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.

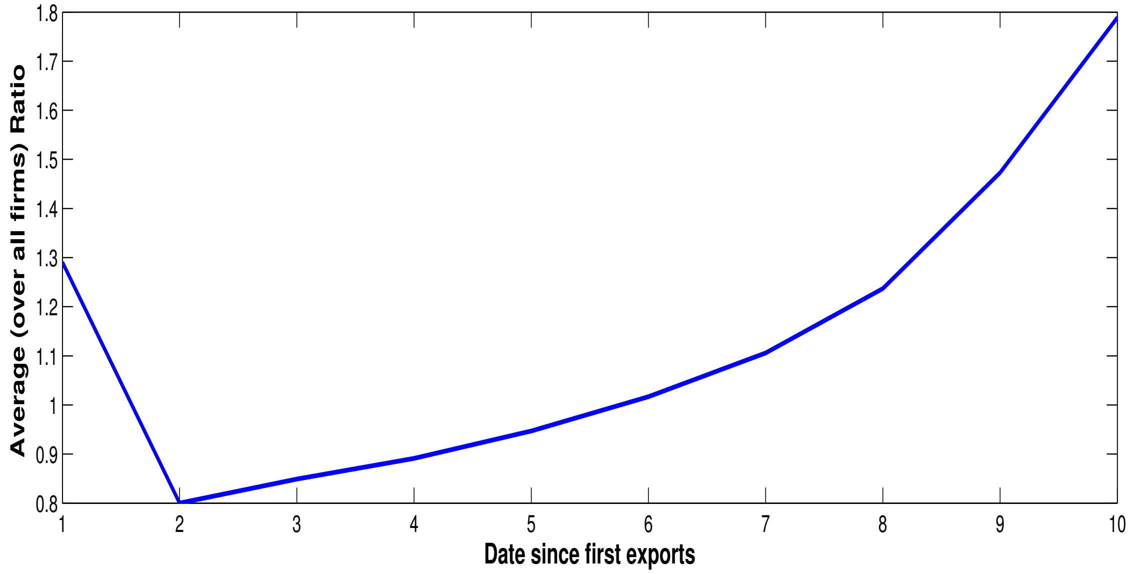

Figure 1: Ratio of quantity (8-digit product) to average quantity of the firm over the period, averaged over all new exporters, consumer non-durable goods

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.

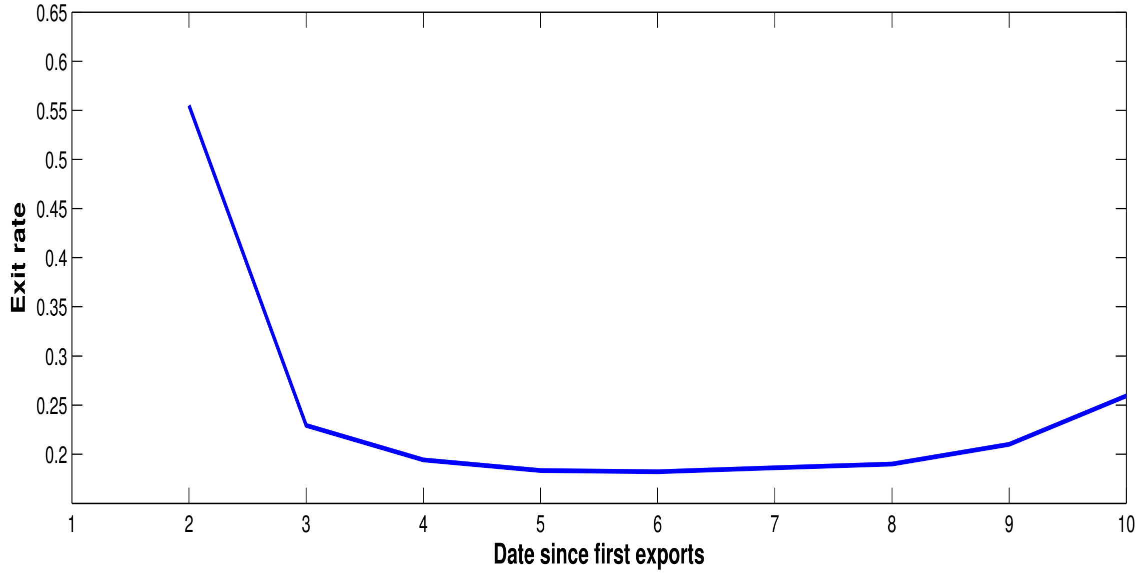

Figure 2: Exit Rate (8-digit product), of all new exporters, consumer non-durable goods

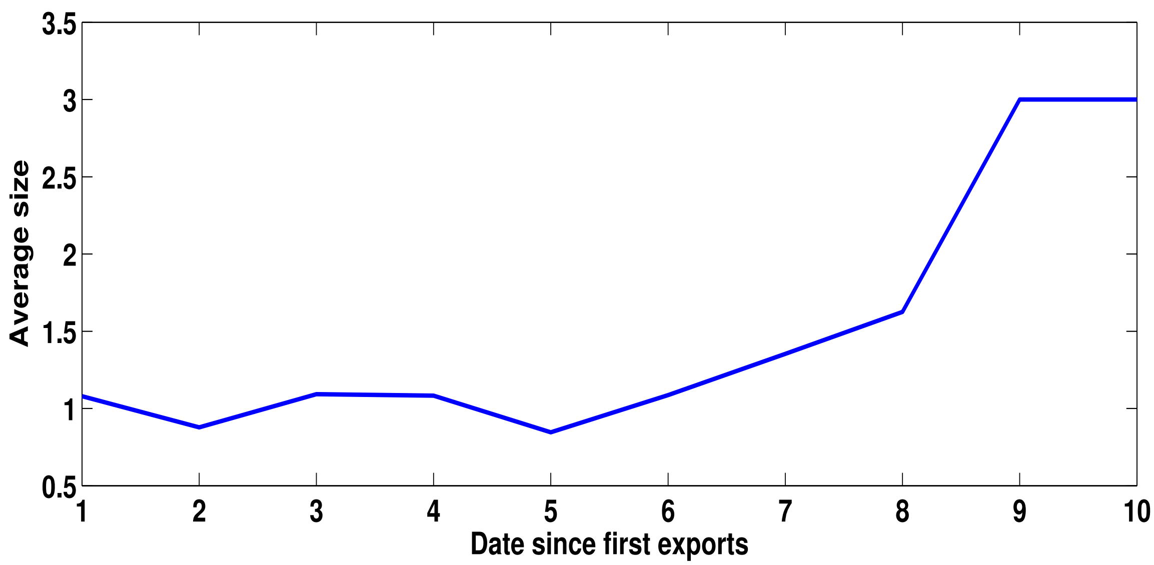

Figure 3: Example of average size evolution, Bulgaria, Beauty or make-up preparations

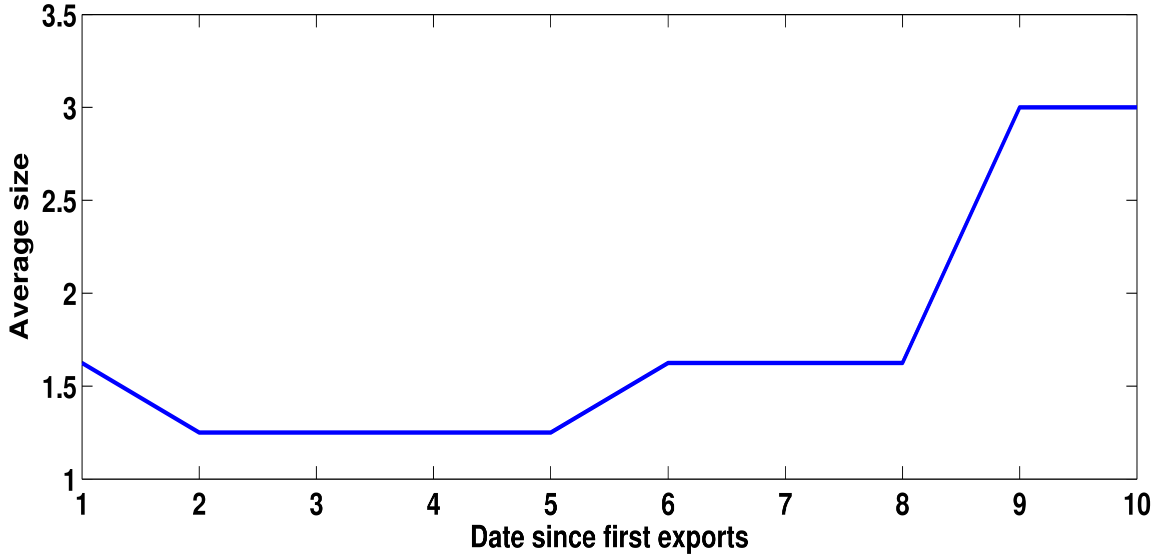

Figure 4: Example of average size evolution,conditional on Survival for at least 8 years, Bulgaria, Beauty or make-up preparations

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^2_N, 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.

