CEPII

CENTRE
D'ÉTUDES PROSPECTIVES
ET D'INFORMATIONS
INTERNATIONALES

No 2013 – 10
Avril

DOCUMENT DE TRAVAIL

A Model of Firm Experimentation under Demand Uncertainty: an Application to Multi-Destination Exporters

Zhanar Akhmetova
Cristina Mitaritonna

TABLE OF CONTENTS

A MODEL OF FIRM EXPERIMENTATION UNDER DEMAND UNCERTAINTY: AN APPLICATION TO MULTI-DESTINATION EXPORTERS

Zhanar Akhmetova
Cristina Mitaritonna

HIGHLIGHTS

- Demand uncertainty about the profitability of exporting in a new market modifies the firm theory: the ability to test a market before full-scale entry makes the decision for the firm an optimal control problem.
- This new model is estimated (Bayesian technics) using French firm-level export data.
- The estimated sunk cost is higher than in Melitz (2003).
- On the policy side, cutting testing costs is better than cutting sunk entry cost to promote exports.

ABSTRACT

Firm level data exhibits that new exporters tend to start small, a large fraction of these drops out by the second year of exporting, and the survivors expand rapidly. To take into account this stylized fact, we propose a theory of firm behavior that assumes demand uncertainty about the profitability of exporting in a new market. The firm can postpone paying the sunk cost of full-scale entry and test the market by observing individual sales to a few consumers. The firm optimally chooses the experimentation intensity, as well as the exit/entry policy. Applying Bayesian econometric techniques, we structurally estimate the model using French firm-level export data. A given geographical regions is viewed as a target market, and countries within the region as consumers. The estimate of the sunk cost is higher than in a model where the sunk cost cannot be postponed, like Melitz (2003). We also perform counterfactual simulations (exchange rate, sunk cost and experimentation cost).

JEL Classification: D21, D83, F14, C11, C33.

Keywords: Demand Uncertainty, Optimal Experimentation, Heterogeneous Producers, New Exporter Dynamics, Structural Estimations, Bayesian Methods.

A MODEL OF FIRM EXPERIMENTATION UNDER DEMAND UNCERTAINTY: AN APPLICATION TO MULTI-DESTINATION EXPORTERS

Zhanar Akhmetova
Cristina Mitaritonna

POINTS CLEFS

- L'incertitude sur la taille du marchés d'exportations modifie la théorie du comportement des firmes: celle-ci pouvant expérimenter à petite échelle le marché avant de payer le coût fixe de la chaîne logistique et marketing, la décision d'exporter devient un choix optimal d'expérimentation.
- Ce nouveau modèle est estimé (technique bayésienne) sur données de firmes françaises.
- La prise en compte de l'expérimentation conduit à une estimation du coût d'entrée plus élevé que Melitz (2003).
- Pour accroître le volume d'exportations, les politiques réduisant le coût d'expérimentation sont plus efficaces que celles réduisant le coût d'entrée.

RÉSUMÉ

Les données de firmes à l'exportation montrent que les nouveaux entrants sur un marché extérieur commencent modestement, qu'une grande proportion de firmes cessent dès la deuxième année et que celles qui persistent accroissent rapidement leurs volumes exportés. Pour prendre en compte ces faits stylisés, nous proposons une nouvelle modélisation des décisions d'une firme faisant face à une incertitude sur le potentiel de marché. Une firme peut décider d'expérimenter à petite échelle un marché et décider sur cette base si elle paye le coût fixe d'une entrée à grande échelle. Elle décide de façon optimale l'intensité de l'expérimentation ainsi que sa durée puis seulement si elle rentre pour de bon sur le marché. Nous précédons à une estimation bayésienne de ce nouveau modèle sur données de firmes françaises : une région du monde est considérée comme un marché cible, les pays de la région en étant des consommateurs. Les coûts d'entrée sont plus importants lorsque le comportement d'apprentissage est pris en compte dans l'estimation que dans le modèle de Melitz (2003) ou un modèle de *passive learning* avec coût convexe. Nous simulons des changements structurels de l'économie (taux de change, coût d'expérimentation, coût d'entrée).

