Firm Experimentation in New Markets

Zhanar Akhmetova

November 2010

Abstract

An important feature of the data on new exporters is the evolution over time of their export size and exit rate. As the exporting age of firms increases, their average export volumes grow and the exit rate decreases. In particular, we observe cases of new exporters expanding slowly initially, and switching to high levels of exports after some time. I propose a quantitative model of active learning that can explain these patterns. In the presence of demand uncertainty and high sunk costs of entry, the firm can postpone paying this cost and learn more about its demand by paying for the testing technology. This is the first model in the trade literature to allow for multiple period and varying sample size experimentation. The duration of the learning stage, the intensity of learning, and the total entry costs are determined endogenously. The model permits us to produce rich comparative statics predictions with respect to market characteristics, firm characteristics, and features of uncertainty.

1 Introduction

For a long time, the field of international trade focused on the study of aggregates - country level export and import volumes, determinants of goods traded, factor content of trade and aggregate factor prices. More recently, the trade literature has occupied itself with the firms that trade - the relationships between firm productivity, size, firm-level factor prices and trade activity. This has been mostly an investigation of long-run, static relationships. The newly available firm-level data suggests, however, that looking at the dynamics of export behavior of individual firms is worthwhile as well. Firms behave differently in different parts of their life cycle. New exporters (firms that start exporting for the first time) exhibit patterns that are unlike those revealed by older exporters, and that cannot be explained by standard models. Instead, models that are tailored to the problems that exporters face in new markets are required. The main feature of these models is the presence of demand uncertainty, and the possibility of learning from exporting activity, so that new exporters start out with little information and have to acquire it as they export. Below we will first lay out the evidence that has attracted attention in recent papers, as well as our own evidence (from French firm-level data), and briefly describe a model that we propose to explain these findings. This model offers unique features, such as a testing technology that allows firms to learn before investing in the sunk cost of entry, and active learning by firms that recognize the information value of their own exports, and optimally choose the duration and intensity of experimentation. It allows for two stages of exporting, where the first stage describes the beginner or new exporter phase, and the second stage captures the old exporter phase of the firm, so it is a complete model of the life cycle of an exporter. We also present the anecdotal evidence we have discovered of test marketing in the domestic and foreign markets.

The first pattern found in the literature is the small initial export volumes (in quantities) of new exporters, and their later expansion over time. This has been documented by Eaton et al. (2007) for Colombia, who find that many new exporters export very small quantities and are likely to drop out of the foreign market in the following few years. This is demonstrated in Figure 1, where we employ French firm-level data (figures using the French firm-level data are borrowed from Akhmetova and Mitaritonna (2010)). This can be explained by the uncertainty that these firms face - they are not sure what the demand for their product is in the foreign market, and start out small. As they learn more about their demand, they either expand (if they find that their demand is higher than expected), or exit the market (if they find that their demand is too low). This also implies that the exit rates will tend to fall over time (although, as we will point out later, not necessarily in a monotonic fashion), as the worst exporters are weeded out gradually, and only the best ones remain. This second pattern was also observed by Eaton et al. (2007), and is reported for the French firm-level data in Figure 2. In Figures 1 and 2, we focus on average export activity as measured by the quantities exported of individual 8-digit products, by destination.

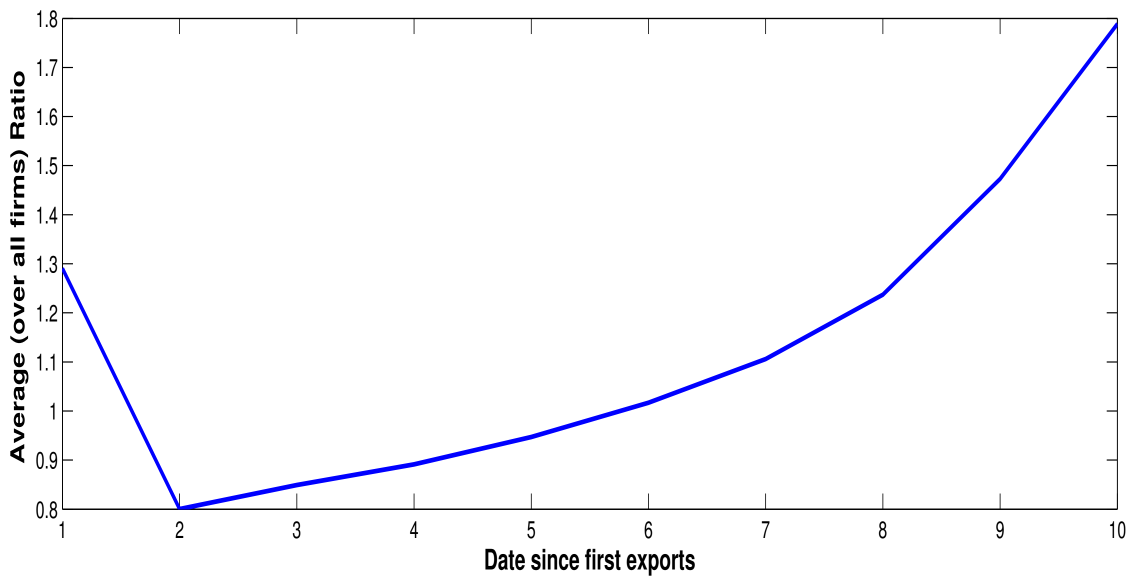

Figure 1: Ratio of quantity (8-digit product) to average quantity of the firm over the period, averaged over all new exporters, consumer non-durable goods

Another dimension along which we can examine these dynamics, is the number of products that firms export. Using the French firm-level data, we can examine the number of

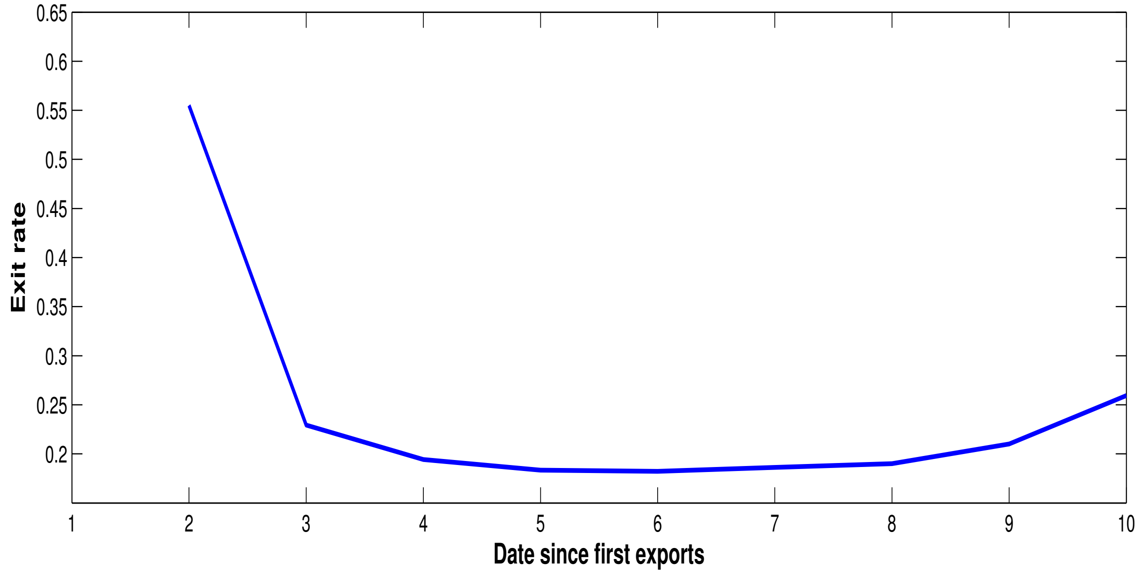

Figure 2: Exit Rate (8-digit product), of all new exporters, consumer non-durable goods

8-digit products within a given 4-digit category that firms export. In Figures 3 and 4 we show the expansion in the average (across all new exporters) number of products in a particular sector (Beauty or make-up preparations and preparations for the care of the skin (other than medicaments), including sunscreen or suntan preparations; manicure or pedicure preparations) and destination (Bulgaria). In these figures, the horizontal axis measures the date since first exporting year (1 for the first year when the firm exports), and the vertical axis measures the average number of products exported. In Figure 3, the average is over all new exporters, and in Figure 4, the average runs only over those new exporters that survived for at least 8 years. Note that we define a firm that exports a positive number of products after 8 years (with possible zeros in between) as a firm that survives for at least 8 years. We present Figure 4, to highlight the fact that the expansion in the average number of products is not (purely) due to the selection effect - less productive and smaller firms exiting in the first years, and is also driven by the expansion of surviving firms.

Notice the interesting feature of Figures 3 and 4 - while the average size grows smoothly initially, it jumps to a high level at year 8 (since first exports), and stays there. This tells us something about the learning dynamics of these firms - it looks like they are learning at first, exporting small volumes, and switch to a larger scale later on.

This evidence is consistent with the hypothesis that firms may start small in a foreign market when they are unsure of their profitability there, in order to collect more information about the market and to either expand or quit the market later, depending on their observations. We propose a theory of firm behavior that assumes demand uncertainty in a foreign market, where the firm is uncertain about the mean of the shift demand parameter that affects its profitability. There is a large sunk entry cost one needs to incur to access the

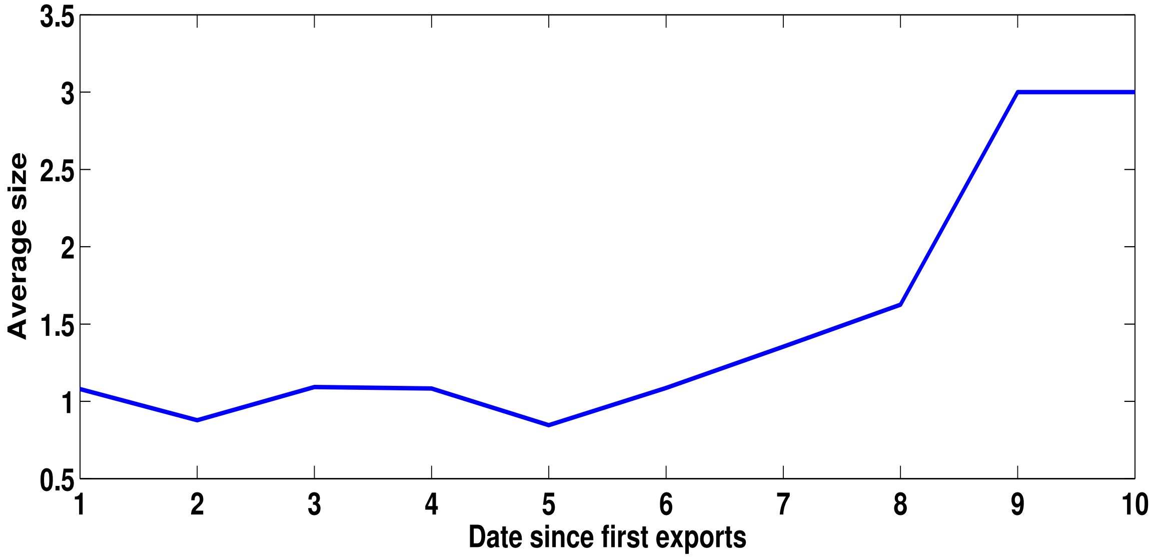

Figure 3: Example of average size evolution, Bulgaria, Beauty or make-up preparations

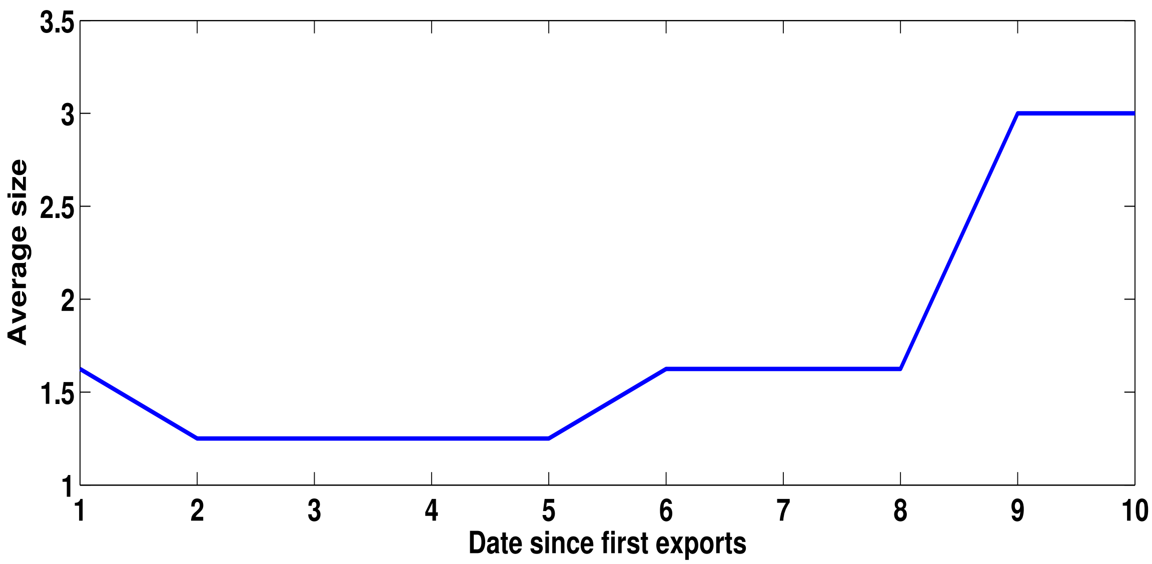

Figure 4: Example of average size evolution,conditional on Survival for at least 8 years, Bulgaria, Beauty or make-up preparations

entire market. While a firm may be unwilling to export to this market under such conditions, we introduce an additional assumption. We allow for a 'testing technology', so that a firm can access a few consumers in the foreign market if it pays some variable testing costs. The sales to the individual consumers serve as noisy signals about the mean demand parameter, so that the firm uses the information about these sales to update its beliefs about demand. Based on these beliefs, the firm decides whether it should incur the large sunk entry cost to access the entire market or quit, or keep experimenting. The firm can keep experimenting as long and as intensively, as it finds optimal.

Essentially, one is assuming two-tiered costs of accessing the foreign market: the variable testing costs to access a few random consumers, and the sunk costs of entry to access the entire market. One can think of the sunk cost as the cost of establishing a distribution and marketing network in the foreign market, and signing long term shipping contracts, and of the testing costs as the costs of accessing a few consumers by temporarily hiring marketing agencies in the foreign country and locating temporary shipping services. Clearly, we need to assume that the cost of accessing the entire market using the testing technology forever is higher than the sunk cost of entry, and that is quite plausible if we think of these costs the way we described above.

That this kind of experimentation takes place when a firm creates a new product and aims to sell it in the domestic market, has been documented in the marketing literature, as test marketing. As defined in the Handbook of Marketing Research (Ferber, 1974), "test marketing is a controlled experiment, done in a limited but carefully selected part of the marketplace, whose aim is to predict the sales or profit consequences, either in absolute or in relative terms, of one or more proposed marketing actions. It is essentially the use of the marketplace as a laboratory and of a direct sales measurement which differentiates this test from other types of market research."

While test marketing is carried out both by small and large firms, examples that get most attention and thus are known to us involve large companies. Procter and Gamble successfully test marketed Tide in October 1946 in six cities: Springfield, Massachusetts; Albany, New York; Evansville, Indiana; Lima, Ohio; Wichita, Kansas; and Sioux Falls, South Dakota. In 1985, Time Inc. announced its test marketing of the Picture Week magazine, in 13 undisclosed markets that represented 10 percent of the national population, "the most extensive test procedure ever mounted by the corporation and perhaps even the industry" (The NYT, July 3, 1985). More recently (July 2009), Coca Cola has been reported to test market a sweetened fizzy milk beverage called Vio. About 200 retailers in New York city were reported to be carrying the drink. These stories tell us that the notion of a representative consumer can be interpreted in many different ways. A consumer can be viewed as an individual, a household, a retail store (as in the story about the drink Vio), a city (as in the story about Tide), or a state/province. Other examples are available from the author upon request.

Hence, it is clear that using the marketplace as a laboratory is certainly not an unusual step in the domestic market. There is no doubt that the same kind of demand uncertainty exists when a firm starts exporting to a new foreign market. Even if a product has been produced and sold in the domestic and other foreign markets previously, the firm may be uncertain about the demand for this product in a new destination. The difficulty of expanding into new export markets is acknowledged by government agencies, such as the US Department of Commerce, who offer extensive assistance to businesses wishing to export. Their services include counseling on the rules and regulations in foreign markets, market research (for example, through trade fairs), and identifying potential buyers and distributors in foreign markets. We talked to one of the trade specialists in the New Jersey office of the Department of Commerce, and asked them if they ever helped or observed a firm engage in test marketing in a foreign market. The agent confirmed that it indeed takes place often, for example, US companies may start exporting in the maquiladora area in Mexico, and later expand to the rest of Mexico, if successful.

This model provides a new way of looking at the dynamics of new exporters. It points out that there may be a state of the firm in-between full-market access and non-exporting, where the firm is granted the chance to learn about demand before making a final decision. The duration of this learning stage is moreover determined endogenously - by the firm and market characteristics, and is a random variable, affected by the draws of demand signals that the firm obtains. Similarly, the total entry cost - which here would be the sum of total testing costs and the (one-time) sunk cost of entry, is endogenous and random.

The model can allow to predict the dynamics of exports by new exporters. One specific application would be the study of the response to trade liberalization - once tariffs fall, new firms will be willing to export, and how they do so can be determined within this model. This dynamics will depend on the firm and market variables, as well as features of uncertainty.

Related literature

We discuss the papers that are most relevant for us. A model where the decision to export is not a binary decision in the sense described above, is Arkolakis (2006). There the firm may also access a few consumers rather than the entire market, through the marketing technology, so that the fixed cost of entry is no longer 'fixed', but endogenously determined. That model is not concerned with learning, however, and the only way that it can produce dynamics in sales is through exogenous shocks to productivity.

Recently, several papers have focused on the role of uncertainty and learning in export markets. In one of the earliest contributions, Rauch and Watson (2003) introduce a model where a firm has to decide whether to contract a supplier in a new destination, under uncertainty about the quality of the producer there. To learn more about this, the firm first places a small order and later either expands (signs a contract with the supplier) or quits the destination. Though this is not a model of export behavior, the idea that a firm might want to start small in a new market is easily applicable to the export setting. Of course, this model lacks such elements as sampling observations over repeated periods of time (rather than over a single period), and choosing an optimal sample size (rather than a predetermined order size). In Eaton et al. (2008), there is uncertainty about the foreign demand, and firms search for buyers and collect information on their sales. If they collect encouraging information, they expand (search for more buyers), or shrink, until finally they exit. This is a model of passive learning, in the sense that the firm does not take optimally into account the effect its sales have on its learning. Moreover, the sunk cost of entry does not play an important part here. Another paper that deals with the learning behavior of exporters is Albornoz et al. (2009). They assume a sunk cost of entry and uncertain demand in a foreign market. The firm has to incur the sunk cost of entry to export, and once it does so, it learns its demand with complete precision. Thus, learning takes place only over one period, so that this model cannot describe complex patterns of firm's growth in a given destination over time. However, this model provides a useful way of thinking about the interaction between export behavior in several markets. That is, if the demand parameters are correlated across markets, export experience in one may result in entry into another. Our model can produce the same predictions if we incorporate correlation between the demand parameters across markets. Nguyen (2008) also introduces demand uncertainty, both in the domestic and foreign market, and correlation between the demand parameters in the two markets. The firm can learn about demand with complete precision once it sells there for at least one period. There are no sunk costs of entry into either market, but there are per period fixed costs of selling. Hence, the firm will exit a market, if the demand there is too low to cover the per period fixed cost. Therefore, there is no need for experimentation within a market, but it is possible to improve the knowledge about demand in any given market by first selling in another market. Finally, a paper that provides interesting evidence of learning by exporters is Freund and Pierola(2008). In their model, the firm is uncertain about the per period cost of exporting, which it can learn once it exports. There is a fixed cost of entry that the firm has to pay once it decides to stay in the market, but it can first export a small quantity (a fixed fraction of the potential total sales) for a smaller fixed cost (cost of trial), to learn the per period cost and make the ultimate decision to stay or not. Hence, again, there is a one-period learning with a predetermined trial sales size (which does not depend in any way on the features of uncertainty, such as the distribution of the uncertain parameter, or on the characteristics of the firm or the market).

To summarize, the previous models of learning lack one or more of the main features introduced here: a sunk cost of entry that can be postponed due to the availability of a testing technology, an endogenously determined learning phase duration and endogenously determined experimentation intensity, and active learning - the acknowledgement by the firm of the information value of its own sales, and optimal choice of sales size accordingly.

The rest of the paper proceeds as follows. We study the model and its solution, in Section 2. In Section 3, we discuss the comparative statics predictions. We contrast the model with alternative models of learning in Section 4. Section 5 concludes.

2 The Model

There are two countries, one Home, another Foreign. There are M consumers in the Foreign country, where $M \in Z^+$. For any foreign consumer i,

$$U_{it} = [\int_{\omega\in\Omega}((e^{\mu})^{\frac{1}{\epsilon}} c_{it}(\omega)^{\frac{\epsilon-1}{\epsilon}} d\omega]^{\frac{\epsilon}{\epsilon-1}},$$

where ω denotes varieties, and Ω the entire set of varieties in the economy. $\mu(\omega) = \bar{\mu}$ with probability p_0, and $\underline{\mu}$ with probability $1 - p_0$. Once $\mu(\omega)$ is drawn for a given variety, it is fixed over time.

Hence,

$$\begin{aligned} c_{it}(\omega) &= C_{it} e^{\mu(\omega)} [\frac{p_t(\omega)}{P_{it}}]^{-\epsilon} \\ &= e^{\mu(\omega)} y_t (p_t(\omega))^{-\epsilon} P_{it}^{\epsilon-1}, \end{aligned}$$

where $p_t(\omega)$ is the price of variety ω, P_{it} is the ideal price index for the differentiated good for consumer i,

$$P_{it} = [\int_{\omega \in \Omega} e^{\mu(\omega)} p_t(\omega)^{1-\epsilon} d\omega]^{\frac{1}{1-\epsilon}},$$

y_t is the total income of consumer i, assumed equal across consumers, and C_{it} is the total consumption of the differentiated good (so that $C_{it}Pit = y_t$).

Labor is the only factor of production, with the constant marginal cost of producing any variety of the differentiated good given by the ratio of wages, w_t, and productivity, ϕ_t. Each firm produces one variety. Hence, the price of any variety is given by

$$p_t(\omega) = \frac{\epsilon}{\epsilon - 1} \frac{w_t}{\phi_t(\omega)},$$

where $\phi(\omega)$ is the productivity of the firm that produces the variety ω.

We can further calculate the steady state value of the ideal price index for any consumer i:

$$\begin{aligned} P_{it} &= [\int_{\omega} e^{\mu(\omega)} p_t(\omega)^{1-\epsilon} d\omega]^{\frac{1}{1-\epsilon}} \\ &= [\int_{\omega} \int_{\phi} e^{\mu(\omega)} [\frac{\epsilon}{\epsilon - 1} \frac{w_t}{\phi_t(\omega)}]^{1-\epsilon} f(\phi, \omega) d\phi d\omega]^{\frac{1}{1-\epsilon}}, \end{aligned}$$

where $f(\phi, \omega)$ is the steady state joint distribution of ϕ and ω. We have to specify this because ϕ and ω are not independent in the steady state - more low-productivity firms (relative to higher-productivity ones) will not produce the varieties with $\mu(\omega) = \underline{\mu}$. In the steady state, the ideal price index is the same for all consumers, even though their variety-specific preferences may be different.

So sales to any consumer i at time t are given by:

$$c_{it}(\omega) = e^{\mu(\omega)} y (p(\omega))^{-\epsilon} P^{\epsilon-1},$$

where all the variables take on their steady state values.

Now drop the ω notation, and denote the variety produced by firm j by j. Operational profits per period of the firm, earned from sales to an individual consumer, can be calculated as

$$\begin{aligned}\pi_{ijt} &= p_{jt}c_{ijt} - w_t\frac{c_{ijt}}{\phi_{jt}} \\ &= (\frac{\epsilon}{\epsilon-1} - 1)c_{ijt}\frac{w_t}{\phi_{jt}} \\ &= \frac{1}{\epsilon-1}\frac{w_t}{\phi_{jt}}e^{\mu_j}y_t(p_{jt})^{-\epsilon}P_{it}^{\epsilon-1} \\ &= e^{\mu_j}y_t\frac{1}{\epsilon-1}[\frac{\epsilon}{\epsilon-1}]^{-\epsilon}[\frac{\phi_{jt}}{w_t}]^{\epsilon-1}P_{it}^{\epsilon-1} \\ &= e^{\mu_j}y\frac{1}{\epsilon-1}[\frac{\epsilon}{\epsilon-1}]^{-\epsilon}[\frac{\phi}{w}]^{\epsilon-1}P^{\epsilon-1},\end{aligned}$$

where again all the variables take on their steady state values.

The expected discounted lifetime profits from selling to consumer i are given by:

$$\begin{aligned}E\pi_{ij} &= E[\int e^{-rt}e^{\mu_j}y\frac{1}{\epsilon-1}[\frac{\epsilon}{\epsilon-1}]^{-\epsilon}[\frac{\phi_j}{w}]^{\epsilon-1}P^{\epsilon-1}dt] \\ &= AG\phi_j^{\epsilon-1}\int e^{-rt}E[e^{\mu_j}]dt \\ &= AG\phi_j^{\epsilon-1}\int e^{-rt}E[e^{\mu_j}]dt \\ &= AG\phi_j^{\epsilon-1}\int e^{-rt}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]dt \\ &= AG\phi_j^{\epsilon-1}\frac{1}{r}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}],\end{aligned}$$

where r is the discount (interest) rate, $AG \equiv y\frac{1}{\epsilon-1}[\frac{\epsilon}{\epsilon-1}]^{-\epsilon}[\frac{P}{w}]^{\epsilon-1}$, the aggregate demand variables, and p is the belief on the part of the firm that $\mu(\omega) = \bar{\mu}$.

The firm has to choose the optimal number of consumers to target subject to the following cost structure: there are two distribution and marketing technologies available. The first technology has zero or negligible sunk costs, but the cost of selling to n consumers is convex in n: $c(n)$ is continuous, strictly increasing and convex. The second technology is linear in n, but to use this technology, the firm has to pay the fixed cost F, which could reflect expenditures on building own distribution and retail centers, and hiring long-term marketing agencies. Thus, the firm will find it optimal to eventually pay the fixed cost F and utilize the linear technology. To derive the optimal behavior of the firm, consider first the learning of the firm.

