Lecture 4: Firm Heterogeneity, Endogenous Markups, and Factor Endowments

Instructor: Thomas Chaney
Econ 357 - **International Trade (Ph.D.)**

1 Introduction

We have seen in the previous lecture how to add firm heterogeneity to the classical Krugman model of trade. We have seen in the Melitz paper how taking into account firm heterogeneity may explain why exposure to trade will induce some increase in aggregate productivity, even without a change in the actual technology of production, through a more efficient reallocation of factors of production. We have seen in Chaney that taking into account firm heterogeneity will change substantially some predictions for the patterns of international trade. Among others, firm heterogeneity will introduce a new margin of adjustment of trade barriers, the extensive, and this barrier behaves differently from the intensive margin traditionally studied. One prediction is that the elasticity of substitution between goods will no longer increase the sensitivity of trade flows to trade barriers, it may actually dampen it.

In this lecture, we will try and refine our approach to international trade with heterogeneous firms to account for some stylized facts that these models were missing. The first important caveat of models based on CES preferences and monopolistic competition is that mark-ups are constant. We will see in Melitz and Ottaviano (2003), as well as in Bernard, Eaton, Jensen and Kortum (2003) how more elaborate models can account for the endogenous determination of mark-ups. These models will give us a better understanding of the adjustments that take place when a country is exposed to trade. In Bernard, Redding and Schott (2004), we will see how the Melitz model can be augmented to include different factors of production, and how we can then relate this model with heterogeneous firms to the Ricardian comparative advantage model.

2 Melitz and Ottaviano (2005)

Melitz and Ottaviano keep the monopolistic competition assumption, but they move away from CES preferences. This will allow them to generate endogenous mark-ups, and derive interesting properties related to market size. Mark-ups respond to the "toughness" of competition (which we will define precisely). In larger markets, competition will be tougher, so that firms charge lower mark-ups, and aggregate productivity is higher. Integration through costly trade will not entirely kill this effect, so that larger countries, even if they are open to trade at some cost, will still be characterized by tougher competition than others. They are also able to describe the effect of some stylized trade policies. Trade liberalization increases import competition and therefore reduces mark-ups, and increases aggregate productivity (as in Melitz).

Main assumptions:

- Firms are heterogeneous in terms of productivity, as in Melitz (their productivity is revealed after they pay a sunk entry cost).
- Preferences give a linear demand system (developed in Ottaviano, Tabuchi and Thisse [2002]), which will generate endogenous mark-ups.
- As in Chaney (2006), there is a homogenous good produced with constant returns to scale so that wages in every country are pinned down.
- As in Chaney (2006), the world is made of asymmetric countries of different size, separated by potentially asymmetric trade barriers. There are only variable trade barriers, and no fixed cost of exporting. Given the linear demand system, no fixed cost is needed to get an endogenous selection of firms into the export market.

Closed economy equilibrium

Preferences and consumer behavior:

- Each consumer c derives utility from a homogenous good (used as the numeraire, as in Chaney), and a continuum of differentiated varieties, $\omega \in \Omega$,

$$U^c = q_o^c + \alpha \int_{\omega\in\Omega} q^c(\omega)\, d\omega - \frac{1}{2}\gamma \int_{\omega\in\Omega} q^c(\omega)^2\, d\omega - \frac{1}{2}\eta \left(\int_{\omega\in\Omega} q^c(\omega)\, d\omega\right)^2 \tag{1}$$

- γ (> 0) indexes the degree of product differentiation between varieties. In the limit of $\gamma = 0$, goods become perfectly homogenous, so that consumers only care about the total amount of differentiated goods they consume, not which specific variety. As γ increases, consumers care more and more about the distribution of consumption over all varieties, so that goods become more and more differentiated.
- α and η (> 0) index the substitution between the differentiated varieties and the homogenous good. Both parameters shift out the demand for differentiated varieties relative to the homogenous good.

- Note that the marginal utility of all goods is bounded from above, so that consumers may not consume all goods, even in the absence of fixed costs. We will use this condition to derive in equilibrium the set of goods that are produced, and define the extensive margin of trade. We assume though that the income is always large enough so that consumers have a positive demand for the numeraire.

- Given these preferences, consumers have a linear demand for each good. If consumers consume a positive quantity of good ω, then the inverse demand function from each consumer is,

$$p(\omega) = \alpha - \gamma q^c(\omega) - \eta Q^c \tag{2}$$

with $Q^c = \int_{\omega \in \Omega} q^c(\omega)\, d\omega$ the aggregate demand for differentiated varieties from consumer c.

- With L identical consumers, we can integrate this expression over all consumers c, rewrite Q^c as a function of the prices $p(\omega)$'s, rearrange and derive the total demand for good ω as,

$$q(\omega) \equiv Lq^c(\omega) = \frac{\alpha L}{\eta N + \gamma} - \frac{L}{\gamma} p(\omega) + \frac{\eta N}{\eta N + \gamma} \frac{L}{\gamma} \bar{p}, \ \forall \omega \in \Omega^* \tag{3}$$

where $\Omega^* \subset \Omega$ is the set of goods that are consumed, N is the measure of consumed varieties, and $\bar{p}$ is the average price of consumed varieties, $\bar{p} = (1/N) \int_{\omega \in \Omega^*} p(\omega)\, d\omega$.

- Since there is a "choke price" for each variety (given the bound on the marginal utility of each variety), we can define the set of goods actually consumed, Ω^*, as the largest subset of Ω such that,

$$p(\omega) \leq \frac{\gamma\alpha + \eta N \bar{p}}{\eta N + \gamma} \tag{4}$$

We can easily see that any price $p(\omega)$ must be below α, the marginal utility of the numeraire good (which we have assumed is consumed), so that $\bar{p} \leq \alpha$.

- With this setting, the demand elasticity that each producer ω is facing is not constant (not equal to γ). As the average price $\bar{p}$ goes down, or as the number of competitor N increases, the environment gets "tougher", and the price elasticity of demand increases. This will be the driving for behind the endogenous adjustment of markups of firms.