Classification JEL : D21, D83, F14, C11, C33.

Mots clés : Incertitude sur la demande, Expérimentation optimale, Producteurs hétérogènes, Dynamique de nouveaux exportateurs, Estimations structurelles, Méthodes bayésiennes

A MODEL OF FIRM EXPERIMENTATION UNDER DEMAND UNCERTAINTY: AN APPLICATION TO MULTI-DESTINATION EXPORTERS[1]

Zhanar Akhmetova *
Cristina Mitaritonna †

1. INTRODUCTION

Recently available firm-level export data has provided crucial insights into the dynamic behavior of individual exporters. In particular, we observe that new exporters in a given market exhibit patterns that cannot be explained by standard models. As documented by Eaton, Eslava, Krizan, Kugler and Tybout (2007) and Ruhl and Willis (2009) for Colombian firms, new exporters tend to start small, a large fraction of these drop out by the second year of exporting, and the survivors expand rapidly. This is demonstrated in Figure 1 for French exporters. On the horizontal axis we measure the date since first exports by a firm of an individual 8-digit-code product to an individual country, where we treat as new exporters all firms who did not export in 1995, the first year in our dataset, but did so later. On the vertical axis in the first panel we measure the quantity of the 8-digit-code product, scaled down by the firm's quantity exported in the first year, and averaged over all firms exporting at that date, conditional on surviving in the market for at least 7 years.[2] The general upward trend in these normalized quantities reveals an expansion in firm exports relative to initial volumes, as the firm exporting age increases. In the second panel, we show the high exit rate of firms in the second year of exporting, and a drop in exit rate afterwards.

We can also examine these dynamics by studying the number of countries that firms export to. Exporting to one destination can provide information about demand in other, similar, destinations. Consider Figure 2, panel 1, where we follow the exports by French firms of beauty or make-up preparations (4-digit code 3304) to the EU (the first fifteen members, excluding France) between 1995 and 2005. The horizontal axis measures the date since first exports to

[1]We thank Marc Melitz, Pinelopi Goldberg, Stephen Redding, Gene Grossman, Esteban Rossi-Hansberg, Oleg Itskhoki, Jan De Loecker, Christopher Sims, Bo Honore, Lionel Fontagne for comments and suggestions, as well as seminar participants at Princeton University, CEPII, UNSW, among others, and the organizers and conference participants of Forum for Research on Empirical International Trade (F.R.E.I.T.) in Slovenia and Japan. This research was financially supported with grants and fellowships from Princeton University, CEPII, and UNSW. Authors are responsible for possible errors and omissions.

*Department of Economics, Australian School of Business. z.akhmetova@unsw.edu.au.
†Centre d'Etudes Prospectives et d'Informations Internationales (CEPII). cristina.mitaritonna@cepii.fr

[2]This way, we make sure that the expansion in the average is not purely due to the selection effect - less productive and smaller firms exiting in the first few years.