The firm does not observe the precise values of quantities sold, i.e. it does not observe the precise value of μ. Instead, it observes the following for an individual consumer indexed by i:

$$X_{ijt} \equiv \int_0^t \ln(\frac{c_{ijt}}{y(p_j)^{-\epsilon}P^{\epsilon-1}})ds + \sigma_x W_{ijt} = \int_0^t \mu_j ds + \sigma_x W_{ijt},$$

so that

$$dX_{ijt} = \mu_j dt + \sigma_x dW_{ijt},$$

where W_{ijt} is a Wiener process, and the firm can update its beliefs about μ_j from these observations. More precisely, denote by p_{jt} the (subjective) probability at time t that firm j's variety has high demand, $\mu_j = \bar{\mu}$. Upon sampling n_{jt} consumers and observing n_{jt} values of dX_{ijt}, and the sample average $\overline{dX_{jt}} \equiv \frac{\sum_{i=1}^{n_{jt}} dX_{ijt}}{n_{jt}}$, the firm updates its beliefs according to:

$$dp_{jt} = p_{jt}(1 - p_{jt})\frac{\bar{\mu} - \underline{\mu}}{\sigma_x}\sqrt{n_{jt}}dW_{jt},$$

where $\frac{\bar{\mu}-\underline{\mu}}{\sigma_x} \equiv \chi$ is the signal-to-noise ratio, and

$$\begin{aligned} dW_{jt} &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\frac{\sum_{i=1}^{n_{jt}} dX_{ijt}}{n_{jt}} - (p_{jt}\bar{\mu} + (1 - p_{jt})\underline{\mu})dt] \\ &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\overline{dX_{jt}} - (p_{jt}\bar{\mu} + (1 - p_{jt})\underline{\mu})dt] \end{aligned}$$

which is an observation-adapted Wiener innovation process , that is, it follows a standard Wiener process relative to the information at time t. How these assumptions can be intuitively translated into what kind of discrete-time data the firm (and the researcher) observe, is explained in the Appendix.

As can be seen from the equation above, when the firm observes a sample average $\overline{dX_{jt}}$ higher than the expected value at time t, $p_{jt}\bar{\mu} + (1 - p_{jt})\underline{\mu}$, it updates its beliefs upwards ($dp_{jt} > 0$), and downwards otherwise. The firm weighs this signal (the difference between observed average and expected average) by the signal-to-noise ratio, χ, and by the sample size n_{jt}. The higher the signal-to-noise ratio, and the higher the sample size, the more weight the firm puts on the new signal. So one can see that the firm would like to sample as many observations as possible, so as to learn faster, subject to the cost constraints.

Consider now the optimal behavior of the firm once it pays the sunk cost F and accesses the linear technology, $\tilde{c}(n) = fn$. Denote this stage as stage 2, and the stage before paying the sunk cost F as stage 1. In the absence of learning, the firm would sell to the maximum number M of consumers, as long as the expected profits covered the total costs, Mf, i.e. as long as $AG\phi_j^{\epsilon-1}\frac{1}{r}[pe^{\bar{\mu}} + (1 - p)e^{\underline{\mu}}] \geq f$. This would give us the threshold value of p_{jt}, below which the firm would quit the market (sell to 0 consumers). However, there is also information value to selling to a non-zero number of consumers, so that the value function of the firm j is given by:

$$V(p_{jt}) = max_{<n_{js}>}E[\int_0^{\infty}(-fn_{js} + n_{js}AG\phi_j^{\epsilon-1}[p_{js}e^{\bar{\mu}} + (1 - p_{js})e^{\underline{\mu}}])e^{-rs}ds|p_{jt}],$$

subject to $dp_{jt} = p_{jt}(1 - p_{jt})\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sqrt{n_{jt}}dW_{jt}$, where dW_{jt} is as above. Note that we do not take into account the possibility that the firm may choose to re-enter the market in the future, since in the model the aggregate variables and firm productivity are constant over time, and once the firm quits is does not get any new signals, and hence p_j also does not change. So once the firm quits, it stays out of the market. Of course, this is not a very realistic result, but it is easy to add the random shocks to aggregate variables and/or productivity to the model to match the observed re-entry rates in the data. However, our focus right now is on the optimal learning strategy of the firm, and how learning affects the value function, and not the effect of exogenous shocks on the behavior of the firm.

The Hamilton-Jacobi-Bellman equation is as follows (we omit the subscripts below):

$$rv(p) = max_{0 \leq n \leq M}[n(-f + AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) + n\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 v''(p)].$$

It is clear that the stage-2-value function is linear in n, so that the optimal size in stage 2 is $n_1^* = M$, as long as the value function is positive. To find the value function $v(p)$ for this case, plug in $n = M$ into the HJB equation:

$$rv(p) = M(-f + AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) + M\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 v''(p).$$

We need to solve this second-order non-linear ODE, subject to the value matching and smooth pasting conditions:

$$v(\underline{\underline{p}}) = 0,$$

$$v'(\underline{\underline{p}}) = 0,$$

and derive the threshold value $\underline{\underline{p}}$, below which the firm will quit the market. For any beliefs p_{jt} above this value the firm will sell to the maximum number of consumers, M. Notice that the value function with $v''(p) = 0$ satisfies the ODE above, which gives us:

$$v(p) = \frac{M}{r}(-f + AG(\phi_j)^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]).$$

Intuitively, the only value of learning to the firm comes from the effect of p_{jt} on the decision to quit the market, where quitting yields a payoff of 0. If we introduced exogenous shocks to the aggregate demand variables or productivity, the firm would possibly want to re-enter the market after quitting, in the case of favorable shocks, so, with the sunk cost of entry F, there would be value to learning resulting from this potential re-entry (and having to pay F again) in the future. However, for simplicity here we do not allow exogenous shocks, and therefore re-entry. Hence, the value function in the second stage is simply the sum of all discounted profits net of fixed costs of exporting. Denote this value function by $\tilde{V}(p)$. The threshold value of p_{jt}, below which the firm would quit the market (sell to 0 consumers) is given by:

$$\underline{\underline{p_j}} = \frac{\frac{f}{AG\phi_j^{\epsilon-1}\frac{1}{r}} - e^{\underline{\mu}}}{e^{\bar{\mu}} - e^{\underline{\mu}}}.$$

Now consider stage 1, when the firm employs the convex costs technology, before paying the sunk cost F. The firm gains profits from selling to n consumers, and has to pay the costs $c(n)$, but also gains information value from updating its beliefs and possibly investing in the linear technology in the future as a result. The function $c(n)$ is twice differentiable on $(0, M)$, and increasing and strictly convex on $[0, M)$. We can assume that $c(0) > 0$, or that $c(0) = 0$, depending on what empirical data we want to match. If there is a fixed cost of maintaining the ability to experiment (think of this for example as the necessity of keeping contracts with marketing agencies), then, given a constant payoff structure over time, whenever it is optimal for a firm to sample zero observations ($n = 0$), it is optimal for it to stop experimenting and make a terminal decision (quit the market). This can be proved in a manner similar to Morgan and Manning (1985) (Proposition 2).

Solution of the problem of Stage 1

This part of the problem is close to the problem of Moscarini and Smith (2001). One way in which our model is different from that of Moscarini and Smith's is that the testing phase in our case actually represents productive activity by the firm, which produces the product and ships it to the sample of consumers, even if not to the entire market. Hence, the firm earns profits in the experimentation phase, and these affect the optimal sample size n.

The value function of firm j takes the form

$$\begin{aligned} V(p_{jt}) &= max_{T,<n_{js}>} E[e^{-rT} K(p_T) \\ &\quad + \int_0^T (-c(n_{js}) + n_{js} AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}])e^{-rs} ds | p_{jt}], \end{aligned}$$

where $K(p) = max[\tilde{V}(p), 0]$ is the payoff to making the terminal decision, $\tilde{V}(p)$ is the value function derived for stage 2 above, T is the stopping time, and p_T is the value of the belief variable at time T. Then the Hamilton-Jacobi-Bellman equation (HJB) becomes

$$\begin{aligned} rv(p) &= max_{0 \le n \le M}[-c(n) + nAG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]] \\ &\quad + \frac{1}{2}(p(1-p)\frac{\bar{\mu} - \underline{\mu}}{\sigma_x})^2 nv''(p)]. \end{aligned}$$

The FOC for the HJB equation give us $n(p) = z(rv(p))$, where $z \equiv g^{-1}$, $g(n) = nc'(n) - c(n)$, and z is strictly increasing. The problem can be transformed into a two-point free boundary value problem

$$v''(p) = \frac{c'(z(rv(p))) - AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]}{\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2}$$

plus the value matching condition:

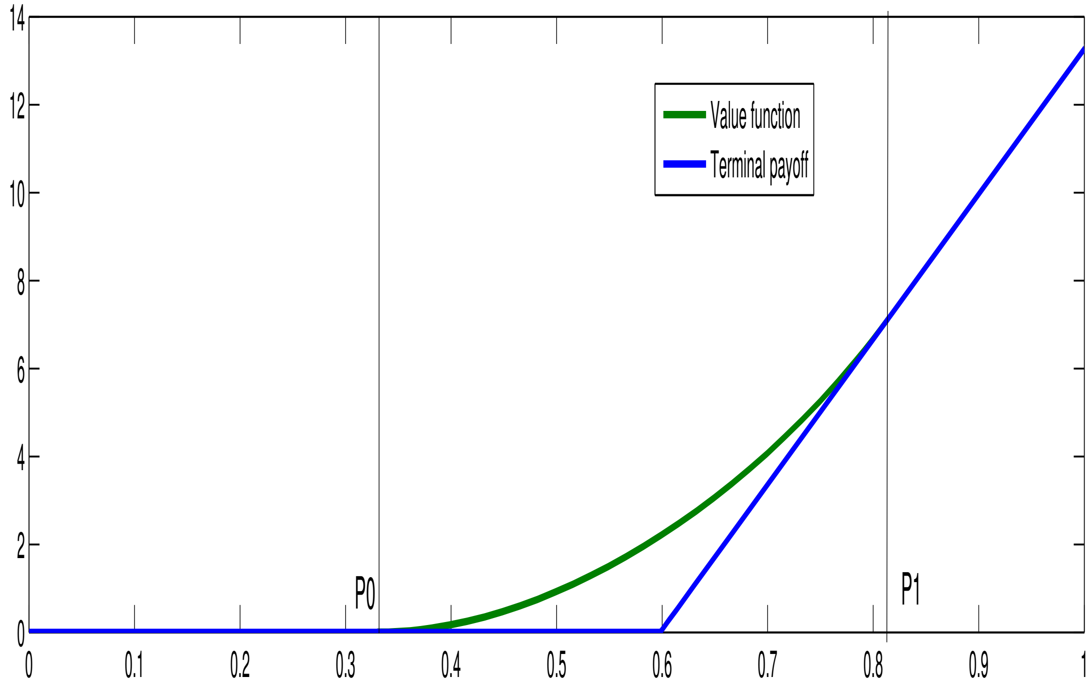

Figure 5: Solution to the problem

The blue line shows the ultimate payoff function, $\pi \equiv \max[0, \tilde{V}(\bar{p}) - F]$, and the green line shows the value function. The value $P0 \equiv \underline{p}$, where the value function is tangent to the zero-line (payoff from quitting), determines the cutoff for quitting, and the value $P1 \equiv \bar{p}$, where the value function is tangent to the profit line after entry, determines the cutoff for entering the market (paying sunk cost F)

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F, v(\underline{p}) = 0.$$

and smooth pasting condition:

$$v'(\bar{p}) = \tilde{V}'(p), v'(\underline{p}) = 0.$$

Since the value of experimentation $v(p)$ should be convex, we require

$$c'(z(rv(p))) > AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}],$$

for all p.

Intuitively, marginal cost of experimentation is the marginal testing cost net of the marginal profits earned from the sample sales:

$$MC = c'(n) - AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]$$

and the marginal benefit of experimentation is the contribution of the sample observations to the udpating of beliefs and to the value function as a result:

$$MB = \frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 v'' > 0,$$

so for an optimal n, where MC and MB are equated, MC should be positive.

It can be shown that if

$$c'(z(0)) > AG\phi_j^{\epsilon-1}pe^{\bar{\mu}} \geq AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}],$$

then $v(p) \geq 0$ and hence

$$c'(z(rv(p))) > AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}],$$

for all p. Moreover, under this assumption we can extend all the proofs of existence and uniqueness of a solution to the free boundary value problem in Moscarini and Smith (2001) (see Appendix).

Export participation condition

We consider only a partial equilibrium in this model, taking the domestic side of the economy as given. The firms that start exporting are assumed to have been producing and selling domestically for some time, and therefore they know their productivity. What matters when they make the decision to export, is the prevailing common belief about the distribution of μ in the foreign market, which is assumed to be the objective probability $p_0 = Prob(\mu = \bar{\mu})$. To find the cutoff for exporting, we need to know how productivity affects the exporting/experimentation behavior. It can be shown that as ϕ rises, both the threshold for switching to stage 2, $\bar{p}$, and the threshold for quitting the market, $\underline{p}$, fall. The firm will be willing to switch to the linear technology (that requires paying the sunk cost) for lower values of p, and will be willing to keep experimenting for lower values of p than before. The optimal experimentation intensity $n(p)$ will also shift up, so that the firm will target more consumers for any given p. This is shown in Figure 6.

Thus, there is a monotonic ranking of cutoff thresholds $\bar{p}$ and $\underline{p}$ over the productivity range. Given the common belief p_0, the lowest productivity exporter will have $\underline{p} = p_0$. Denote this productivity level as $\underline{\phi}$. All firms with productivity below this cutoff will have $\underline{p}$ higher than p_0, that is it will not be worthwhile for them to start exporting, even as experimenters. Thus, we have

$$\underline{p}(\underline{\phi}) = p_0.$$

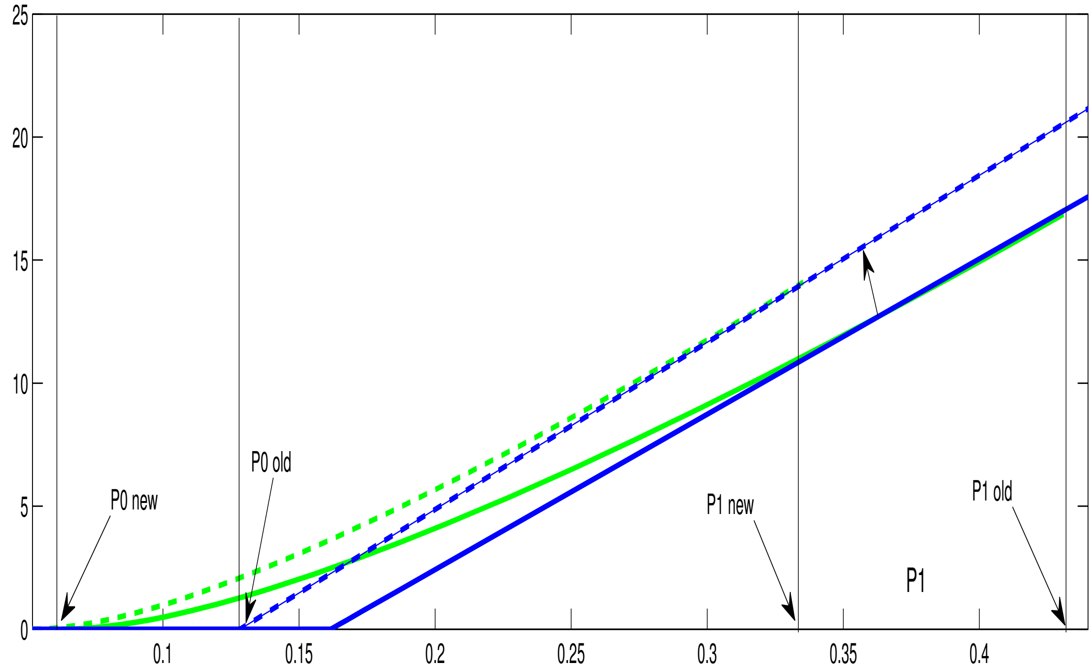

Figure 6: Comparing the solutions to the problem for two levels of productivity

The blue line shows the ultimate payoff function, $\pi \equiv \max[0, \tilde{V}(\bar{p}) - F]$, and the green line shows the value function. The dotted blue line shows the new ultimate payoff function, for higher productivity, and the dotted green line shows the new value function. The value function shifts up, and the thresholds fall - the thresholds for entering is lower, and the threshold for quitting is lower for the higher productivity firm.

Notice that this export participation condition is very different from the one we usually work with, where the expected lifetime discounted export profits of the lowest productivity exporter are just high enough to cover the sunk entry cost F. Here we cannot apply this rule, since the value of exporting is no longer given by expected lifetime profits for the lowest productivity exporter, but by the total value of exporting - which also includes the information value of export sales. Therefore, we must find the firm whose expected value is exactly zero at the initial belief p_0, which would be the firm whose $\underline{p}$ is p_0.

Similarly, it is not possible to tell whether a firm will forego experimenting or not by simply comparing its expected lifetime discounted profits with the sunk entry cost F. In fact, the firm that satisfies the equality between its expected lifetime discounted export profits with the sunk entry cost F ($\frac{M}{r}AG\phi_j^{\epsilon-1}[p_0e^{\bar{\mu}}+(1-p_0)e^{\underline{\mu}}] = F$) will choose to experiment first, since its $\bar{p}$ will be above p_0. Even though its expected profits from entry exceed F, it chooses to experiment first because of the information value it gains. As productivity increases, one of the two things can happen. Either the profit line for second stage (discounted expected lifetime profits from full-scale exporting) will cross the origin, and all firms with productivity above this level of productivity, denote it as ϕ_1 (where $\frac{M}{r}AG\phi_1^{\epsilon-1}e^{\underline{\mu}} = F$) will choose to enter right away, or for some $\phi < \phi_1$, the threshold for entering $\bar{p}$ will be just equal to p_0, and all firms with productivity at or above this level, call it ϕ_2, will enter right away. So, we get

$$\frac{M}{r}AG\phi_1^{\epsilon-1}e^{\underline{\mu}} = F,$$

$$\bar{p}(\phi_2) = p_0,$$

and the productivity of the highest productivity experimenter is given by

$$\bar{\phi} = min[\phi_1, \phi_2].$$

We show the two firms - the lowest productivity exporter, and the highest productivity experimenter in Figure 7.

So, we can conclude that in the presence of a testing technology, the cutoffs for experimenting and entering right away compare as follows with the cutoff for exporting in a standard Melitz-type model: $\underline{\phi} < \tilde{\phi} < \bar{\phi}$, where $\underline{\phi}$ is the cutoff for exporting (by experimenting first), and $\bar{\phi}$ is the cutoff for entering right away in the experimentation model, and $\tilde{\phi}$ is the cutoff for exporting in the standard model.

3 Comparative statics predictions

We now consider the comparative statics with respect to the main parameters.

- As ϕ rises, both the threshold for switching to stage 2, $\bar{p}$, and the threshold for quitting the market, $\underline{p}$, fall. The firm will be willing to switch to the linear technology (that requires paying the sunk cost) for lower values of p, and will be willing to keep experimenting for lower values of p than before. The optimal experimentation intensity $n(p)$ will also shift up, so that the firm will target more consumers for any given p.

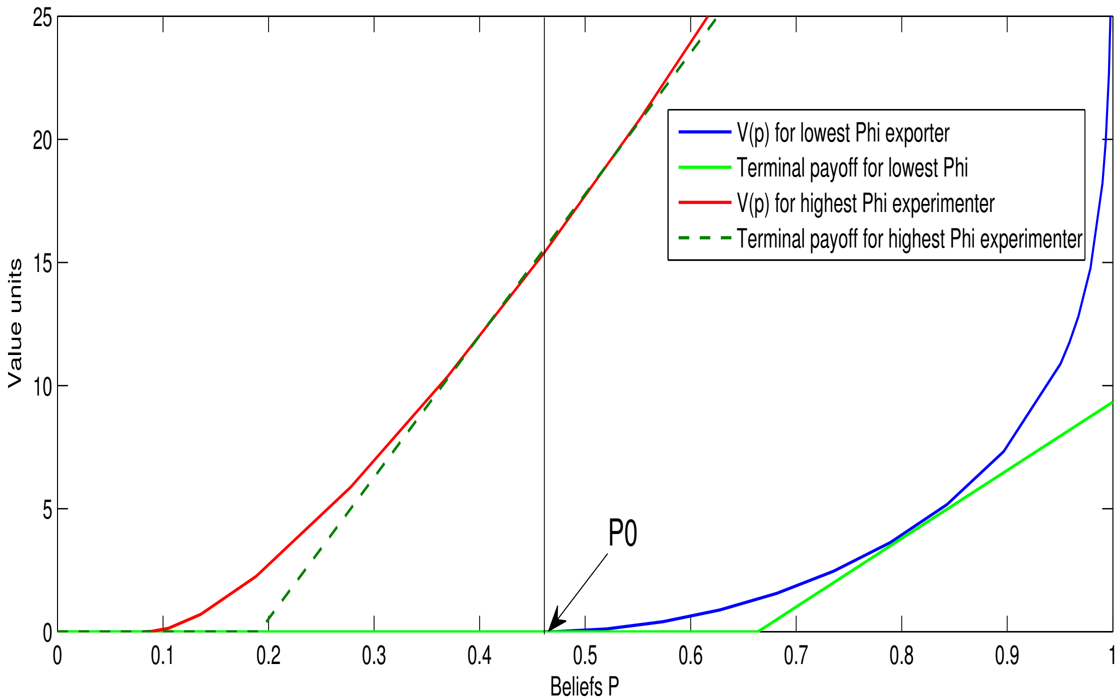

Figure 7: The cutoffs for exporting (experimenting) and entering right away

The blue curve shows the value function for the lowest productivity exporter, and the green line shows the terminal payoff function for this exporter. Since the threshold for quitting $\underline{p}$ for this firm is exactly at p_0, this firm is just indifferent between exporting (through experimenting first) and not exporting. The red curve shows the value function for the highest productivity experimenter, and the dotted green line - the terminal payoff function for this firm. Since its threshold for entering the market is exactly p_0, it is indifferent between experimenting and entering at the initial belief p_0. All firms with productivity in-between these two levels of ϕ will experiment, and all firms with productivity above the higher level of ϕ will enter right away.

- The same holds for higher AG, i.e. all aggregate variables. As y and P increase, or w falls, the terminal payoff function and value function shift up. Both the threshold for switching to stage 2, $\bar{p}$, and the threshold for quitting the market, $\underline{p}$, fall. The optimal intensity schedule $n(p)$ shifts up.

- As σ_x falls, so that the signal-to-noise ratio increases, the value function for all p shifts up, so that the optimal experimentation level n rises for all p, and the thresholds $\bar{p}$ and $\underline{p}$ shift out. Thus, in a market where the noise in observations for any given individual consumer is less dispersed, a given firm will experiment more intensively, and will require a better observed history of sales growth rates to enter, or a worse history - to quit. That is, experimentation will become more valuable to the firm, since any given amount of experimentation effort now results in more belief updating (remember the expression for the Marginal Benefit of experimentation).

- Similarly, as the testing cost function $c(n)$ grows less convex and initially (for n close to 0) weakly lower and less steep, the value function $v(p)$ rises for all p, so that optimal n rises for all p, and the thresholds shift out. The intuition is the same as above.

- As F rises, so that the final payoff to entry falls, the firm will experiment less intensively, for any given value of beliefs, and the threshold for entering the market will increase, that is, the firm will require a more successful history of sales, ceteris paribus, to enter the market. The firm will also quit the market for higher values of p_t, than otherwise, since its expected profits from experimentation (which is directly related to the payoff of entry) is now lower.

We consider the implications of these comparative statics results for the observable features of new exporters' behavior - export size, export volatility (over time), and experimentation phase duration. These results tell us how various characteristics of the firms and the market will affect the size dynamics of new exporters, their export sales volatility, and duration of experimentation phase. The volatility of exports is affected through two channels - the relative length of the two stages (experimentation, where volatility is high, and full-scale, where volatility is zero in the model), and the experimentation intensity. The higher is $n(p)$, the larger is the variance $Var(dp)$, that is beliefs change much more with more observations sampled, and hence, the larger is the change in n in the next period.

Notice that the duration of the experimentation phase becomes endogenous in this model - the experimentation phase is not only multi-period, unlike in other models of learning, but its duration is random, firm-specific, and can be determined in expected value by the productivity of the exporter, the size of the market, the magnitude of the sunk cost, and testing costs, and the signal-to-noise ratio in the market. These effects are also not straightforward and should be determined numerically for each set of values of the parameters, since the duration of the experimentation phase depends on two things: the thresholds for quitting and entering the market, and the intensity of experimentation ($n(p)$). The higher the intensity of experimentation $n(p)$, the faster beliefs p attain the thresholds for quitting or entering.

We have already discussed the implications of the comparative statics results with respect to productivity for the cutoffs for exporting above. Consider now the implications for volatility. We should see firms with higher productivity that start exporting full-scale right away, and whose export volatility is low. There will be firms with lower productivity that will not start with full-scale exports, and will export small volumes, and exhibit high volatility initially. Some of these, however, will later switch to higher export volumes (full scale), and become more stable. The higher the productivity of these firms, the more likely they are to do so, and the sooner they will do so. We show a result from the empirical paper by Akhmetova and Poncet (2009) in Table 1. In the regression, the dependent variable is the standard deviation of growth rates of quantity exported by 8-digit product and destination for all exporting firms in the manufacturing sector.The sample period is 1995-2005. We define a dummy for a volatile exporter, where a volatile exporter is defined as a new exporter whose standard deviation of growth of quantity exported is higher than the standard deviation of the mean growth rate over the relevant product and destination. The result shows that more productive exporters on average are less volatile (the coefficient on productivity is negative), however, conditional on being a volatile exporter (which in our model means the firm is an experimenter), higher TFP translates into higher volatility (the coefficient on the interaction between TFP and the volatile exporter dummy is positive), since, as predicted in the model, higher productivity implies higher experimentation intensity $n(p)$ and higher volatility for experimenters.

Table 1: Volatility as depending on productivity

Independent Variable	Dependent variable		
	StDev of growth	StDev of growth	StDev of growth
Volatile exporter	51.108	0.775	1.917
	(8.52)**	(38.37)**	(56.22)**
Log-tfp	5.966	-0.179	-0.306
	(1.01)	(7.60)**	(7.64)**
Volatile exporter* Log-tfp	6.381	0.367	0.675
	(1.15)	(14.81)**	(17.52)**
Constant	-8.682	1.665	2.052
	(1.01)	(60.27)**	(41.09)**
Restr. On Std growth	NO	< 11	< 26
Observations	236680	211390	223922
R-squared	0.13	0.34	0.33

Robust z-statistics in parentheses.

*significant at 5%; ** significant at 1%.

in the last two columns we bounded the standard deviation of growth rates from above

- by 11 in the second column, and by 26 in the second column

Second, the market size y, the toughness of the market as measured by P, and any other variables that affect average profitability in the market (for example, tariff rates) will affect the optimal export/experimentation behavior. The higher the profitability in the market, the more likely are the exporters to forego the experimentation phase, and to transition from the experimentation stage into a full-scale phase, if they do experiment. While they experiment, however, they will do so more intensively, and thus volatility of their (observed) exports will be higher. Thus, we will observe a more prompt switch to a full scale, fewer (relatively) firms with high export volatility, but higher volatility among the volatile firms, in the larger markets, more liberalized markets, and less competitive markets.