- We can derive a simple expression for the indirect utility of each consumer c,

$$U^c = I^c + \frac{1}{2}\left(\eta + \frac{\gamma}{N}\right)^{-1} (\alpha - \bar{p})^2 + \frac{1}{2}\frac{N}{\gamma}\sigma_p^2 \tag{5}$$

with I^c the income of consumer c, and $\sigma_p^2 = (1/N)\int_{\omega\in\Omega^*} (p(\omega) - \bar{p})^2 \, d\omega$ the variance of prices.

  - Welfare rises with a decrease in average prices, $\bar{p}$.
  - Welfare also rises with an increase in the variance of price, σ_p^2, as consumers reoptimize their consumption across varieties.
  - Finally, as in the CES case, consumers exhibit "love for variety", as welfare increase with the number of varieties available, N (holding the distribution of prices, $\bar{p}$ and σ_p^2, constant).

Production and firm behavior:

- Labor is the only factor of production. The unit labor requirement in the homogenous sector is one. This good is used as the numeraire. We assume that all countries will be producing at least some of the homogenous good so that the wages in all countries is equalized to one, as in Helpman and Krugman, and in Chaney.

- There is a sunk cost (in labor units) of starting production, f_E. Once this cost is paid, each firm receives a random productivity draw. Production is constant returns to scale, with unit labor cost c, drawn from a (known) distribution $G(c)$ over $[0, c_M]$.

- Monopolistic competition means that firm maximize profits choosing price or quantity, taking as given the residual demand for their good (i.e. the prices set by their competitors, $\bar{p}$ and N). Using the expression for demand in Eq. (3), total profits earned by a firm with cost c selling quantity q is

$$\begin{aligned}\pi(q) &= (p(q) - c)\, q \\ &= \left(\frac{\alpha\gamma}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} - \frac{\gamma}{L} q - c\right) q\end{aligned}$$

The optimal quantity to maximize this profit must be such that,

$$2\frac{\gamma}{L} q = \frac{\alpha\gamma}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} - c$$

Plugging back the price from Eq. (3), which gives $\frac{\alpha\gamma + \eta N \bar{p}}{\eta N + \gamma} = p + \frac{\gamma}{L} q$, we get the optimal pricing strategy,

$$q(c) = \frac{L}{\gamma}(p(c) - c) \tag{6}$$

- Firms with too high a cost, i.e. a cost c above the threshold $c_D = \frac{\alpha\gamma + \eta N \bar{p}}{\eta N + \gamma}$ have zero demand, and exit immediately. A firm with a cost c_D is exactly indifferent between staying in business or exiting, $p(c_D) = c_D$. We assume that the upper bound on cost, c_M, is always have enough so that in equilibrium there are some firms in the differentiated sector ($c_D < c_M$). The threshold c_D summarizes all the information that is needed to describe the behavior of the firms that stay in business.

$$\begin{aligned}p(c) &= \frac{1}{2}(c_D + c) && \text{(price)}\\ \mu(c) &= \frac{1}{2}(c_D - c) && \text{(mark-up)}\\ q(c) &= \frac{L}{2\gamma}(c_D - c) && \text{(quantity)}\\ r(c) &= \frac{L}{4\gamma}\left(c_D^2 - c^2\right) && \text{(revenue)}\\ \pi(c) &= \frac{L}{4\gamma}(c_D - c)^2 && \text{(profits)}\end{aligned}$$

We get the nice following properties:

  - Lower cost firms charge lower prices.

- Lower cost firms earn higher revenues and profits.
- Lower cost firms set higher mark-ups. So unlike the CES, not the entire productivity gain is passed on to consumers, part of it is retained as higher mark-ups.

Free entry condition:

- To determine the general equilibrium of this economy, we need to solve for the total number of entrants, and the cost threshold. To do so, we use the free entry condition. The expected profits of a potential entrant are $\int_0^{c_D} \pi(c)\, dG(c) - f_E$. These expected profits must be driven down to zero from free entry. On top of it, we know that the threshold for survival is given by $c_D = \frac{\alpha\gamma + \eta N \bar{p}}{\eta N + \gamma}$, which we can rearrange to get the zero cutoff profit condition. So the equilibrium is given by two conditions, the free entry condition, (FE), and the zero cutoff profits condition, (ZCP):

$$\frac{L}{4\gamma}\int_0^{c_D} (c_D - c)^2\, dG(c) = f_E \qquad (FE)$$

$$N = \frac{2\gamma}{\eta}\frac{a - c_D}{c_D - \bar{c}} \qquad (ZCP)$$

with $\bar{c} = \int_0^{c_D} c\,dG(c)/G(c_D)$ the average cost conditional on survival.

- We can already describe a few properties of the equilibrium:
  - Average productivity, $\bar{c}$, is higher when sunk costs are lower.
  - Average productivity, $\bar{c}$, is higher when varieties are closer substitute (low γ).
  - Average productivity, $\bar{c}$, is higher in larger markets (high L).
  - The demand parameters α and η do not affect the selection of firms, it only affects the total number of firms.
  - Competition is "tougher" in larger markets, so that average prices in such markets are lower, and all firms respond by charging lower mark-ups.

- We assume for simplicity that the cost is drawn from a Pareto distribution with a scaling parameter $k \geq 1$,

$$G(c) = \left(\frac{c}{c_M}\right)^k, \; c \in [0, c_M] \qquad (7)$$

This is exactly the same distribution as in Chaney (2006). There, the labor productivity, $\varphi = 1/c$, was drawn from a Pareto over $[1, +\infty)$ with a scaling parameter γ, so it's exactly the same as here replacing $k = \gamma$ and $c_M = 1$. As in Chaney (2006), k is an inverse measure of the dispersion of labor productivity, or labor cost. A high k means that most of the cost draws are concentrated around c_M, whereas $k = 1$ corresponds to a uniform distribution over $[0, c_M]$. With this specific functional form, we get the simple closed form solutions,

$$c_D = \left(\frac{\gamma\phi}{L}\right)^{\frac{1}{k+2}} \tag{FE}$$

$$N = \left(\frac{2\,(k+1)\,\gamma}{\eta}\right)\frac{\alpha - c_D}{c_D} \tag{ZCP}$$

with $\phi = 2\,(k+1)\,(k+2)\,c_M^k f_E$. We can describe with specific functional assumption several properties of the equilibrium.