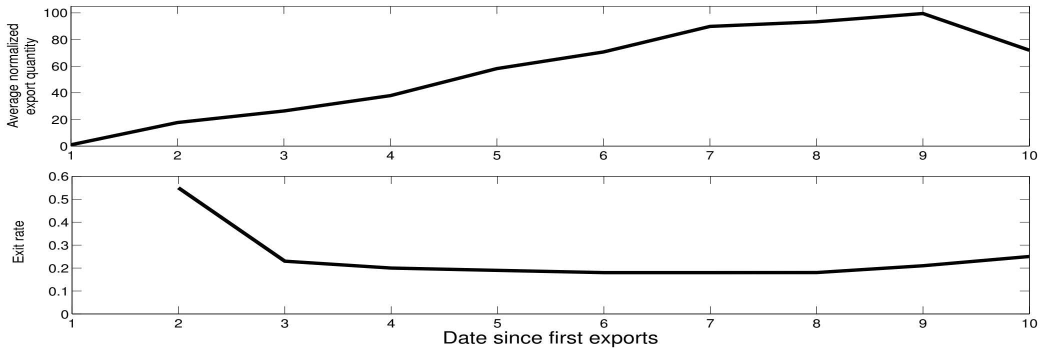

Figure 1 – First panel: ratio of export quantity (by 8-digit-code-product-country pair) to the initial export quantity of the firm, averaged over all new exporters, that stayed in the market for at least 7 years. Second panel: exit rate of new exporters. Consumer non-durable goods.

this region, and the vertical axis - the number of export destinations within this region, averaged over all firms exporting at that date.[3] In panel 2, we study the exports of footwear with outer soles of rubber, plastics, leather or composition leather and uppers of leather (4-digit code 6403) to the region of Eastern Europe, broadly defined. In both graphs, the average grows over time and reaches a stable level by date 8.

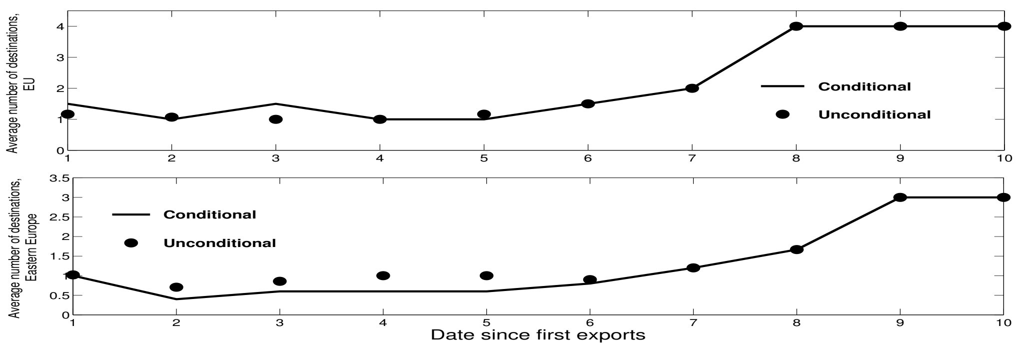

Figure 2 – First panel: exports by French firms of beauty or make-up preparations to the first fifteen members of the EU, excluding France. Second panel: exports by French firms of footwear with outer soles of rubber, plastics, leather or composition leather and uppers of leather to Eastern Europe.

This empirical evidence suggests that firms start small in a foreign market when they are unsure of their profitability there, in order to collect more information about the market. The firm then either expands or quits the market, depending on the observed export performance. This would explain the high initial exit rate of new exporters and its decrease in later years, and the gradual

[3]The dots show the average over all firms exporting at that date, while the solid line tracks the average over all exporters that survived for at least 7 years in the market.

expansion in exports of successful new exporters. Moreover, Figure 2 indicates that there may be an optimal scale that firms switch to, once they finish learning and decide to stay in the market.

We propose a theory of firm behavior that assumes demand uncertainty in a foreign market, where the firm is uncertain about a demand shift parameter that affects its profitability. There is a large sunk entry cost one needs to incur to access the entire market (Melitz (2003); Bernard, Jensen, Redding and Schott (2007); Das, Roberts and Tybout (2007)). The firm can postpone its full-scale entry and instead access a few consumers in the foreign market at a certain variable cost, called testing cost. The sales to individual consumers serve as noisy signals about the demand parameter, and the firm uses this information to update its beliefs about demand. Based on these beliefs, the firm decides whether it should incur the sunk entry cost to access the entire market (full-scale entry, or simply entry in what follows), quit the market, or keep experimenting, and if so, how many consumers to access. 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.[4] A consumer can be viewed as an individual, a household, a retail store, a city, a state/province, or a country.[5] We will apply our model to a setting where the firm views a geographical region as a target market, and countries within the region - as individual consumers. We carry out a structural estimation of the model, using French firm-level export data and applying Bayesian techniques. These estimates can be used to evaluate the duration of experimentation and total entry costs, i.e. the sum of testing costs and sunk entry costs, incurred by new exporters. We carry out simulation exercises, that highlight the new dynamics generated in this model. If currency appreciates temporarily in only one country within a region, exports to other countries within the same region will also increase, unlike in a standard model. Cutting testing costs is not equivalent to cutting the sunk entry cost in terms of the effect on export volumes. Lower testing costs will lead to better selection of high-demand producers into full-scale exporting, and higher total export volumes. Thus, cutting testing costs is a better way of achieving higher export volumes, than cutting the sunk entry cost, ceteris paribus.