Consider the effect of a higher signal-to-noise ratio: as the signal-to-noise ratio increases, so does the optimal $n(p)$ schedule. Hence, there will be more (proportionally) new exporters that are highly volatile, and these will be more volatile than previously. The effect on the duration of the experimentation phase is ambiguous. The intensity of experimentation $n(p)$ falls, which speeds up the convergence of beliefs. However, the thresholds for quitting and entering the market shift out as the signal-to-noise ratio increases, and the firm will actually keep experimenting at values of beliefs p where it previously would have made a terminal decision. Thus, the expected duration of experimentation may go up or down, depending on the relative magnitudes of these effects.

Analogously, the total costs of entry, as measured by the sum of testing costs incurred and the sunk cost of entry (in the case of successful experimentation) may rise or fall with the signal-to-noise ratio. Hence, even though the conditions of market access improve as the signal-to-noise ratio increases, and therefore the firms learn much more from testing, the firm now will pay more in expected value to enter the market. This is an interesting result - we cannot judge how easily accessible the market is just by surveying the exporters about the cost of entry into the market, unless we ask them specifically what they mean by the costs of entry, and how they carry out the process of entry. If there is a testing phase, higher total entry costs in a market may not mean that this market is tough to enter - on the contrary, the exporters may optimally pay the higher cost in this market in response to higher information value.

In markets with higher F, we expect to see less intensive experimentation, and a higher threshold on beliefs for entering the market ($\bar{p}$). This tells us that fewer firms will start with full-scale exports right away, that firms will take much longer to switch to the full-scale phase from the experimentation phase, and they will be less volatile in the experimentation phase. One proxy of F could be the distance to the market, so that we expect to see lower initial volumes of exports, and slower duration of this phase in more distant markets.

4 Alternative models of learning

Here we solve alternative models of learning in the same framework, and compare their predictions and their implications for the estimates of exporting costs.

Melitz model

While there are two levels of demand, $\bar{\mu}$ and $\underline{\mu}$, the firms know exactly what their demand is once they start exporting. That is, they only know the probability of a high demand before they enter, but once they pay F, they are told what their demand is (before they even export). Hence, there is no learning. There is a convex periodic cost of exporting. Every firm therefore exports a constant optimal number of products, that depends on its productivity, and demand level, μ_j:

$$n^* = (c')^{-1}(AG\phi_j^{\epsilon-1}e^{\mu_j}),$$

where all the variables are as defined above.

Passive learning with linear costs of exporting

If there is demand uncertainty in the market in the same sense as presented in the experimentation model, but the firm has to pay the sunk cost right away, and the exporting costs are linear in n, then we have the same problem of the firm as in stage 2 of the experimentation model. That is, the value function of the firm j is given by:

$$V(p_{jt}) = max_{<n_{js}>}E[\int_0^\infty(-fn_{js} + n_{js}AG\phi_j^{\epsilon-1}[p_{js}e^{\bar{\mu}} + (1-p_{js})e^{\underline{\mu}}])e^{-rs}ds|p_{jt}],$$

subject to $dp_{jt} = p_{jt}(1-p_{jt})\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sqrt{n_{jt}}dW_{jt}$, where dW_{jt} is as above.
The Hamilton-Jacobi-Bellman equation is as follows (we omit the subscripts below):

$$rv(p) = max_{0\leq n\leq M}[n(-f + AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) + n\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2v''(p)].$$

The value function is linear in n, so that the optimal size is $n_1^* = M$, as long as the value function is positive. Plug in $n = M$ into the HJB equation:

$$rv(p) = M(-f + AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) + M\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2v''(p).$$

The value function with $v''(p) = 0$ satisfies the ODE above, which gives us:

$$v(p) = \frac{M}{r}(-f + AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]).$$

The threshold value of p_{jt}, below which the firm would quit the market (sell to 0 consumers) is given by:

$$\underline{\underline{p_j}} = \frac{\frac{f}{AG\phi_j^{\epsilon-1}\frac{1}{r}} - e^{\underline{\mu}}}{e^{\bar{\mu}} - e^{\underline{\mu}}}.$$

Thus, the only response to demand signals (beliefs) takes the form of exit by the firm if beliefs fall below this threshold, and beliefs do not affect the export size.

Passive learning with convex costs of exporting

Assume that the firm has to pay the sunk cost right away, to start exporting, and that there are convex, strictly increasing fixed costs of exporting in every period, $c(n)$. The firm faces the same kind of uncertainty as before, and its beliefs about μ_j evolve in the same fashion as before, but now the firm ignores the effect of its size n. That is, it ignores the effect of n_t on $\text{Var}(dp_t)$, shown in the main theoretical part. This means that the firm engages only in passive learning, and maximizes only current profits in every period. Formally, the firm's value function is

$$V(p_{jt}) = max_{<n_{js}>}E[\int_0^\infty(-c(n_{js}) + n_{js}AG\phi_j^{\epsilon-1}[p_{js}e^{\bar{\mu}} + (1-p_{js})e^{\underline{\mu}}])e^{-rs}ds|p_{jt}],$$

subject to

$$dp_{jt} = p_{jt}(1-p_{jt})\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sqrt{n_{jt}}dW_{jt},$$

where $\frac{\bar{\mu}-\underline{\mu}}{\sigma_x} \equiv \chi$ is the signal-to-noise ratio, dX_{ijt} is the i-th observation at time t of firm j, $\overline{dX_{jt}} \equiv \frac{\sum_{i=1}^{n_{jt}} dX_{ijt}}{n_{jt}}$ and

$$\begin{aligned} dW_{jt} &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\frac{\sum_{i=1}^{n_{jt}} dX_{ijt}}{n_{jt}} - (p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt] \\ &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\overline{dX_{jt}} - (p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt] \end{aligned}$$

which is an observation-adapted Wiener innovation process , that is, it follows a standard Wiener process relative to the information at time t.

The Hamilton-Jacobi-Bellman equation is

$$\begin{aligned} rv(p) &= \max_{0\leq n\leq M}[-c(n) + nAG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] \\ &\quad + \frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 nv''(p)]. \end{aligned}$$

Now, we impose the assumption that the firm ignores the presence of n in the second term in the HJB equation, and the FOC for the HJB equation gives us

$$AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] = c'(n),$$

$n^*(p) = [c']^{-1}(AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) \equiv q(p)$.

The SOC is satisfied due to the convexity of $c(n)$.The problem then becomes

$$v''(p) = \frac{rv(p) + c(q(p)) - q(p)AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]}{\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 q(p)},$$

with the free boundary conditions

$$v(\underline{p}) = 0,$$

$$v'(\underline{p}) = 0,$$

for some $\underline{p}$.

Moreover, there is an upper bound on the value function: the best the firm can hope for is that at some point in the future, say, $s > t$, its beliefs p will converge to 1 and stay there. In that case, its value function at time s will be simply

$$\frac{AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r},$$

where $\tilde{n} \equiv (c')^{-1}(AG\phi_j^{\epsilon-1}e^{\bar{\mu}})$.

The best scenario would be the one where $s = t + 1$, that is, the firm's beliefs converge to 1 already in the next period. Therefore, the value function is bounded above by

$$AG\phi_j^{\epsilon-1}e^{\mu_j}n^*(p) - c(n^*) + (1-r)\frac{AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r},$$

since

$$\int_0^\infty e^{-rs}[AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})]ds = \frac{AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r}.$$

Notice that $v(p)$ is undefined at $p = 1$ (from the ODE for $v(p)$). Instead, the value function goes in the limit to this maximum value as $p \to 1$:

$$\lim_{p\to 1} v(p) = \frac{AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r}.$$

This gives us the following observations:

$$\lim_{p\to 1} v''(p) = \infty,$$

$$\lim_{p\to 1} v''(p)p^2(1-p)^2 = 0.$$

which provide additional conditions on the threshold $\underline{p}$ and the value function $v(p)$. We need to solve the BVP (plus the two limits) above to find the optimal $v(p)$ and the threshold for exiting the market $\underline{p}$. This can be done easily numerically.

The gist of the solution derived is that the firm learns every period, decides whether to exit or not, and if it stays, sets n to maximize current profits. However, now, unlike in the model of Passive Learning with linear costs, the firm sets export size optimally in response to beliefs - if its beliefs improve, it expands, and contracts, if its beliefs deteriorate. This happens because the firm now expects higher sales per consumer, which can cover the higher marginal cost of exporting, but not because the firm wants to learn more (as in the active learning models).

Active learning with convex costs throughout, and no postponing sunk entry cost F

Assume the following environment for the firm. The firm has to pay the sunk cost of entry right away to access the market - i.e. in order to sell a non-zero quantity in the market, the firm pays sunk cost F. The firm then gains access to the marketing and distribution technology that is convex in the number of consumers it sells to, say $c(n)$, where n is again the number of consumers. The demand for the product and production technology are just as above, that is, only the foreign market distribution technology is different this time. The firm, therefore, again is uncertain about its demand - whether it has drawn a high or low μ - $\bar{\mu}$ or $\underline{\mu}$. The firm has to decide on the optimal n, and whether to export or not (quit). This time the firm recognizes the information value of its sale. The value function of the firm is:

$$V(p_{jt}) = max_{<n_{js}>} E[\int_0^\infty (-c(n_{js}) + n_{js} AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}])e^{-rs} ds | p_{jt}],$$

subject to

$$dp_{jt} = p_{jt}(1-p_{jt})\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sqrt{n_{jt}} dW_{jt},$$

where $\frac{\bar{\mu}-\underline{\mu}}{\sigma_x} \equiv \chi$ is the signal-to-noise ratio, $\overline{dX_{jt}} \equiv \frac{\sum_{i=1}^{n_{jt}} dX_{ijt}}{n_{jt}}$ and

$$\begin{aligned} dW_{jt} &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\frac{\sum_{i=1}^{n_{jt}} dX_{ijt}}{n_{jt}} - (p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt] \\ &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\overline{dX_{jt}} - (p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt] \end{aligned}$$

which is an observation-adapted Wiener innovation process , that is, it follows a standard Wiener process relative to the information at time t.

The Hamilton-Jacobi-Bellman equation is

$$\begin{aligned} rv(p) &= \max_{0 \le n \le M}[-c(n) + nAG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] \\ &\quad + \frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 nv''(p)]. \end{aligned}$$

The FOC for the HJB equation give us $n^{**}(p) = z(rv(p))$, where $z \equiv g^{-1}$, $g(n) = nc'(n) - c(n)$, and z is strictly increasing. The problem then becomes

$$v''(p) = \frac{c'(z(rv(p))) - AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]}{\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2}$$

plus the value matching condition:

$$v(\underline{p}) = 0,$$

and smooth pasting condition:

$$v'(\underline{p}) = 0.$$

Just as in the Passive Learning with convex costs model above, the value function is bounded above by

$$\frac{AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r},$$

where $\tilde{n} \equiv (c')^{-1}(AG\phi_j^{\epsilon-1}e^{\bar{\mu}})$.

The value function goes in the limit to this maximum value as $p \to 1$:

$$\lim_{p\to 1} v(p) = \frac{AG\phi_j^{\epsilon-1}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r}.$$

This gives us the following conditions:

$$\lim_{p\to 1} v''(p) = \infty,$$

$$\lim_{p\to 1} v''(p)p^2(1-p)^2 = 0.$$

These two equations impose additional conditions on the optimal n_{js}, and $\underline{p}$, apart from the BVP above.

This is intuitive: the information value comes from two sources - first, the value of not exiting the market unnecessarily, while future sales may reveal that staying is actually profitable (this value is especially high when p is close to 1), and second, the value of correctly setting the optimal sales size n. As p goes to 1, the first source of information value becomes less important, since p is very high, and the second source becomes less important, since it is very likely that p will converge to 1 and stay there, in which case the optimal size is simply $\tilde{n}$. Hence, the information value $\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 n v''(p)$ goes to 0. This is depicted in Figure 19.

Contrasting the four models of learning

We would like to compare the predictions of the four models of learning, given the same structural parameters (such as $F, c(n), f, \bar{\mu}, \underline{\mu}, \sigma_x$).

First, compare the predictions for export size n. Clearly, the interesting comparison here is that between the export size under Passive Learning with convex costs, and Active Learning. Export size for the same (convex) cost schedule $c(n)$ will be lower under Passive

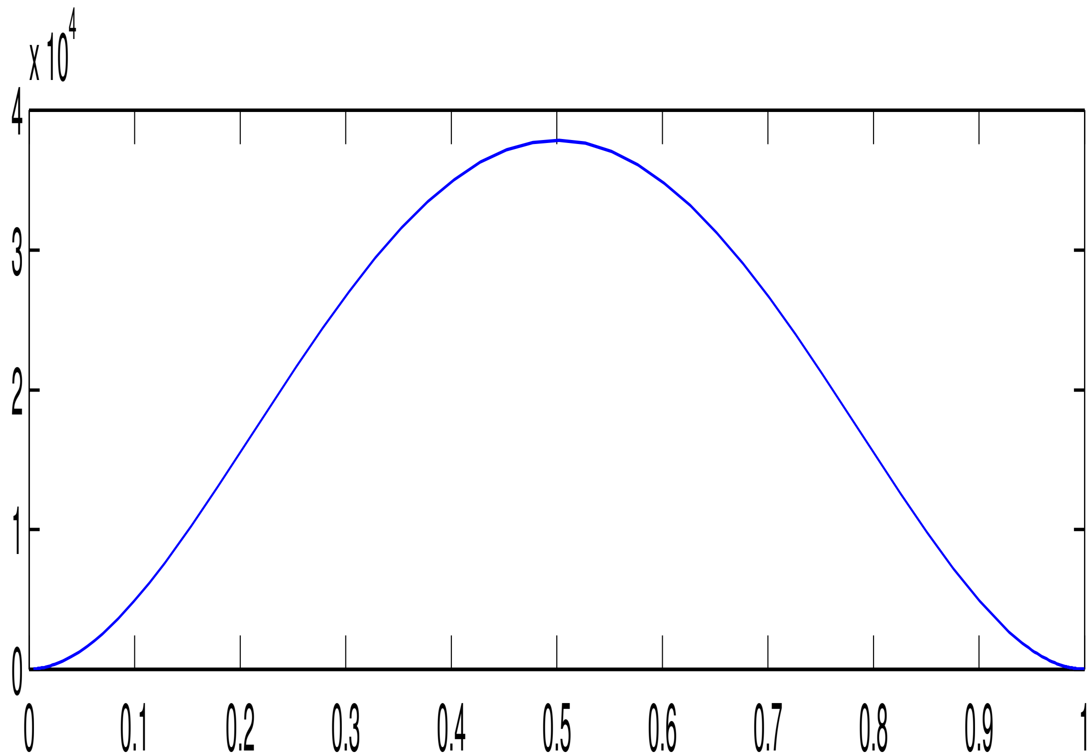

Figure 8: The information value

The information value, given by the expression $\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2 v''(p)$, attains a maximum for some $p \in (0,1)$, and goes to 0 as p goes to 1.

Learning with convex costs than, for any given beliefs value p, under Active Learning. From the FOC for PL,

$$c'(n^{pl}) = \tilde{\pi}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}],$$

and from FOC and the HJB equation for AL,

$$\begin{aligned} n^{al}c'(n^{al}) - c(n^{al}) &= rv^{al}(p), \\ &= -c(n^{al}) + n^{al}\tilde{\pi}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] \\ &\quad + \frac{1}{2}(p(1-p)\chi)^2 n^{al}[v^{al}]''(p). \end{aligned}$$

Simplifying the last equation, we get

$$\begin{aligned} c'(n^{al}) &= \tilde{\pi}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] \\ &\quad + \frac{1}{2}(p(1-p)\chi)^2[v^{al}]''(p) \geq \tilde{\pi}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] \\ &= c'(n^{pl}), \end{aligned}$$

since $(p(1-p)\chi)^2[v^{al}]''(p) \geq 0$ for all $p \in [\underline{p}^{al}, 1)$ for the solution $v^{al}(p)$. Since $c(n)$ is convex, we obtain

$$n^{al}(p) \geq n^{pl}(p)$$

for all p where both solutions are defined .

Also, in both models the solution $v(p)$ satisfies the condition

$$\lim_{p \to 1} v(p) = \frac{\tilde{\pi}e^{\bar{\mu}}\tilde{n} - c(\tilde{n})}{r}.$$

where $\tilde{n} \equiv (c')^{-1}(AG\phi_j^{\epsilon-1}e^{\bar{\mu}})$.

Therefore, both the value functions are bounded above by and go to this value as p goes to 1. Knowing that in the Passive Learning model the firm (by assumption) ignores the effect of its export size n on learning, and sets it only to maximize current profits in every period, we can conclude that $v^{al}(p) > v^{pl}(p)$ for all p where both value functions are defined. Combining this with the boundary conditions for $v^{pl}(p)$, we get

$$v^{al}(\underline{p}^{pl}) > v^{pl}(\underline{p}^{pl}) = 0,$$

that is, to satisfy the boundary condition on $v^{al}(p)$, we need

$$\underline{p}^{al} < \underline{p}^{pl}.$$

Thus, we can compare the solutions to the PL and AL models: in the AL model, the threshold for quitting is lower, and the export size and the value function are higher for all values p, where the PL solution is also defined. By the same token, the export size is higher in the experimentation stage in the Experimentation model than the export size in the Passive Learning model (with convex costs), given the same exporting costs $c(n)$ in the experimentation stage.

This also implies that volatility of exports will be higher under the active learning models, than under the passive learning ones. Starting from the same initial beliefs, the firm will set lower size n under Passive Learning. Since $Var(dp_t)$ is proportional to n, that is, the variance in beliefs is higher the higher the sample size, we will get smaller changes in future beliefs and future export size, under the Passive Learning assumption.

Another way to compare the models is to consider the implications for the estimates of export costs they would provide. Given the same data, one would require lower exporting cost schedule $c(n)$ to match the same export sizes of firms in the Passive Learning model with convex costs of exports, than in the Active Learning models (single-stage or two-stage), because of the observation we just made that the export size is lower in the former under the same cost schedule. So to generate the same export sizes in both models, one needs a lower cost schedule in the Passive Learning model.

Second, the estimate of the sunk cost of entry F will in general be higher in the Experimentation model than in the Passive Learning or single-stage Active Learning model, or a Melitz type with no learning model. This can be seen intuitively by thinking of the way we measure sunk entry costs. When we measure sunk entry costs in standard models, we take note of the lowest productivity exporter in the market, and evaluate this firm's expected lifetime discounted profits from exporting. We use the free entry condition (or export participation condition) to estimate F as equal to this value. Suppose the lowest productivity exporter has not yet incurred the sunk entry cost F, and is only testing the market. Then this firm's lifetime discounted profits may be actually much lower than the sunk entry cost F, at the initial belief p_0. So we will underestimate F based on observing these lowest productivity exporters. We need to fit the experimentation model to correctly evaluate exporting costs.

5 Conclusion

We present a model with demand uncertainty and sunk costs of entry. The firm is able to learn more about the demand before incurring the sunk cost of entry, by using a costly testing technology that allows to sample individual sales observations with some noise, and update its beliefs about the parameter of interest. Thus, the decision for new exporters is no longer a binary choice between entry and non-entry, but an optimal control and optimal stopping problem where the scale and duration of the initial testing phase is optimally chosen, depending on the characteristics of the firm, the destination and the product. In particular, we present interesting comparative statics results with respect to the features of uncertainty, such as the signal-to-noise ratio, where a higher signal-to-noise ratio in the market may either prolong or shorten the duration of the experimentation phase. This also gives us the observation that the expected total costs of entry (the sum of testing costs and sunk entry cost) may be higher in markets with higher signal-to-noise ratio. Therefore, measuring total costs of entry is not sufficient to judge how easy it is to enter the market. We show that the zero-cutoff profit condition for exporters is very different in this model, than in standard Melitz type models, or in other models of learning.

Overall, the model proposed is very intuitive, very flexible, in that it allows two stages, that can accommodate different phases of the life cycle of an exporter, and allows for endogenously determined duration and intensity of the experimentation phase. We believe that it can provide a framework for numerous further applications.

References

Akhmetova Z. and C. Mitaritonna (2010), Structural Estimation of the Model of Firm Experimentation, Working Paper.

Akhmetova Z. and S.Poncet (2009), New Exporter Dynamics, Working Paper, CEPII.

Albornoz F., H.F.C.Pardo, G.Corcos and E.Ornellas (2009), Sequential Exporting, CEP Discussion Papers.

Arkolakis C. (2006), Market Penetration Costs and the New Consumers Margin in International Trade, NBER Working Papers.

Bernard A.B., J.B.Jensen, S.J.Redding and P.K.Schott (2009), The Margins of US Trade, American Economic Review, Vol. 99, No. 2 (May, 2009),pp. 487-93.

Bolton P. and C. Harris (1999), Strategic Experimentation, Econometrica, Vol. 67, No. 2 (March, 1999), pp. 349-374.

Eaton, J., S. Kortum, and F. Kramarz (2008), An Anatomy of International Trade: Evidence from French Firms, NBER Working Papers.

Eaton J., M.Eslava, C.J.Krizan, M.Kugler and J.Tybout (2007), Export Dynamics in Colombia: Firm-Level Evidence, Working Paper, BANCO DE LA REPBLICA.

Eaton J., M.Eslava, C.J.Krizan, M.Kugler and J.Tybout (2008), A Search and Learning Model of Export Dynamics, Working Paper.

Ferber, R., (Ed., 1974), Handbook of Marketing Research, New York, McGraw-Hill book company.

Freund C. and M.D.Pierola (2008), Export Entrepreneurs: Evidence from Peru, Policy Research Working Paper Series, World Bank.

Moscarini G. and L.Smith (2001), The Optimal Level of Experimentation, Econometrica, Vol. 69 (Nov., 2001), No. 6, pp. 1629-1644.

Nguyen D.X. (2008), Demand Uncertainty: Exporting Delays and Exporting Failures, Discussion Paper, Dept. of Economics, University of Copenhagen.

Rauch, J.E. and J.Watson (2003), Starting Small in an Unfamiliar Environment, International Journal of Industrial Organization, Vol. 21(7), pp. 1021-1042.

Ruhl K.J. and J.L.Willis (2008), New Exporter Dynamics, Working Paper, NYU Stern School of Business.

Segura-Cayuela and Vilarrubia (2008), Uncertainty and Entry into Export Markets, Banco de Espaa Working Papers.

Appendix

Verifying some comparative statics results of Moscarini and Smith (2001)

To study the effect of including profits in the testing phase in the value function on the behavior of the solution, modify the value function by introducing an indicator Δ in front of the testing phase profits, which is assumed to be continuous, and range from 0 to 1. If Δ is 0, testing profits do not enter the value function, if it is 1, they do so fully.

$$\begin{aligned} V(p_0) &= max_{T,<n_t>} E[e^{-\delta T}\pi(p_T) \\ &+ \int_0^T (-c(n_t) + \Delta n_t A \frac{p_t \bar{\mu} + (1-p_t)\underline{\mu}}{\bar{\mu} + \underline{\mu}}) e^{-\delta t} dt | p_0]. \end{aligned}$$

Take the derivative of the value function with respect to Δ, using the envelope theorem:

$$V_\Delta(p_0) = E[\int_0^{T^*} n_t^* A \frac{p_t \bar{\mu} + (1-p_t)\underline{\mu}}{\bar{\mu} + \underline{\mu}}) e^{-\delta t} dt | p_0] > 0,$$

where T^* and n_t^* are the optimal values of the stopping time and sample size, respectively. Hence, the value function increases for all p_t when Δ goes from 0 to 1. This implies, by the Theorem in Moscarini and Smith (2001), that in the case with testing profits in the value function, compared with the case without, $n(p_t) = f(rv(p_t))$ goes up for all p_t, and $\bar{p}$ goes up and $\underline{p}$ goes down. Hence, the firm experiments more intensively and the thresholds for it to take a terminal decision expand, when we allow the firm to gather the profits in the experimentation phase.

To make sure that all the comparative statics of the solution with respect to key parameters derived in Moscarini and Smith still hold, we check if all the proofs of these results can be replicated.

Consider the comparative statics with respect to $A \equiv y \frac{1}{\epsilon-1}[\frac{\epsilon}{\epsilon-1}]^{-\epsilon}[\frac{\phi_j}{w}]^{\epsilon-1} P^{\epsilon-1}$, which will effectively give us the comparative statics with respect to any of ϕ, P, and w. The derivative of the value function (with Δ set to 1) with respect to A, using the envelope theorem, is

$$\begin{aligned} V_A(p_0 | \Delta = 1) &= E[\int_0^{T^*} n_t^* \frac{p_t \bar{\mu} + (1-p_t)\underline{\mu}}{\bar{\mu} + \underline{\mu}} e^{-\delta t} dt | p_0] \\ &+ \mathtt{Prob}(p_T = \bar{p} | p_0, n^*, T^*) E[e^{-rT^*} | p_T = \bar{p}, p_0] * \\ &* (\bar{p} \frac{\bar{\mu}}{(\bar{\mu} + \underline{\mu})\delta} + (1-\bar{p}) \frac{\underline{\mu}}{(\bar{\mu} + \underline{\mu})\delta}) > 0. \end{aligned}$$

Hence, we already know that the value function increases for all p_0 as A rises, and since $n(p_0)$ is increasing in $v(p_0)$, so does the experimentation intensity n for each p.

Let $A_1 > A_0$, and denote the corresponding thresholds for quitting as $\underline{p}_1$ and $\underline{p}_0$, and for entering as $\bar{p}_1$ and $\bar{p}_0$, respectively. Since the value function of the problem for $A = A_1$, evaluated at $\underline{p}_0$ is, by linear approximation,

$$V(\underline{p}_0|A_1) \simeq V(\underline{p}_0|A_0) + V_A(\underline{p}_0|A_0)(A_1 - A_0) > 0,$$

by value matching ($V(\underline{p}_0|A_0) = 0$), and $V(p)$ being increasing, we get $\underline{p}_1 < \underline{p}_0$. That is, the threshold for quitting goes down as A increases. This result is the same as in Moscarini and Smith (2001).