Size matters:

- Larger markets are characterized by:
  - a higher average firm size, and average profits (the market size effect dominates the indirect effect of lower prices and mark-ups).
  - higher average mark-ups (the direct effect of tougher competition outweighs the selection effect on more productive firms with lower mark-ups).
  - invariant average profitability ($\bar{\pi}/\bar{r}$) (both average profits and average sales increase proportionally).
  - a lower variance of cost, prices and mark-ups (the support of the distribution of cost draws shrinks with tougher competition).
  - a higher variance of firm size (the direct market size magnification effect dominates the reduction in the support of productivity draws).
- These results are consistent with stylized facts from the IO literature, mainly Campbell and Hopenhayn (2002) and Syverson (2004, 2005). Campbell and Hopenhayn look at the retail sector, and find that larger markets have higher average size (measured

in sales or employment), as well as more dispersed sizes. Syverson looks at sectors where real output (quantities) is measurable (cement and concrete), so that he can recover unit prices (this is a unique example of reliable unit price data at the firm level!). He finds larger plants, higher productivity, and tougher competition (less dispersed productivity as well as a higher lower bound for productivity) in larger markets.

Open economy equilibrium

- We now consider opening up trade between two of these economies, home and foreign (denominated with a *), with respective size L and L^* (potentially asymmetric), and separated by some trade barriers, modelled as iceberg transport cost, τ. We could have asymmetric transport costs, and more than two countries too, the model would still be (almost) as tractable. This tractability mainly comes from the assumption of a freely tradable homogenous sector with constant returns to scale.

- The price that a firm charges in a given market (domestic or export) depends on the local demand. The unit cost for a firm with cost draw c is c for the domestic market, τc for the export market. Because of the constant returns to scale assumption, firms decide separately how much to produce on each market, home and foreign. As in the autarky case, we can easily derive the "choke" price on each market, that is the price below which a firm would have zero demand, and therefore the cost threshold for domestic production,

$$p_D = \frac{\alpha\gamma + \eta N \bar{p}}{\eta N + \gamma} = c_D \tag{8}$$
$$p_D^* = \frac{\alpha\gamma + \eta N^* \bar{p}^*}{\eta N^* + \gamma} = c_D^*$$

as well as the "choke" price for export and the cost threshold for exports,

$$p_X = \frac{p_D^*}{\tau} = c_X \tag{9}$$
$$p_X^* = \frac{p_D}{\tau} = c_X^*$$

- When considering whether or not to enter, a firm compare the cost of entry, f_E, to the benefit from entering, i.e. the expected discounted sum of future profits. Those profits include both profits earned from domestic sales, and potentially profits earned from foreign sales. The free entry condition imposes that these two are equalized,

$$\int_0^{c_D} \pi_D(c)\,dG(c) + \int_0^{c_X} \pi_X(c)\,dG(c) = f_E \tag{10}$$
$$\int_0^{c_D^*} \pi_D^*(c)\,dG(c) + \int_0^{c_X^*} \pi_X^*(c)\,dG(c) = f_E$$

If we use the specific functional form of Pareto distributed cost shocks, after rearranging, we get the simple closed solutions for the open economy equilibrium,

$$c_D = \left(\frac{1}{1+\tau^{-k}}\frac{\gamma\phi}{L}\right)^{\frac{1}{k+2}} \left(< c_D^{Autarky}\right) \qquad (FE)$$
$$c_D^* = \left(\frac{1}{1+\tau^{-k}}\frac{\gamma\phi}{L^*}\right)^{\frac{1}{k+2}} \qquad (FE^*)$$
$$N = \left(\frac{2(k+1)\gamma}{\eta}\right)\frac{\alpha - c_D}{c_D} \left(> N^{Autarky}\right) \qquad (ZCP)$$
$$N^* = \left(\frac{2(k+1)\gamma}{\eta}\right)\frac{\alpha - c_D^*}{c_D^*} \qquad (ZCP^*)$$

Note that in that case of an open economy, N represents the total number of varieties consumed in the home market. It is made of the sum of domestic producers, and exporters from the foreign country: if N_E (N_E^*) is the number of entrants in the home country (foreign), of which only a subset survive on the domestic market/export, then $N = G(c_D)\,N_E + G(c_X^*)\,N_E^*$. We can solve for the total number of entrants in each country (see the paper). In the special case of the Pareto distribution, we can see that the total number of varieties consumed increases.

- We can immediately see that opening up to trade reduces the cost cutoff c_D. Only the most productive firms still survive in the open economy equilibrium, and labor is reallocated towards the most productive firms. As in Melitz (2003), the least productive firms disappear, and their workers are reallocated towards the more productive survivors. On top of that, only the most productive among survivors export to the foreign market, inducing a further reallocation of workers towards more productive firms. In this model, all variables of interest depend only on the cutoff cost c_D.

The impact of trade on prices, mark-ups, sizes, welfare:

- We can describe precisely the impact of opening up to trade. All the effects are qualitatively similar to increasing the size of the economy in autarky.
  - the increased competition from foreign exporters induces a reduction in mark-ups. Note that only the most productive firms survive, which charge higher mark-ups than less productive firms. But since they face increased competition, they have to reduce their mark-ups, and this effect dominates.
  - prices go down, both because only the most productive firms stay, and because all firms reduce their mark-ups.
  - average firm size, firm profits, and product variety increase.
- We also see that if trade is costly ($\tau > 1$), trade does not entirely integrate markets. This is obvious from the fact that size still matters:
  - the larger country has a lower cost cutoff, higher average productivity and product variety, and lower mark-ups and prices (consumers benefit from all these combined effects).
  - with that specific functional form, the size of one's trading partner does not affect domestic variables. Even though a larger trading partner represents increased export opportunities, this is offset by increased competition. Similarly, even though a larger trading partner represents an increased import competition, exit in the long run reduces the number of entrants and offsets the competition effect.