In general, recognizing the experimentation stage is important for the correct prediction of the effects of policies aimed at increasing exports. Any shock will not have an immediate impact in

[4]We need to assume that the discounted cost of accessing the entire market using the testing technology forever is higher than the sunk cost of entry. This assumption is plausible, since temporary, short-term arrangements with distributors, shipping and marketing agencies are likely to be more expensive on average than long-term arrangements or in-house services dedicated to the exporting activity. This is supported by Klompmaker, Hughes and Haley (1976), who mention various direct and indirect costs of test marketing, and in particular, 'higher media expenses because of low volumes, ..., and higher trade allowances to obtain distribution'.

[5]For example, Heineken test marketed Heineken Premium Light in Phoenix, Dallas, Providence, and Tampa in 2005, before launching it nationally in 2006. US companies often start exporting in the maquiladora area in Mexico and later expand to the rest of Mexico, if successful. Austria serves as a testing ground for cell phone companies looking to export to German-speaking countries.

the form of full-scale exports by new exporters. Instead, a transition period during which some firms will learn about demand should be expected. The duration and intensity 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. Our model helps evaluate these costs, using structural estimates.

We also show that the estimate of the sunk entry cost will be higher when fitting our model to the data than when fitting an alternative model (e.g. Melitz (2003)), where the sunk entry cost cannot be postponed. Suppose we have a dataset of firms, such that the firms with productivity below some level are not exporting and the rest are exporting. The usual way one infers the magnitude of the sunk entry cost from such a dataset is locating the lowest productivity exporter and calculating the value from exporting for this firm. This value should then be exactly equal to the sunk entry cost, so that this firm is indifferent between exporting and not exporting. In our case, the lowest productivity exporter has not incurred the sunk entry cost and is instead only testing the market. We know for sure that their expected value from exporting is lower than the sunk entry cost. Thus, the standard estimate of the sunk entry cost will be too low. This is proven rigorously in what follows.

1.1. Related Literature

Test marketing is an activity well-documented in the marketing literature, but there are few formal models. Hitsch (2006) builds a model of test marketing, where the firm launching a new product and uncertain about its demand has to choose an optimal exit and advertising policy. In that model, advertising is positively related to expected demand (through expected higher profits), but the level of advertising does not affect the precision or speed of learning.