Next, we want to see what happens to the threshold for entering, $\bar{p}$. Consider

$$\begin{aligned} V_p(\bar{p}_0|A_1) &\simeq V_p(\bar{p}_0|A_0) + V_{pA}(\bar{p}_0|A_0)(A_1 - A_0) \\ &= \frac{A(\bar{\mu} - \underline{\mu})}{(\bar{\mu} + \underline{\mu})\delta} + V_{pA}(\bar{p}_0|A_0)(A_1 - A_0). \end{aligned}$$

If $V_{pA}(\bar{p}_0|A_0) > 0$, then $V_p(\bar{p}_0|A_1) > \frac{A(\bar{\mu}-\underline{\mu})}{(\bar{\mu}+\underline{\mu})\delta}$, and since the smooth pasting condition has to be satisfied by the new solution, and by convexity of the value function $V(p)$, we will get $\bar{p}_1 < \bar{p}_0$. To check if this is true:

$$\begin{aligned} V_{pA}(p) &= V_{Ap}(p) \\ &= E[\int_0^{T^*} (n_t^* \frac{\bar{\mu} - \underline{\mu}}{\bar{\mu} + \underline{\mu}} + \frac{dn_t^*}{dp_0} \frac{p_t \bar{\mu} + (1 - p_t)\underline{\mu}}{\bar{\mu} + \underline{\mu}}) e^{-\delta t} dt] \\ &+ E[n_{T^*}^* \frac{p_{T^*}\bar{\mu} + (1 - p_{T^*})\underline{\mu}}{\bar{\mu} + \underline{\mu}} e^{-\delta T^*}|p_0] \frac{dE(T^*)}{dp_0} \\ &+ \frac{dP(p_T = \bar{p}|p_0)}{dp_0} E[e^{-rT^*}|p_T = \bar{p}, p_0] \\ &+ \texttt{Prob}(p_T = \bar{p}|p_0) \frac{dE[e^{-rT^*}|p_T = \bar{p}, p_0]}{dp_0}] * \\ &* (\bar{p} \frac{\bar{\mu}}{(\bar{\mu} + \underline{\mu})\delta} + (1 - \bar{p}) \frac{\underline{\mu}}{(\bar{\mu} + \underline{\mu})\delta}). \end{aligned}$$

For p_0 small enough, $\frac{dE(T^*)}{dp_0} > 0$, and for p_0 close to 1, $\frac{dE(T^*)}{dp_0} < 0$. However, even if p_0 is close to 1,

$$\frac{dP(p_T = \bar{p}|p_0)}{dp_0} > 0,$$

and

$$\frac{dE[e^{-rT^*}|p_T = \bar{p}, p_0]}{dp_0} > 0.$$

Hence, if the last term is large enough relative to the preceding term, the cross-derivative should be positive. That is, if the optimal experimentation intensity is not too large compared to the entire market size (normalized by us to 1), so that the profits in the second to last term are small compared to the profits in the last term, then the threshold for entering the market $\bar{p}$ falls as A increases. This is intuitive: since A is a measure of profits from an average consumer in the market, if the optimal testing sample size is large and approaches the entire market size, then the firm will have little incentive to rush towards a terminal decision to enter the market, and even less so as A rises, so that the profits in the testing phase dominate the pure testing costs ($c(n)$).

Interpreting the assumptions about the demand signals of the firms

The firm j does not observe the precise values of quantities sold, i.e. it does not observe the precise value of μ. Instead, it receives an imperfect signal from an individual consumer indexed by i:

$$X_{ijt} \equiv \int_0^t \ln\left(\frac{c_{ijs}}{y(p_j)^{-\epsilon}P^{\epsilon-1}}\right)ds + \sigma_x W_{ijt} = \int_0^t \mu_j ds + \sigma_x W_{ijt},$$

so that

$$dX_{ijt} = \mu_j dt + \sigma_x dW_{ijt},$$

where W_{ijt} is a Wiener process. We now assume that the firm (and the researcher) observe the signals only at discrete time intervals, $t \in 0, 1, 2, 3, 4, \ldots$. We can approximate the signal process as

$$\Delta X_{ijt} \equiv \ln\left(\frac{c_{ijt}}{y(p_j)^{-\epsilon}P^{\epsilon-1}}\right)\Delta t + \sigma_x \Delta W_{ijt},$$

and with $\Delta t = 1$, and denoting $\Delta X_{ijt} \equiv \alpha_{ijt}$, and $\Delta W_{ijt} \equiv \eta_{ijt} \sim N(0,1)$, we get

$$\alpha_{ijt} = \ln c_{ijt} - \ln y + \epsilon p_j - (\epsilon - 1)\ln P + \sigma_x \eta_{ijt}.$$

Now, our assumption will be that the way this happens is that the firm observes not the true quantities ($\ln c_{ijt}$), but noise-contaminated ones:

$$\ln \tilde{c}_{ijt} \equiv \ln c_{ijt} + \sigma_x \eta_{ijt},$$

so that

$$\ln \tilde{c}_{ijt} - \ln y + \epsilon p_j - (\epsilon - 1)\ln P = \mu_j + \sigma_x \eta_{ijt} \equiv \alpha_{ijt},$$

$$\ln \tilde{c}_{ijt} = \ln y - \epsilon p_j + (\epsilon - 1)\ln P + \alpha_{ijt} = \ln c_{ijt} + \alpha_{ijt},$$

where $\alpha_{ijt} \sim N(0,1)$. Therefore, we can interpret the assumptions made about the demand signals as saying that in discrete time the firm observes log quantities sold with some normally distributed error.

Structural Estimation of the Model of Firm Experimentation

Zhanar Akhmetova, Cristina Mitaritonna

November 2010

Abstract

We present the structural estimation of the model of firm experimentation of Akhmetova (2010). We make the following empirical interpretation of the model. We assume that a firm can learn about the demand for its good by observing its exports of individual subcategories of the good. That is, sales of individual 8-digit products within a given 4-digit category (according to the ComTrade classification) can provide information about the demand for that 4-digit good. We estimate the model in this formulation using the French firm level export data, and infer the values of testing costs and sunk cost of entry, as well as the parameters characterizing demand uncertainty, for various 4-digit sectors and destinations. We also estimate alternative models, such as Melitz type model with no learning and a model with passive learning, and compare the results. We carry out a policy experiment, where iceberg trading costs fall as a result of trade liberalization by the export destination, and study the dynamic response of French firms.

1 Introduction

Recently available data suggests that demand uncertainty and learning play an important role in the dynamics of exports. The first pattern found in the literature is the small initial export volumes (in quantities) of new exporters, and their later expansion over time. This is demonstrated in Figure 1 for French exporters, where on the horizontal axis we measure the date since first exports (one for those exporters who just started exporting), and on the vertical axis we measure the (averaged over all exporters) ratio of quantity exported at that date to the firm-average quantity exported. This ratio gives us a sense of how the firm evolves over time. The general upward trend in this ratio reveals an expansion over time in firm exports, as the firm exporting age increases. This can be explained by the uncertainty that these firms face - they are not sure what the demand for their product is in the foreign market, and start out small. As they learn more about their demand, they either expand (if they find that their demand is higher than expected), or exit the market (if they find that their demand is too low). This also implies that the exit rates will tend to fall over time, as the worst exporters are weeded out gradually, and only the best ones remain. We document this pattern in Figure 2. In Figures 1 and 2, we focus on average export activity as measured by the quantities exported of individual 8-digit products, by destination.

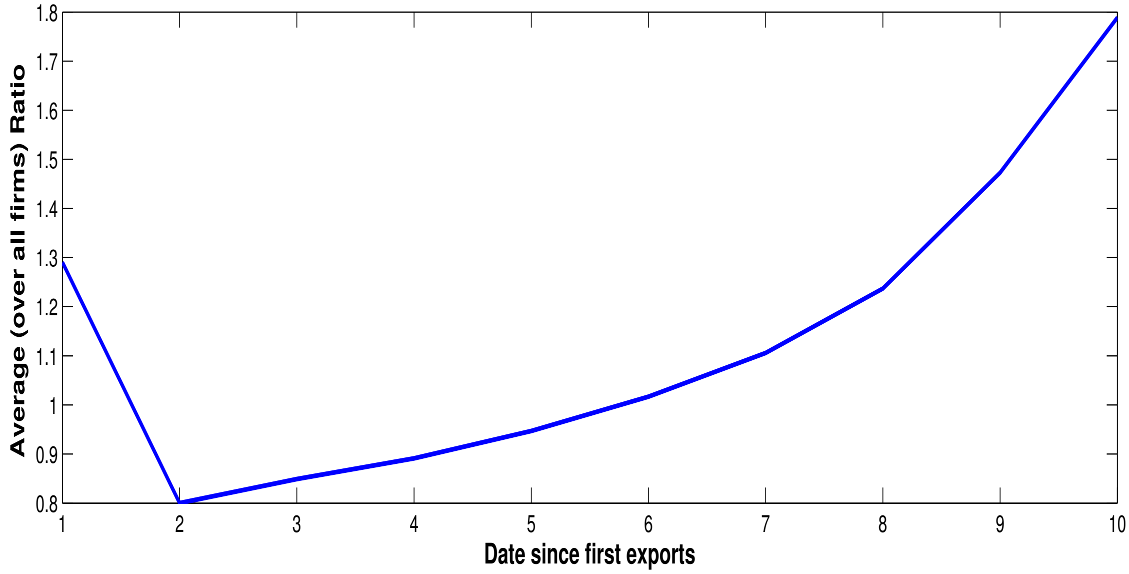

Figure 1: Ratio of quantity (8-digit product) to average quantity of the firm over the period, averaged over all new exporters, consumer non-durable goods

Another dimension along which we can examine these dynamics, is the number of products that firms export. We can examine the number of 8-digit products within a given 4-digit category that firms export. In Figures 3 and 4 we show the expansion in the average (across all new exporters) number of products in a particular sector (Beauty or make-up preparations and preparations for the care of the skin (other than medicaments), including sunscreen or suntan preparations; manicure or pedicure preparations) and destination (Bulgaria). In these figures, the horizontal axis measures the date since first exporting year (1 for the first year when the firm exports), and the vertical axis measures the average number of products exported. In Figure 3, the average is over all new exporters, and in Figure 4, the average runs only over those new exporters that survived for at least 8 years. Note that we define a firm that exports a positive number of products after 8 years (with possible zeros in between) as a firm that survives for at least 8 years. We present Figure 4, to highlight the fact that the expansion in the average number of products is not (purely) due to the selection effect - less productive and smaller firms exiting in the first years, and is also driven by the expansion of surviving firms.

Notice the interesting feature of Figures 3 and 4 - while the average size grows smoothly initially, it jumps to a high level at year 8 (since first exports), and stays there. This tells us something about the learning dynamics of these firms - it looks like they are learning at first, exporting small volumes, and switch to a larger scale later on.

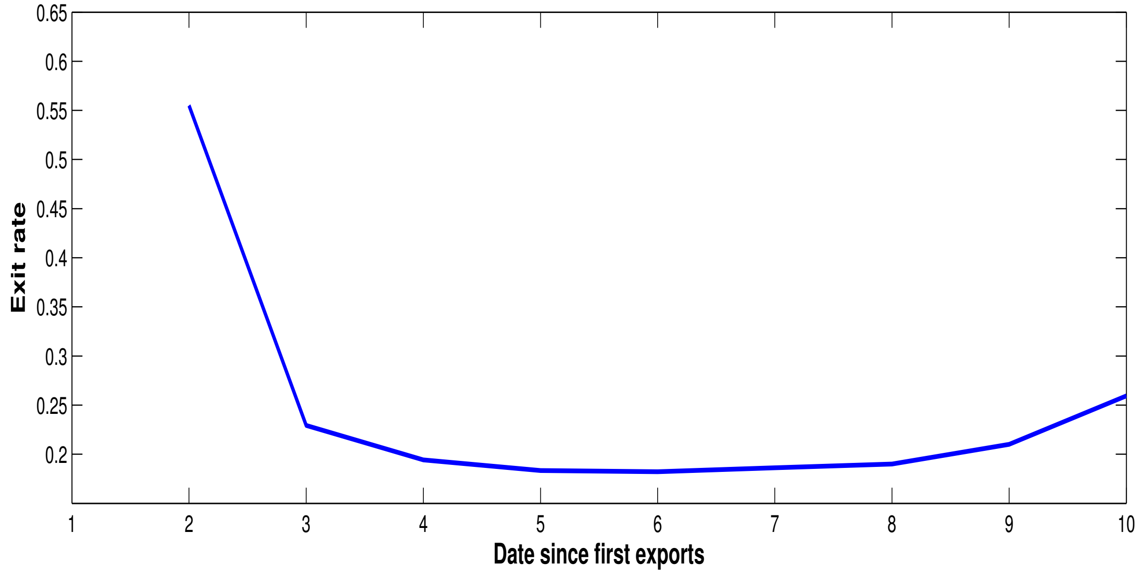

Figure 2: Exit Rate (8-digit product), of all new exporters, consumer non-durable goods

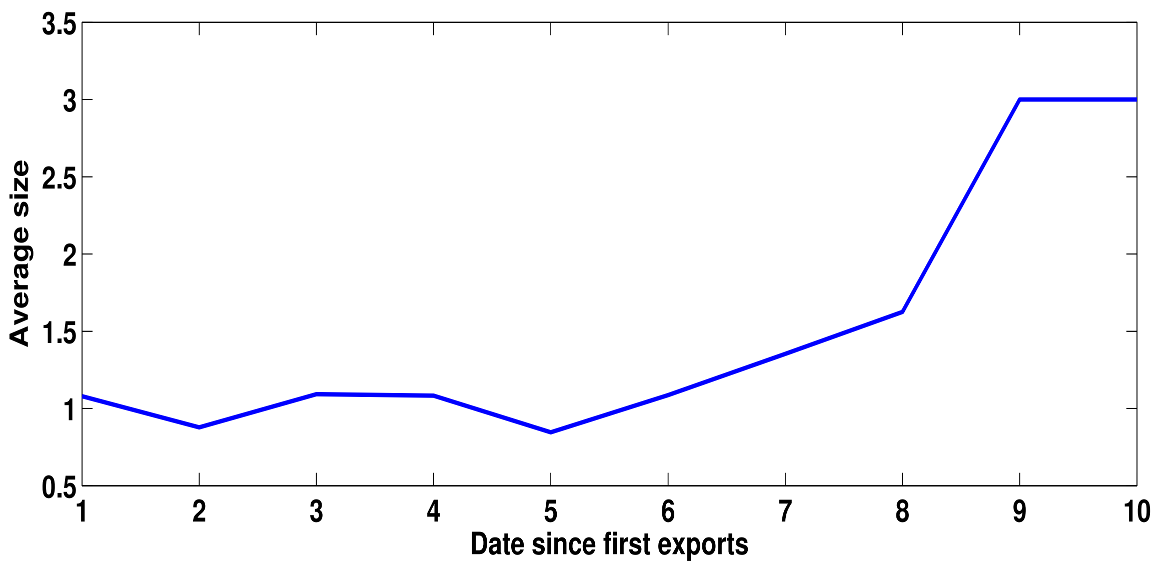

Figure 3: Example of average size evolution, Bulgaria, Beauty or make-up preparations

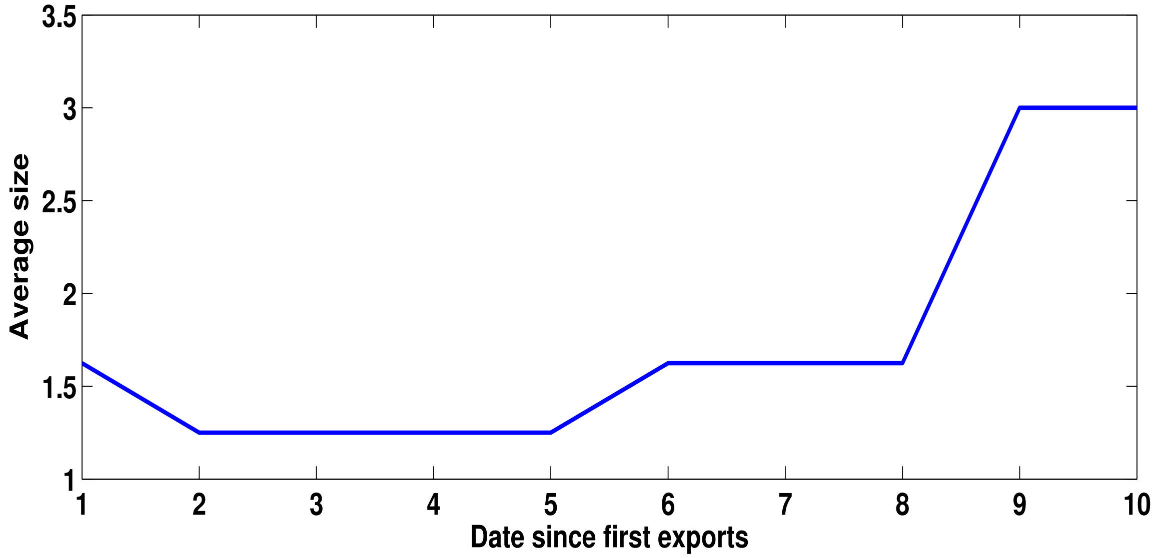

Figure 4: Example of average size evolution,conditional on Survival for at least 8 years, Bulgaria, Beauty or make-up preparations

Akhmetova (2010) presents a model of experimentation, where firms may learn about market demand before paying the sunk entry cost, by sampling a few consumers in the market. We propose the following interpretation of the model. Suppose a firm starts exporting a particular product for the first time in a new destination. In particular, suppose the firm is considering exporting a specific 4-digit code category (according to the ComTrade classification). There are several various 8-digit products within any 4-digit category, and if the firm produces several of these, it has to choose how many of these products to export to the new destination. Initially, it might choose to sell a few products, learn more about the demand for its 4-digit good in this market, and later to expand to the maximum range of products. For example, if a firm produces suitcases, wallets, handbags and backpacks - all in the 4-digit category of 'trunks, bags, cases. etc.', then the firm might learn from its sales of just handbags how popular it is in the new destination. So here the sales of individual 8-digit products play the role of statistical observations, and the number of products can be viewed as the sample size n.

Thus, here we apply the model along a different dimension, then the one taken initially, where individual consumers in a country served as the statistical observations, and the number of consumers accessed - as the sample size. We plan to estimate the model in its original formulation in the near future. However, one difficulty in doing that lies in the lack of data on consumers accessed, and individual consumer demand. Since we observe only data on total quantities sold of 8-digit products, we cannot tell how many consumers (that could be individual cities or regions in a country) are targeted in any period. Hence, we cannot distinguish the two sources of changes in total quantity exported - a change in the number of consumers accessed and a change in the sales to each individual consumer.

The two variables, sample size n and demand signals α are unidentified. We can potentially identify these two using Bayesian estimation techniques, and we plan to do so.

We will proceed as follows. We will describe the data first. We will then lay out the estimation procedure, the details of which are also presented in the Appendix. Next, we will showcase the results of the estimation for several sectors and destinations, and compare these estimates with those of alternative models. We will finally carry out a policy experiment where the export destination reduces its import tariffs, and study the response of French firms.

2 Description of the Data

The data on export flows comes from the database collected by the French Customs. It provides records of French firms' exports (both quantities in tons and export values in euros) aggregated by firm, year, product (identified by an 8-digit code, NC8, which is equivalent to the 6-digit classification of ComTrade in the first 6 digits, with the two last digits appended by the French Customs) and destination country, over the 1995-2005 period. To obtain information on the characteristics of firms, we use a second data source, the French Annual Business Surveys (Enqutes Annuelles d'Entreprises) for the manufacturing sector over the same period, provided by the French ministry of Industry. These contain information on firms that have more than 20 employees, and the variables listed are the address, the identification number of the firm (siren), sales, production, number of employees, and wages. Therefore, we are able to obtain measures of productivity for the firms with employment above 20, even though these measures are not product-specific (and so a single measure is assumed to apply to all the products of a multi-product firm).

There are some measurement issues with the exports data. Within the EU, French customs collect information on the product (NC8 categories) exported by firms when the annual cumulated value of all shipments of a firm (in the previous year) is above 100,000 euros from 2001 onwards. This threshold was 99,100 euros in 2000 and 38,100 euros before 2000. For non-EU exports, all shipments above 1,000 euros are reported. Since the reporting procedure is much clearer, more stable over time and less restrictive for the non-EU exports, we restrict our attention to the exports to non-EU countries in the analysis that follows. Thus, we count the number of non-EU destinations (EU here defined as the 15 first members and Switzerland), as the number of destinations in the world market that the firm accesses. This seems to be justified, since the EU destinations represent a different market for France, due to the free trade area, and some similarities in tastes and preferences, so that French firms may not need to learn or may need to learn very little about demand in these destinations. The European countries that became EU members later are much more different from France in terms of preferences, and they became members mostly in 2004, which is close to the end of our sample period. Hence, we can safely include them in our analysis.

We consider only consumer non-durable products, and calculate the number of 8-digit products that a firm exports within a given 4-digit category to a given destination. We define a class as a combination of destination country and 4-digit category. The estimation is then carried out for each class individually. This allows us to estimate separately the features of export behavior and demand in each 4-digit category and destination. We focus on new exporters, which are defined as firms that start exporting to the given class in or after 1996, since the sample we have runs over 1995-2005. That is, all firms that export in 1995 are considered 'old exporters'.

3 Estimation Procedure

The estimation of the model relies on fitting the time series of export size (number of products), denoted by n, over all new exporters over 1996-2005, and the time series of firm beliefs, denoted by p. Given the profitability of a firm, the cost parameters, and the parameters of demand - $\bar{\mu}, \underline{\mu}, \sigma_x$, the model predicts the optimal decision rule (experimentation versus entry or exit), and the optimal export size ($n(p)$ in the experimentation phase, and full-scale export size M in the post-entry stage). We can therefore predict the time series of n for each firm and year, as well as their optimal exit decisions, and compare this with the times series of n observed, and the observed exit behavior.

Since there are a lot of unknowns in the model, we employ Bayesian techniques, which are very suitable for such situations. We show below how we estimate $\bar{\mu}, \underline{\mu}, \sigma_x$ and beliefs p, based on the time series of quantities of individual 8-digit products (recall that 8-digit products serve as individual statistical observations here). We will argue that relying only on this piece of data is good enough for updating the posterior of these variables, that is, using the information on the number of products n as well will not greatly modify their posteriors.

Hence, we can treat the estimation of the other parameters, namely, M, testing cost function $c(n)$ and sunk entry cost F independently. Given the estimates of $\bar{\mu}, \underline{\mu}, \sigma_x$, and beliefs p, we look for cost parameters that provide the best fit between the optimal n^* and the observed n, as well as the observed exit of new exporters and that predicted by the model - recall that in the model, an experimenter will quit whenever its beliefs fall below a threshold $\underline{p}$. For that, we need a measure of firm profitability. We employ a measure of profitability which relies directly on the estimated productivity of firms (TFP), which we obtain from the domestic balance sheet data of French firms.

We assume that the optimal scale (M) is the same for all firms. Of course, it can be argued that the optimal scale is different for different firms - after all, an entire literature is devoted to the study of the decision of the firm with respect to the number of products to produce. Considerations like economies/diseconomies of scope, i.e. decreasing/increasing marginal costs of producing additional products, and product cannibalization (driving away demand from older products by newly introduced ones), are among the ones discussed in this case. However, we abstract away from this issue and assume a single optimal scale for all firms in a given class for now. The model was originally developed for studying a general case of learning by a firm, by sampling over a few observations at first, and later expanding to some optimal scale, which is fixed and common across all firms in the market. The application to the number of products exported by the firm as the sample of observations therefore should be viewed exactly as such - as an application of the experimentation model, rather than as a model explaining all possible nuances of the behavior of multi-product firms. It is possible, however, that we will return to this issue, and enrich the model in this direction in the future.

Estimating the preference shocks

An important part of the empirical procedure is estimating the preference shocks α_{jt}^k for each destination k, firm j and time t. This can be done as follows: from the CES formulation, the quantity observed in a destination k at time t of an 8-digit product k (within the given 4-digit category) produced by firm j is

$$c_{jt}^k = e^{\alpha_{jt}^k} y(p_j)^{-\epsilon} P^{\epsilon-1},$$

In the data, the variables y, p_j, and P variables might differ across different 8-digit products k within the same 4-digit category of the same firm j. Therefore, we re-write:

$$c_{jt}^k = e^{\alpha_{jt}^k} y_t^k {p_{jt}^k}^{-\epsilon} {P_t^k}^{\epsilon-1},$$

so that in logs

$$\ln c_{jt}^k = \ln y_t^k - \epsilon \ln p_{jt}^k + (\epsilon - 1) \ln P_t^k + \alpha_{jt}^k.$$

$j = 1, ..., J, t = 1, ..., T, k = 1, ..., K(j)$, where $K(j)$ is the number of 8-digit products exported by firm j within the given class. By assumption, the α_{jt}^k in the above equation do not have a zero mean, and, moreover, have a mean that varies from one firm to another (each firm has μ_j that is either $\bar{\mu}$ or $\underline{\mu}$). Therefore, what we can do first is run the above regression with firm-fixed effects (to account for the firm-specific means in residuals), and obtain estimates of ϵ. As a proxy for prices we use the unit values (export values in euros divided by the export quantities). Since the unit values may be correlated with the demand shocks, we instrument for this using firm-productivities. We calculated TFP a la Ackerberg, Caves and Frazer, to address the issue of endogeneity of production inputs in the production function and avoid the collinearity issue arising in Olley-Pakes or Levinsohn-Petrin. For more details on the estimation of TFP, please, see the Appendix.

Thus, the regression used to estimate the elasticities is:

$$\ln c_{jt}^k = -\epsilon \ln uv_{jt}^k + \sum_1^{KT} g_t^k D_t^k + \sum_1^J \beta^j D_j + e_{jt}^k,$$

where uv_{jt}^k are the unit values and are instrumented for with firm's TFP, and D_t^k are product-time fixed effects. The product-time fixed effects are meant to capture the variation in the aggregate variables y and P across k and time t. The D_j are the firm-fixed effects. We assume that productivity ϕ is the same across all products of the firm: $\phi_{jt}^k \equiv \phi_{jt}$ for all products k.

We summarize the estimates of ϵ over all classes that we obtained as a result of these regressions in the Appendix. We must note that for some of these classes, we obtained values of ϵ that are smaller than 1. We then do not consider these particular classes in further analysis, since the assumption that $\epsilon > 1$ is crucial in the model. We have 192 classes that we can work with, as a result.

Once we obtain the estimates of ϵ, we calculate the following for each firm j, product k and time t:

$$v_{jt}^k \equiv \ln c_{jt}^k + \epsilon \ln uv_{jt}^k,$$

which by assumption contains the aggregate factors, common to all firms, and the firm-product specific demand shocks α_{jt}^k. Now, we would like to filter out the aggregate factors, common to all firms and products, and obtain demand signals α_{jt}^k. Since α_{jt}^k have firm-specific means, we cannot treat these as a usual error term, and just run a regression of v_{jt}^k on time-fixed effects to control for aggregate demand. Moreover, since there are a lot of unknowns involved - we do not know the values of $\bar{\mu}$, $\underline{\mu}$, σ_x, p_0 (the objective distribution of μ in the population), and what μ each firm has drawn, we apply Gibbs sampling. We assume that there are two (unobserved) aggregate factors, each of which follows an AR(1) process, and employ Gibbs sampling (with Kalman filtering) to deduce these aggregate factors, the demand shocks and $\bar{\mu}$, $\underline{\mu}$ and σ_x.