Aggregate exports:

- Finally, from this model, we can derive gravity type predictions for bilateral trade flows. With N_E the total number of entrants in the domestic country, we get total exports from home to foreign,

$$X = \left(\frac{c_M^{-k}}{2(k+1)\gamma} \right) N_E L^* c_D^{*k+2} \tau^{-k}$$

As in the Melitz/Chaney or the Eaton and Kortum models, total exports result from both the extensive and the intensive margins of trade. As in Eaton and Kortum (2002) and Chaney (2006), the parameter driving the substitutability between varieties (γ here) does not affect the sensitivity of aggregate exports to trade barriers. Only the distribution of productivity shocks (indexed by k) matters.

3 Bernard, Eaton, Jensen and Kortum (2003)

We have seen with the Melitz and Ottaviano (2004) paper how we can remove the assumption of CES preferences, but keep the monopolistic competition framework, in a model with heterogeneous firms. We keep most of the predictions of the initial Melitz (2003) model, and in addition, we derive prediction for the endogenous determination of mark-ups. Among other, the model predicts that larger countries will be characterized by a "tougher competition", and therefore lower mark-ups. Consumers benefit from opening up to trade for several reasons then: first, as in Krugman, they get access to a wider range of differentiated goods; second, as in Melitz, opening up to trade induces in increase in aggregate productivity; last, and unlike the previous models we saw, they benefit from a reduction in prices, part of it due to lower mark-ups charged by firms.

We will see now how we can instead keep the CES preferences assumption, but instead remove the monopolistic competition assumption. Eaton and Kortum (2002) consider the case of perfect competition. Perfect competition actually gives results that are almost identical to monopolistic competition: instead of all firms charging exactly their marginal cost, they charge a constant mark-up (Dixit-Stiglitz mark-up) over their marginal cost. BEJK instead consider Bertrand competition. We will see that once heterogeneous firms compete in prices, since the price they charge depends on the price of their direct competitors, mark-ups will endogenously respond to changes in the "toughness" of competition, which will be the case when countries open up to trade.

Set-up

Preferences:

- As in Krugman, we assume that consumers maximize CES preferences. Unlike the Melitz/Chaney model, we assume a fixed set of differentiated goods of mass 1. Consumers in country n consume a quantity $q_n(\omega)$ of each variety ω, and derive a utility,

$$U_n \equiv \left(\int_0^1 q_n(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \tag{11}$$

Note that BEJK call each ω an industry, whereas in Dizit-Stiglitz, each ω was a variety (within the same industry). This is more labelling differences than anything else. However, in BEJK, as in Eaton and Kortum (2002), there is some competition within each "industry" ω, so in each industry, there are potentially many firms competing to produce the exact same good (industry here, variety in Dixt-Stglitz).

- This gives a simple isoelastic demand system. If good ω has a price $p_n(\omega)$ in country n, total expenditure by consumers in n is X_n, then the demand for each good ω is,

$$x_n(\omega) = \left(\frac{p_n(\omega)}{P_n} \right)^{1-\sigma} X_n \tag{12}$$
$$\text{with } P_n = \left(\int_0^1 p_n(\omega)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}$$

Fréchet distributions:

- In each country, there are many different firms. Each of these firms get a random productivity draw. In each country, only the best technology will be used, so that only the minimal cost is used. The distribution of the lowest cost of producing good ω in country i, $z_i^1(\omega)$ (note that the superscript 1 denotes the best draw), is drawn from a Fréchet distribution, F_i:

$$\begin{aligned} F_i(z_1) &= \Pr\left[z_i^1(\omega) \leq z_1\right] \\ &= \exp\left(-T_i z_1^{-\theta}\right) \end{aligned} \tag{13}$$

- Because of the nature of competition, we will need the distribution not only of the highest productivity, but also on the joint distribution of the highest and the second highest productivity. As proved in Appendix A.1, if the joint distribution of the highest productivity and of the second highest productivity, is,

$$\begin{aligned} F_i(z_1, z_2) &= \Pr\left[z_i^1(\omega) \leq z_1; z_i^2(\omega) \leq z_2 \mid 0 \leq z_2 \leq z_1\right] \\ &= \left[1 + T_i\left(z_2^{-\theta} - z_1^{-\theta}\right)\right] \exp\left(-T_i z_2^{-\theta}\right) \end{aligned} \tag{14}$$

then the distribution of the highest productivity is exactly as given in Eq. (13). This proposition can be proven by looking at the different possible orders of $z_i^1(\omega)$, $z_i^2(\omega)$, z_1 and z_2, and the respective probabilities of these different orders. Note that setting $z_1 = z_2$ returns the initial Fréchet distribution from Eq. (13).

- T_i scales up the technology of all goods in country i. It is a measure of the absolute advantage of country i. The parameter θ, which we assume is the same in all countries, is an inverse measure of the heterogeneity in productivity between different sectors. It will index the strength of comparative advantages between countries.

Why the Fréchet?

- Kortum (1997) and Eaton and Kortum (1999) derive the distribution of the leading-edge efficiencies from a dynamic model with endogenous innovation. If one randomly draws technologies from some distribution, and the market only keeps the best draws, the distribution of the best draws (suitably normalized), if it exists, will be extremal. There are 3 types of extreme value distributions (Gumbel, Fréchet and Weibull). The Fréchet is one of the 3 exteme type distribution, and the only one in which heterogeneity does not vanish in the limit. So we have some reasons to believe that the distribution of the efficiency frontier in a given country may be Fréchet.

- The second property (which is a direct corrolary) of the Fréchet is that the maximum of a family of Fréchet distributions is still a Fréchet. This propoerty will be usefull when we consider international trade: in a given country, consumers will only consume goods that are the cheapest available.

