I will only describe the patterns of entry of French exporters into foreign countries. Preferences are CES with an elasticity of substitution, σ. The cost of selling q units of good in country c for a French firm with productivity φ is

$$c_c(q) = \frac{w_F \tau_{F,c}}{\varphi} q + f_{F,c}$$

where w_F is the French wage, $\tau_{F,c}$ and $f_{F,c}$ are respectively the variable and fixed cost of exporting to country c for a French firm. Productivities are distributed Pareto,

$$\Pr(\Phi \leq \varphi) = 1 - \varphi^{-\gamma}$$

Chaney (2008) proves that all firms with a productivity above $\overline{\varphi}_{F,c}$ export to country c, with $\overline{\varphi}_{F,c}$ defined as,

$$\overline{\varphi}_{F,c} = \lambda \left(\frac{Y}{Y_c}\right) \left(\frac{w_F \tau_{F,c}}{\theta_c}\right) f_{F,c}^{1/(\sigma-1)}$$

with λ and θ_c constants, Y the world *GDP*, and Y_c the *GDP* of country c. There is a strict ordering of the productivity thresholds faced by French exporters. Arranging countries in increasing productivity thresholds, any firm with a productivity above $\overline{\varphi}_{F,c}$ exports at least to all markets $c' \leq c$.

Prediction regarding the p.d.f. $f(M)$: The fraction of firms that export to exactly M markets is simply given by the probability of receiving a productivity φ above $\overline{\varphi}_{F,M}$ but strictly below $\overline{\varphi}_{F,M+1}$. A firm with such a productivity will export to all countries $c \leq M$, but not to any country $c > M$. Using the assumption of Pareto distributed productivity shocks, I can derive a simple expression for the fraction of firms that export to exactly M countries, $f(M)$,

$$f(M) = (\lambda w_F)^{-\gamma} \left[\left(\frac{Y}{Y_M}\right)^{-\gamma} \left(\frac{\tau_{F,M}}{\theta_M}\right)^{-\gamma} f_{F,M}^{-\gamma/(\sigma-1)} - \left(\frac{Y}{Y_{M+1}}\right)^{-\gamma} \left(\frac{\tau_{F,M+1}}{\theta_{M+1}}\right)^{-\gamma} f_{F,M+1}^{-\gamma/(\sigma-1)} \right]$$

The only prediction of this model is that this density is non negative (it may be zero if the knife-edge case where two productivity thresholds are exactly equal arises). It is easy to see from this formula that one can make any $f(M)$ either arbitrarily small (by having $\left(\frac{Y}{Y_{M+1}}\right)\left(\frac{\tau_{F,M+1}}{\theta_{M+1}}\right) f_{F,M+1}^{1/(\sigma-1)}$ arbitrarily close to $\left(\frac{Y}{Y_M}\right)\left(\frac{\tau_{F,M}}{\theta_M}\right) f_{F,M}^{1/(\sigma-1)}$), or arbitrarily close to 1 (by having $\left(\frac{Y}{Y_M}\right)\left(\frac{\tau_{F,M}}{\theta_M}\right) f_{F,M}^{1/(\sigma-1)}$ arbitrarily small compared to $\left(\frac{Y}{Y_{M+1}}\right)\left(\frac{\tau_{F,M+1}}{\theta_{M+1}}\right) f_{F,M+1}^{1/(\sigma-1)}$). There is no reason a priori that the function $f(M)$ is even decreasing in M: if the thresholds of entry for two relatively accessible countries are arbitrarily close, and the thresholds of entry into two relatively inaccessible countries are arbitrarily distant, then $f(M)$ will be upward sloping.

One may argue that country sizes (the Y_M's) are approximately Pareto distributed. One may further argue that the fixed export costs are approximately linear in country size, or at least linear in logs. In such a case, and in the absence of variable trade barriers, the thresholds of entry into different foreign markets would be Pareto distributed. Formally, in such a case, one can write $\overline{\varphi}_M = \alpha M^{\beta}$, so that $f(M) = \alpha \left(M^{-\beta\gamma} - (M+1)^{-\beta\gamma} \right)$. While this relationship does not exhibit the curvature in a log-log scale that we see in the data, the right choice for $\beta\gamma$ would make the predicted line close to the empirical distribution $f(M)$.