We have

$$\alpha_{jt}^k = \mu_j + \sigma_x \eta_{jt}^k, \eta_{jt}^k \sim N(0,1) iid,$$

$\mu_j = \bar{\mu}$ with probability p_0, and $\underline{\mu}$ with probability $1 - p_0$,

the transition equations for the aggregate factors, denoted by A_1 and A_2,

$$A_{t1} = \rho_1 A_{(t-1)1} + CA_1 + \nu_{t1},$$

$$A_{t2} = \rho_2 A_{(t-1)2} + CA_2 + \nu_{t2},$$

$\rho_1 \in (-1, 1)$, $\rho_2 \in (-1, 1)$, $CA_1 \in R, CA_2 \in R$ are constants, and ν_{t1} and ν_{t2} are normal error terms with means zero and covariance matrix V. The observation equation is

$$v_{jt}^k = D_1 A_{t1} + D_2 A_{t2} + \alpha_{jt}^k.$$

It can be seen from the equations above that the unobserved α_{jt}^k and A_{t1}, A_{t2} can be identified up to a constant, and therefore we normalize $\underline{\mu} = 0$. Denote $\Theta_1 \equiv \bar{\mu}, \sigma_x, p_0$, the main parameters characterizing demand uncertainty. The details of the estimation of Θ_1 are reported in the Appendix. We summarize the posterior distributions $\bar{\mu}$, σ_x and p_0, as well as signal-to-noise ratio χ in the tables and graphs in the Appendix, as well.

Once we have generated draws of the values of $\bar{\mu}, \sigma_x, p_0$, as well as the aggregate factors A_{t1}, A_{t2}, and demand shocks α_{jt}^k, from their respective posteriors, we can calculate the beliefs of any given firm j at any time t, using the discrete time Bayes updating rule:

$$Prob_{jt}(\mu_j = \bar{\mu}) \equiv P_{jt} = \frac{fn(\frac{\bar{\alpha}_{jt}-\bar{\mu}}{\frac{\sigma_x}{\sqrt{n}_{jt}}})P_{j(t-1)}}{fn(\frac{\bar{\alpha}_{jt}-\bar{\mu}}{\frac{\sigma_x}{\sqrt{n}_{jt}}})P_{j(t-1)} + fn(\frac{\bar{\alpha}_{jt}-\underline{\mu}}{\frac{\sigma_x}{\sqrt{n}_{jt}}})(1-P_{j(t-1)})},$$

where fn is the standard normal density, and n_{jt} is the sample size at time t, which in this application is the number of 8-digit products within a given class that the firm exports.

Measure of firm heterogeneity (and profitability)

We treat all the relevant variables (aggregate factors and firm productivity) as stationary, with constant variance, and assume that the firm uses their expected values to make the optimal export/experimentation decision. Hence, we need to measure the expected value of the profits of a firm with productivity ϕ_j to predict its behavior.

The profits from an 8-digit k product of a firm j are given by

$$\begin{aligned}
\pi^k_{jt} &= \frac{1}{\epsilon-1}\frac{w_t}{\phi_{jt}}e^{\mu_j}y^k_t(p^k_{jt})^{-\epsilon}[P^k_t]^{\epsilon-1} \\
&= e^{\mu_j}y^k_t\frac{1}{\epsilon-1}[\frac{\epsilon}{\epsilon-1}]^{-\epsilon}[\frac{\phi_{jt}}{w_t}]^{\epsilon-1}[P^k_t]^{\epsilon-1},
\end{aligned}$$

and in logs,

$$\begin{aligned}
\ln\pi^k_{jt} &= \ln\frac{1}{\epsilon-1} - \epsilon\frac{\epsilon}{\epsilon-1} + \ln y^k_t - (\epsilon-1)\ln w_t + (\epsilon-1)\ln\phi_{jt} + (\epsilon-1)\ln P^k_t + \mu_j \\
&\equiv C + (\epsilon-1)\ln\phi_{jt} + A^k_t + \mu_j,
\end{aligned}$$

where C summarizes the constant terms, and A_t summarizes the aggregate terms (y, w, P).

We assume that $\ln\phi_{jt}$ follows an AR(1) process for every firm j:

$$\ln\phi_{jt} = C^j_\phi + \rho_\phi\ln\phi_{j(t-1)} + e_{jt},$$

where $\rho_\phi \in (-1,1)$, $e_{jt} \sim N(0,\sigma_\phi)$, and C^j_ϕ is a firm fixed effect. This is consistent with the assumption that log-TFP follows a first-order Markov process made when estimating TFP (in the Appendix). Then $\ln\pi_j$ is normally distributed, and given

$$\begin{aligned}
E(\ln\pi_{jt}) &= E[C + (\epsilon-1)\ln\phi_{jt} + A^k_t + \mu_j] \\
&= C + (\epsilon-1)E(\ln\phi_{jt}) + E(A^k_t) + \mu_j \\
&= C + (\epsilon-1)\frac{C^j_\phi}{(1-\rho_\phi)} + E(A^k_t) + \mu_j,
\end{aligned}$$

and

$$\begin{aligned} Var(\ln \pi_{jt}) &= (\epsilon-1)^2 Var(\ln \phi_{jt}) + Var(A_t^k) \\ &= (\epsilon-1)^2 \frac{\sigma_\phi^2}{(1-(\rho_\phi)^2)} + \sigma_A^2, \end{aligned}$$

we obtain

$$\begin{aligned} E(\pi_j) &= exp[E(\ln \pi_j) + \frac{Var(\ln \pi_j)}{2}] \\ &= exp[C + (\epsilon-1)\frac{C_\phi^j}{(1-\rho_\phi)} + E(A_t^k) + \mu_j + \frac{(\epsilon-1)^2 \frac{\sigma_\phi^2}{(1-(\rho_\phi)^2)} + \sigma_A^2}{2}] \\ &= e^{\mu_j} exp[\frac{C_\phi^j}{(1-\rho_\phi)}] exp[(\epsilon-1)^2 \frac{\sigma_\phi^2}{2(1-(\rho_\phi)^2)}] exp[\frac{\sigma_A^2}{2}] \bar{A}, \end{aligned}$$

where $\bar{A}$ is subsumes $E(A_t^k)$ and the constants C and $(\epsilon - 1)$. We take the time series of ϕ for all firms, evaluate $\rho_\phi, \sigma_\phi, C_\phi^j$, and calculate the corresponding $E(\phi_{jt})$ for each firm j. To measure heterogeneity between firms, it suffices to focus on $\frac{C_\phi^j}{(1-\rho_\phi)}$.

We show in the Appendix that normalizing the profits by any constant results in the estimates of cost variables that are scaled by the same constant. Therefore, we can scale the profits π_j^k:

$$\hat{\pi}_j \equiv \frac{E(\pi_j)}{exp[(\epsilon-1)^2 \frac{\sigma_\phi^2}{2(1-(\rho_\phi)^2)}] exp[\frac{\sigma_A}{2}] md\bar{A}} = e^{\mu_j} \frac{exp[\frac{C_\phi^j}{(1-\rho_\phi)}]}{md} \equiv e^{\mu_j} \tilde{\pi}_j,$$

where md is the median of $exp[\frac{C_\phi^j}{(1-\rho_\phi)}]$ over all exporters in the given class, and $\tilde{\pi}_j \equiv \frac{exp[\frac{C_\phi^j}{(1-\rho_\phi)}]}{md}$ is the measure of heterogeneity we will focus on in the estimation.

Estimating the cost parameters

Given the parameters of the model, one can solve the second-order free boundary value problem of the firm outlined in the theoretical part, where we now treat the equation for the value function as a discrete time equation. That is, we input annual values for all variables in the equation, and solve for the (annual) values of the value function, and corresponding annual sample size. To remind the reader, that BVP was

$$v''(p) = \frac{c'(z(rv(p))) - AG\phi_j^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]}{\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2},$$

where $n(p) = z(rv(p))$, $z \equiv g^{-1}$, $g(n) = nc'(n) - c(n)$,
plus the value matching condition:

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F, v(\underline{p}) = 0.$$

and smooth pasting condition:

$$v'(\bar{p}) = \tilde{V}'(p), v'(\underline{p}) = 0.$$

$\tilde{V}(p)$ is the value function in the second stage (post-entry),

$$\tilde{V}(p) = \frac{M}{r}(-f + AG(\phi_j)^{\epsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]),$$

and it is simply lifetime discounted profits from selling to the entire market (of size M) using the linear exporting technology.

There are three states that a new exporter can be in: experimentation (pre-entry), full-scale export (post-entry), and non-exporting (post-exit). In the model, once the firm exits the market, it does not re-enter, since we assume a stationary equilibrium with static aggregate variables and firm productivity. In the data we also assume that when a firm exports no products in the class under consideration after some point up to 2005, it is in the non-exporting state. However, when we observe a firm exporting no products in one year, and at least one product within the given class later on, this is considered as just a temporary zero in the time series of the number of products of this firm. That is, we simply set $n_{jt} = 0$ for that year for that firm. We denote the states as follows: experimentation phase as $S_t = 0$, full-scale export phase as $S_t = 1$ and non-exporting phase as $S_t = 2$.

Relying on the BVP, given the profitability of a firm, as measured by $AG(\phi_j)^{\epsilon-1}$, or by $\tilde{\pi}_j$ (explained above), the cost parameters, and the beliefs of the firm, one can calculate the optimal testing size, and thresholds $\bar{p}$ and $\underline{p}$ and states for this firm. Thus, we can predict which firms are experimenters - these are firms that have profitability below the cutoff value for entering the market right away, or equivalently, as outlined in the model (Akhmetova (2010)),

$$\frac{M}{r}\tilde{\pi}_j e^{\underline{\mu}} < F,$$

and

$$\bar{p}(\tilde{\pi}_j) > p_0.$$

We also require that the export participation condition holds:

$$\underline{p}(\tilde{\pi}_j) <= p_0.$$

Given that a firm is an experimenter, we can predict the states it is in at every t:

$$P_{jt} \in (\bar{p}, \underline{p}) \Longrightarrow S_{jt} = 0,$$

$$n_{jt} = z(rv(P_{jt})|\tilde{\pi}_j),$$

$$P_{jt} \geq \bar{p} \Longrightarrow S_{jt} = 1,$$

$$n_{jt} = M,$$

$$P_{jt} \leq \underline{p} \Longrightarrow S_{jt} = 2,$$

$$n_{jt} = 0.$$

Thus, the model gives us predictions with respect to the optimal export size of firms. Note that the full scale M enters not only through setting the size of exporters in the second stage (post-entry), but also by affecting the solution to the BVP - it shows up in the value matching and smooth pasting conditions. Therefore, we cannot estimate M by simply averaging out the export sizes of all firms in the second stage. We need to evaluate the likelihoods instead.

The cost parameters that we are interested in are the sunk entry cost F and the coefficients of the testing cost function $c(n)$:

$$c(n) = \gamma_0 + \gamma_1 n + \gamma_2 n^2 + \gamma_3 n^3 + \gamma_4 n^4 + \gamma_5 n^5.$$

From now on, denote the set of cost parameters $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ by Γ. Additionally, the sunk entry cost F, the exogenous death rate δ (which we assume in the empirical part, to accommodate exit of exporters for reasons other than a drop in beliefs), the discount rate r, which we assume equals δ plus the bank interest rate, the fixed exporting cost in the second stage, f, and M, the full-scale export size (the number of products exported in the full-scale phase), are also unknown. Together, Γ, and F,f, r,δ and M form the set Θ_2.

Recall that in the theoretical part, we assumed a linear exporting technology for firms in the second stage (post-entry), more precisely, $c^2(n) = fn$. As long as the profits from selling a single 8-digit product in the presence of low demand ($\mu_j = \underline{\mu}$) are higher than f for the lowest productivity exporter, every firm will export M products in the second stage, and will exit only if hit by an exogenous death shock. Under this assumption, F and f are not separately identifiable: for any F and f, the same kind of behavior will be produced by $F' \equiv F + \frac{f}{\delta}$ and $f' = 0$. That is, paying f in every period, with only δ as a probability of exit is equivalent to paying $\frac{f}{\delta}$ upfront upon entry into second stage. Therefore, we normalize $f = 0$ and focus on estimating F.

Above we have shown how to generate draws of $\Theta_1 \equiv \bar{\mu}, \sigma_x, p_0$ conditional on observed quantities. However, the observed behavior of firms in terms of export size (number of products exported) also provides information for the posterior of Θ_1. Therefore, one should incorporate this into updating the posterior of Θ_1. We did consider this effect, and found that the posterior of Θ_1, conditional on just quantities, are tight enough that the information about firm behavior does not change these much. In other words, picking different draws from the posterior of Θ_1, conditional on just quantities, does not result in very different time series for the beliefs of the firms. Therefore carrying out, say, Metropolis-Hastings sampling to update the posterior of Θ_1 further, conditional on the observed behavior of the firms (in response to beliefs), will only affect the first-stage posteriors marginally. Hence, we set the median values of the posterior distributions of Θ_1 as the values of $\bar{\mu}, \sigma_x, p_0$, calculate the beliefs of all firms, and proceed to estimate the cost parameters.

There are some additional parameters, that we introduce to accommodate the data better. The model predicts the optimal experimentation sample size (number of products) for each experimenting firm, given its profitability in the market, and the testing costs and entry cost F. We still need to allow for some error in the observed number of products sold, since it is impossible to match perfectly the predicted n_{jt}^* and the observed n_{jt} for all firms and periods t:

$$\ln n_{jt} = \ln n_{jt}^* + e_{jt},$$

where e_{jt} is a mean zero normal error term, with variance σ_N^2, that we will also estimate. In the experimentation phase, $n_{jt}^* = n(p_{jt})$, and in the full-scale export phase, $n_{jt}^* = M$, the full scale. We view the error term as the discrepancy caused by managerial error, delays and disruptions in international delivery, considerations affecting the choice of the number of products to export that are outside the model, etc. Of course, the goal of estimation is to minimize this discrepancy.

Additionally, the model predicts that in the experimentation stage, the firm quits as soon as its beliefs p_{jt} fall below some level $\underline{p}_j$. In the data, we cannot match these perfectly, and therefore introduce $P_q \equiv Prob(quit|p_{jt} \leq \underline{p}_j)$, which is common to all firms. Thus, the probability that a firm will not quit, given its beliefs fall below the quitting threshold, is $1 - P_q$. We restrict P_q to be between 0.7 and 1.

We carry out Metropolis-Hastings and Gibbs sampling for the second level parameters, $\Theta_2 \equiv \Gamma, F, r, \delta, M$, given values for parameters Θ_1, the time series for the beliefs of the firms, P_{jt}, number of products n_{jt}, observed profitability measures $\tilde{\pi}_j$, and observed exit by firms. In the Appendix, we explain the procedure.

4 Results

In the Appendix, we show the histograms for estimated elasticities ϵ, and demand uncertainty parameters $\bar{\mu}, \sigma_x, \chi \equiv \frac{\bar{\mu}}{\sigma_x}$ (recall that we normalized $\underline{\mu} = 0$). Here we show some interesting scatter graphs, where it is clear that ϵ is positively correlated with $\bar{\mu}$ and σ_x, but more so with $\bar{\mu}$ so that it is also positively correlated with the signal-to-noise ratio χ. It looks like the more differentiated sectors (as measured by ϵ) have a correspondingly larger gap between lowest and highest demand ($\bar{\mu} - \underline{\mu}$). It is interesting that these sectors have a higher signal-to-noise ratio at the same time.

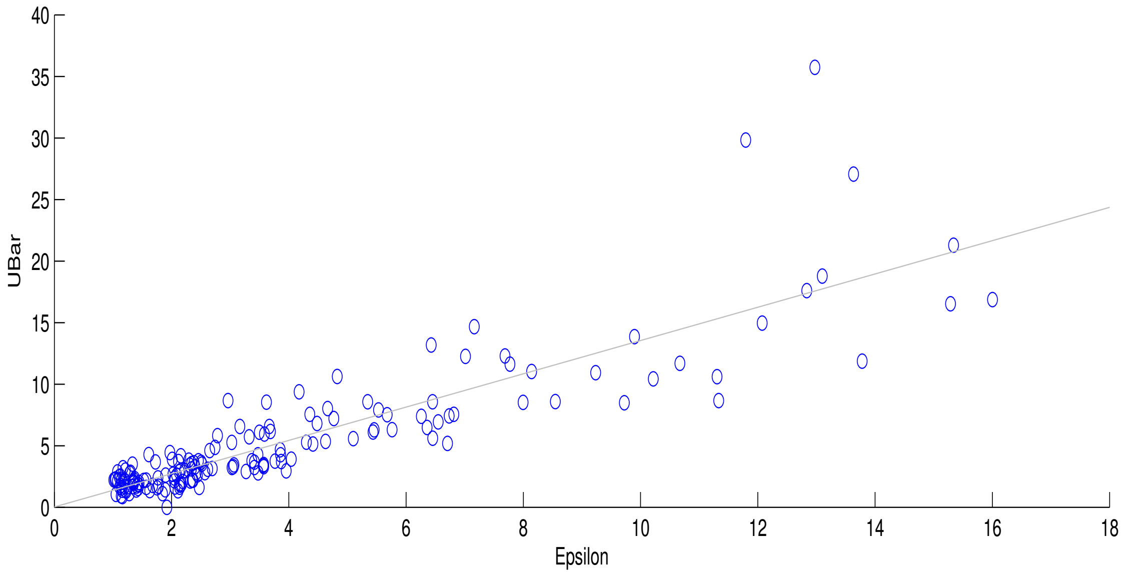

Figure 5: Correlation between ϵ and $\bar{\mu}$

As a general look at the relationships predicted by the model, we run two simple regressions: first, a regression of the number of products exported on estimated beliefs (with firm-class fixed effects), and second, a logistic regression of the exit of firms on beliefs (controlling for firm-class again, so that we in a way follow each firm in every class over time, and check how its beliefs affect its behavior - exit and export size). In Table 1, we show in how many classes we find regression coefficients as expected - a positive coefficient for the number of products, and a negative coefficient for the exit. Out of 192 classes, we find 98 such classes. Next, in Tables 2 and 3, we report the regression coefficients, where we run the regression over all classes that exhibit the relationships as desired.

Table 1: Tabulating the classes according to regression results

	Exit coeff.positive	Exit coeff.negative	Total
Size coeff. neg.	15	48	63
Size coeff. pos.	31	98	129
Total	46	146	192

We present in this paper the results of the structural estimation for two classes: Australia, Womens' or girls' suits, ensembles,jackets, etc. (code 6204), and Estonia, Trunks, suitcases, vanity cases, executive-cases, briefcases, etc. (code 4202). Note that even though Estonia

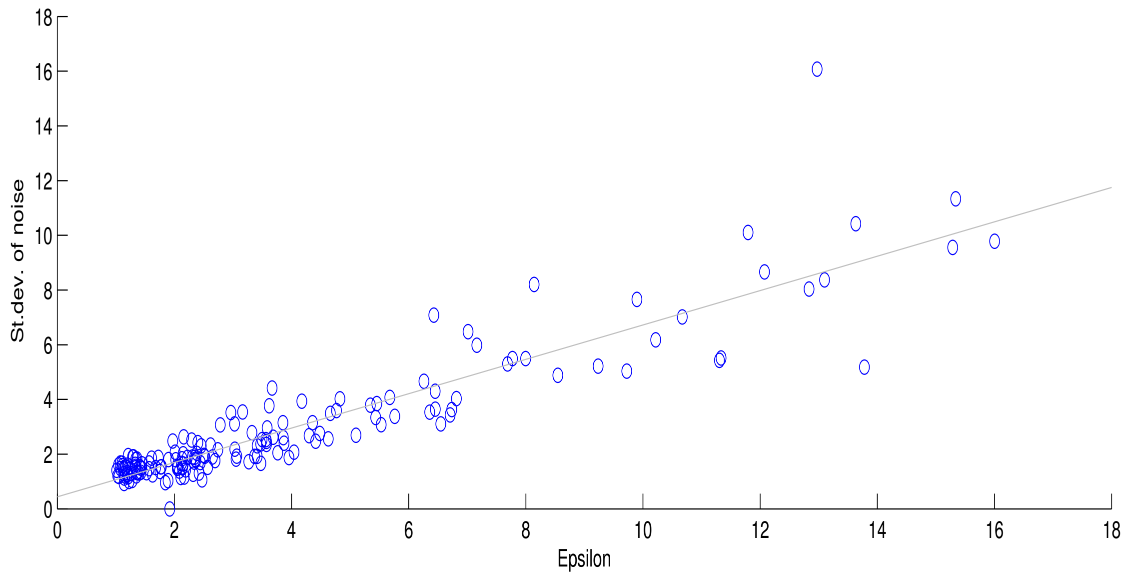

Figure 6: Correlation between ϵ and standard deviation of noise

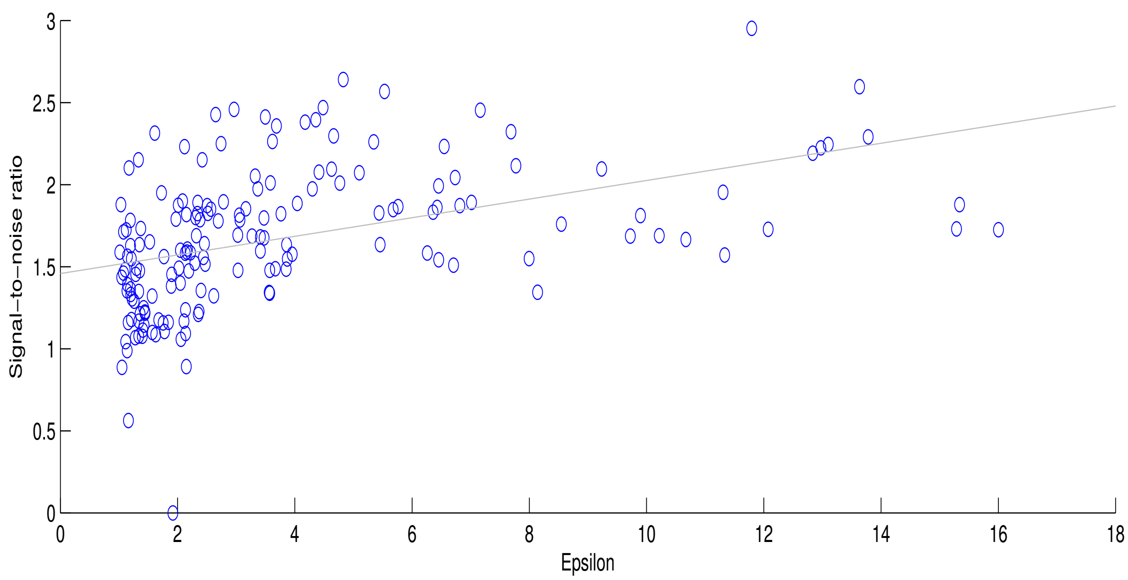

Figure 7: Correlation between ϵ and signal-to-noise ratio

Table 2: Regression of number of products on beliefs

	Number of Products
Beliefs P	0.377
	(12.97)**
Constant	1.46
	(149.44)**
Observations	65713
Number of group(firm prod)	15851
R-squared	0
Absolute value of t statistics in parentheses	
* significant at 5%; ** significant at 1%	

Table 3: Logistic regression of exit on beliefs

	Exit
P	-1.27
	(17.73)**
Observations	44667
Absolute value of z statistics in parentheses	
* significant at 5%; ** significant at 1%	

joined the EU in 2004, it did not follow the trade policy trajectory of other new European members that started liberalizing with respect to the EU around 1994-1996. Estonia was already engaging in free trade with most countries, the EU in particular, in 1995. So we do not have to worry about the potential shift in tariffs in Estonia vis-a-vis France over 1995-2005. We present the histograms of the maximum number of products and the average number of products for new exporters in these classes in the Appendix.

In the table below we summarize the estimates of the main parameters (means of/draws from their posterior distributions).

First of all, one can see that the signal-to-noise ratio in the first class (Australia, women's jackets) is much lower than that in the second class (Estonia, trunks, etc.). This will show up later on when we present the simulations of the evolution over time of new exporters, given these estimates. Next, the estimated costs are quite high - the testing costs for the two classes are depicted in Figures 8 and 9. We present the costs as estimated (recall that the units of measurement are median profits, as explained in the section on the measures of profitability). We also scale these down by the demand portion that does not appear in our measure of profitability, but still should be taken into account - since total profits of a firm will be given by the product of our measure $\tilde{\pi}$ and $exp(\bar{\mu})$. Therefore, we also present in the same graphs the cost functions scaled down by $exp(\bar{\mu})$. This should give us a better picture.

Similarly, in Table 4, we show the estimated sunk cost estimates, as well as the scaled down estimates ($\frac{F}{p_0 * exp(\bar{\mu}) + (1-p_0) * exp(\underline{\mu})}$). Even though the raw estimates F are very different for the two classes, when scaled by the demand shift parameter, they are very similar - around

Table 4: Estimates (means/medians of the posteriors for all parameters, other than the cost parameters Γ (these are draws from the posterior)

Class	Australia, Women's jackets	Estonia, Trunks, etc.
$P0$	0.3304	0.213
$\bar{\mu}$	1.4248	2.8927
σ_x	1.6345	1.8289
χ	0.8717	1.5817
M	3	4
F	24.5	50.2
$\frac{F}{p0*exp(\bar{\mu})+(1-p0)*exp(\underline{\mu})}$	11.5	10.8
δ	0.2	0.1
r	0.2	0.2
P_q	0.7	0.7
γ_0	0.0874	0.2169
γ_1	2.1626	9.6126
γ_2	-1.2028	-3.4783
γ_3	0.8	1.5
γ_4	0.001	0.1
γ_5	0.001	0.01
σ_N	2.01	1.63

10-11. Thus, we estimate the sunk cost of entry F at around 10-11 times the expected (at initial beliefs) median profit in each class.

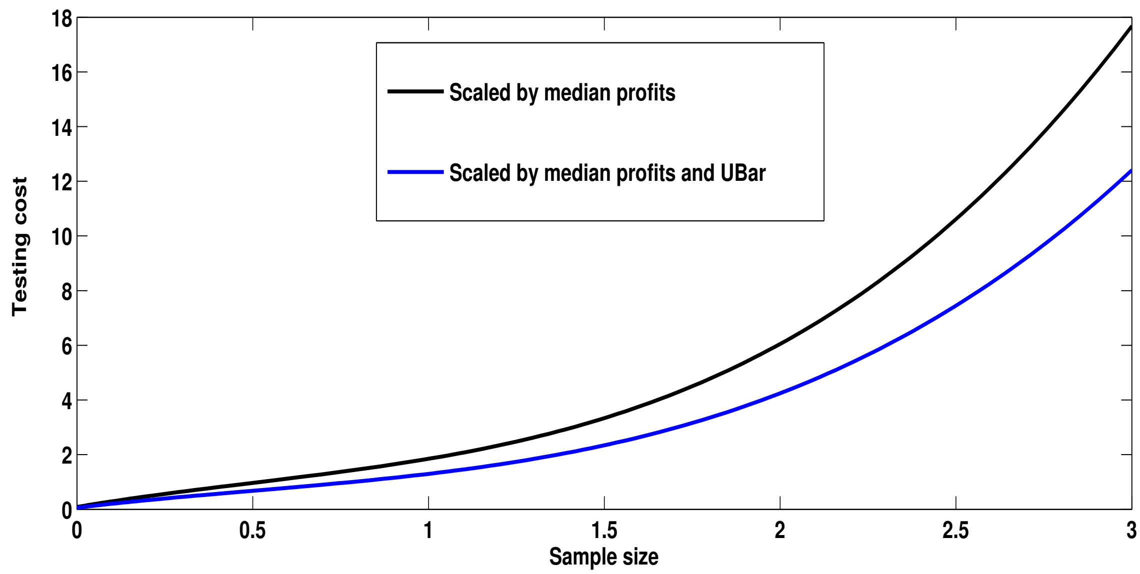

Figure 8: Testing cost schedule, Australia, Women's jackets, etc.