Production and trade costs:

- Firms face constant returns to scale technology, so that a firm with productivity z will produce z units per unit of labor. If a firm in country i wants to export goods towards country n, it faces an iceberg transportation cost, $\tau_{in} \geq 1$. We impose that those bilateral trade barriers satisfy the triangular inequality, $\tau_{kn} \leq \tau_{ki}\tau_{in}$, $\forall k, i, n$. So the cost of shipping the cheapest version of good ω from country i to country n, if the wage in country i is w_i, is,

$$c_{in}^1(\omega) = \frac{\tau_{in} w_i}{z_i^1(\omega)} \tag{15}$$

- The cheapest version of good ω in country n, looking at all potential source countries, is

$$c_n^1(\omega) = \min_i \left\{ c_{in}^1(\omega) \right\} \tag{16}$$

- Knowing that the cost depends on the productivity draw, on the domestic wage, and on the trade barriers, we can derive the distribution of the cheapest cost for good ω from country i in country n,

$$\begin{aligned} G_{in}^1(c_1) &= \Pr\left[c_{in}^1(\omega) \leq c_1\right] \\ &= \Pr\left[z_{in}^1(\omega) \geq \frac{\tau_{in} w_i}{c_1}\right] \\ &= 1 - \exp\left(-T_i (\tau_{in} w_i)^{-\theta} c_1^{\theta}\right) \end{aligned} \tag{17}$$

- The distribution of the cheapest cost for good ω in country n (from potentially any country in the world),

$$\begin{aligned} G_n^1(c_1) &= \Pr\left[c_n^1(\omega) \leq c_1\right] \\ &= \Pr\left[\min_i \left\{c_{in}^1(\omega)\right\} \leq c_1\right] \\ &= 1 - \Pr\left[\min_i \left\{c_{in}^1(\omega)\right\} > c_1\right] \\ &= 1 - \prod_i \left[1 - G_{in}^1(c_1)\right] \\ &= 1 - \exp\left(-\Phi_n c_1^{\theta}\right) \\ \text{with } \Phi_n &= \sum_i T_i (\tau_{in} w_i)^{-\theta} \end{aligned} \tag{18}$$

  Note that Φ_n is some aggregate measure of the productivity that country n has access to: it's some average of the productivity (both absolute and comparative) of each country in the world (T_i and w_i), scaled by the trade barriers into country n (the $\tau'_{in}s$).

- From this, we can derive the probability that country i is the cheapest provider of a given good ω in country n. Given that we have a continuum of goods, the law of large numbers holds, and this probability is exactly the share of goods that are imported from i

to n,

$$\begin{aligned}
\pi_{in} &= \Pr\left[c^1_{in}(\omega) = c^1_n(\omega)\right] \qquad (19) \\
&= \Pr\left[c^1_n(\omega) \leq \min_{j\neq i}\left\{c^1_{jn}(\omega)\right\}\right] \\
&= \int_0^\infty \prod_{k\neq i}\left[1 - G^1_{kn}(c_1)\right] dG^1_{in}(c_1) \\
&= T_i(\tau_{in}w_i)^\theta \int_0^\infty \prod_k \left[1 - G^1_{kn}(c_1)\right] \theta c_1^{\theta-1} dc_1 \\
&= T_i(\tau_{in}w_i)^\theta \int_0^\infty \exp\left[-\Phi_n c_1^\theta\right] \theta c_1^{\theta-1} dc_1 \\
&= \frac{T_i(\tau_{in}w_i)^\theta}{\Phi_n}
\end{aligned}$$

This share of imported good from i (and we'll see later that this is also the share of imports in nominal terms) depends on the productivity of country i (scaled by the trade barriers between i and n), relative to the productivity of all other trading partners of n.

Bertrand competition:

- We now assume that in each sector ω, firms compete Bertrand. If firms were identical, even with only two firms in a sector, prices would be equalized to marginal costs. With firm heterogeneity, firms can charge some mark-up. In equilibrium, firms either set a price equal to the second lowest cost, i.e. a price equal to the cost of their closest competitor, or the Dixit-Stiglitz mark-up if the second lowest price is not binding (which may happen with some probability). The price in country n is therefore,

$$p_n(\omega) = \mu(\omega) \times c^1_n(\omega) = \min\left\{c^2_n(\omega); \bar{\mu}c^1_n(\omega)\right\} \qquad (20)$$

$$\text{with } \bar{\mu} = \begin{cases} \frac{\sigma}{\sigma-1}, & \sigma > 1 \\ \infty, & \sigma \leq 1 \end{cases}$$

So we can see this paper as an extension of the Melitz model, where we allow more than one firm within each sector. The Dixit-Stiglitz outcome will only only occur with some probability, when the first and the second draw are sufficiently far apart. Otherwise, the mark-up depends on the realization of the first and second draw of productivity (or cost).

- If country i is the cheapest provider of good ω in country n, it means it has the lowest cost, and therefore we know that the lowest cost for that good ω in country n must be such that,

$$c_n^2(\omega) = \min\left\{\min_{j\neq i}\left\{c_{jn}^1(\omega)\right\}, c_{ni}^2(\omega)\right\} \tag{21}$$

- From this result and the distribution of the highest and the second highest productivities in Eq. (14), we get the joint distribution of the lowest and the second lowest costs from country i in country n. It is convenient to work with the complementary distributions for a moment,

$$\begin{aligned} G_{in}^c(c_1, c_2) &= \Pr\left[c_{in}^1(\omega) \geq c_1, c_{in}^2(\omega) \geq c_2\right] \\ &= \Pr\left[z_{in}^1(\omega) \leq \frac{\tau_{in}w_i}{c_1}, z_{in}^2(\omega) \leq \frac{\tau_{in}w_i}{c_2}\right] \\ &= \left[1 + T_i(\tau_{in}w_i)^{-\theta}\left(c_2^\theta - c_1^\theta\right)\right]\exp\left(-T_i(\tau_{in}w_i)^{-\theta}c_2^\theta\right) \end{aligned} \tag{22}$$