However, this argument abstracts entirely from the existence of variable trade barriers. Such a model would make the counter-factual prediction that the number of French exporters is log-proportional to country size. Eaton, Kortum and Kramarz (2011) show that French exporters tend to cluster in countries that are geographically close to France, whether large or small, and not in potentially much larger countries that are faraway from France. In other words, they show evidence that variable trade barriers do play an important role in shaping the entry of French exporters into foreign markets, or that fixed export costs are not proportional to country size. For instance, in 1986, $17,695$ French firms export to Belgium, with a *GDP* of \$144 billion and $7,608$ export to the U.S., with a *GDP* of \$4.43 trillions. So Belgium, which is 31 times smaller than the U.S. attracts more than twice as many French exporters as the U.S. Even comparing two English speaking countries that are not contiguous to France, the U.K. with a GDP of \$570 billion is almost 8 times smaller than the U.S., but receives 30% *more* French exporters, not 87% *less*. These departures from a linear relationship between country size and number of French exporters are not restricted to these 3 countries, but occur systematically. Empirically, there is

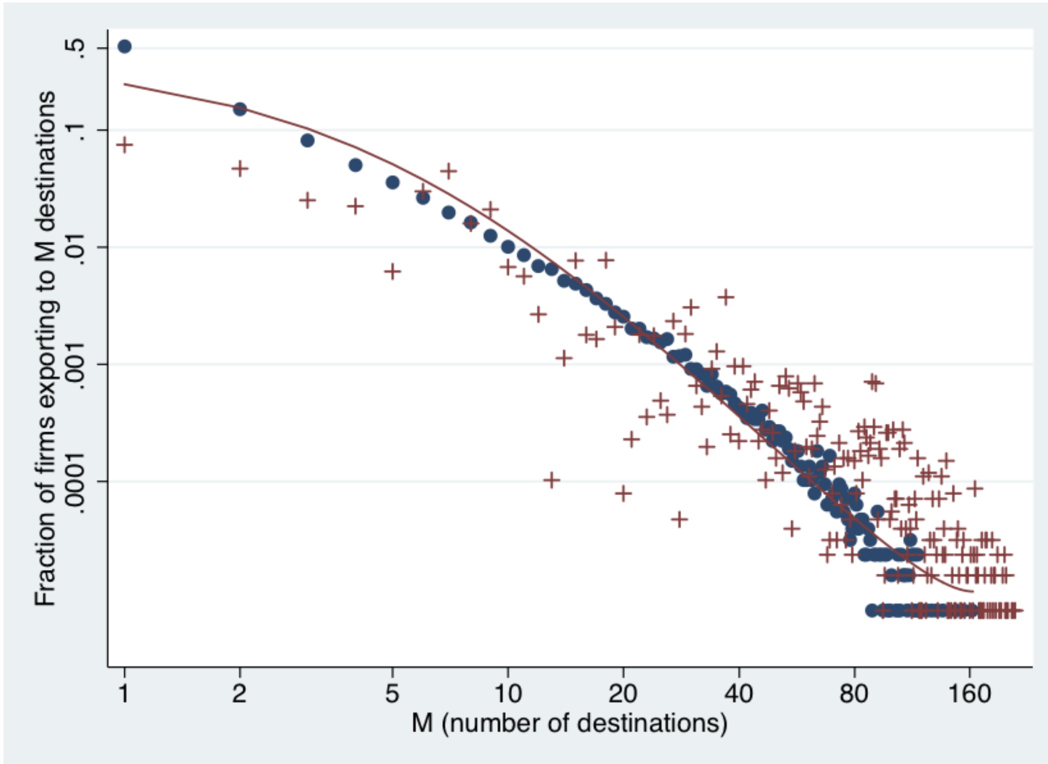

Figure 4: Network versus Melitz/Chaney model, $f(M)$ versus M.

Notes: $f(M)$ is the fraction of firms exporting to M destinations; blue dots: data, all French exporters in 1992; red line: calibrated network model; red plus signs: calibrated Melitz/Chaney model.

no systematic correlation between country size (measured as GDP) and the distance from France. So there is no reason to believe that the thresholds of entry into different markets are themselves Pareto distributed.