Given the parameters, we estimate that the first 6 deciles of profitability $\tilde{\pi}$ in each class experiment before investing in the sunk entry cost F. We show in Figures 10 and 11 the optimal testing schedules for each range - the lowest range has the lowest schedule, and has the highest threshold for quitting (which is exactly equal to p_0), and the highest threshold on beliefs for entry into the market.

We started the paper with the motivating evidence of exporters expanding over time in export size, and their exit rates falling as they age. We would like to replicate these patterns using our estimates - simulate the dynamics of new exporters. We start with 100 new exporters, and assign each one randomly into a profitability range, as well as demand range - high or low (with probability of high demand given by the estimated p_0). We then draw demand signals for each firm according to the estimated distribution of demand signals. We update the beliefs of each exporter, and calculate their optimal response - whether they experiment or enter or quit, and if they experiment, their optimal sample size $n(p)$. We show the evolution of exit rates, and of average number of products (averaged over exporters). We do not allow for any errors in these simulations (such as deviations from the optimal size, or not quitting when beliefs P are below the threshold for quitting).

We pointed out earlier that the signal-to-noise ratio for Australia was much lower than that for Estonia. This shows up in these simulations: the convergence rate for the first class (Australia, women's jackets, etc.) is much lower, as their average size goes to 3 (the optimal size) only by date 13 since the beginning of exporting. For the same reason, the exit rate peaks very late in the period - at date 7 since first exports, since it takes that long for the

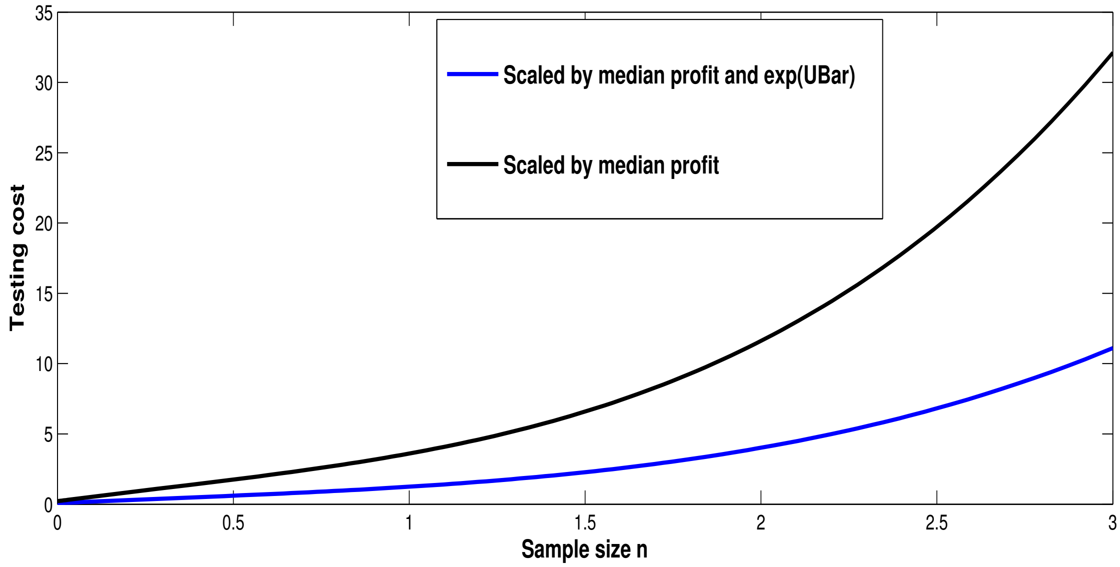

Figure 9: Testing cost schedule, Estonia, Trunks, cases, etc

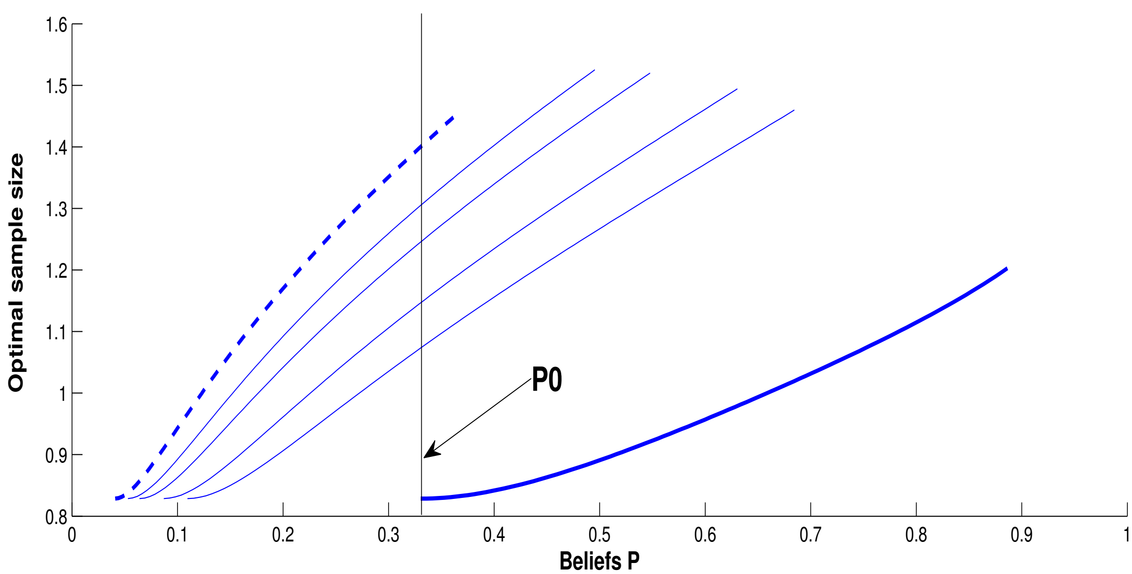

Figure 10: Optimal sample size schedules for experimenter ranges of productivity, from lowest (bold) to highest (dashed), Australia, Women's jackets, etc.

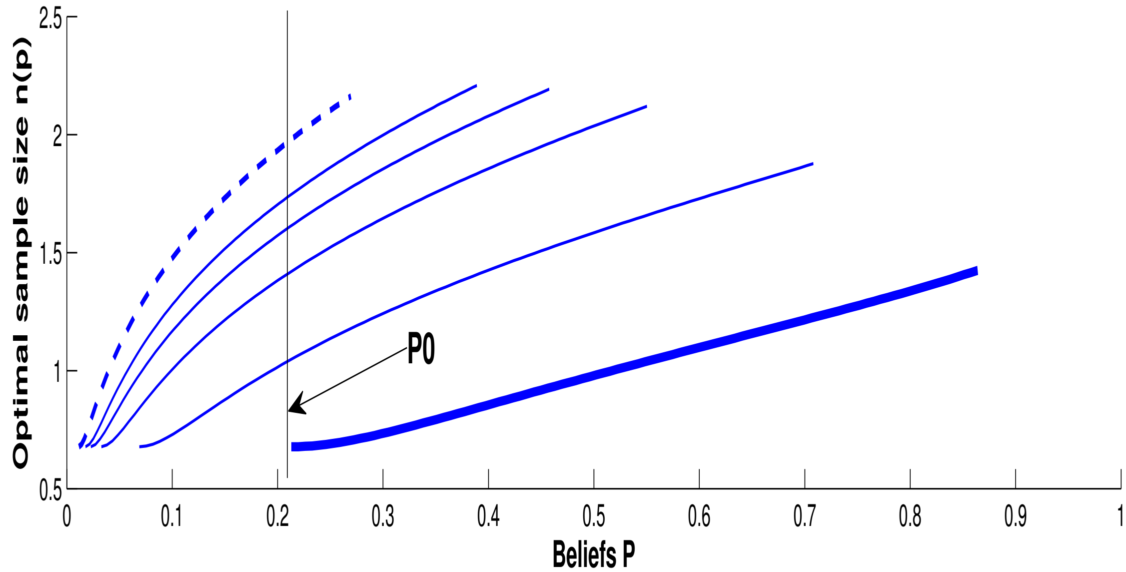

Figure 11: Optimal sample size schedules for experimenter ranges of productivity, from lowest (bold) to highest (dashed), Estonia, Trunks, cases, etc.

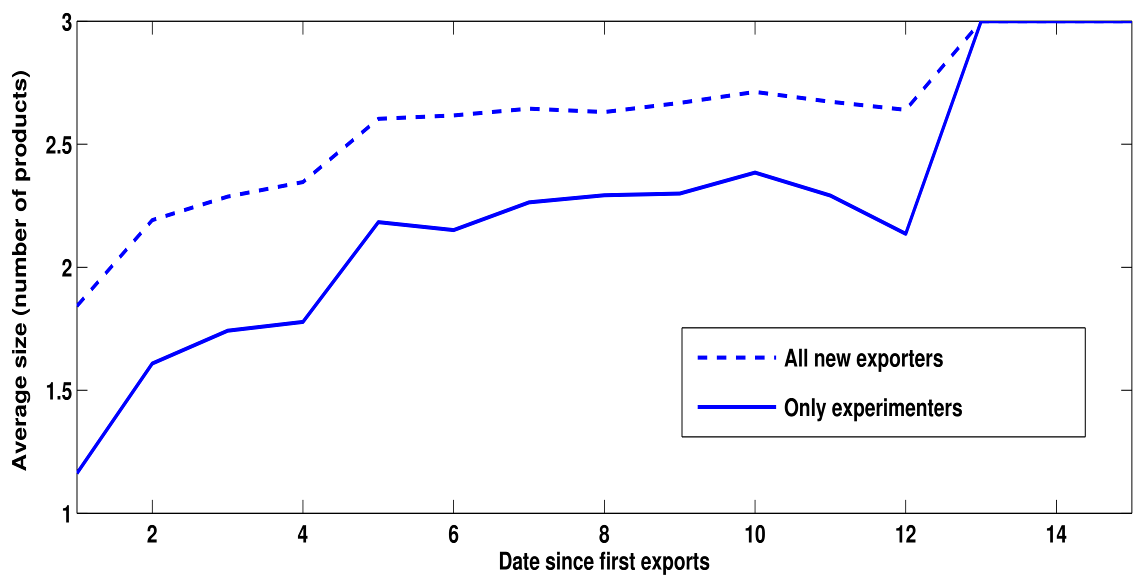

Figure 12: Simulated average size, over 100 exporters, Australia, Women's jackets, etc.

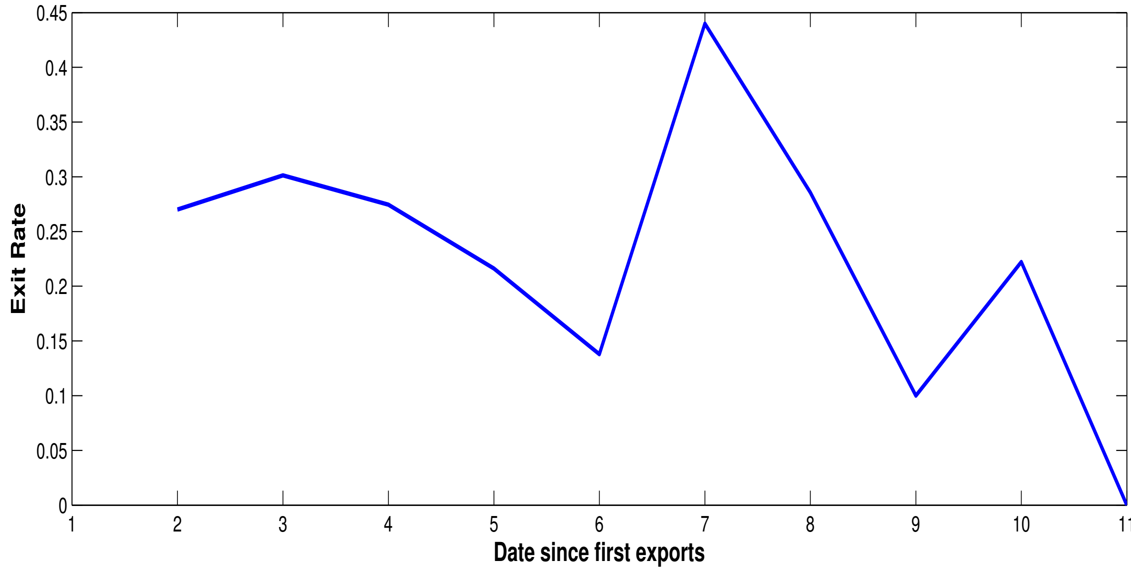

Figure 13: Simulated exit rate, over 100 exporters, Australia, Women's jackets, etc.

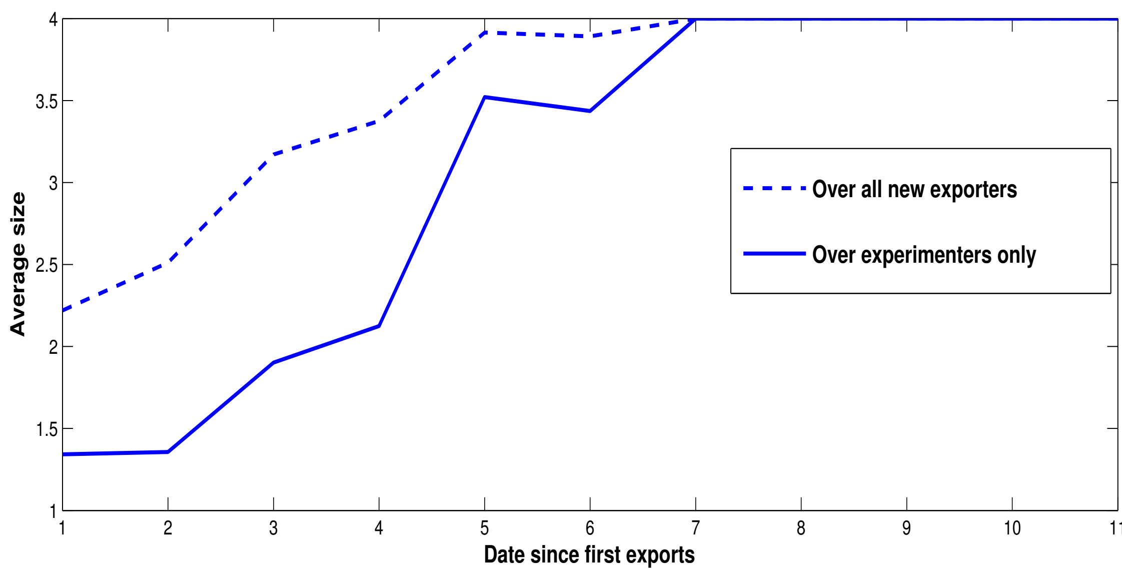

Figure 14: Simulated average size, over 100 exporters, for Estonia, Trunks, cases, etc.

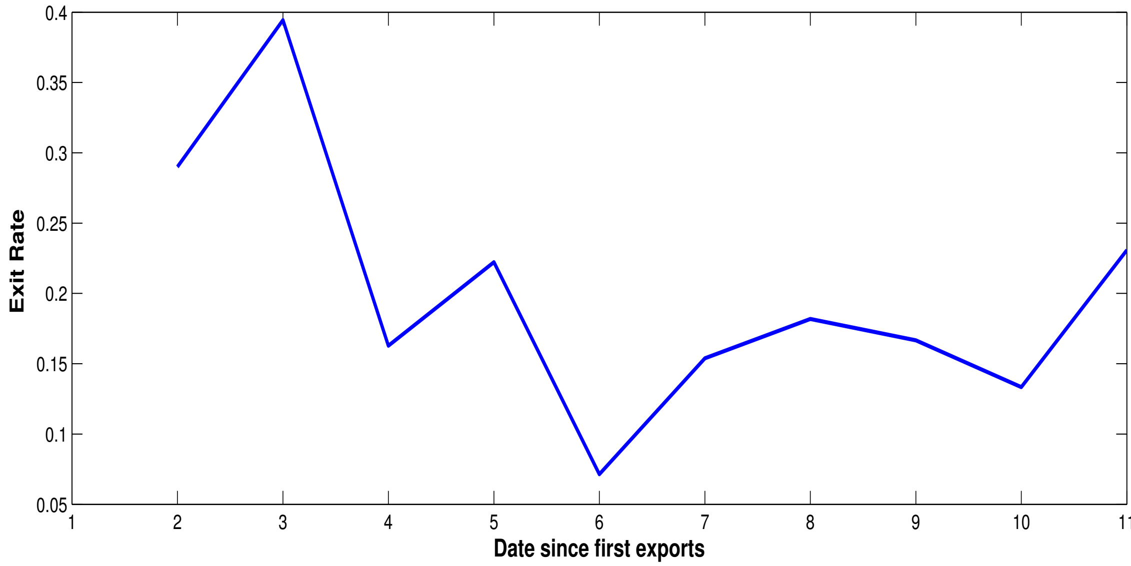

Figure 15: Simulated exit rate, over 100 exporters, Estonia, Trunks, cases, etc.

low demand exporters to learn that they have drawn a low μ, and exit the market. On the contrary, in the second class, where the signal-to-noise ratio is much higher, the convergence to the optimal scale happens by date 7 since first exports, and the exit rate peaks at date 3 - when most low demand exporters learn that they are indeed low demand sellers, and if they are not happy with that, quit the market.

We can also produce many (500) simulated patterns of export size, and average these out. We obtain the following graph for Estonia, Trunks, etc. (see Figure 16). Clearly, averaging over many possible trajectories smoothes out the evolution of size substantially - so we do not observe the sharper switches we could see in a specific simulation exercise.

Trade liberalization exercise

Here we carry out the following exercise. Suppose the destination country cuts down tariffs for French exporters, by 20 percent. This tariff cuts applies only to France, and assume that France is a small exporter in this market. In that case, the aggregate variables (in particular, the ideal price index) will not be affected by the behavior of French exporters, and we can focus only on the direct effect of lower tariffs on the profits of French exporters.

To simulate the response of French firms, we need to know how the thresholds for exporting will be affected, and in particular, how many (and what productivity) exporters will start exporting. First of all, we have the following expression for the profits earned by the firm j in the foreign market

$$\pi_j \quad = \quad p_j c(p_j^*) - c(p_j^*) mc$$

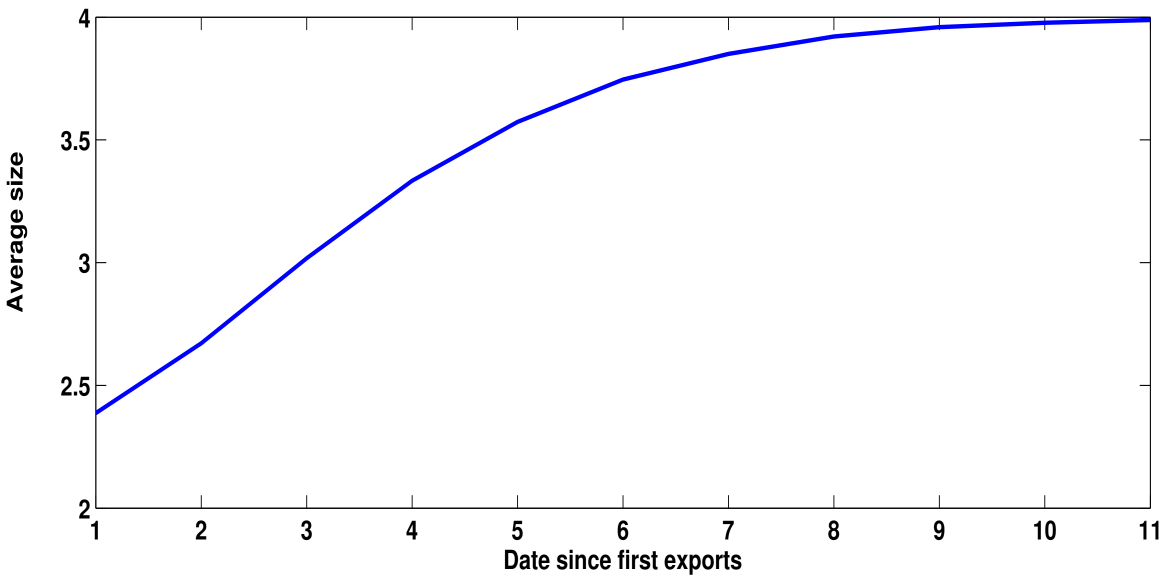

Figure 16: Smoothed out simulated average size, Estonia, Trunks, cases, etc.

$$
\begin{aligned}
&= (p_j^*)^{-\epsilon} A(p_j - mc) \\
&= (p_j \tau)^{-\epsilon} A(p_j - mc),
\end{aligned}
$$

where A incorporates all the aggregate demand variables, mc is the marginal cost of production, p_j is the price set by the firm (as a markup times the marginal cost), and $p_j^* \equiv p_j \tau$, the price faced by the consumers in the foreign market, which is the price p_j the multiplied by the tariff rate, and hence consumption is a function of this price. Thus, as tariffs fall by 20 percent, from say, 1.25τ to τ, the profits of a firm increase from $(p_j 1.25\tau)^{-\epsilon} A(p_j - mc)$ to $(p_j \tau)^{-\epsilon} A(p_j - mc)$, that is they increase by a factor of $(1.25)^{\epsilon}$.

If we keep scaling profits by the same scale factors as before (recall we scaled all profits by $exp((\epsilon-1)^2 \frac{\sigma_\phi^2}{2(1-(\rho_\phi)^2)}) exp(\frac{\sigma_A}{2}) md\bar{A}$ above), we will get the old $\tilde{\pi}$ shifted for each firm by this factor $(1.25)^{\epsilon}$. In Estonia, the estimated elasticity is 2.1315, and in Australia - 2.138.

This will give us the new threshold for starting exporting in this destination - we simply divide the old threshold (the $\tilde{\pi}$ of the minimum productivity new exporter in the market) by this factor, and calculate the underlying productivity. Denote the old productivity threshold by ϕ_1, and the new one by ϕ_2. All the firms with productivity between these two thresholds for exporting (before and after trade liberalization) will be new exporters.

To know how many of these there will be, we need to know the distribution of productivities in this sector. For that, we use the data on all the producers in the given sector in France, fit a Pareto distribution to their productivities, and calculate the (fitted) cumulative densities of the two productivity thresholds. Fit

$$\ln(1 - CF(\phi)) = k(\ln \phi_m - \ln \phi),$$

where ϕ_m is the minimum productivity in the sector, k is the shape parameter, and CF is the cumulative distribution.

In the dataset, each firm is assigned an industrial sector (NAF 2), according to their main activity. The exporters in a particular 4-digit category (such as Trunks, cases, etc., or Women's jackets, etc.) may come from different sectors. For example, the exporters in Estonia, Trunks, Cases, etc. belong to such sectors as (in decreasing frequency) Chemical Industry (production of artificial and synthetic fabrics), around 30 percent, Apparel and Furs, Leather products and shoes, and others, and exporters in Australia, Women's jackets, etc., belong to such sectors as Apparel and Furs (over 85 percent), Chemical Industry (mostly production of artificial and synthetic fabrics), Textile Industry, and Leather products and shoes. Hence, we pick the Chemical Industry (production of artificial and synthetic fabrics) as the domestic sector for the exporters of Trunks, Cases, etc, and Apparel and Furs as the domestic sector for the exporters of Women's jackets, etc. We obtain shape parameters of 1.258 and 1.69 for Trunks and Women's jackets, respectively.

We find that in Estonia, Trunks, etc., the productivity threshold for exporting falls from 23.5 to 15.7, which means that the firms with cumulative density of productivity in this sector between 0.5034 and 0.7016 will start exporting in response to the cut in tariffs. Thus, around 20 percent of all producers in the sector will become new exporters. If we count the number of producers in the sector that have productivity below the original cutoff for exporting (23.5), and take a share of $0.2/0.7 = 0.29$ of these, we get around 55 new exporters. We simulate these new exporters in Figure 17. We simulate only over the new exporters, who are the low productivity firms in this sector - therefore, the initial size is very small, around 1, and is stays close to that until all new exporters either quit or switch to the full scale. It takes 7 years for the firms to enter the entire market.

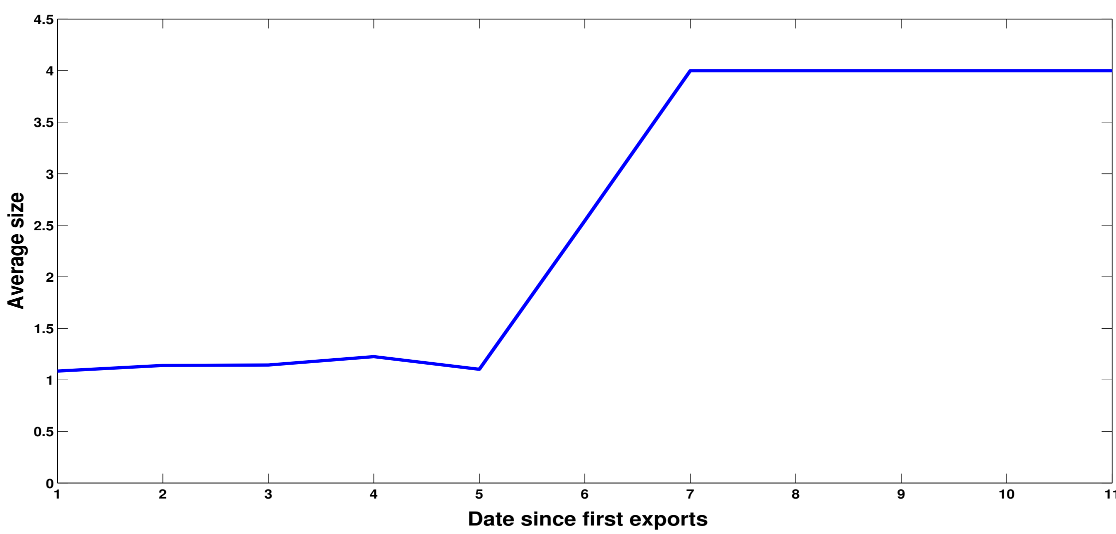

Figure 17: Response to tariffs cuts in Estonia, Trunks, cases, etc.

Estimates for Melitz model and Passive Learning

Here we compare the estimates of the experimentation model with those of the alternative models - Passive Learning with convex costs and Melitz model with no learning. As can be seen in Figure 25, the exporting cost is lower in the Passive Learning model than the testing cost estimated in the Experimentation model. This is as expected (see the theoretical comparison). To evaluate the sunk cost of entry F in the Passive Learning model, we look at the value at the initial belief p_0 of the lowest productivity exporter (depicted in Figure 26). It is found to be around 1.5, much lower than the F found above -around 50. Similarly, we find a much lower sunk cost (and lower exporting costs) for the Melitz model, around 0.7. Figures 25 and 26 can be found in the Appendix. We also show the optimal export size schedules for the 9 ranges of profitability in the Passive Learning model, in Figure 27.