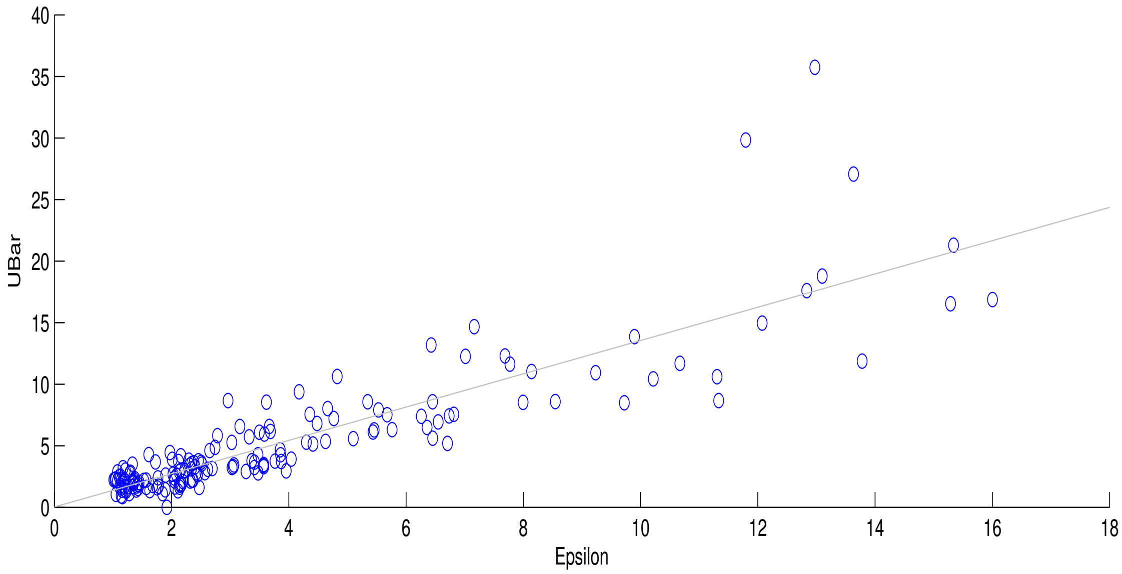

Figure 5: Correlation between ϵ and $\bar{\mu}$

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

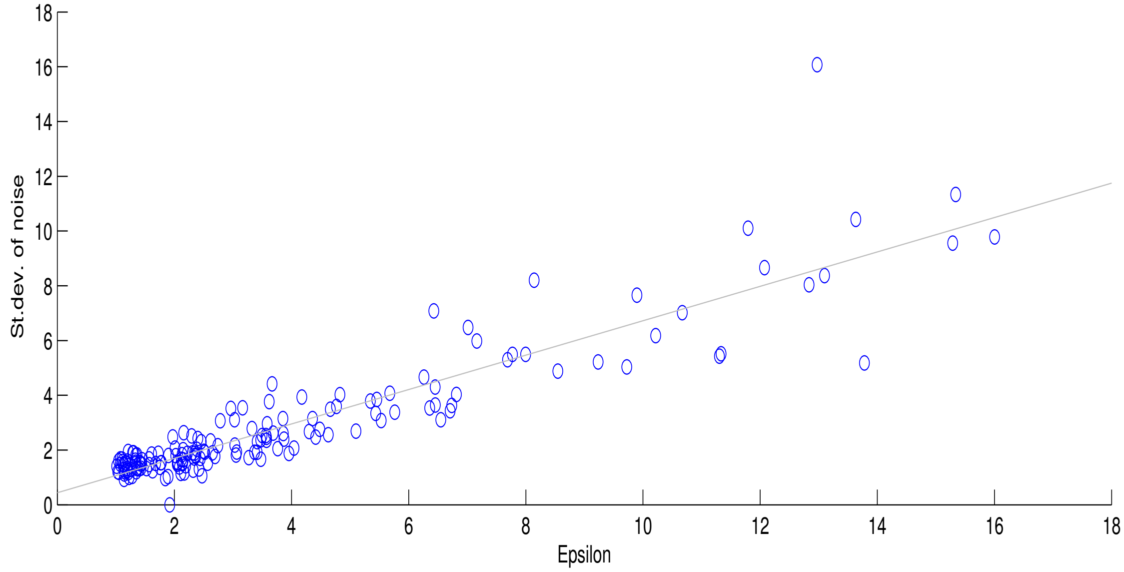

Figure 6: Correlation between ϵ and standard deviation of noise

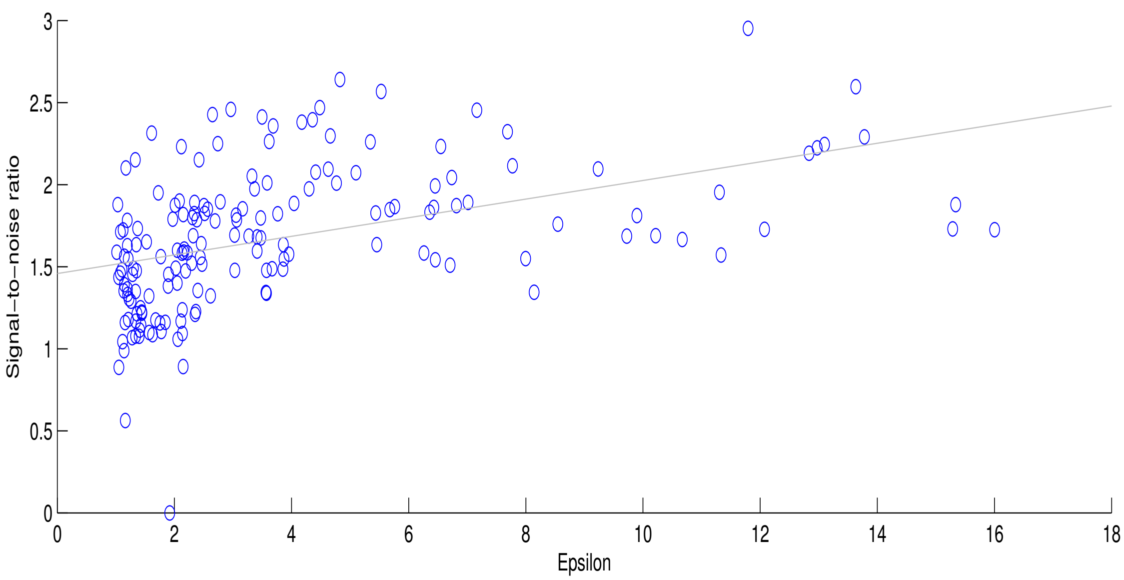

Figure 7: Correlation between ϵ and signal-to-noise ratio

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.