There is a burgeoning literature on the issues of new exporters, market penetration and learning under demand uncertainty. In Arkolakis (2010), the firm may access a few consumers, rather than the entire market, through a marketing technology, so that the cost of entry is no longer 'fixed', but is endogenously determined. We borrow the assumption of diminishing returns to marketing from his paper. Several works have addressed the issue of firm learning about profitability in a foreign market. Horstmann and Markusen (1996) study the choice by a multinational firm to service a new foreign market through direct investment or contracting with a local sales agent, when facing uncertainty in demand. In Rauch and Watson (2003), a firm decides whether to learn about the quality of a supplier in a new destination by placing a small order or to invest in a big order right away. In this multi-period model, the firm can choose the (costly) probability with which supplier's quality will be perfectly revealed in the next period. In Eaton, Eslava, Krizan, Kugler and Tybout (2008), there is uncertainty about the foreign demand for a firm's product. In each period, a firm chooses search intensity, which determines the probability with which a new encounter with a buyer will occur. When the encounter takes place, it is either a success or a failure, and the firm updates its beliefs about the probability of success in the foreign market. Thus, similarly to Rauch and Watson (2003), the firm chooses the probability with which the signal arrives, but not the precision of this signal. In Freund and Pierola (2010), the firm is uncertain about the periodic overhead cost of exporting, which it can learn perfectly once it exports. There is a sunk 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 predetermined fraction of the total sales) for a smaller cost (a predetermined fraction of the actual sunk cost), to learn the periodic overhead exporting cost and make the ultimate decision to stay or not. Segura-Cayuela and Vilarrubia (2008) provide a model of informational spillovers, where in equilibrium the most productive firms pay the sunk entry cost and enter the market first, learn the value of periodic fixed exporting costs there, and through the decision to stay in the market or exit indirectly help other firms update their beliefs about exporting costs.

Several papers deal with exports by firms to multiple destinations and provide support to the empirical application in this paper. In the two-period model of Albornoz, Pardo, Corcos and Ornelas (2012), the firm can export to two destinations, in each of which it faces a sunk cost of entry and uncertainty in demand, as well as unit trade costs. Once the firm exports to a market, it learns its demand there immediately and with complete precision. Due to correlation in demand and differences in unit trade costs across destinations, optimal sequential entry strategy is generated, where the firm enters one market first, and enters the other market only if successful in the former. Nguyen (2012) also introduces demand uncertainty, in domestic and foreign markets, and imperfect positive correlation among the demand parameters across markets. The firm can learn about demand in a destination with complete precision once it sells there for at least one period. There are periodic fixed costs of selling, so the firm will exit a market, if the demand there is too low to cover that cost. While differences in trade costs and market sizes can produce sequential entry patterns, as in Albornoz, Pardo, Corcos and Ornelas (2012), delayed exporting to some destinations can be optimal even with symmetric countries.

Our model has a few distinct features. There is a sunk cost of entry that can be postponed due to the availability of a testing technology. This technology allows the firm to observe sales to individual consumers with some measurement error. The firm updates its beliefs based on the average of these observations. The higher the sample size, the lower the variance of the sample average, and the higher the speed of learning. The firm realizes the effect the number of consumers has on the precision of information and sets experimentation intensity (number of consumers accessed) accordingly. We call this active learning, as opposed to passive learning, - the recognition by the firm of the information value of exporting and the ability to speed up the learning process by increasing experimentation intensity. As a result, the firm has to choose both optimal stopping time and experimentation intensity in every period while testing the market. The rest of the paper is organized as follows. Section 2 presents and solves the model. Section 3 describes the structural estimation and discusses the results and simulations. Section 4 concludes. Detailed derivations, proofs and estimation steps, as well as a few graphs, can be found in the Appendix.

2. The Model

2.1. Consumer Preferences

We study the optimal behavior of firms at Home wishing to sell in the Foreign market. There are M consumers in the Foreign market, where $M \in Z^{+}$. For any foreign consumer k, utility from consuming quantities q^k_{jt} at time t is given by

$$U^k_t = \left[\int_j ((e^{\mu_j})^{\frac{1}{\varepsilon}} (q^k_{jt})^{\frac{\varepsilon-1}{\varepsilon}} dj\right]^{\frac{\varepsilon}{\varepsilon-1}},$$

where j denotes varieties, $\varepsilon > 1$ is the elasticity of substitution, and μ_j is the variety-specific demand shift parameter. Utility-maximizing quantity demanded of a variety j is given by

$$q^k_{jt} = Q^k_t e^{\mu_j} \left[\frac{h_{jt}}{P_t}\right]^{-\varepsilon} = e^{\mu_j} y_t (h_{jt})^{-\varepsilon} (P_t)^{\varepsilon-1},$$

where h_{jt} is the price of variety j, P_t is the aggregate price index for the differentiated good, $P_t = [\int_j e^{\mu_j}(h_{jt})^{1-\varepsilon} dj]^{\frac{1}{1-\varepsilon}}$, y_t is the total income of consumer k, assumed equal across consumers, and Q^k_t is the total consumption of the differentiated good by consumer k (so that $Q^k_t P_t = y_t$, i.e. $Q^k_t = Q_t \equiv \frac{y_t}{P_t}, \forall k$).