To further illustrate this point, I calibrate the Melitz/Chaney model so as to match exactly the number of French exporters in every foreign market. I use the same data on all French exporters in 1992 that I used in Section 3. Given the precise ordering of foreign markets predicted by the model, I can rank foreign markets in decreasing order of accessibility for French exporters. The number of firms that export to exactly M markets is then the difference between the number of firms that export to the M^{th} and $(M+1)^{th}$ market. As can be seen visually on Figure 4, the Melitz/Chaney model cannot replicate well the empirical distribution of the number of foreign markets accessed by individual firms.

I can also describe the predictions of the Melitz/Chaney model regarding the geographic dispersion of exports. In the Melitz/Chaney model, more productive firms are able to enter both more

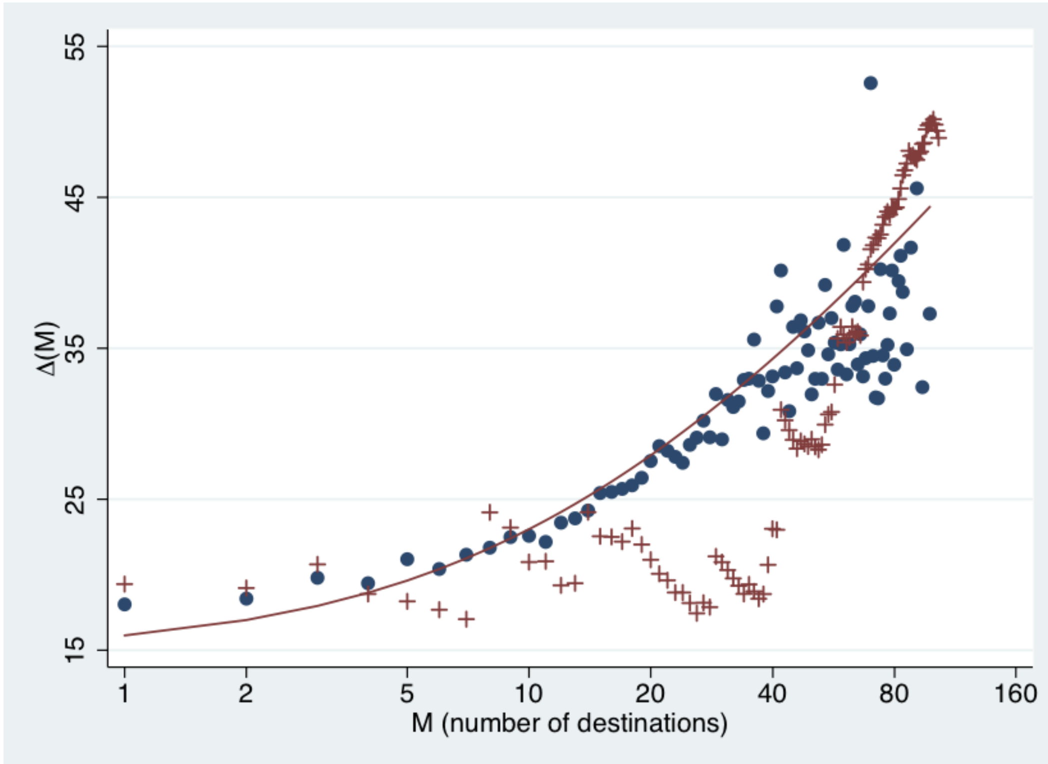

Figure 5: Network versus Melitz/Chaney model, $\Delta(M)$ versus M.

Notes: $\Delta(M)$ is the average (squared) distance of exports, among firms exporting to M destinations; blue dots: data, all French exporters in 1992; red line: calibrated network model; red plus signs: calibrated Melitz/Chaney model.

markets, and less accessible markets. Since less accessible markets tend to be geographically more remote, this model predicts that the geographic dispersion of exports tends to be larger among firms that export to many markets. However, to generate such a prediction, the Melitz/Chaney model again relies on a series of exogenous parameters that this model has nothing to say about a priori. Therefore, this model does not deliver any precise predictions regarding the shape of the relationship between geographic dispersion and the number of markets accessed. To illustrate this point, I calibrate the geographic dispersion of exports, $\Delta(M)$, in the Melitz/Chaney model. I control for country size in the same way as I do when constructing the empirical measure of $\Delta(M)$ presented in Section 3. I order foreign markets in decreasing order of accessibility for French exporters, controlling for market size as follows: call N_M the number of French firms that export to market M, where M is defined such that $\frac{N_1}{GDP_1} > \frac{N_2}{GDP_2} > \frac{N_3}{GDP_3} > \ldots$ The Melitz/Chaney model predicts that any firm that exports to exactly M markets will export to all countries $c \leq M$. As can be seen visually on Figure 5, the Melitz/Chaney model can only predict that the geographic dispersion of exports tends to increase with the number of foreign markets accessed, but it has nothing to say about the specific shape of this relationship. Not controlling for market size, the empirical fit of the Melitz/Chaney model would be substantially worse.