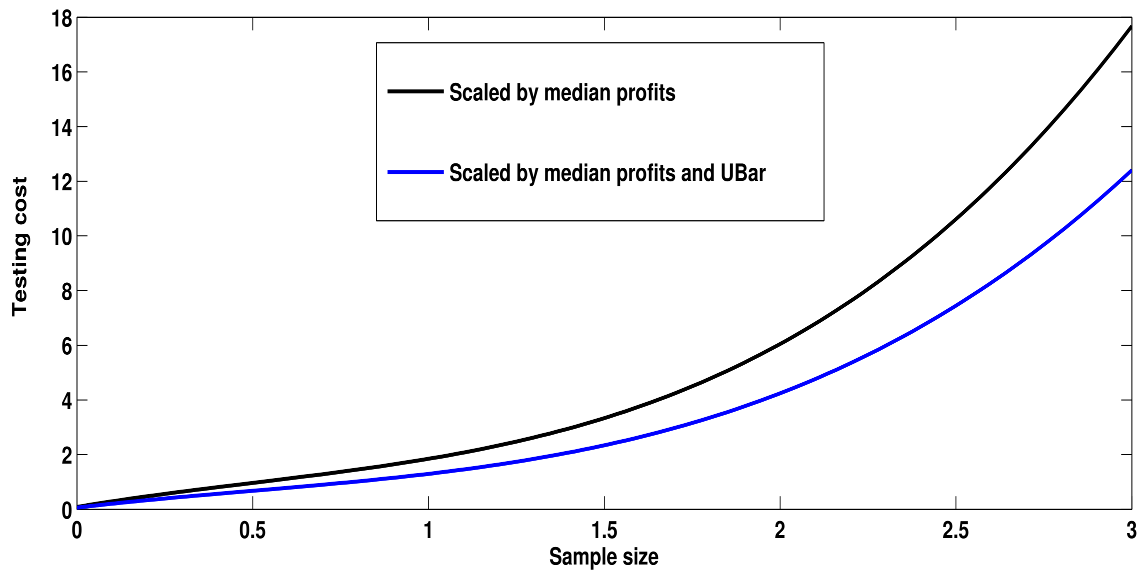

Figure 8: Testing cost schedule, Australia, Women's jackets, etc.

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

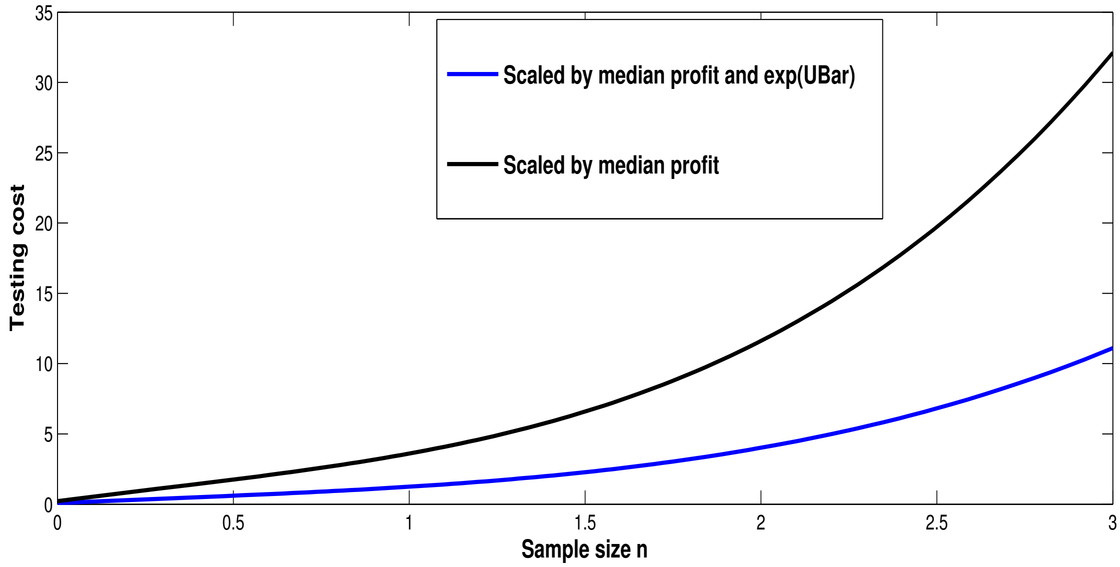

Figure 9: Testing cost schedule, Estonia, Trunks, cases, etc

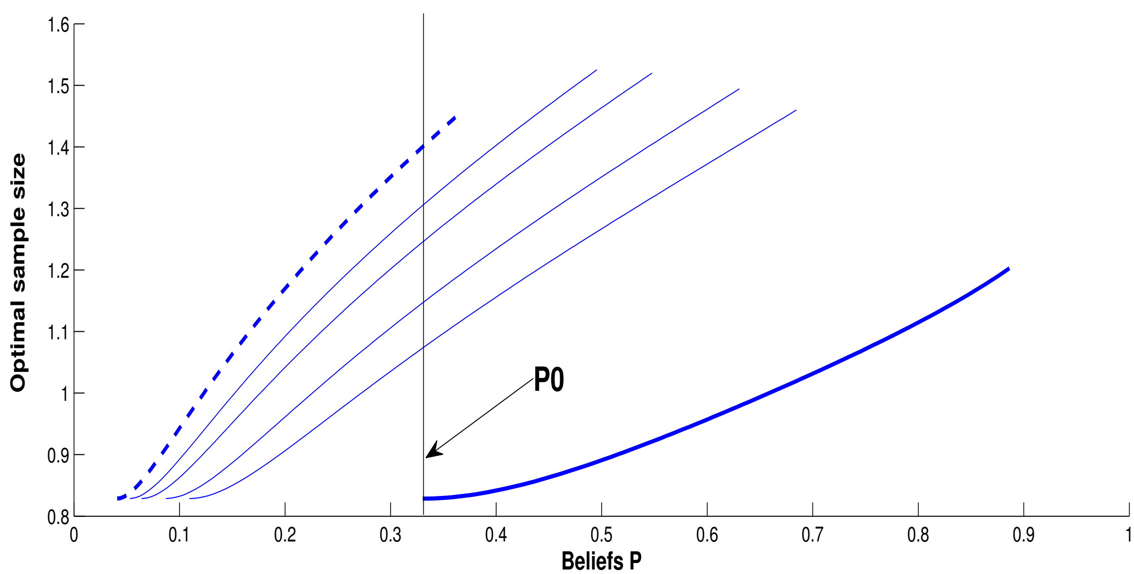

Figure 10: Optimal sample size schedules for experimenter ranges of productivity, from lowest (bold) to highest (dashed), Australia, Women's jackets, etc.

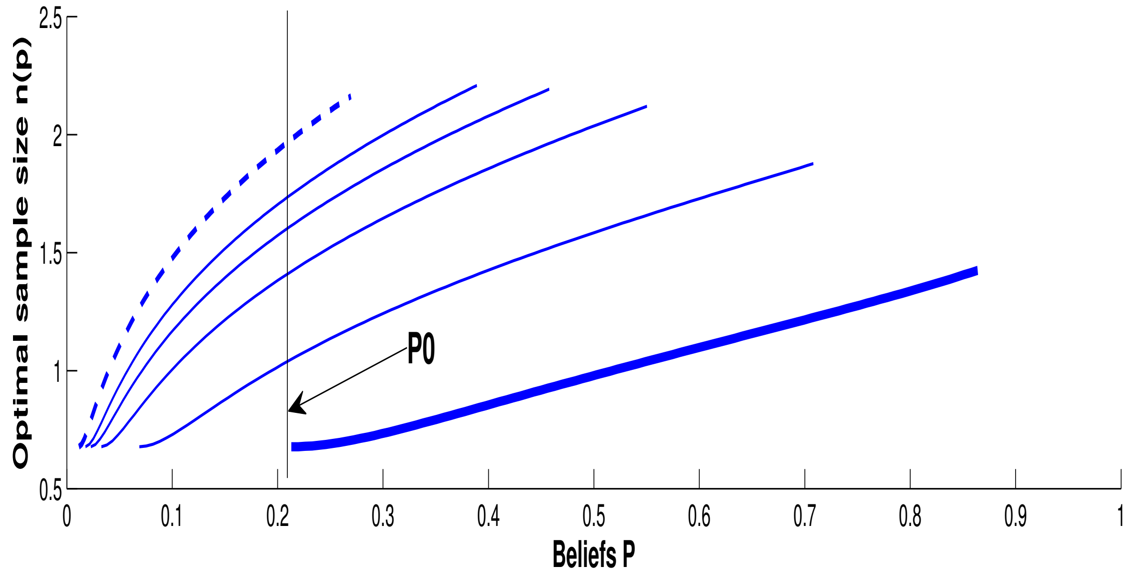

Figure 11: Optimal sample size schedules for experimenter ranges of productivity, from lowest (bold) to highest (dashed), Estonia, Trunks, cases, etc.