2.2. Firm's Problem

Each firm produces one variety. We index firms with j, so that firm j produces variety j. Firm j draws μ_j from a distribution, where $\mu_j = \bar{\mu}$ with probability p_0, and $\underline{\mu}$ with probability $1 - p_0$. μ_j is the same across all consumers for a given j, and $\bar{\mu}, \underline{\mu}, p_0$ are known to the firm, but μ_j is not. Once μ_j is drawn for a given firm and variety j, it is fixed over time.

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_{jt}. Gross profits from selling to a single consumer at any time t are given by

$$\pi^k_{jt} = q^k_{jt} h_{jt} - q^k_{jt} \frac{w_t}{\phi_{jt}} = e^{\mu_j} y_t (h_{jt})^{-\varepsilon} (P_t)^{\varepsilon-1} (h_{jt} - \frac{w_t}{\phi_{jt}}).$$

To maximize these profits, the firm sets the price as a constant mark-up over the marginal cost:

$$h_{jt} = \frac{\varepsilon}{\varepsilon - 1} \frac{w_t}{\phi_{jt}},$$

where ϕ_{jt} is the productivity of firm j.

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 k:

$$X_{jt}^k \equiv \int_0^t \ln\left(\frac{q_{js}^k}{y(h_j)^{-\varepsilon}P^{\varepsilon-1}}\right)ds + \sigma_x W_{jt}^k = \int_0^t \mu_j ds + \sigma_x W_{jt}^k,$$

so that

$$dX_{jt}^k = \mu_j dt + \sigma_x dW_{jt}^k,$$

where W_{jt}^k is a Wiener process, σ_x is a constant known to the firm. It is shown in Appendix 5.1 that the discrete-time analog of this assumption would be the case of the firm observing the log-quantities exported per consumer with some observation noise:

$$\widetilde{\ln q_{jt}^k} \equiv \ln q_{jt}^k + \sigma_x \eta_{jt}^k,$$

where $\eta_{jt}^k \sim$ iid $N(0,1)$, so that σ_x is the standard error of the observation noise. The firm can update its beliefs about μ_j from these observations. More precisely, as is shown in Appendix 5.2, upon sampling n_{jt} consumers and observing n_{jt} values of dX_{jt}^k, and the sample average $\overline{dX_{jt}} \equiv \frac{\sum_{k=1}^{n_{jt}} dX_{jt}^k}{n_{jt}}$, the firm updates its beliefs according to:

$$dp_{jt} = p_{jt}(1-p_{jt})\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sqrt{n_{jt}}d\tilde{W}_{jt} \equiv p_{jt}(1-p_{jt})\chi\sqrt{n_{jt}}d\tilde{W}_{jt}, \quad (1)$$

where p_{jt} is the belief on the part of the firm that $\mu_j = \bar{\mu}$, given all information up to time t:

$$p_{jt} \equiv Prob[\mu_j = \bar{\mu}|I_t],$$

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

$$\begin{aligned} d\tilde{W}_{jt} &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}\left[\frac{\sum_{k=1}^{n_{jt}} dX_{jt}^k}{n_{jt}} - (p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt\right] \\ &\equiv \frac{\sqrt{n_{jt}}}{\sigma_x}[\overline{dX_{jt}} - (p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt], \end{aligned} \quad (2)$$

where $\tilde{W}_{jt}$ is an observation-adapted Wiener innovation process, that is, it follows a standard Wiener process relative to the information at time t.