From this, and from Eq. (21), we get the complementary joint distribution of the lowest and the second lowest cost in country n, unconditional of the origin country,

$$\begin{aligned} G_n^c(c_1, c_2) &= \Pr\left[c_n^1(\omega) \geq c_1, c_n^2(\omega) \geq c_2\right] \\ &= \Pr\left[\begin{array}{c} \text{"both the lowest and the second lowest draws are above } c_2\text{"} \\ \text{or} \quad \text{"the lowest cost is in } [c_1, c_2] \text{ and the second lowest above } c_2\text{"} \end{array}\right] \\ &= \Pr\left[\begin{array}{c} \text{for all } i's, \text{"}c_{in}^1(\omega) \geq c_2 \text{ and } c_{in}^2(\omega) \geq c_2 \text{ in all } i's\text{"} \\ \text{or, for all } i's, \left\{\begin{array}{l} \text{"}c_{kn}^1(\omega) \geq c_2 \text{ and } c_{kn}^2(\omega) \geq c_2 \text{ in all } k \neq i\text{"} \\ \text{and "}c_1 \leq c_{in}^1(\omega) < c_2 \text{ and } c_{in}^2(\omega) \geq c_2 \text{ in } i\text{"} \end{array}\right. \end{array}\right] \\ &= \prod_i\left[1 - G_{in}(c_2, c_2)\right] \\ &\quad + \sum_i\left\{\left(\left[1 - G_{in}(c_1, c_2)\right] - \left[1 - G_{in}(c_2, c_2)\right]\right)\prod_{k\neq i}\left[1 - G_{kn}(c_2, c_2)\right]\right\} \\ &= \prod_i e^{-T_i(\tau_{in}w_i)^{-\theta}c_2^\theta} \\ &\quad + \sum_i\left\{T_i(\tau_{in}w_i)^{-\theta}\left(c_2^\theta - c_1^\theta\right)e^{-T_i(\tau_{in}w_i)^{-\theta}c_2^\theta}\prod_{k\neq i}e^{-T_k(\tau_{kn}w_k)^{-\theta}c_2^\theta}\right\} \\ &= \exp\left(\Phi_n c_2^\theta\right) + \Phi_n\left(c_2^\theta - c_1^\theta\right)\exp\left(\Phi_n c_2^\theta\right) \end{aligned} \tag{23}$$

And finally, we recover the joint distribution of the lowest and second lowest costs, unconditional on the country of origin,

$$\begin{aligned} G_n(c_1, c_2) &= \Pr\left[c_n^1(\omega) \leq c_1, c_n^2(\omega) \leq c_2\right] \\ &= 1 - G_n^c(0, c_2) - G_n^c(c_1, c_1) + G_n^c(c_1, c_2) \\ &= 1 - \exp\left(\Phi_n c_1^\theta\right) + \Phi_n c_1^\theta \exp\left(\Phi_n c_2^\theta\right) \end{aligned} \tag{24}$$

Distribution of mark-ups:

- Now that we have the joint distribution of the lowest and second lowest costs in country n, we can describe the distribution of mark-ups in country n. For all $\mu's$ such that $1 \leq \mu \leq \bar{\mu}$,

$$\begin{aligned} \Pr\left[\mu_n(\omega) \leq \mu | c_n^2(\omega) = c_2\right] &= \Pr\left[\frac{c_2}{\mu} \leq c_n^1(\omega) \leq c_2 | c_n^2(\omega) = c_2\right] \\ &= \frac{\int_{c_2/\mu}^{c_2} \frac{\partial^2 G_n}{\partial c_1 \partial c_2}|_{c_1,c_2} dc_1}{\int_0^{c_2} \frac{\partial^2 G_n}{\partial c_1 \partial c_2}|_{c_1,c_2} dc_1} \\ &= \frac{c_2^\theta - (c_2/\mu)^\theta}{c_2^\theta} \\ &= 1 - \mu^{-\theta} \end{aligned} \tag{25}$$

So the distribution of the mark-ups in country n, conditional on the second lowest cost being c_2, is a Pareto distribution that does not depend on c_2. This property is specific to the functional form we assumed for the distribution of costs, and it is quite convenient. The unconditional distribution will therefore be the same (we just integrate that probability for all realizations of c_2, which is integrating a constant over the support of c_2, and we get exactly that same constant). This was the distribution of mark-ups in country n conditional on the mark-up being below the Dixit-Stiglitz mark-up. The unconditional distributions of mark-ups in country n is then this Pareto distribution, truncated from above by $\bar{m}$,

$$H(\mu) = \Pr\left[\mu_n(\omega) \leq \mu\right] = \begin{cases} 1 - \mu^{-\theta} & , \ 1 \leq \mu \leq \bar{\mu}, \\ 1 & , \ \mu > \bar{\mu} = \frac{\sigma}{\sigma - 1} \end{cases} \tag{26}$$

- We have the following properties for the distribution of mark-ups:
  - Note that the distribution of mark-ups is the same in all countries, irrespective of the cost of trading with that country. This results differs radically from the prediction in Melitz and Ottaviano (2005).

- The reason for that is that while reducing trade barriers will increase the number of potential competitors in sector ω and therefore lower mark-ups. At the same time however, this is exactly offset by the exit of domestic firms who used to charge the lowest mark-ups.
- Note also that the distribution of mark-ups only depends on the heterogeneity parameters, (inverse) heterogeneity in preferences, σ, and (inverse) heterogeneity in productivity between firms, θ. A higher heterogeneity in productivity between firms, lower θ, will increase the probability of high mark-ups, as there are relatively more dispersion between firms (and therefore more distance between the lowest and the second lowest cost draw, on average). If agents see goods as more differentiated, lower σ, firms are more likely to charge a high mark-up, as mark-ups are truncated at a higher point.