We can also contrast the dynamics of exports in these alternative models. In the Passive Learning (with convex costs) model, the dynamics of average size (over all exporters) is very random - as can be seen in Figure 28, the average size may either converge to some high value, such as 3 or 4, or drop to a low value below 2 or 1. This depends on the draws of demand signals that the firms get , as well as the exogenous death shocks - if these hit the high productivity/high demand firms first, the remaining firms will be low productivity/low demand exporters, exporting low volumes. However, if we simulate many trajectories like these (around 500), and average these out (at each date since first exports, take the average over all simulated trajectories), we obtain a stable expansion in average size of new exporters over time. This is depicted in Figure 29. On the other hand, in the experimentation model, the predictions are very clear - the export size will be volatile and small initially, and will reach the full scale later on.

We also show the smoothed out (over many simulations) average size in the Melitz model - it is very stable around 1.8-2, since beliefs do not affect firm behavior at all in this model.

5 Conclusion

This paper shows that the model of experimentation (as well as alternative models of learning) can be applied to the data, and can be used to produce interesting dynamics of exporters. We argue that we need to take into account learning by firms to be able to explain the patterns of exports evolution over time.

We structurally estimate the model of firm experimentation in new markets of Akhmetova (2010). We carry out a particular application, where the firm learns about the demand for its 4-digit good by observing sales of the 8-digit subcategories of that good. We show the importance of the parameters in the model, such as the testing and sunk costs, and the signal-to-noise ratio, for the implied dynamics of new exporters. We also solve and estimate several alternative models, and compare the results. It is clear that the estimates of exporting costs are very sensitive to the assumptions made about uncertainty, learning, and the sequence of important steps in exporting, such as investing in the sunk cost of entry. We also show how different the dynamics of exports are in the alternative models - the predictions of the experimentation model regarding the evolution of average export volumes over time of new exporters are much more straightforward. We carry out a trade liberalization exercise, and produce a sample estimate of the length of the experimentation stage. As future exercises, we will study the numerical effect of changing testing costs and the signal-to-noise ratio in a sector on the duration of the experimentation stage, total costs of entry (testing costs plus sunk cost of entry), and exit rates.

References

Akhmetova Z. (2010), Firm Experimentation in New Markets, Working Paper.

Appendix

Estimation Details

Showing that a shift in profits results in a proportional shift in estimated costs

One important question is how a shift in the profits of the firms in the dataset affects the estimates of costs, the $c(n)$ schedule, the sunk entry cost F and the fixed cost of exporting f. That is, suppose we changed the units of account, and instead of units, expressed all profits in thousands of euros, or expressed all profits in dollars instead of euros. It is easy to show that this would shift all the costs proportionally - by the same proportionality factor.

To show this, we show that multiplying all profits and all the costs by the same factor λ results in the multiplication of the value function by λ and no change in the thresholds $\bar{p}$, $\underline{p}$, and the optimal testing schedule $n(p)$. Given original costs $c(n), F, f$, and profits $\tilde{\pi}_j, j = 1, ..., J$, multiply all these by λ:

$$\hat{c}(n) \equiv c(n)\lambda, \hat{F} \equiv F\lambda$$

$$\hat{f} \equiv f\lambda, \hat{\tilde{\pi}}_j \equiv \tilde{\pi}_j\lambda, j = 1, ..., J.$$

Recall that the solution to the original problem for a fixed firm j is given by $n(p) = z(rv(p))$, where $z \equiv g^{-1}$, $g(n) = nc'(n) - c(n)$, and z is strictly increasing, and $v(p)$ is the solution of the two-point free boundary value problem

$$v''(p) = \frac{c'(z(rv(p))) - \tilde{\pi}_j[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]}{\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2}$$

plus the value matching condition:

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F, v(\underline{p}) = 0.$$

and smooth pasting condition:

$$v'(\bar{p}) = \tilde{V}'(p), v'(\underline{p}) = 0.$$

Now, since $\hat{c}(n) \equiv c(n)\lambda$, $\hat{c}'(n) = c'(n)\lambda$, and $\hat{g}(n) = g(n)\lambda$. Hence, $\hat{z}(x) \equiv z(\frac{x}{\lambda})$, $\hat{n}(p) = \hat{z}(r\hat{v}(p)) = z(\frac{r\hat{v}(p)}{\lambda})$. Also, the value function in the second stage, $\hat{\tilde{V}}(p)$ is now given by

$$\begin{aligned} \hat{\tilde{V}}(p) = \frac{M}{r}(-\hat{z} + \lambda\tilde{\pi}_j[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) \\ = \lambda\tilde{V}(p). \end{aligned}$$

Now by simple substitution we can show that

$$\hat{v}(p) \equiv v(p)\lambda$$

satisfies the new BVP:

$$\hat{v}''(p) = \frac{\hat{c}'(\hat{z}(r\hat{v}(p))) - \lambda\tilde{\pi}_j[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]}{\frac{1}{2}(p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x})^2}$$

plus the value matching condition:

$$\hat{v}(\bar{p}) = \hat{\hat{V}}(\bar{p}) - \hat{F}, \hat{v}(\underline{p}) = 0.$$

and smooth pasting condition:

$$\hat{v}'(\bar{p}) = \hat{\hat{V}}'(p), \hat{v}'(\underline{p}) = 0.$$

Thus, proportionally shifting both the profits and the cost parameters results in the same experimentation behavior, and therefore, will produce the same values of likelihood, given data, as the original profits and cost parameters. That is, we only need to multiply by λ all the values we generate originally as draws from the posterior of these cost parameters to obtain the draws from the new posterior. This observation allows us to carry out estimation for normalized profits, and later rescale to match the monetary values in the dataset.

Estimating demand parameters

We would like to update the posterior of the parameters $\bar{\mu}$, $\underline{\mu}$, σ_x, and p_0, as well as $A_{t1}, A_{t2}, t = 1, ..., T$. It can be seen from the equations above that the unobserved α_{jt}^k and A_{t1}, A_{t2} can be identified up to a constant, and therefore we normalize $\underline{\mu} = 0$. We set the following priors:

- $\bar{\mu} \sim N(\mu_h, \sigma_{\mu_h})$,
- $\frac{1}{\sigma_x^2} \sim Gamma(g_1, g_2)$,
- $p_0 \sim Beta(b_1, b_2)$.
- $\rho_1 \sim N(0, 1)$, $\rho_2 \sim N(0, 1)$, and these are independent.
- $D_1, D_2 \sim N(0, 1)$.
- CA_1 and CA_2 have normal priors with mean 0.

Denote $\Theta_1 \equiv \bar{\mu}, \sigma_x, p_0$, the main parameters characterizing demand uncertainty. The joint posterior of these parameters, aggregate factors A_{t1}, A_{t2}, and the other relevant paramaters, given $v_{jt}^k, j = 1, ..., J, t = 1, ..., T, k = 1, ..., K(j)$, is given by the product of their priors and their joint likelihood:

$$f(\Theta_1, \alpha_{jt}^k, A_{t1}, A_{t2}, \sigma_x^2, \rho_1, \rho_2, CA_1, CA_2 | v_{jt}^k) = \frac{f(\Theta_1) f(\alpha_{jt}^k, A_{t1}, A_{t2}, \sigma_x^2, \rho_1, \rho_2, CA_1, CA_2) L(\Theta_1 | v_{jt}^k)}{f(v_{jt}^k)}.$$

Since the joint posterior is quite complicated, we apply Gibbs sampling.

Denote $H_j = I(\mu_j = \bar{\mu})$, that is, H_j is 1 if firm j holds a high μ, and 0 otherwise, and $\tilde{H} \equiv H_j, j = 1, ..., J$. Given p_0, the prior $P(H_j = 1)$ is simply p_0. We treat $\tilde{H}$ as another unobserved parameter. The conditional distributions of the main parameters of interest are as follows:

- $f(\bar{\mu} | \underline{\mu}, \sigma_x, p_0, \tilde{H}, \alpha_{jt}^k)$ is a normal with mean $\frac{m_h \sigma_\alpha^2 + N_h \bar{\mu}_h \sigma_{\mu_h}^2}{\sigma_\alpha^2 + N_h \sigma_{\mu_h}^2}$ and variance $\frac{\sigma_\alpha^2 + N_{ah} \sigma_{\mu_h}^2}{\sigma_\alpha^2 \sigma_{\mu_h}^2}$, where $\bar{\mu}_h \equiv \frac{\sum_{j: H_j = 1} \sum_t \sum_k \alpha_{jt}^k}{N_{ah}}$, and N_{ah} is the number of observations over all high-μ firms in the given combination $\tilde{H}$.

- $f(\frac{1}{\sigma_x} | \bar{\mu}, \underline{\mu}, p_0, \tilde{H}, \alpha_{jt}^k)$ is Gamma with hyperparameters $g_1 + \frac{N}{2}$ and $g_2 + \frac{SSR}{2}$, where SSR is the sum of squared residuals, $e_{jt}^k = \alpha_{jt}^k - \bar{\mu}$ if $H_j = 1$ and $e_{jt}^k = \alpha_{jt}^k - \underline{\mu}$ if $H_j = 0$.

- $f(p_0 | \bar{\mu}, \underline{\mu}, \sigma_x, \tilde{H}, \alpha_{jt}^k)$ is Beta with hyperparameters $b_1 + N_h$ and $b_2 + N_l$, where N_h is the number of high-μ firms and N_l is the number of low-μ firms in the given combination $\tilde{H}$.

- The conditional distribution of H_j is given by $P(H_j = 1 | \bar{\mu}, \underline{\mu}, \sigma_x, p_0, \alpha_{jt}^k)$. To calculate this, we apply the discrete analog of the Bayesian updating equation used in the theoretical part:

$$Prob(\mu_j = \bar{\mu}) \equiv P_j = \frac{fn(\frac{\bar{\alpha}_j - \bar{\mu}}{\frac{\sigma_x}{\sqrt{n_j}}}) P_0}{fn(\frac{\bar{\alpha}_j - \bar{\mu}}{\frac{\sigma_x}{\sqrt{n_j}}}) P_0 + fn(\frac{\bar{\alpha}_j - \underline{\mu}}{\frac{\sigma_x}{\sqrt{n_j}}})(1 - P_0)},$$

  where $\bar{\alpha}_j$ is the average over all t and k of the previously generated draws of α_{jt}^k for firm j, and n_j is the number of these draws. fn is the standard normal density

- To evaluate the posteriors of A_{t1}, A_{t2}, conditional on the rest of the unknowns, we apply the Kalman smoother. Once we have draws from the posterior of A_{t1}, A_{t2}, and D_1, D_2, we can calculate α_{jt}^k as the residuals in

$$v_{jt}^k = D_1 A_{t1} + D_2 A_{t2} + \alpha_{jt}^k.$$

Using these α_{jt}^{k}, we can then proceed iterating and updating the posteriors of Θ_1 and other parameters. We draw the values from the conditional posterior distributions of the parameters Θ_1, one at a time, iterating 500 times, until convergence.

Estimating TFP

To carry out TFP estimation we use only data on domestic sales of the firm, i.e. $R_{jt} = \hat{R}_{jt} - X_{jt}$, where $\hat{R}_{jt}$ is total revenues of the firm, and X_{jt} is its export revenues. Thus, we evaluate the firm's TFP from domestic production only, which allows us to abstract away from the additional complications of a demand system for al the markets of the firm (domestic and foreign). We assume that the inputs used for domestic production only are the same fraction of total inputs used as the fraction of domestic sales in total sales.

An extensive literature is devoted to the estimation of TFP. We explain how we deal with some issues that have been brought up so far below. Begin with the pair of production and demand equations:

$$Q_{jt} = L_{jt}^{\alpha_l} M_{jt}^{\alpha_m} K_{jt}^{\alpha_k} e^{\omega_{jt}+u_{jt}},$$

$$Q_{jt} = Q_{st}[\frac{P_{jt}}{P_{st}}]^{-\epsilon} e^{\eta_{jt}},$$

where L, M, K are inputs - labor, intermediate inputs and capital, respectively, ω_{jt} is unobserved productivity shock, u_{jt} is an error term, P_{jt} is firm j-s price, Q_{jt} is firm j-s quantity produced and sold, η_{jt} is unobserved demand shock, Q_{st} and P_{st} are industry-wide output and price index, respectively. j indexes firms, and t indexes time (years in our case). We observe only total domestic revenues (and not physical output):

$$R_{jt} \equiv Q_{jt}P_{jt} = Q_{jt}Q_{jt}{}^{-\frac{1}{\epsilon}} Q_{st}{}^{\frac{1}{\epsilon}} P_{st} e^{\eta_{jt}\frac{1}{\epsilon}},$$

so that upon dividing both sides by P_{st} and taking logs:

$$\tilde{r}_{jt} = \frac{\epsilon - 1}{\epsilon} q_{jt} + \frac{1}{\epsilon} q_{st} + \frac{1}{\epsilon} \eta_{jt},$$

and plugging in the equation for the production function:

$$\tilde{r}_{jt} = \beta_l l_{jt} + \beta_m m_{jt} + \beta_k k_{jt} + \beta_s q_{st} + \omega_{jt} + \eta_{jt} + u_{jt},$$

the small letters denote the logarithms of the capitalized variables (e.g. $q_{jt} \equiv \ln Q_{jt}$), and all the coefficients are reduced form parameters combining the production function and demand parameters.

We next denote $\tilde{\omega}_{jt} \equiv \omega_{jt} + \eta_{jt}$ and re-write:

$$\tilde{r}_{jt} = \beta_l l_{jt} + \beta_m m_{jt} + \beta_k k_{jt} + \beta_s q_{st} + \tilde{\omega}_{jt} + u_{jt}.$$

Here we follow Levinsohn and Melitz (2006) in treating both the unobserved productivity and demand shocks jointly, so that we do not identify these two sources of firm profitability independently. Since we later use the estimates of $\tilde{\omega}$ to estimate demand shocks α in the foreign markets, we need to consider how this assumption affects that part of estimation. For $\tilde{\omega}$ to be a good instrument for unit values, we need $\tilde{\omega}$ to be correlated with the unit values which it is : as shown in the main part of the paper, we assume that

$$Q^*_{jt} = Q^*_{st}[\frac{P^*_{jt}}{P^*_{st}}]^{-\epsilon} e^{\alpha^*_{jt}},$$

where stars denote foreign market variables. Assume for now that there are constant returns to scale in this industry, so that average cost is equal to marginal cost. Denote by T the combined input $L^{\alpha_l}_{jt} M^{\alpha_m}_{jt} K^{\alpha_k}_{jt}$, and by c_T the average cost of this input: $c_T \equiv \frac{wL+rK+mM}{L^{\alpha_l}_{jt} M^{\alpha_m}_{jt} K^{\alpha_k}_{jt}}$, which is constant for all output levels with constant returns to scale. Then the foreign price is a product of the firm's average cost of production and mark-up and possibly iceberg trade cost τ:

$$P^*_{jt} = \tau P_{jt} = \tau \frac{c_T}{e^{\omega_{jt}}},$$

and since $\tilde{\omega}_{jt} = \omega_{jt} + \eta_{jt}$, P^*_{jt} and $\tilde{\omega}$ are clearly correlated. We also assume that η_{jt}, the domestic demand shocks, are not correlated with foreign demand shocks, which is necessary to have no correlation between $\tilde{\omega}$ and the error term in the regression

$$q^*_{jt} = q^*_{st} + \epsilon p^*_{st} - \epsilon p^*_{jt} + \alpha^*_{jt},$$

which is the equation (in different notation) we relied on in the main part of the paper to estimate ϵ and later generate $\alpha - s$.

Now returning to the task at hand - estimating TFP (combined with domestic demand shocks, i.e. $\tilde{\omega}$). Since our goal is only estimation of TFP, it suffices for us to control for industry-wide output with industry-time fixed effects. Therefore, we have

$$\tilde{r}_{jt} = \beta_l l_{jt} + \beta_m m_{jt} + \beta_k k_{jt} + \tilde{\omega}_{jt} + \sum_{t=1}^{T} \beta_{Dt} D_{st} + \epsilon_{jt}.$$

Next issue to deal with is the identification of the variable inputs' coefficients. To do that, we take note of the ACF (Ackerberg, Caves and Frazer) critique of the Levinsohn-Petrin and Olley-Pakes estimation approach, and estimate all the input coefficients in the second stage. We use value added in the first stage regression, so that we only estimate the coefficients on labor and capital in the production function. We do use the data on intermediate inputs, however, as the control for (unobserved) productivity. We use the equation for intermediate inputs

$$m_{jt} = m_t(k_{jt}, \tilde{\omega}_{jt}),$$

so that assuming monotonicity in the function $m_t(.)$, we can invert:

$$\tilde{\omega}_{jt} = \psi_t(m_{jt}, k_{jt}).$$

Similarly, we assume that the optimal quantity of labor is chosen once current productivity is observed by the firm, so that

$$l_{jt} = l_t(k_{jt}, \tilde{\omega}_{jt}),$$

and once we utilize the expression for $\tilde{\omega}_{jt}$ above,

$$l_{jt} = l_t(k_{jt}, \psi_t(m_{jt}, k_{jt})).$$

Inserting these into the expression for value added:

$$\begin{aligned}
\tilde{v}a_{jt} &= \beta_l l_{jt} + \beta_k k_{jt} + \psi(m_{jt}, k_{jt}) + \sum_{t=1}^{T} \beta_{Dt} D_{st} + \epsilon_{jt} \\
&= \beta_l l_t(k_{jt}, \psi(m_{jt}, k_{jt})) + \beta_k k_{jt} + \psi(m_{jt}, k_{jt}) + \sum_{t=1}^{T} \beta_{Dt} D_{st} + \epsilon_{jt} \\
&= \Psi_t(m_{jt}, k_{jt}) + \sum_{t=1}^{T} \beta_{Dt} D_{st} + \epsilon_{jt},
\end{aligned}$$

where $\Psi_t(m_{jt}, k_{jt})$ is a polynomial in capital and intermediate inputs, one for each time period (year in our case):

$$\begin{aligned}
\Psi(m_{jt}, k_{jt}) \equiv\ & \sum_{t=1}^{T} q_{0t} D_t + \sum_{t=1}^{T} q_{1kt} D_t k_{jt} + \sum_{t=1}^{T} q_{1mt} D_t m_{jt} + \sum_{t=1}^{T} q_{2mkt} D_t k_{jt} m_{jt} \\
& + \sum_{t=1}^{T} q_{2kkt} D_t k_{jt}^2 + \sum_{t=1}^{T} q_{2llt} D_t l_{jt}^2 + \sum_{t=1}^{T} q_{3kkmt} D_t k_{jt}^2 m_{jt} \\
& + \sum_{t=1}^{T} q_{3kmmt} D_t k_{jt} m_{jt}^2 + \sum_{t=1}^{T} q_{3kkkt} D_t k_{jt}^3 + \sum_{t=1}^{T} q_{3mmmt} D_t m_{jt}^3.
\end{aligned}$$

Assuming that productivity follows a first-order Markov process:

$$\tilde{\omega}_{jt} = E[\tilde{\omega}_{jt} | \tilde{\omega}_{j(t-1)}] + \nu_{jt},$$

where ν_{jt} is uncorrelated with k_{jt}, and given values of β_l and β_k, one can estimate the residual ν_{jt} (unobserved innovation to productivity) non-parametrically from

$$\tilde{\omega}_{jt} = \tilde{v}a_{jt} - \beta_l l_{jt} - \beta_k k_{jt} - \sum_{t=1}^{T} \beta_{Dt} D_t,$$

$$\tilde{\omega}_{jt} = z_0 + z_1\tilde{\omega}_{j(t-1)} + z_2\tilde{\omega}^2_{j(t-1)} + z_3\tilde{\omega}^3_{j(t-1)} + \nu_{jt}.$$

Next, use the moments

$$E[\nu_{jt}(\beta_k, \beta_l)k_{jt}] = 0,$$

$$E[\nu_{jt}(\beta_k, \beta_l)l_{j(t-1)}] = 0,$$

to identify the coefficients on capital and labor.

Finally, $\tilde{\omega}_{jt}$ is given by

$$\tilde{\omega}_{jt} = \tilde{v}a_{jt} - \beta_l l_{jt} - \beta_k k_{jt} - \sum_{t=1}^{T} \beta_{Dt} D_t.$$

MCMC Sampling for Θ_2

We apply Metropolis-Hastings and Gibbs sampling to update the posteriors of $\Gamma, F, r, \delta, M, \sigma_N, P_q$. We often employ the Metropolis-Hastings step to generate draws from the conditional densities, since these are in most cases not standard distributions we can generate draws from. In calculating the likelihoods of the data, given the parameters, we need to know the states of the firms - experimenter, full-scale exporter, non-exporter. We explained above that we treat as a non-exporter (a firm that exited the market) any firm that displays zeros in the number of products from some year t up to 2005. Thus, the non-exporter state is fully observed. Given the costs ($c(n)$ and F), we can calculate the cutoff thresholds for quitting and entering for experimenters - $\bar{p}$ and $\underline{p}$. These imply that firms with profitabilities above a certain threshold of productivity (such that $\bar{p}$ for these productivities is below p_0, the initial belief for all firms) will enter right away, and start exporting at a full scale, and firms below this threshold will experiment. Of course, we need to make sure that the $\underline{p}$ for the lowest present exporter is at least as high as p_0 (and in fact, it should be equal to p_0). This is the "cutoff profit" condition in this model, stemming from the fact that a firm will only start exporting in the market if the initial belief p_0 is not lower than its threshold for quitting the market - where its expected value from exporting (experimenting) is exactly 0.

Once we know the thresholds for quitting and entering for experimenters, we can also predict their states throughout their history, given their beliefs p_{jt}. If the belief of firm j at time t exceeds $\bar{p}_j$, it should enter the market, so that $n^*_{js} = M, s = t, ..., T$. If the belief of firm j at time t falls below $\underline{p}$, it should quit, with probability P_q, or stay and export $n_{jt} = n^*(\underline{p})$ (by assumption), with probability $1 - P_q$. If the belief of the firm is in-between $\bar{p}$ and $\underline{p}$, it experiments, with optimal $n^*(p_{jt})$. Recall the notation for states: experimentation phase is denoted by $S_t = 0$, full-scale export phase by $S_t = 1$, and non-exporting phase by $S_t = 2$. Also note that in the full-scale export state, $S_{jt} = 1$, $n^*(p_{jt}) = M$. Denote $E^1_{jt} = [j, t|S_{jt} = 2, S_{j(t-1)} = 1]$, that is, all j, t where the firm exits the market from State

1, $R^1_{jt} = [j,t : S_{jt} = 1, S_{j(t-1)} = 1]$, that is, all j,t where the firm remains in the market in State 1, $Q^0_{jt} = [j,t : S_{jt} = 2, S_{j(t-1)} = 0, p_{jt} \leq \underline{p}]$, that is, all j,t where the firm exits the market from State 0, given its beliefs are below $\underline{p}$, $Q^1_{jt} = [j,t : S_{jt} = 2, S_{j(t-1)} = 0, p_{jt} > \underline{p}]$, that is, all j,t where the firm exits the market from State 0, given its beliefs are above $\underline{p}$, $R^0_{jt} = [j,t : S_{jt} < 2, S_{j(t-1)} = 0, p_{jt} \leq \underline{p}]$, that is, all j,t where the firm remains in the market in State 0, given its beliefs are below $\underline{p}$, which happens with probability P_q.

Solving the problem of the firm, given all the parameters, requires solving a free boundary value problem for each possible profitability level. Since that would be extremely demanding computationally, we discretize the space of profitabilities into 10 ranges, by deciles. Another simplification that speeds up the estimation procedure is discretizing the parameters $\gamma_3, \gamma_4, \gamma_5 \in \Gamma$.

To summarize, we do the following. We set the set of feasible M, given the data. We evaluate the marginal likelihood of M by generating the posterior draws of $\Gamma, F, r, \delta, \sigma_N, P_q$, given M, and calculating the harmonic mean of likelihoods:

$$L(M|n_{jt}, p_{jt}) \sim \left(\frac{\sum L(\Gamma, F, r, \delta, \sigma_N, P_q|n_{jt}, p_{jt}, M)^{-1}}{ns}\right)^{-1},$$

where ns is the number of draws of $\Gamma, F, r, \delta, \sigma_N, P_q$ from their posterior conditional on M. We then apply the Metropolis-Hastings step to update the distribution of M.

Given the value of M, we apply Gibbs sampling to update the posteriors of r, δ, σ_N, P_q, by evaluating their marginal likelihood, conditional on the rest of the subset, as the harmonic mean of likelihoods over posterior draws of the cost parameters Γ, F:

$$L(\theta|n_{jt}, p_{jt}, \theta^{-1}, M) \sim \left(\frac{\sum L(\Gamma, F|n_{jt}, p_{jt}, r, \delta, \sigma_N, P_q, M)^{-1}}{ns}\right)^{-1},$$

where θ is the parameter being considered, θ^{-1} is the rest of the parameters (out of r, δ, σ_N, P_q), and ns is the number of draws of the rest of the parameters that we use to evaluate this.

The following is the likelihood of the data given all the parameters:

$$\begin{aligned} f(n_{jt}, p_{jt}|c(n), F, \sigma_N, P_q, \delta, r, M) \\ &= \prod_{jt} fn\left(\frac{\ln n_{jt} - \ln n^*(p_{jt})}{\sigma_N}|S_{jt}\right) * \\ &\prod_{R^0_{jt}}(1-P_q)\prod_{E^1_{jt}}\delta\prod_{R^1_{jt}}(1-\delta) * \\ &\prod_{Q^0_{jt}}((1-\delta)P_q+\delta)\prod_{Q^1_{jt}}\delta\prod_{R^0_{jt}}(1-\delta)(1-P_q), \end{aligned}$$

where fn denotes the standard normal distribution.

To update the posterior of σ_N, it suffices to observe the deviation of observed export sizes from the optimal ones: given the cost parameters, M, and r, we can fix the states for all firms, and we can predict the optimal sizes for firms in both stages. Hence, the posterior of σ_N is simply an inverse gamma distribution:

$f(\frac{1}{\sigma_N}|n_{jt}, p_{jt}, c(n), F, M, \delta, r)$ is Gamma with hyperparameters $g_1 + \frac{N}{2}$ and $g_2 + \frac{SSR}{2}$, where SSR is the sum of squared residuals, $n_{jt} - n^*(p_{jt})$, where $n^*(p_{jt})$ is given by the optimal experimentation schedule if $S_{jt} = 1$, and $n^*(p_{jt}) = M$ if $S_{jt} = 0$, N is the number of these residuals, and g_1, g_2 are the prior parameters.