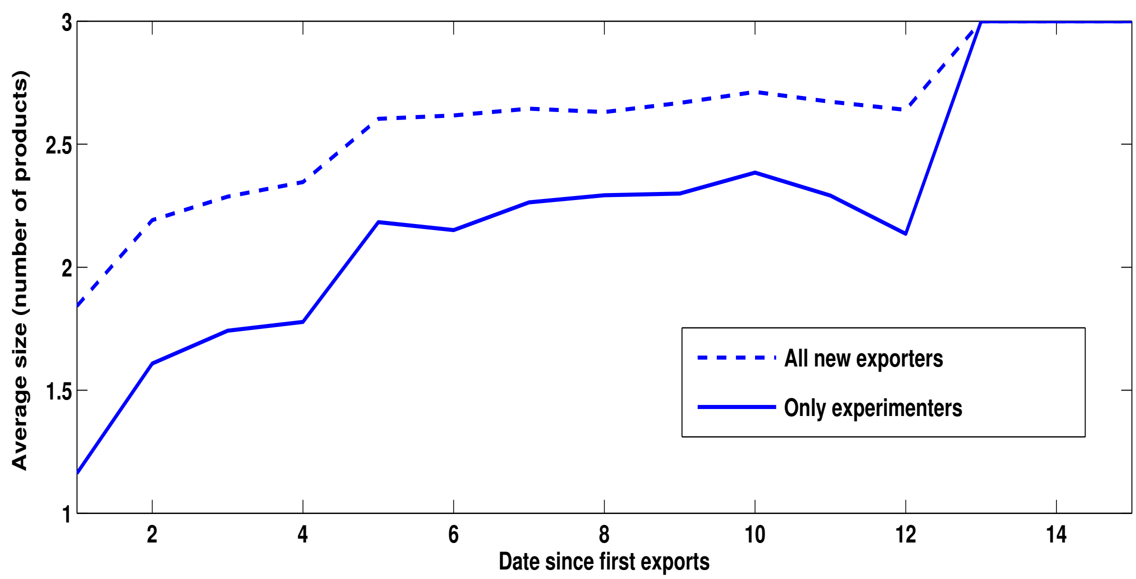

Figure 12: Simulated average size, over 100 exporters, Australia, Women's jackets, etc.

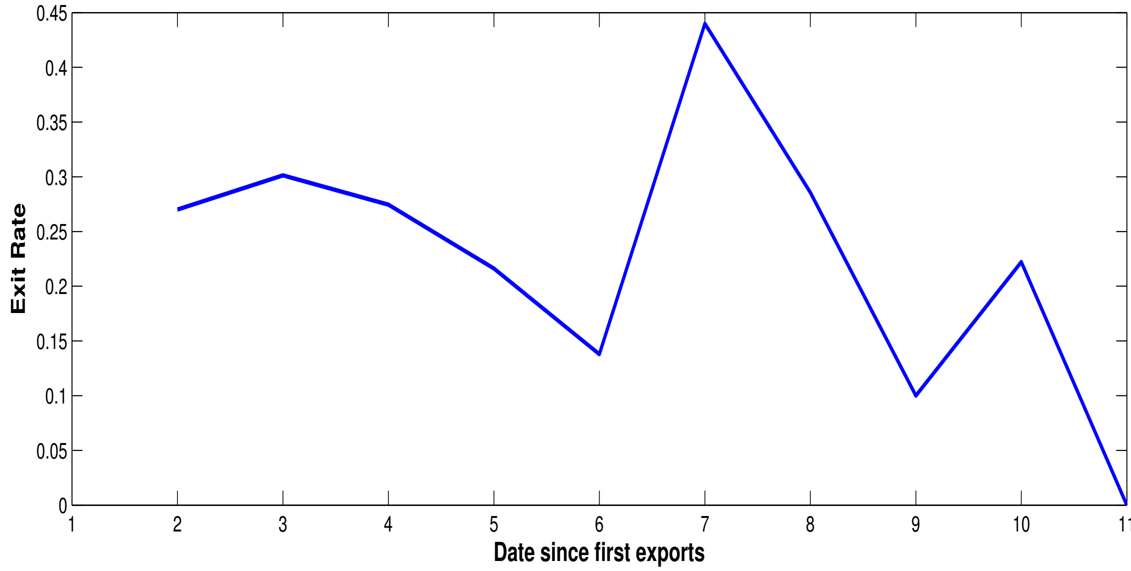

Figure 13: Simulated exit rate, over 100 exporters, Australia, Women's jackets, etc.

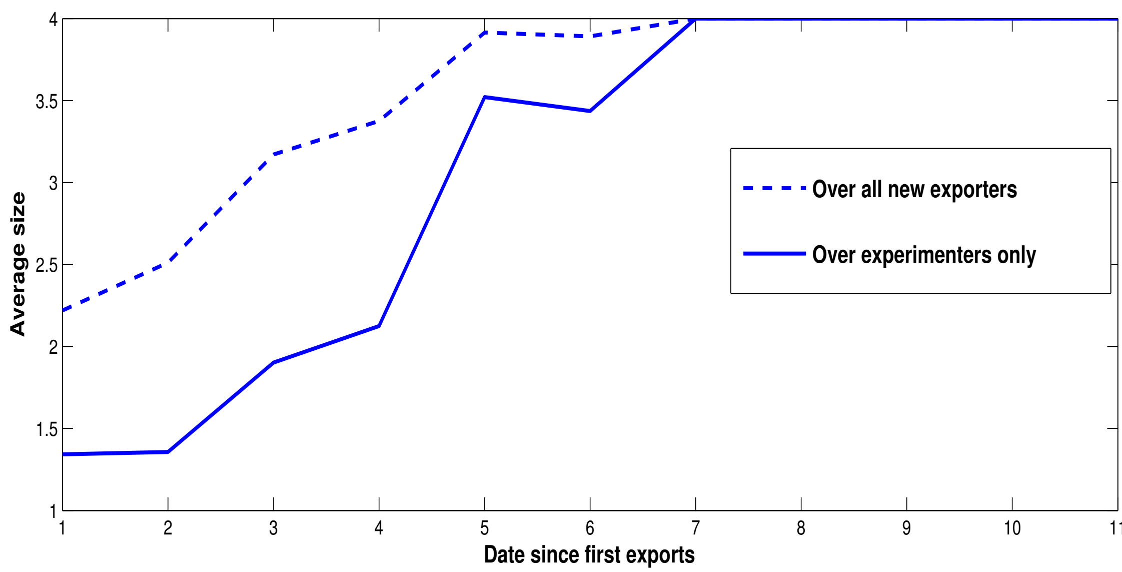

Figure 14: Simulated average size, over 100 exporters, for Estonia, Trunks, cases, etc.

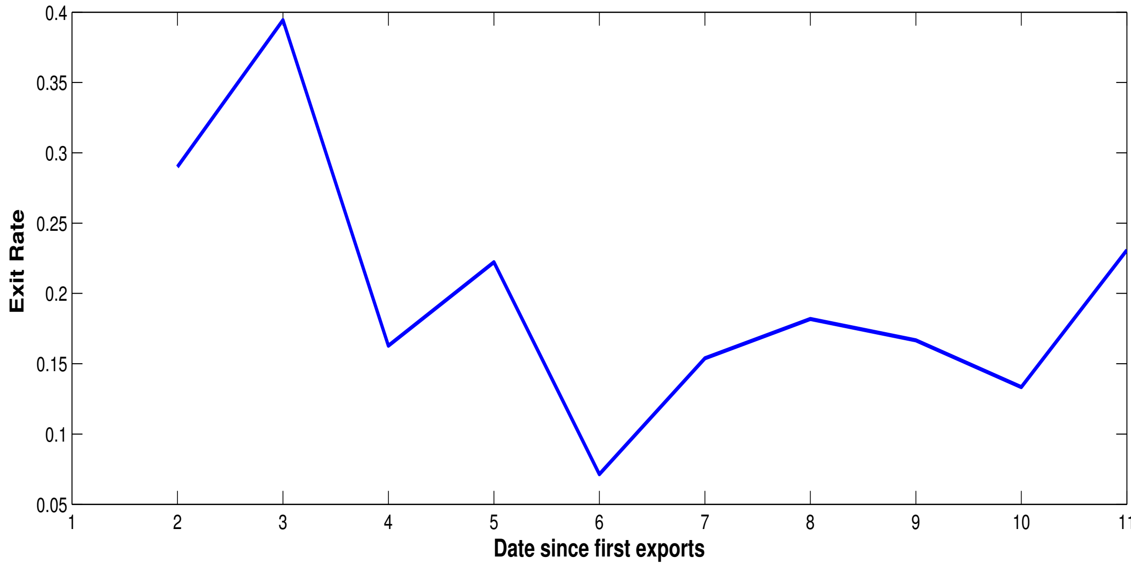

Figure 15: Simulated exit rate, over 100 exporters, Estonia, Trunks, cases, etc.