As can be seen from (2), when the firm observes a sample average $\overline{dX_{jt}}$ higher than its expected value at time t, $(p_{jt}\bar{\mu} + (1-p_{jt})\underline{\mu})dt$, it updates its beliefs upwards ($dp_{jt} > 0$), and downwards otherwise. The firm weighs this signal (the difference between the observed average and the 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. The larger the number of consumers sampled, the faster the firm learns the true value of μ_j.

The firm has to choose the optimal number of consumers accessed subject to the following exporting 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 has a linear in n cost function, $\tilde{c}(n) = fn$, but to use this technology, the firm has to pay the sunk cost F, which could reflect expenditures on building own distribution and retail centres, and upfront costs of long-term contracts with shipping and marketing agencies. It is intuitive and is discussed in section 2.3 that once the firm possesses the linear cost technology, it will export to the entire market. The present value of the cost of selling to the entire market of size M using the testing technology forever is assumed to be higher than that using the linear technology: $\frac{c(M)}{r} > F + \frac{Mf}{r}$. Thus, the firm will find it optimal to eventually pay the sunk cost F and utilise the linear cost technology, conditional on staying in the market. Before it does so, however, it may wish to learn more about demand by employing the testing technology. The firm's decision-making timeline is shown in Figure 3.

We solve the dynamic problem of the firm using backward induction. In what follows, we carry out a partial equilibrium analysis where all aggregate variables and firm's productivity take their steady-state values. Thus, the expected gross profits from selling to a single consumer k, with respect to the information at time t, are given by:

$$
\begin{aligned}
E[\pi_j^k|p_t] &= E\left[e^{\mu_j} y \frac{1}{\varepsilon - 1} \left[\frac{\varepsilon}{\varepsilon - 1} \right]^{-\varepsilon} \left[\frac{\phi_j}{w} \right]^{\varepsilon - 1} P^{\varepsilon - 1} | p_{jt} \right] \\
&= D\phi_j^{\varepsilon - 1} [p_{jt} e^{\bar{\mu}} + (1 - p_{jt}) e^{\underline{\mu}}],
\end{aligned}
$$

where we substitute all variables with their steady state values, and $D \equiv y\frac{1}{\varepsilon-1}[\frac{\varepsilon}{\varepsilon-1}]^{-\varepsilon}[\frac{P}{w}]^{\varepsilon-1}$, a composite of the aggregate demand variables.

2.3. Solution of the Problem of Stage 2

Consider first the optimal behavior of the firm once it pays the sunk cost F and accesses the linear technology with the cost function $\tilde{c}(n) = fn$. Denote this stage as stage 2, and the stage before paying the sunk cost F as stage 1. We show in Appendix 5.3 that the optimal size in stage 2 is $n^* = M$, as long as beliefs p exceed the threshold

$$
\underline{\underline{p_j}} = \max \left\{ \frac{\frac{f}{D\phi_j^{\varepsilon-1}} - e^{\underline{\mu}}}{e^{\bar{\mu}} - e^{\underline{\mu}}}, 0 \right\},
$$