Note of course that the Melitz/Chaney model is not only meant to explain the extensive margin of international trade (the number of foreign markets accessed by exporters), but it also delivers a series of predictions on the intensive margin of international trade (the size of firm level exports), on how the size of sales in the domestic market helps predict which markets a firm enters, and how much it exports there. As shown by Eaton, Kortum and Kramarz (2011), these predictions are strongly supported by the data.

Comparison with Bernard, Eaton, Jensen and Kortum (2003)

Whereas in Bernard, Eaton, Jensen and Kortum (2003), the formula for the fraction of firms that export to country m is almost identical to the formula in Chaney (2008), there is not a strict hierarchy of foreign markets in terms of accessibility to French exporters. As a consequence, the formula for the fraction of firms that export to exactly M markets is substantially more complicated.

In the interest of clarity, I will therefore solve a simple special case with 3 countries (potentially asymmetric in labor productivity), and symmetric trade barriers. All the intuition derived in this special case carry over to the case of many countries with asymmetric bilateral trade barriers.

Set-up: As a reminder, the set-up in Bernard, Eaton, Jensen and Kortum (2003) is as follows. There is a continuum of sectors that produce differentiated goods. The distribution of labor productivity z of the most productive firm in country m in any of those sectors is Fréchet,

$$\Pr\left(Z_m \leq z\right) = \exp\left(-T_m z^{-\theta}\right)$$

The parameter θ is the same across countries, but T_m differs across countries. Productivity draws are independent across countries. Firms face no extra cost of selling domestically, but they face an iceberg cost τ when exporting to any foreign country.

Prediction regarding the p.d.f. $f(M)$: Let us set country 1 as France, and compute the fraction of French firms that sell in 1, 2 or 3 markets.

All firms with a productivity $z_1 > \max\{\tau z_2, \tau z_3\}$ sell in all three markets. With independent Fréchet distributions, the probability of such an event occurring, and therefore the fraction of firms from 1 that export to exactly three markets, $f(3)$, is given by,

$$f(3) = \frac{T_1}{T_1 + \tau^\theta T_2 + \tau^\theta T_3}$$

A firm in country 1 sells in exactly two markets if either it is the best in country 1 and 2, but not in 3, or it is the best in country 1 and 3, but not in 2. Formally, a firm with productivity z_1 sells in exactly two markets if either $\{z_1/\tau > z_2 \text{ and } z_1/\tau < z_3 < z_1\}$ or $\{z_1/\tau > z_3 \text{ and } z_1/\tau < z_2 < z_1\}$ are true. The respective probabilities of each of those mutually exclusive events are,

$$\begin{aligned} \Pr\{z_1/\tau > z_2 \text{ and } z_1/\tau < z_3 < z_1\} &= \frac{T_1}{T_1 + \tau^\theta T_2 + T_3} - \frac{T_1}{T_1 + \tau^\theta T_2 + \tau^\theta T_3} \\ \Pr\{z_1/\tau > z_3 \text{ and } z_1/\tau < z_2 < z_1\} &= \frac{T_1}{T_1 + T_2 + \tau^\theta T_3} - \frac{T_1}{T_1 + \tau^\theta T_2 + \tau^\theta T_3} \end{aligned}$$

The fraction of firms from 1 that export to exactly two markets, $f(2)$, is given by,

$$f(2) = \frac{T_1}{T_1 + \tau^\theta T_2 + T_3} - \frac{T_1}{T_1 + \tau^\theta T_2 + \tau^\theta T_3} + \frac{T_1}{T_1 + T_2 + \tau^\theta T_3} - \frac{T_1}{T_1 + \tau^\theta T_2 + \tau^\theta T_3}$$

The formula for the fraction of firms that sell in exactly one market (in the home market necessarily) is even more complicated, and contains a total of 16 terms. I will spare the reader and skip this formula, concentrating instead on exporters only. Note the slight abuse of notation due the fact that there is a non empty set of firms which, despite being the most productive among home firms, do not sell in any market (not even at home). The exact fractions of firms selling to exactly 2 and 3 markets are the above formulas divided by the same number, the probability of selling in at least one market.