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

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

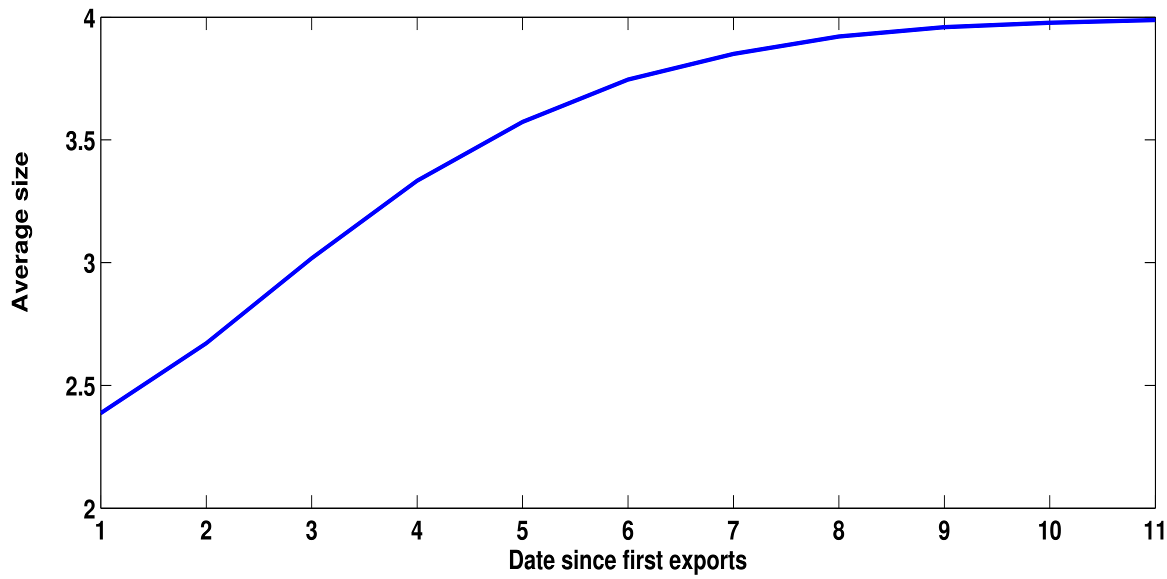

Figure 16: Smoothed out simulated average size, Estonia, Trunks, cases, etc.

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

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.

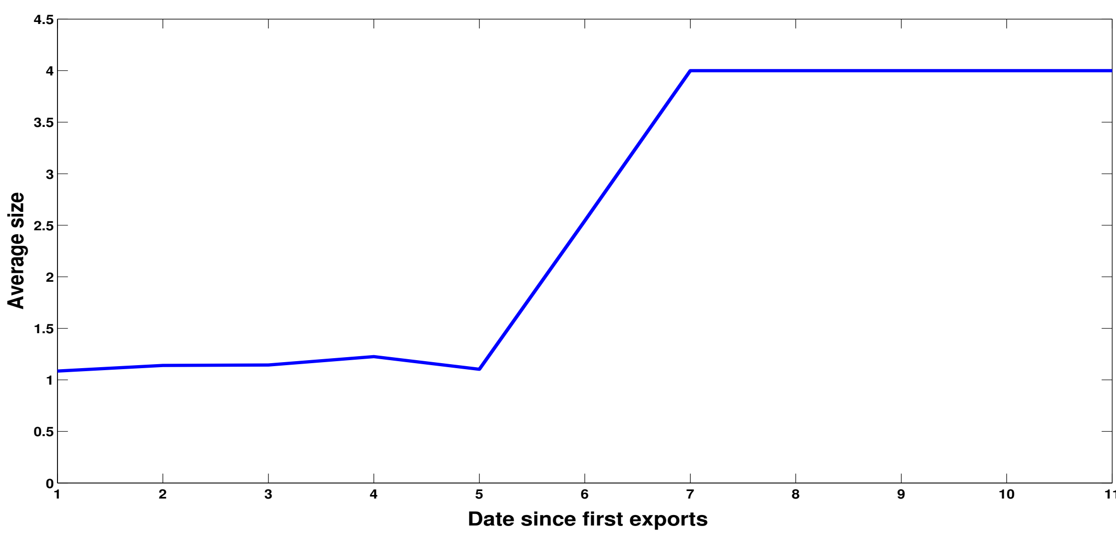

Figure 17: Response to tariffs cuts in Estonia, Trunks, cases, etc.