Measured productivity:

- We now know the distribution of mark-ups in a given country. The next question is, what is the distribution of mark-ups for a given firm with productivity z_1? This mark-up is exactly the typical measure of labor productivity that we would observe in the data: $\frac{\text{output}(z_1)}{\text{workers}(z_1)} = \mu(z_1)$. This mark-up does not only depend on the firm's characteristics, it depends on what the competitors of that firm are doing. We can derive the distribution of the mark-up of a firm with productivity z_1. For example, let's look at the mark-up of a firm from a country i that is the cheapest provider of a good in country n, and assume that this firm is not charging the

Dixit-Stiglitz mark-up:

$$\begin{aligned}
H_{in}(\mu \mid z_1) &= \Pr\left[\mu_{in}(\omega) \leq \mu \mid z_{in}^1(\omega) = z_n^1(\omega) = z_1\right] \qquad (27)\\
&= \Pr\left[\frac{c_n^2(\omega)}{c_n^1(\omega)} \leq \mu \mid c_{in}^1(\omega) = c_1 = \frac{\tau_{in} w_i}{z_1}\right]\\
&= \Pr\left[c_1 \leq c_n^2(\omega) \leq \mu c_1 \mid c_{in}^1 = c_1\right]\\
&= \frac{\int_{c_1}^{\mu c_1} \frac{\partial^2 G(c_1,c_2)}{\partial c_1 \partial c_2} dc_2}{\int_{c_1}^{\infty} \frac{\partial^2 G(c_1,c_2)}{\partial c_1 \partial c_2} dc_2}\\
&= \frac{\exp\left[-\Phi_n c_1^\theta\right] - \exp\left[-\Phi_n (\mu c_1)^\theta\right]}{\exp\left[-\Phi_n (\mu c_1)^\theta\right]}\\
&= 1 - \exp\left[-\Phi_n \left(\mu^\theta - 1\right) c_1^\theta\right]\\
&= 1 - \exp\left[-\Phi_n \left(\mu^\theta - 1\right) (\tau_{in} w_i)^\theta z_1^{-\theta}\right]
\end{aligned}$$

We now have to truncate that distribution to account for the fact that with some probability, the firm will charge the Dixit-Stiglitz mark-up,

$$\begin{aligned}
H_{in}(\mu \mid z_1) &= \Pr\left[\mu_{in}(\omega) \leq \mu \mid z_{in}^1(\omega) = z_n^1(\omega) = z_1\right] \qquad (28)\\
&= \begin{cases} 1 - \exp\left[-\Phi_n \left(\mu^\theta - 1\right) (\tau_{in} w_i)^\theta z_1^{-\theta}\right] & , \ 1 \leq \mu \leq \bar{\mu}, \\ 1 & , \ \mu > \bar{\mu} = \frac{\sigma}{\sigma - 1} \end{cases}
\end{aligned}$$

- The measured productivity of a firm has the following properties:
  - A plant with a higher efficiency will charge a higher mark-up on average than another firm (its distribution of mark-ups first-order stochastically dominates the other's). This result is similar to Melitz and Ottaviano (2005). In this model, it comes from the fact that the distance between the best efficiency and the second best efficiency increases with the level of efficiency (in a statistical sense). Therefore, the actual measure of productivity (realized $\mu(z_1)$) is (on average) an increasing of a firm's intrinsic productivity.
  - The easier it is for a firm to access a given country (τ_{ij} low or w_i low), the higher the mark-up it will charge (on average). On the other hand, the "tougher" the competition it faces (the easier it is for its competitors to compete, i.e. the lower Φ_n), the lower the mark-up it charges.

– The two measures of heterogeneity (heterogeneity in tastes, σ, and heterogeneity in productivity, θ) affect the distribution of the mark-up of a single firm in the same way that they affect the whole distribution of mark-ups: more heterogeneity (lower σ or θ) implies higher mak-ups, on average.

Efficiency and exporting:

- It is easy to see that the model predicts that exporters will typically be more productive than non-exporters. The model also predicts that all exporters will also sell on their domestic market, whereas only a fraction of domestic firms are also exporters.
- To sell domestically, a domestic producer of good ω must be more efficient than any of its foreign competitors,

$$z_i^1(\omega) \geq z_k^1(\omega) \frac{w_i}{\tau_{ki} w_k}, \ \forall k \neq i$$

  To be able to export towards country n on the other hand, this same producer must be more efficient than any other foreign competitor after incurring the trade barrier τ_{in},

$$z_i^1(\omega) \geq z_k^1(\omega) \frac{\tau_{in} w_i}{\tau_{kn} w_k}, \ \forall k \neq i$$

  From the triangular inequality that we imposed on trade barriers, $\tau_{kn} \leq \tau_{ki}\tau_{in}$, it is harder to export than to sell domestically,

$$z_k^1(\omega) \frac{w_i}{\tau_{ki} w_k} \leq z_k^1(\omega) \frac{\tau_{in} w_i}{\tau_{kn} w_k}$$

  Exporters also have larger domestic sales than non exporters because exporters tend to be more productive, and more productive firms tend to be larger: more productive firms charge lower mark-ups, and therefore have larger market shares.

Aggregate exports:

- We have seen that prices of a given commodity ω in a given country n do not systematically vary with the country of origin. Therefore, the share of country n's export from country i is directly given by the share of goods that country n imports from country i, given in Eq. (19),

$$\frac{X_{in}}{X_n} = \pi_{in} = \frac{T_i \left(\tau_{in} w_i\right)^{\theta}}{\Phi_n} \tag{29}$$

- As in Chaney (2006), or Eaton and Kortum (2002), this share only depends on the relative trade barriers between i and n (scaled by i's productivity) and the trade barriers from all countries and n.
- As in Eaton and Kortum (2002) and Chaney (2006), the sensitivity of trade flows to trade barriers does not depend on the elasticity of substitution between goods, but only on the measure of firm heterogeneity θ (equivalent to the γ parameter of the Pareto distribution in Chaney). The reason why σ drops out of the exports expression is somehow similar to the mechanism described in Chaney (2006). When trade barriers between i and n (τ_{in}) increase, several things happen.

  First, the extensive margin of trade adjusts. Because the trade barriers are higher, there are some goods for which country i is not the cheapest producer anymore. How many of these goods there are does not depend on σ, it only depends on the distribution of productivity shocks, governed by θ. How much market shares each of these exporters had prior to losing their edge does depend on σ, but it does not depend on average.