From the formulas above, it is easy to see that the distribution $f(M)$ does not have to be downward sloping. As trade barriers become large ($\tau \to +\infty$), no firm is able to export anywhere, $f(2) = f(3) = 0$, and all firms sell at home, $f(1) = 1$. On the opposite extreme, when trade barriers vanish ($\tau \to 1$), any firm that survives will sell in all three markets, $f(3) = \frac{T_1}{T_1+T_2+T_3}$, but no firm sells to exactly 1 or 2 markets, $f(1) = f(2) = 0$. In the first case, the distribution $f(M)$ is decreasing in M, whereas in the second case, it is increasing in M. The fraction of firms able to enter all 3 markets monotonically decreases with the level of trade barriers, τ, but the fraction of firms that sell to exactly 2 markets is not monotone in the level of trade barriers. $f(2)$ increases with τ for τ small, and decreases after. For different levels of trade barriers, the fraction of firms that sell to exactly 2 markets will be alternatively larger or smaller than the fraction of firms selling to 3 markets.

As in the Melitz/Chaney model, the distribution $f(M)$ in the Bernard, Eaton, Jensen and Kortum model not only can take any shape, but does not even have to be downward sloping. The specific shape of this distribution depends on assumptions regarding the distribution of exogenous parameters (country sizes or relative productivities, T_m's, and bilateral trade barriers, τ_{nm}'s), about which the model has nothing to say.

D Additional mathematical proof

Proposition 6 *The variance of the distance from a firm's contacts increases with the firm's number of contacts. The variance of the squared distance from a firm's contacts, for a firm with m contacts, $\sigma^2(m)$, is given by,*[35]

$$\sigma^2(m) = \Delta(m)^2 \left(\frac{\sigma_g^2 + \Delta_g^2}{\Delta_g \Delta(m)} + \frac{3}{1 + \frac{1}{m\pi}} - 1 \right)$$

where $\Delta_g \equiv \sum_{x \in \mathbb{Z}} x^2 g(x)$ is the average squared distance of contacts, $\sigma_g^2 \equiv \sum_{x \in \mathbb{Z}} x^4 g(x) - \left(\sum_{x \in \mathbb{Z}} x^2 g(x)\right)^2$ is the variance of that squared distance.

Proof. The Fourier transform $\hat{f}_t$ of the distribution of contacts of a firm of age t, with $\alpha = \gamma\mu\pi$ and $\hat{g}$ the Fourier transform of the probability distribution of the geographic distribution of random contacts g, is given by,

$$\hat{f}_t = \left((1 + \alpha\hat{g})^t - 1\right)/\pi$$

The total number of contacts is given by,

$$m_t = \left((1 + \alpha)^t - 1\right)/\pi$$

The characteristic function of the probability distribution of the number of contacts for a firm of age t is given by,

$$\hat{g}_t = \frac{\hat{f}_t}{m_t} = \frac{(1 + \alpha\hat{g})^t - 1}{(1 + \alpha)^t - 1}$$

[35] As for Propositions 1 and 2, I present this proposition for γ small only to get a more elegant expression. The (inelegant) general formula for time intervals of any length is given in the proof below. Note also that the m's only take a discrete set of values (corresponding to m_1, m_2, ...etc). The formula for $\sigma^2(.)$ for those values is exact and neither an approximation nor a limit.