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

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

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

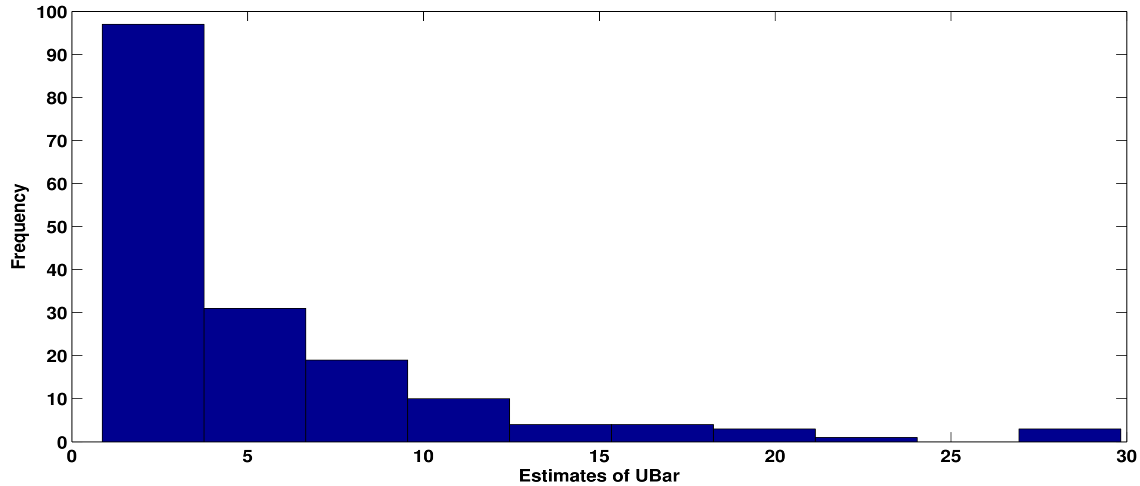

Figure 18: Summary of $\bar{\mu}$

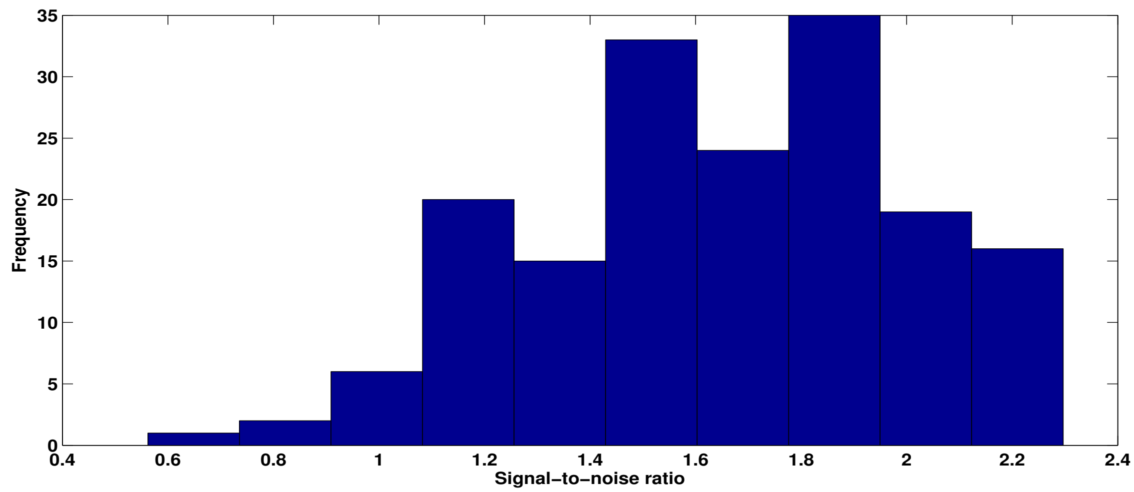

Figure 19: Summary of the signal-to-noise ratio

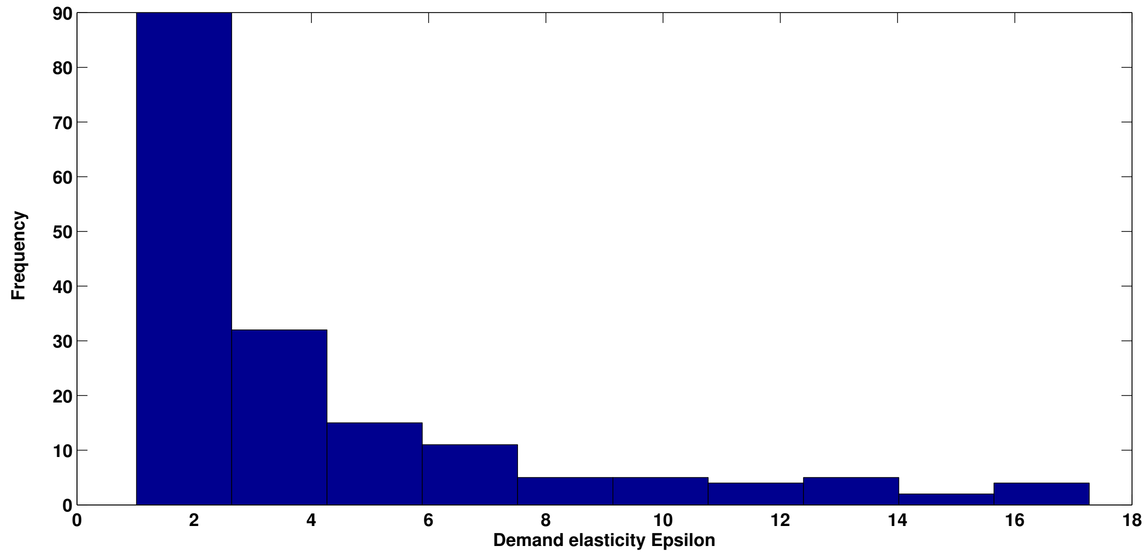

Figure 20: Summary of elasticities

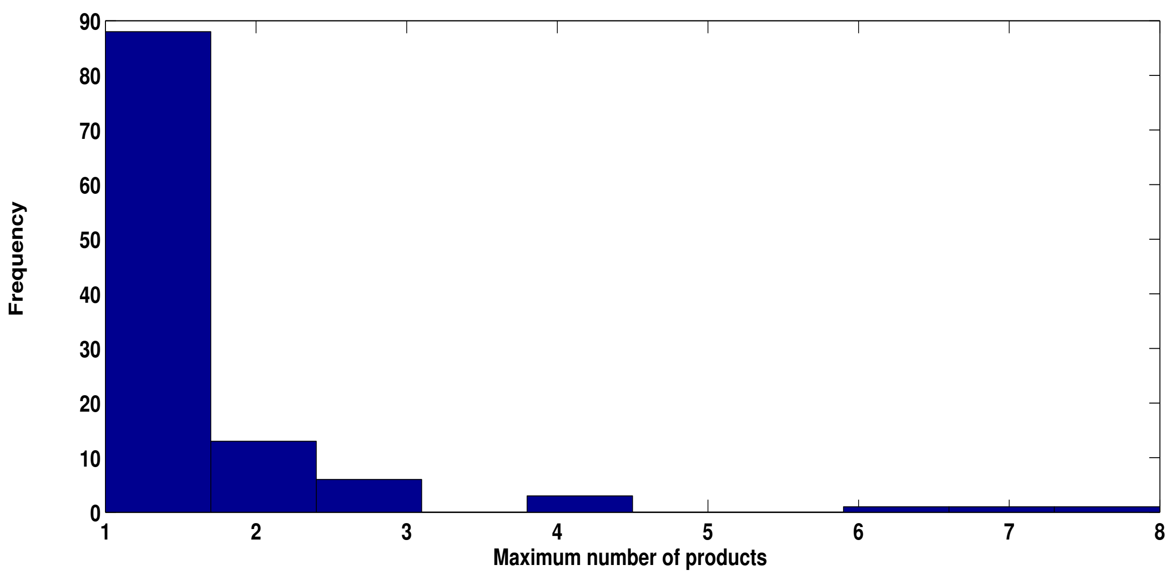

Figure 21: Histogram of maximum number of products, new exporters, Australia, women's dress, etc.

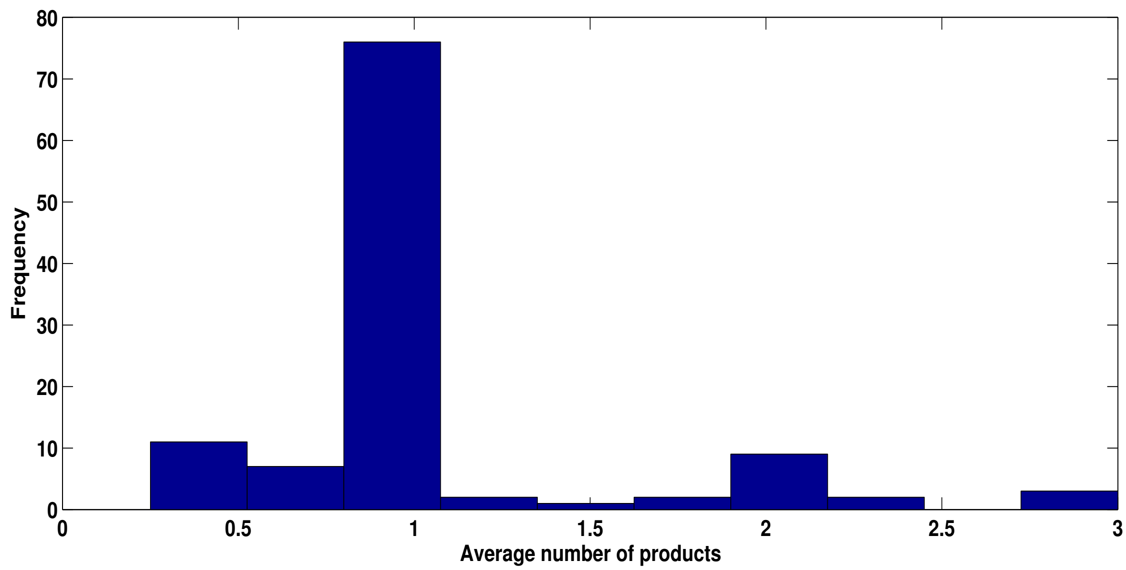

Figure 22: Histogram of average number of products, new exporters, Australia, women's dress, etc.

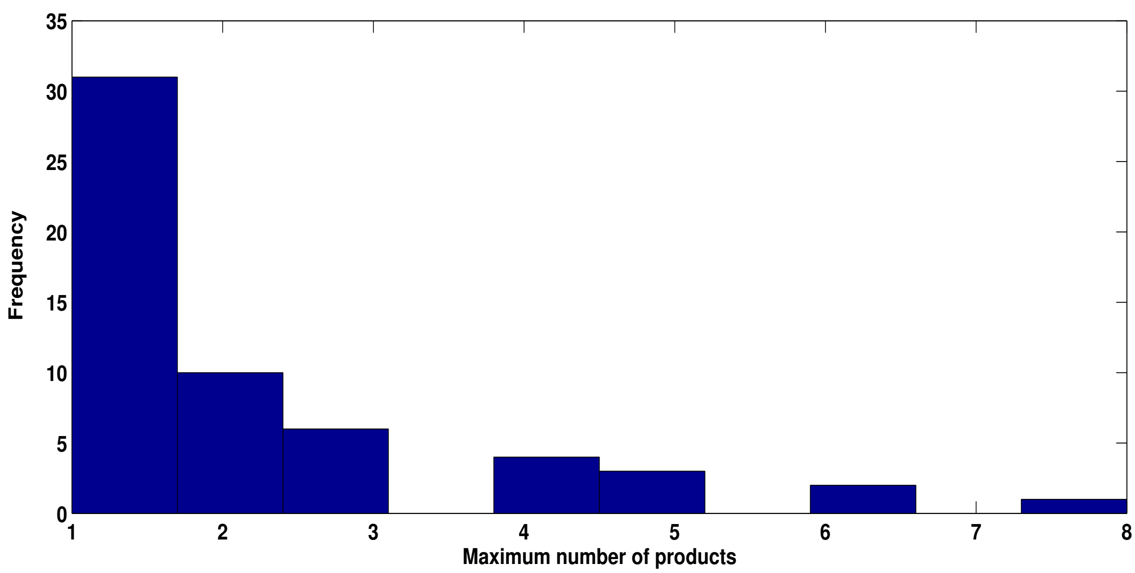

Figure 23: Histogram of maximum number of products, new exporters, Estonia, Trunks, cases, etc.