Figures

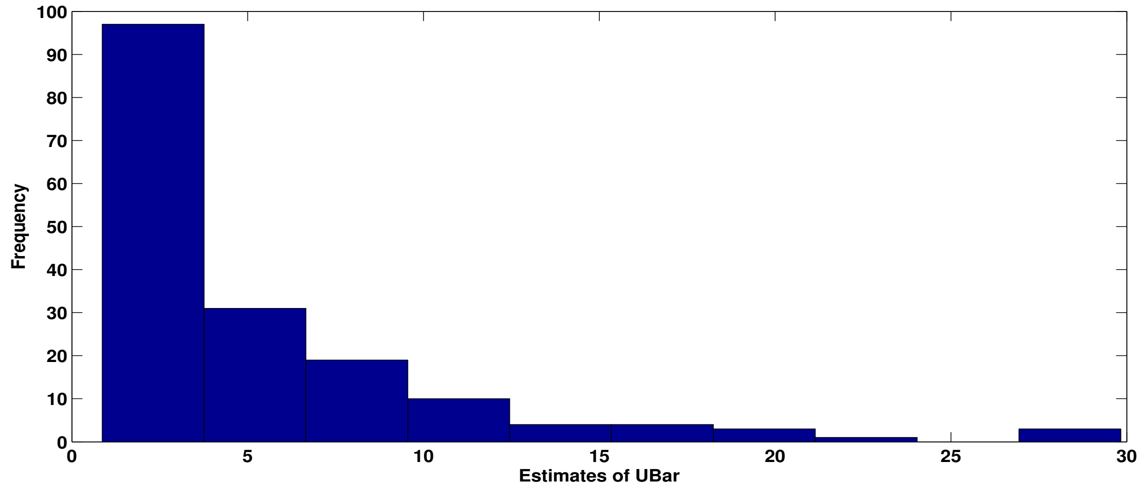

Figure 18: Summary of $\bar{\mu}$

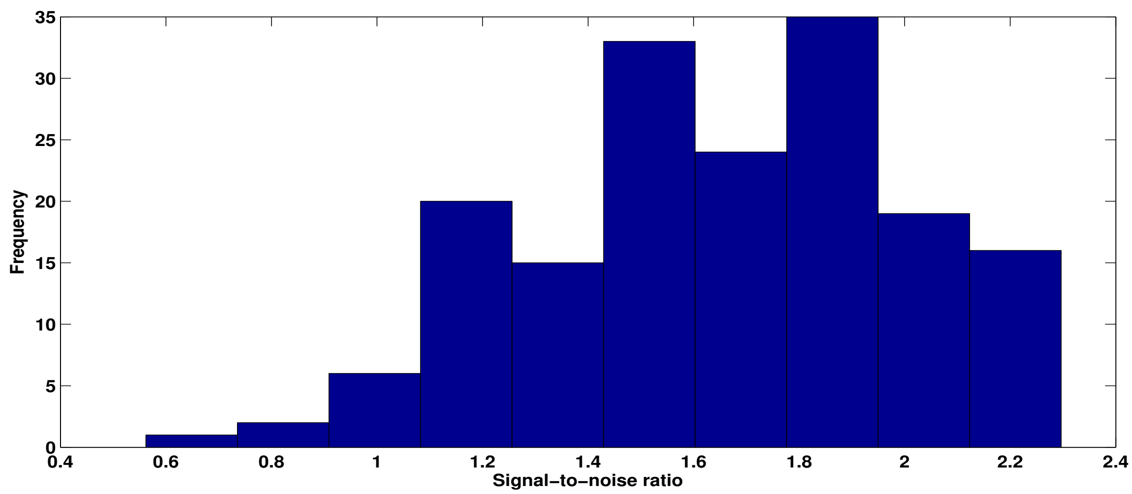

Figure 19: Summary of the signal-to-noise ratio

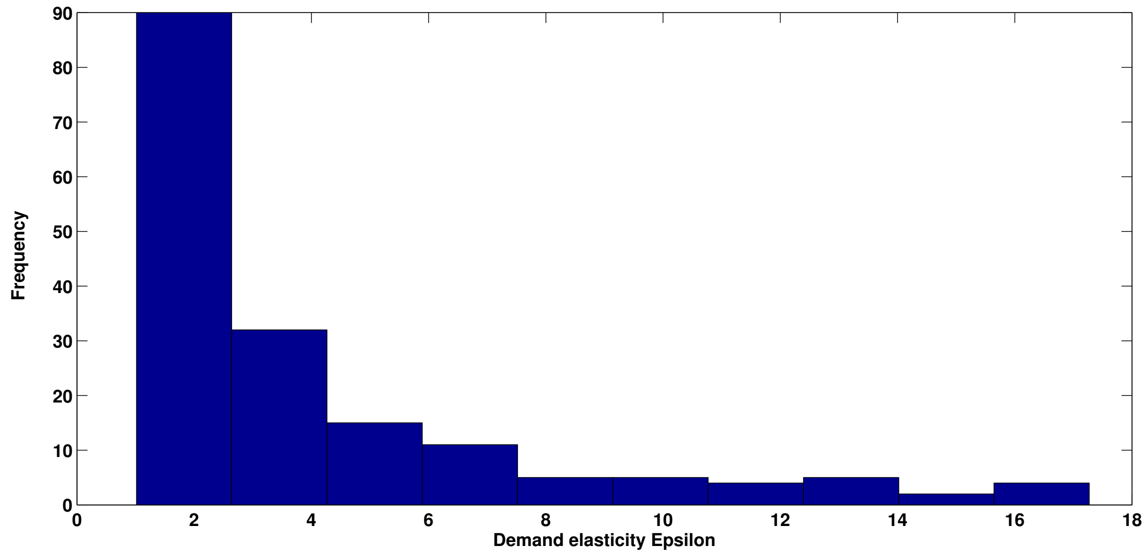

Figure 20: Summary of elasticities

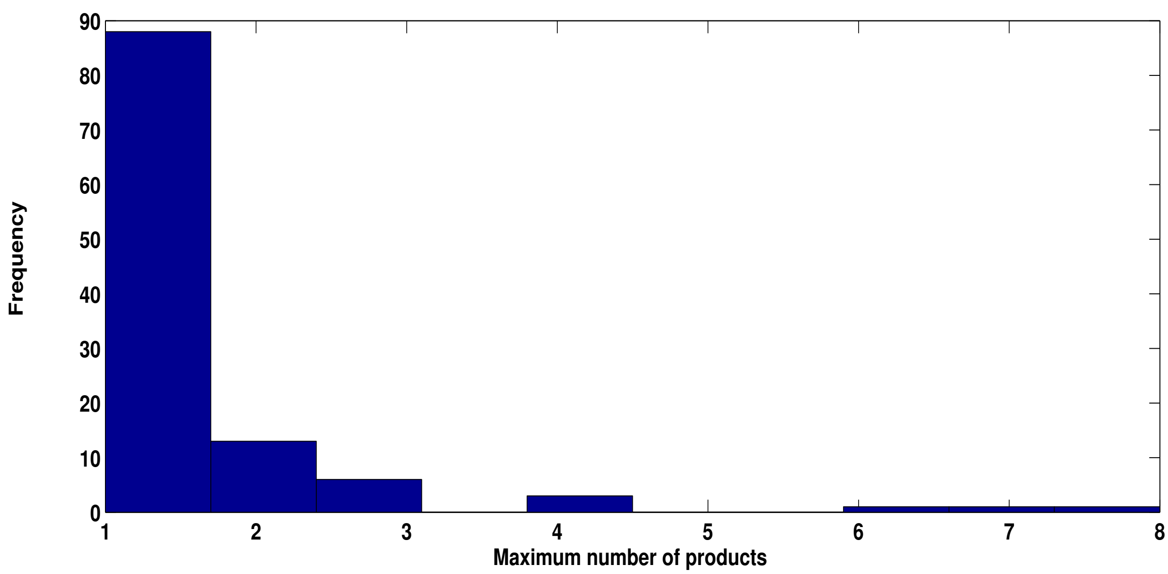

Figure 21: Histogram of maximum number of products, new exporters, Australia, women's dress, etc.

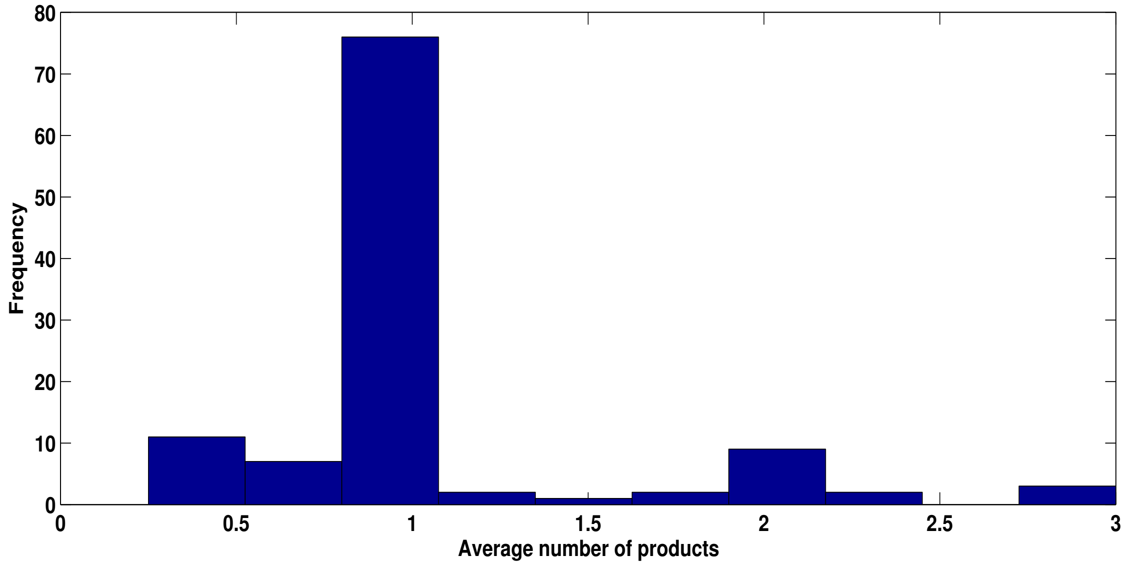

Figure 22: Histogram of average number of products, new exporters, Australia, women's dress, etc.

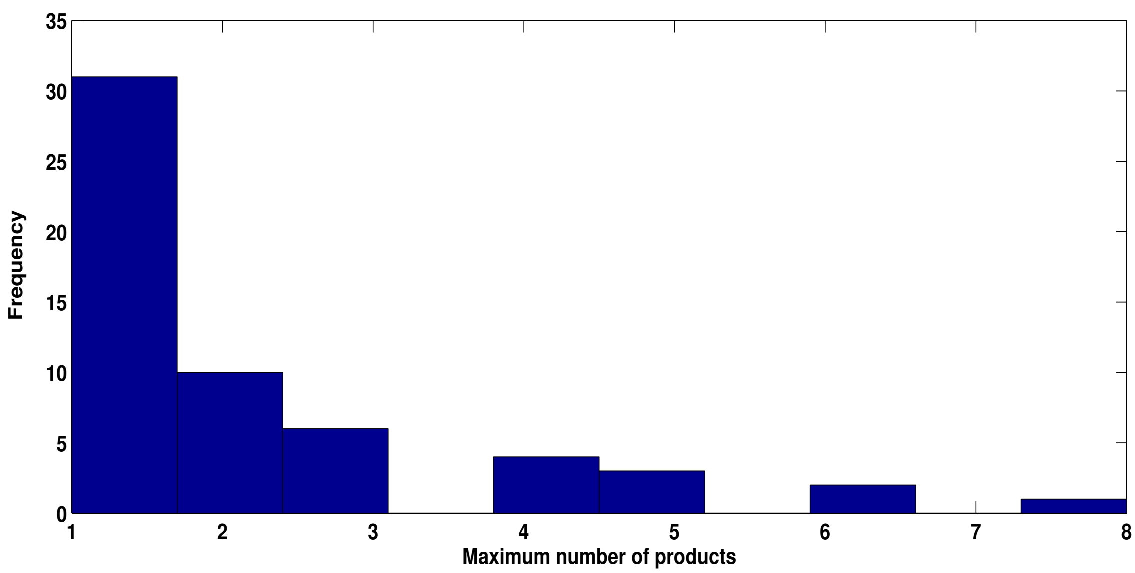

Figure 23: Histogram of maximum number of products, new exporters, Estonia, Trunks, cases, etc.

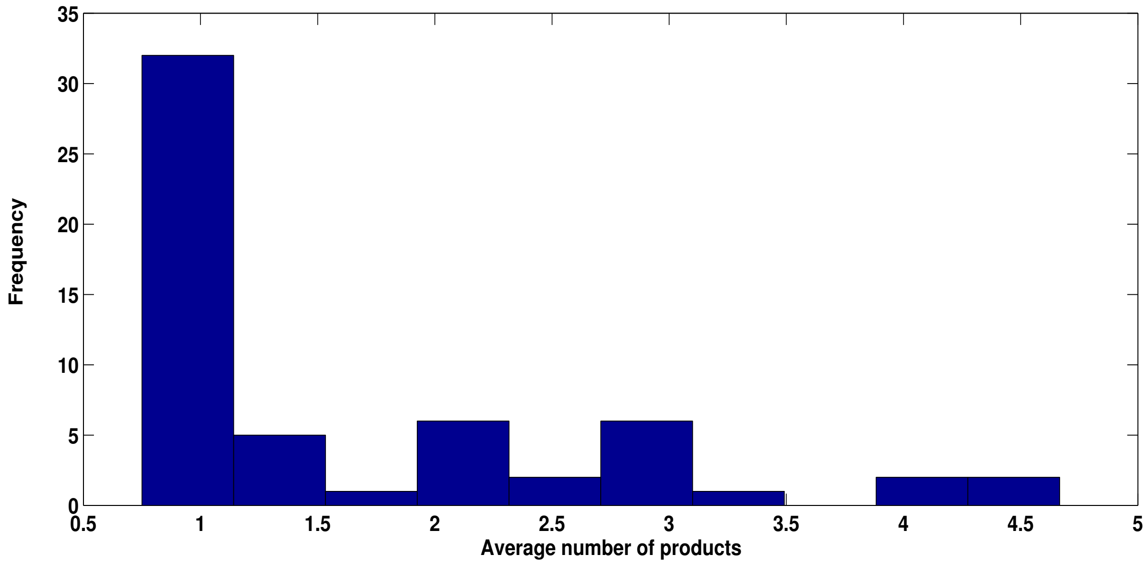

Figure 24: Histogram of average number of products, new exporters, Estonia, Trunks, cases, etc.

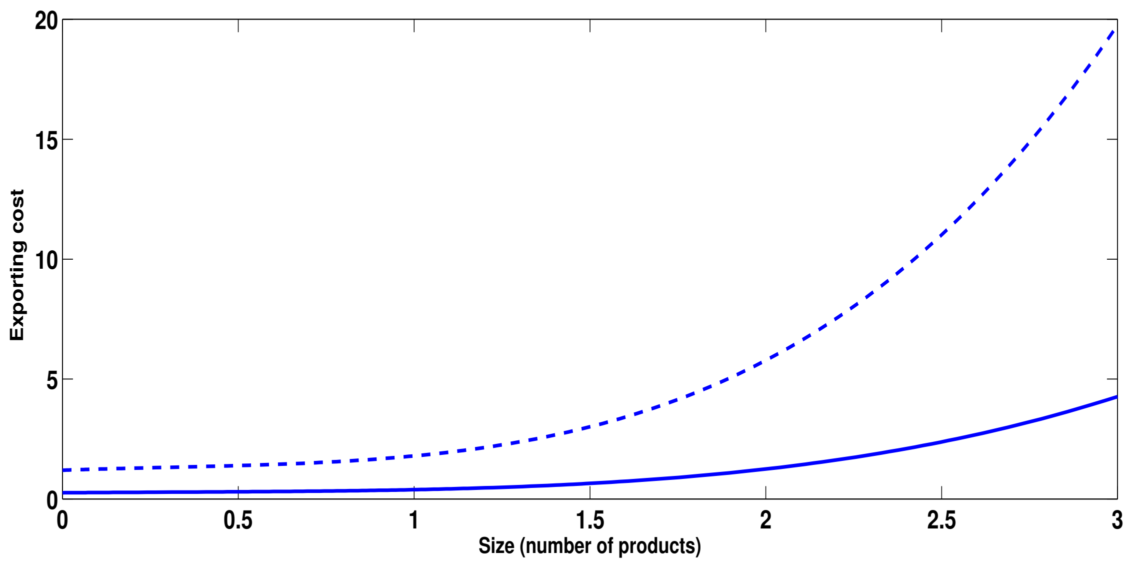

Figure 25: Exporting cost for the Passive Learning model, Estonia, Trunks, cases, etc.

The dashed line shows the estimated exporting cost (scaled by median profitability), and the solid line shows the costs scaled by median profitability and $exp(\bar{\mu})$.

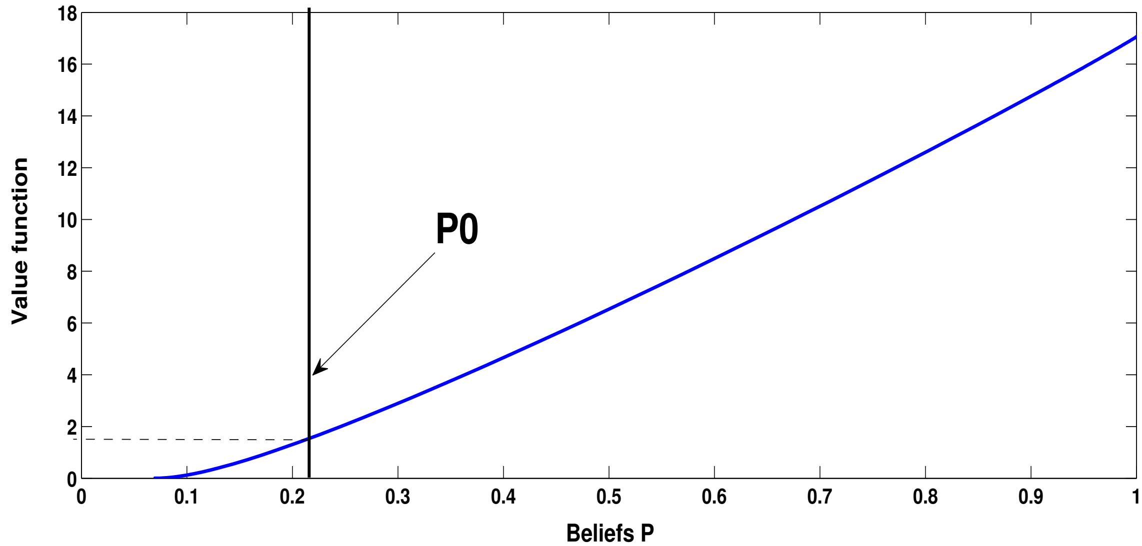

Figure 26: The value function for the lowest productivity exporter, and the estimate of sunk cost F, Estonia, Trunks, cases, etc.

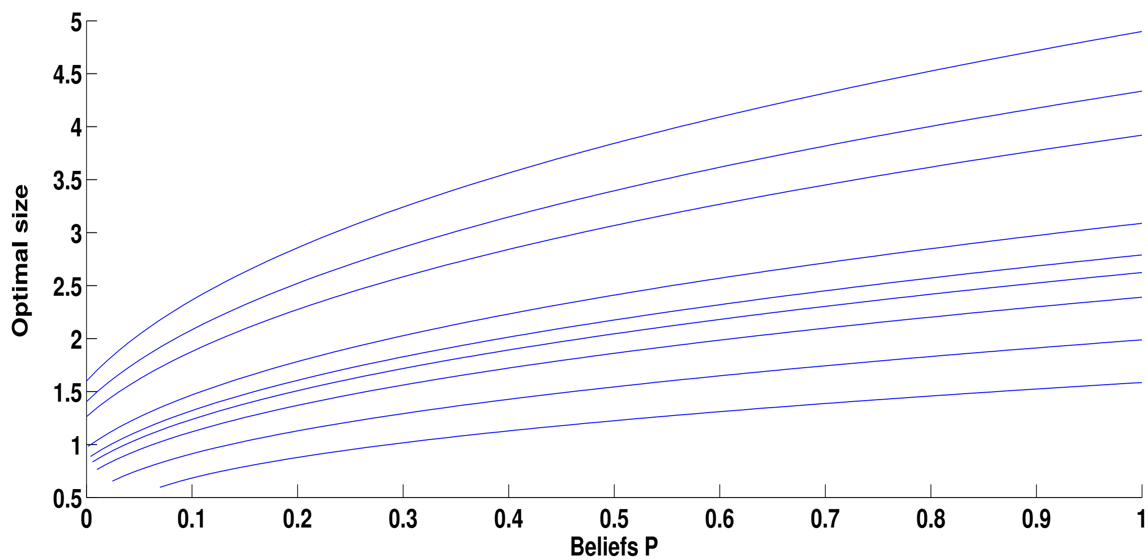

Figure 27: Optimal size schedules, by ranges of profitability, Passive Learning with convex costs, Estonia, Trunks, cases, etc.

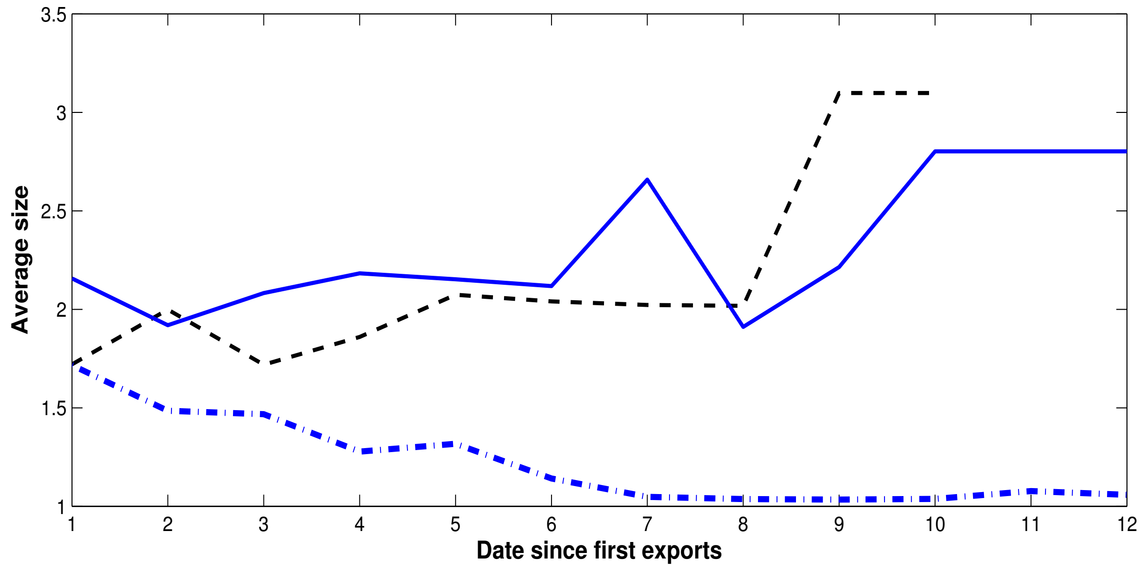

Figure 28: Simulated Average Size Dynamics, 3 Examples, Passive Learning, Estonia, Trunks, cases, etc

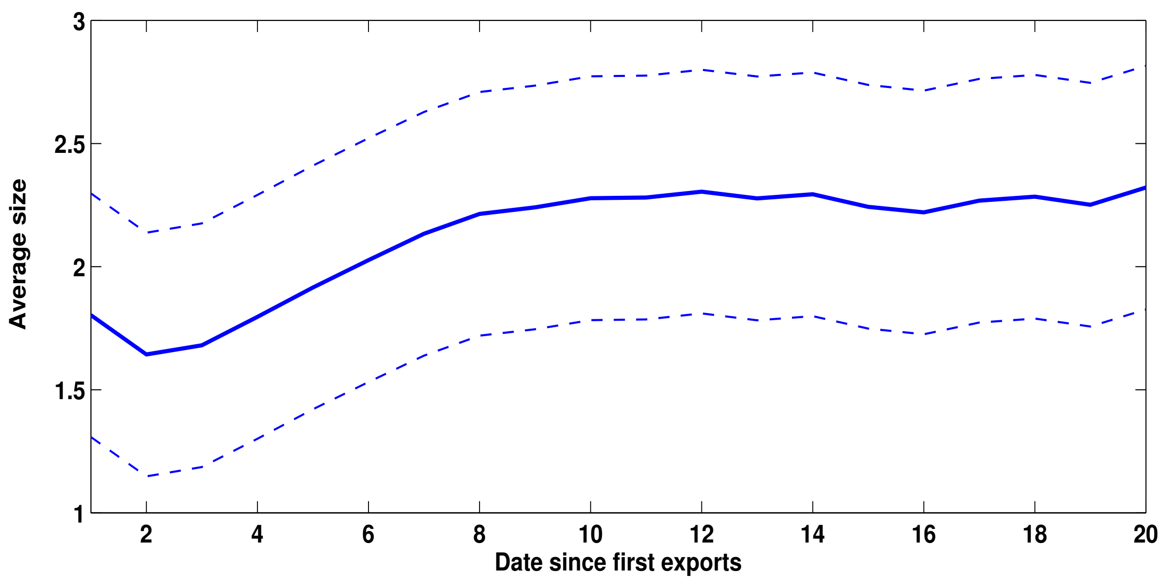

Figure 29: Smoothed out (over 500 simulations) average size in Passive Learning, Estonia, Trunks, cases, etc.

The curve in the middle is the smoothed out average, and the bands around it allow for one standard error deviation.

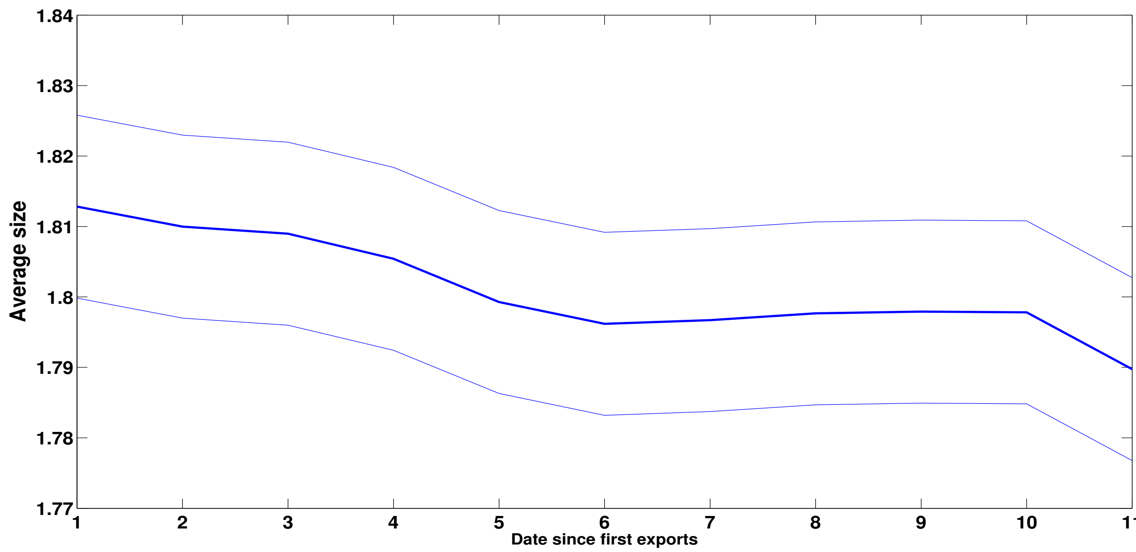

Figure 30: Smoothed out (over 500 simulations) average size in Melitz model, Estonia, Trunks, cases, etc.

The curve in the middle is the smoothed out average, and the bands around it allow for one standard error deviation.