To characterize the various moments of this distribution, I simply have to calculate the various derivatives of $\hat{g}_t$.

$$\begin{aligned}
\hat{g}_t^{(1)} &= \frac{\alpha t \hat{g}' \left(1+\alpha \hat{g}\right)^{t-1}}{\left(1+\alpha\right)^t - 1} \\
\hat{g}_t^{(2)} &= \alpha t \frac{\hat{g}" \left(1+\alpha\right)^{t-1} + \alpha \left(t-1\right) \hat{g}'^2 \left(1+\alpha \hat{g}\right)^{t-2}}{\left(1+\alpha\right)^t - 1} \\
\hat{g}_t^{(3)} &= \alpha t \frac{\hat{g}^{(3)} \left(1+\alpha \hat{g}\right)^{t-1} + 3\alpha \left(t-1\right) \hat{g}' \hat{g}" \left(1+\alpha \hat{g}\right)^{t-2} + \alpha^2 \left(t-1\right)\left(t-2\right) \hat{g}'^3 \left(1+\alpha \hat{g}\right)^{t-3}}{\left(1+\alpha\right)^\tau - 1} \\
\hat{g}_t^{(4)} &= \alpha t \frac{\hat{g}^{(4)} \left(1+\alpha \hat{g}\right)^{t-1} + \cdots + 3\alpha \left(t-1\right) \hat{g}"^2 \left(1+\alpha \hat{g}\right)^{t-2} + \cdots + \cdots + \ldots}{\left(1+\alpha\right)^t - 1}
\end{aligned}$$

where the "…" terms will correspond to odd moments that are zero because of the symmetry of the distribution g. From those derivatives, I can derive both the average (squared) distance from a firm's contacts, and the variance of this (squared) distance,

$$\begin{aligned}
\Delta_t = \sum_{x \in \mathbb{Z}} x^2 g_t\left(x\right) = \mathbb{E}\left[X_t^2\right] = \hat{g}_t^{(2)}\left(0\right) &= \frac{\alpha t \left(1+\alpha\right)^{t-1}}{\left(1+\alpha\right)^t - 1} \mu_g^{(2)} \\
\mathbb{E}\left[X_t^4\right] = \hat{g}_t^{(4)}\left(0\right) &= \frac{\alpha t \left(1+\alpha\right)^{t-1}}{\left(1+\alpha\right)^t - 1} \left(\mu_g^{(4)} + \frac{3\alpha\left(t-1\right)}{1+\alpha} \left(\mu_g^{(2)}\right)^2\right) \\
\sigma_t^2 = \mathbb{E}\left[X_t^4\right] - \left(\mathbb{E}\left[X_t^2\right]\right)^2 &= \frac{\alpha t \left(1+\alpha\right)^{t-1}}{\left(1+\alpha\right)^t - 1} \left(\mu_g^{(4)} + \frac{3\alpha\left(t-1\right)}{1+\alpha} \left(\mu_g^{(2)}\right)^2 - \frac{\alpha t \left(1+\alpha\right)^{t-1}}{\left(1+\alpha\right)^t - 1} \left(\mu_g^{(2)}\right)^2\right)
\end{aligned}$$

with $\mu_g^{(k)} = \sum_{x \in \mathbb{Z}} x^k g\left(x\right)$ the k^{th} moment of g. Note that the standard deviation of the (squared) distance from a firm's contacts grows asymptotically as fast as the average (squared) distance itself,

$$\lim_{t \to +\infty} \frac{\sigma_t}{\Delta_t} = \sqrt{2}$$

Now, I can use the relationship between the age t of a firm and number of contacts m_t to express everything as a function of a firm's contacts, m.

$$t\left(m\right) = \frac{\ln\left(1+\pi m\right)}{\ln\left(1+\alpha\right)}$$

from which I get,

$$\begin{aligned}
m &= \frac{1}{\pi}\left(\left(1+\alpha\right)^t - 1\right) \\
\left(1+\alpha\right)^t - 1 &= \pi m \\
\left(1+\alpha\right)^{t-1} &= \frac{\pi m + 1}{1+\alpha}
\end{aligned}$$

Using the above expression for σ_t^2, I get,

$$\sigma^2(m) = \frac{\alpha}{(1+\alpha)\ln(1+\alpha)}\left(1+\frac{1}{\pi m}\right)\ln(1+\pi m)$$
$$\left(\mu_g^{(4)} + \frac{3\alpha\ln(1+\pi m)}{(1+\alpha)\ln(1+\alpha)}\left(\mu_g^{(2)}\right)^2 - \frac{3\alpha}{1+\alpha}\left(\mu_g^{(2)}\right)^2 - \frac{\alpha\ln(1+\pi m)\left(1+\frac{1}{\pi m}\right)}{(1+\alpha)\ln(1+\alpha)}\left(\mu_g^{(2)}\right)^2\right)$$