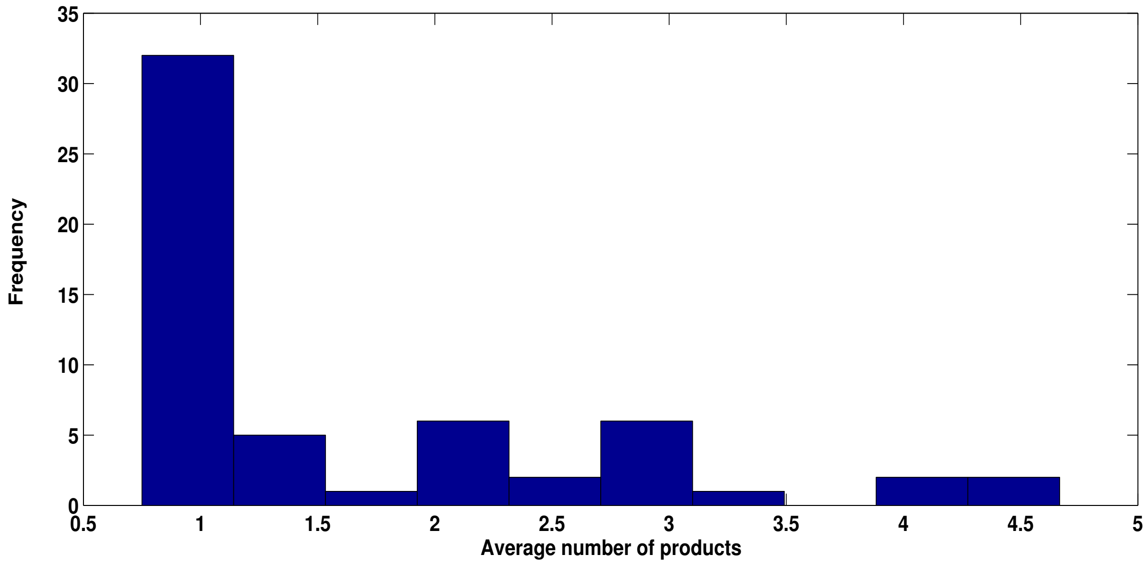

Figure 24: Histogram of average number of products, new exporters, Estonia, Trunks, cases, etc.

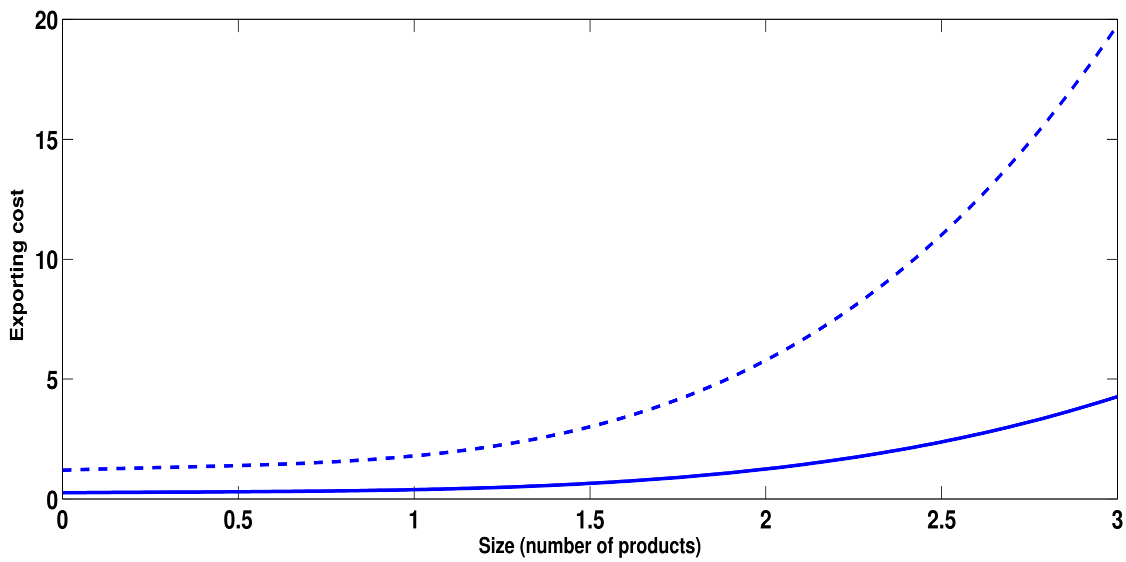

Figure 25: Exporting cost for the Passive Learning model, Estonia, Trunks, cases, etc.

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

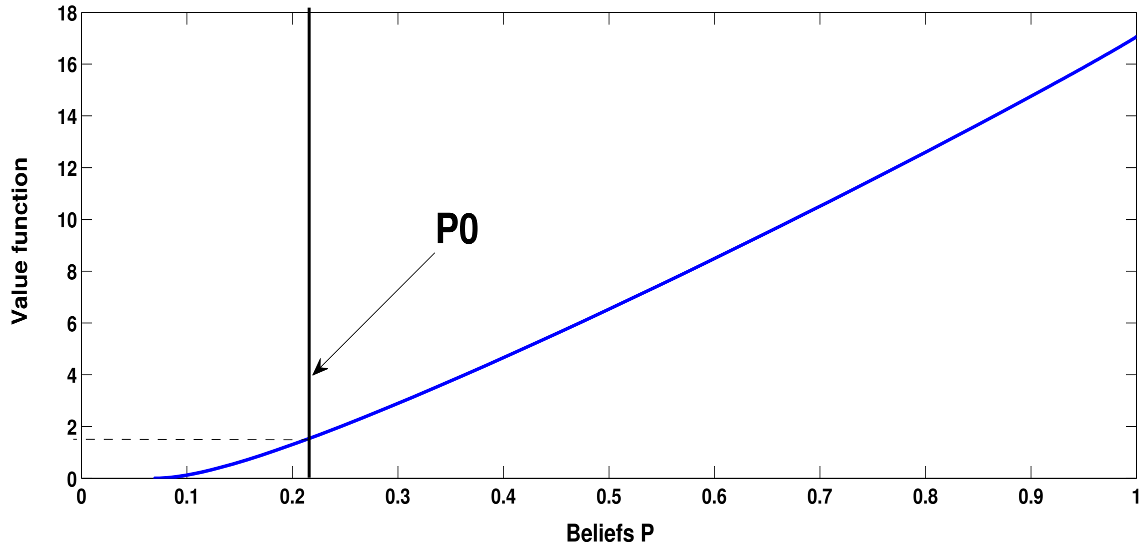

Figure 26: The value function for the lowest productivity exporter, and the estimate of sunk cost F, Estonia, Trunks, cases, etc.