  Second, there are some goods for which i is still the cheapest provider. However, because the cost for i's exporters has increased, the difference between the lowest cost (from i) and the second lowest (from some other country, unaffected by this change in τ_{in}) has shrunk. For some fraction of goods for which the second lowest cost was not binding, and therefore for which the exporter from i was charging the Dixit-Stiglitz mark-up, the second highest price is still not binding, so that i's exporter is still charging the Dixit-Stiglitz mark-up, and hence increases its price. How much market share does i's exporter lose to other sectors, this depends on the elasticity of substitution σ, as consumers substitute towards other sectors: the bigger σ, the larger the loss in market share. For some other i's exporters, the second lowest cost becomes binding, so that the firm switches from the Dixit-Stiglitz mark-up to the constrained mark-up (therefore reduces its mark-up). These types of i's exporters increase their price in n, but less so that the previous category of exporters. How much they increase their price depends on σ (which determins the Dixit-Stiglitz mark-up). How much market share they lose to other sectors depends on σ as well. Finally, for some of i's exporters, the second lowest cost was binding and is still binding, so that their price is unchanged.

  Third, there are goods for which the best producer in i either remains the (potential) second cheapest provider, or was the second cheapest provider, but no longer is. Because the cost has changed for those (potential) exporters, the price charged by the cheapest provider (no matter where this guy comes from) will change. How much market share gain that price change induces depends on the elasticity of substitution σ (the bigger σ, the bigger the gains in market share of these cheapest suppliers).

However, because we have a continuum of goods, and because we know from the distribution of mark-ups in Eq. (26) that the distribution of prices in country n is independent of any trade barriers, those impacts of changing τ_{in} on the intensive margin of trade exactly cancel out. The only margin of adjustment, on average, is the extensive margin. This margin only depends on the distribution of productivity shocks, driven by θ.

- "Absolute advantages" (captured by T_i) increase the export of i to any country in the world, if we do not solve for wages in general equilibrium at least.
- "Comparative advantages" (captured by θ) dampen the impact of trade barriers: more dispersed comparative advantages, i.e. θ lower, imply that trade barriers do not have much of an impact on trade flows: no matter how large trade barriers are, there is always a fraction of goods for which i is extremely good, so it always exports at least some of these goods.

Welfare:

- We can compute welfare as the inverse of the price index in country n:

$$P_n^{-1} = \gamma \Phi_n^{1/\theta}$$

with γ a constant[1]. We immediately see that the cheaper it is to import goods from the rest of the world increase welfare (Φ_n lower, either because trade barriers towards n are lower, or because n's trading partners have a better technology/lower wages). Interestingly, welfare in country n does not depend on wages in country n, and therefore does not depend on the trade barriers that potential exporters from n face.

[1] $\gamma = \left[\frac{1+\theta-\sigma+(\sigma-1)\bar{\mu}^{-\theta}}{1+\theta-\sigma}\Gamma\left(\frac{1+2\theta-\sigma}{\theta}\right)\right]^{1/(\sigma-1)}$.

Calibration and empirical exercises

BEJK go on to calibrate their model on actual firm level data (US firms), to test some of the predictions of their model, and to do some simulation exercises.

Parameters:

- From Eq. (29), using bilateral trade flows and aggregate output data between the US and 47 other countries, BEJK infer trade barriers measures, $T_i (\tau_{in} w_i)^\theta / \Phi_n$'s for all i, n's,

$$\frac{T_i (\tau_{in} w_i)^\theta}{\Phi_n} = \frac{X_{in}}{X_n}$$

  The data on bilateral trade flows come from Feenstra, Lipsey and Bower (1997). Data on aggregate output come from UNIDO (1999), completed with data from the World Bank.

- From Eaton and Kortum, they use $\theta = 8.28$ and $\sigma = 6$ (they experiment with other values too).

- They take $500,000$ draws from joint Fréchet distributions (for the highest and second highest productivity) for each 47 countries (we have 500,000 sectors). From these productivity draws, they know which firm sells in which country, what price it charges, what its total size and measured productivity is...

Matching the data:

- They compare their simulated US economy (they also have a simulated economy for all other countries, that they needed to determin which US firm survives, where it exports, and what price it sets), to actual data on US firms. These are Census data on 200,000 firms in 1992.

- They overpredict the **fraction of exporters** (21% in the data, 51% in their simulation). Weighting by size gets them closer to the actual fraction.

- They underpredict the **dispersion of productivity** between firms.

- They get the **exporters' size advantage** about right (overpredict it for large σ's).

- They get about right the **fraction of revenue from exports**: most exporters export only a small fraction of their output. They underestimate the fraction of export-oriented firms when they take large σ's.

Simulations:

- They go on simulating counterfactual worlds. To do so, they need however to solve for the general equilibrium wages and prices in these counterfactual worlds. They do so by assuming a homogenous sector freely tradable with constant marginal labor productivity W_n in country n. They expand the model to account for labor and intermediate inputs in production.

- They submit their simulated economy to 3 types of shocks:

  (i) Globalization: a reduction of trade barriers of 5%.

  (ii) Autarky: prohibitive trade barriers.

  (iii) Exchange rate appreciation: a 10% rise in US wages relative to wages in other countries.

- **Globalization:** reducing trade barriers leads to an increase of aggregate productivity (4%). As in Melitz, part of the story is reallocation of labor towards more productive firms (8% of US firms exit, 4.6% of gross industrial job creation, 7.3% gross industrial job destruction), another driving force is a fall in the cost of intermediates.

- **Autarky:** a rise in trade barriers reduces aggregate productivity by 9%, mainly because of an increase in the cost of intermediates, but also due to the entry of inefficient US firms.

- **Exchange rate appreciation:** this loss of competitiveness increases US productivity by 4%, as plants substitute intermediate for workers (the price of intermediate goods rises less than wages), and as less productive US firms exit. Those two together induce an 18% fall in industrial employment. 24% of US exporters stop exporting.