For α small (which corresponds to γ small since $\alpha = \gamma\mu\pi$), the following first order approximation is exact,

$$\lim_{\alpha\to 0}\frac{\alpha}{(1+\alpha)\ln(1+\alpha)} = 1$$
$$\lim_{\alpha\to 0}\frac{3\alpha}{1+\alpha} = 0$$

Using the notation $\mu_g^{(2)} = \Delta_g$ and $\mu_g^{(4)} - \left(\mu_g^{(2)}\right)^2 = \sigma_g^2$, using the expression for the average (squared) distance, $\Delta(m)$ from Proposition 2, I simplify further, and I get the proposed expression,

$$\sigma^2(m) = \Delta(m)^2\left(\frac{\sigma_g^2 + \Delta_g^2}{\Delta_g\Delta(m)} + \frac{3}{1+\frac{1}{\pi m}} - 1\right)$$

■

E Instructions for replicating the simulation

In this section, I present detailed instructions for replicating the simulations in the main paper.

Simulating a population of firms that export to various countries

To simulating the model for each of the three choices for the set $\mathcal{S}$, I follow almost exactly the same algorithm as for the simulation in the SMM estimation, except that I simulate 250 cohorts instead of 360. I follow steps 1 through 5 of the simulation algorithm described in Section 3.1 on page 23, with the parameters $\gamma = .02$, $\mu = 1$, $\pi = 1$ and $\lambda = 3.5$.

$\mathcal{S}_{cities}$: placing 8766 equidistant locations on a circle

I simply place 8766 exactly equidistant locations on a circle of circumference $\mathcal{C} = 40,000$ km's, which corresponds approximately to the circumference of the globe. The distance between two locations is the length of the shorter of the two arcs between them.

$\mathcal{S}_{sphere}$: placing 8766 approximately equidistant locations on a sphere

The algorithm for placing $N = 8766$ locations on a sphere is as follows.

First, I place $n = 166$ equidistant locations along the equator.

Second, I place $(n-2)/4 = 41$ equidistant parallels in both the Northern and Southern hemispheres. Those parallels have latitudes $\pm i2\pi/166$ for $i = 1, \cdots, 41$. I place a location along the Greenwich meridian (longitude zero) at each of those $2 \times 41 = 82$ parallels.

Third, for each parallel $i = 1, \cdots, 41$ in the northern and southern hemispheres, I place a number $n \cos(i2\pi/n)$ rounded up to the nearest integer of equidistant locations along parallel i. Note that the distance between two neighboring locations along a parallel is not exactly the same as along the equator for two reasons. First, $n \cos(i2\pi/n)$ may not be an integer, so there is a rounding up error. Second, even if $n \cos(i2\pi/n)$ were an integer, the distance along a parallel between two neighbors would be the same as along the equator, but that distance is not exactly equal to the shortest path between the two neighbors.[36]

I now have $N = 166 + 2\sum_{i=1}^{41} \left\lfloor n \cos(i2\pi/n) + \frac{1}{2} \right\rfloor = 8766$ approximately equidistant locations covering the entire globe, with their corresponding longitudes and latitudes. Using a transformation of the spherical law of cosines that is not subject to rounding error when calculated on a computer, I calculate $||i-j||$, the shortest path between locations i and j with coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2) as,

$$||i-j|| = \frac{\mathcal{C}}{\pi} \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

where $\mathcal{C}$ is the circumference of the sphere. I set $\mathcal{C} = 40,000$ km's, which should have been the exact circumference of the globe had the French revolutionaries done a more proper job at defining the metric system in 1795.

$\mathcal{S}_{cities}$: placing 8766 actual cities on a sphere

For the set $\mathcal{S}_{cities}$, I use the same formula with the same $\mathcal{C}$ for calculating distances between 8766 the actual cities in my dataset, using their actual coordinates.

[36] Think of two locations on the same parallel, very near the South pole, but diametrically placed, with longitudes 0^0 and 90^0 for instance. Weather permitting, it is about 36% faster $(1 - 2/\pi \approx .36)$ to walk over the South pole to go from one to the other (if the parallel has circumference $2\pi r$, the distance would be approximately $2r$), rather than walking all the way along the parallel (for a distance of πr). $2r$ is approximately 36% shorter than πr. This is fitting for the 36^{th} footnote of this paper. For locations away form the poles, the difference is small.