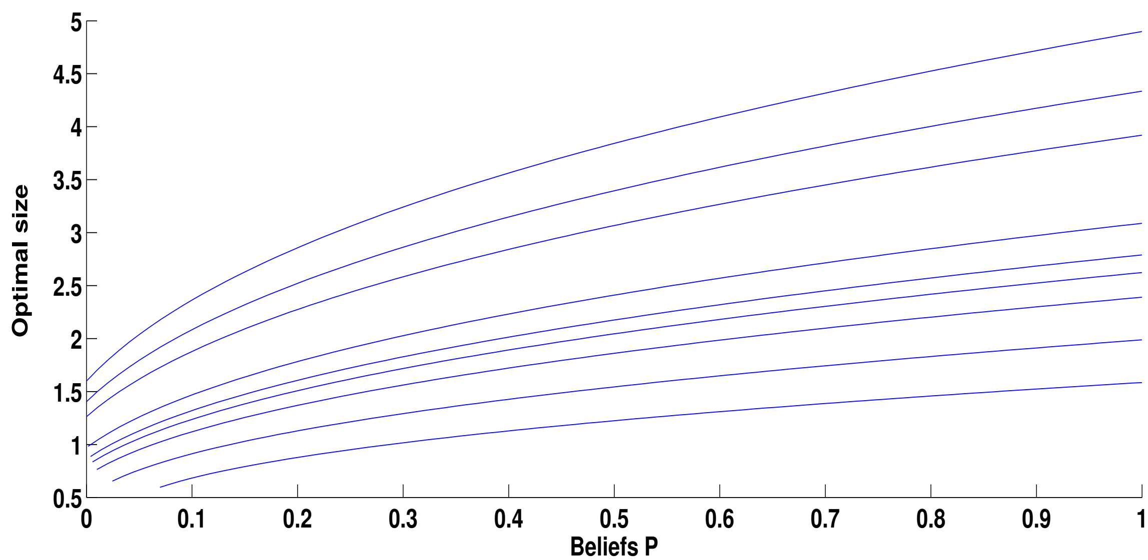

Figure 27: Optimal size schedules, by ranges of profitability, Passive Learning with convex costs, Estonia, Trunks, cases, etc.

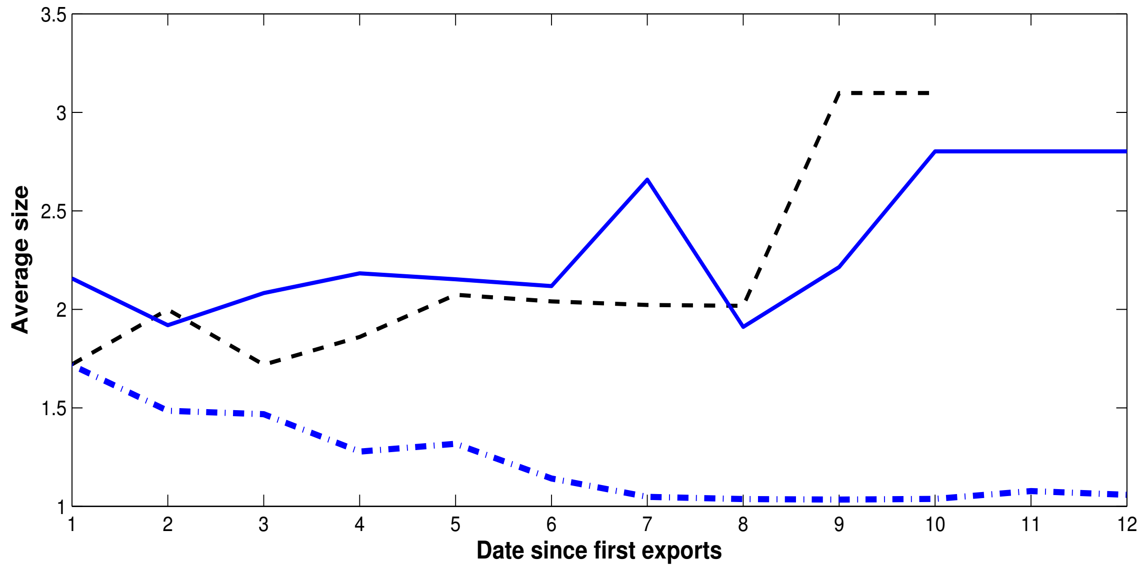

Figure 28: Simulated Average Size Dynamics, 3 Examples, Passive Learning, Estonia, Trunks, cases, etc

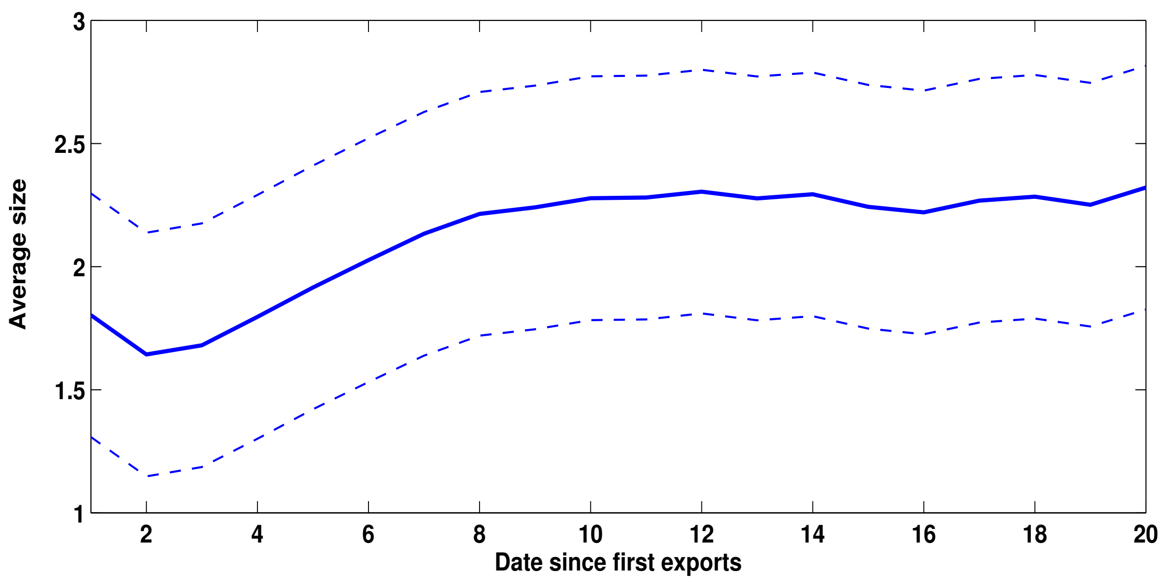

Figure 29: Smoothed out (over 500 simulations) average size in Passive Learning, Estonia, Trunks, cases, etc.

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

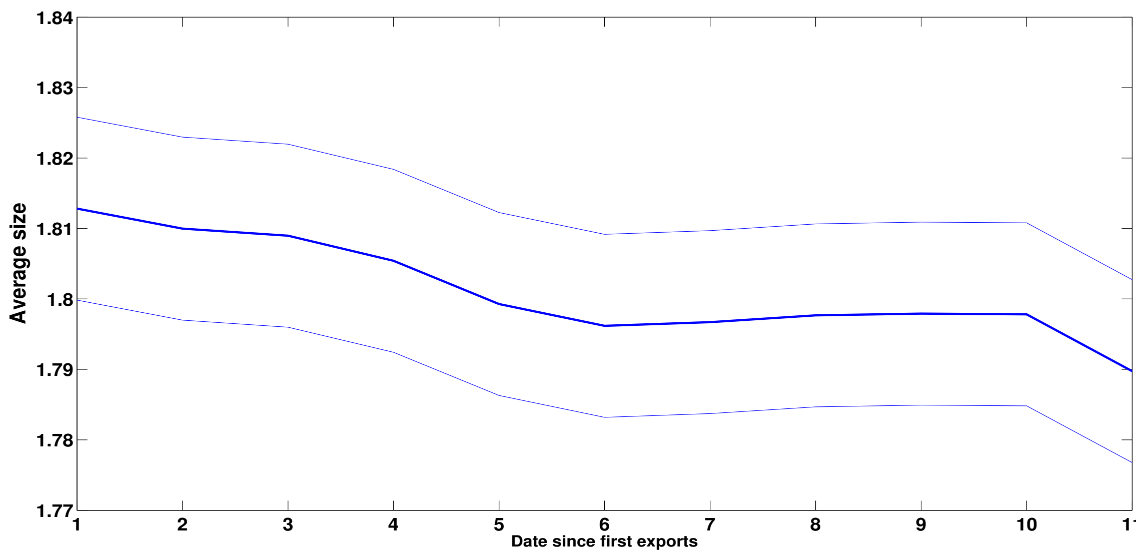

Figure 30: Smoothed out (over 500 simulations) average size in Melitz model, Estonia, Trunks, cases, etc.

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