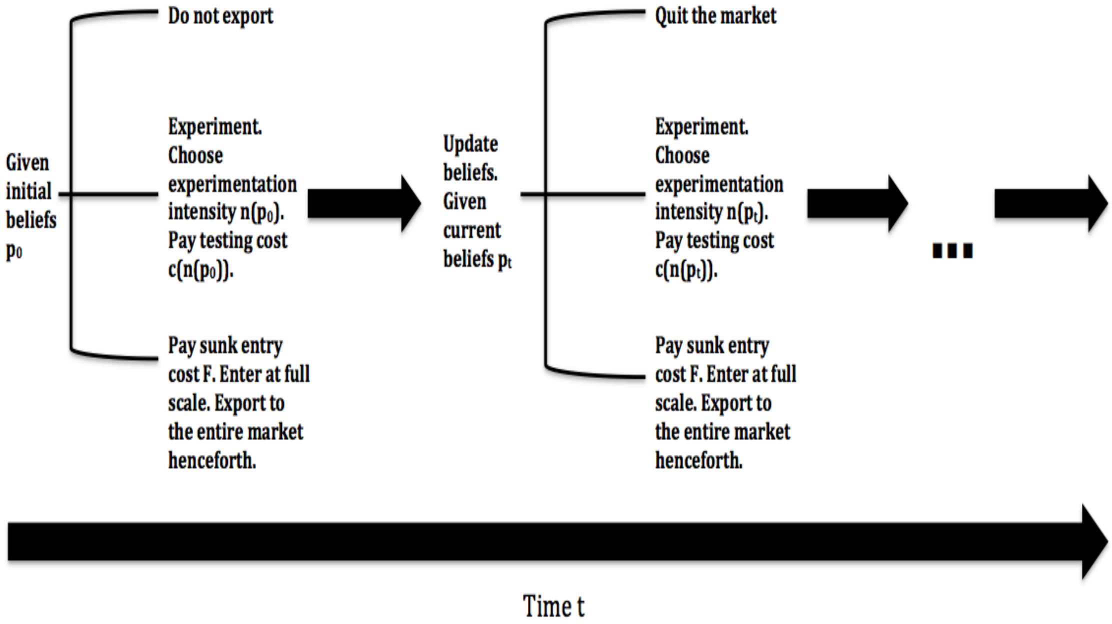

Figure 3 – Timeline of the firm's decision-making.

and 0 otherwise. The value function in the second stage is simply the expected present value of profits from selling to the entire market net of fixed costs of exporting. Denote it by $\tilde{V}(p)$. We will use this value function in the next section.

$$\tilde{V}(p) = \frac{M}{r}(-f + D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]). \tag{3}$$

2.4. Solution of the Problem of Stage 1

Now consider stage 1, when the firm employs the convex costs technology. The firm gains profits from selling to n consumers and has to pay the cost $c(n)$, 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 will assume throughout that $c(0) > 0$, which is a necessary condition for all the results of Moscarini and Smith (2001), which we apply extensively here. That is, there is a fixed cost of maintaining the ability to experiment (think of this, for example, as the fixed periodic cost of the contracts with marketing agencies and warehouses).

One way in which our model is different from that of Moscarini and Smith (2001) 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 (Appendix 5.4.1).

The value function of firm j is found by maximizing the sum of discounted total profits net of testing costs in the experimentation stage and discounted payoff to full-scale entry, net of sunk entry cost, in the event of entry:

$$V(p_{jt}) = \sup_{T,<n_{js}>} E[e^{-rT}K(p_T) + \int_t^T \left(-c(n_{js}) + n_{js}D\phi_j^{\varepsilon-1}\left[p_{js}e^{\bar{\mu}} + (1-p_{js})e^{\underline{\mu}}\right]\right)e^{-r(s-t)}ds|p_{jt}],$$

subject to (1),(2), where $K(p) \equiv \max\{\tilde{V}(p) - F, 0\}$ is the payoff to making the terminal decision, that is, deciding between quitting and investing in the linear technology, $\tilde{V}(p)$ is defined in (3), T is the stopping time, when the firm makes the terminal decision, and p_T is the value of the belief variable at time T. The Hamilton-Jacobi-Bellman equation (HJB) for the control problem is

$$rv(p) = \max_{0\leq n\leq M}\left[-c(n) + nD\phi_j^{\varepsilon-1}\left[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}\right] + \frac{1}{2}(p(1-p)\chi)^2 nv''(p)\right],$$

subject to the value matching conditions:

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F = \frac{M}{r}(-f + D\phi_j^{\varepsilon-1}[\bar{p}e^{\bar{\mu}} + (1-\bar{p})e^{\underline{\mu}}]) - F, \quad v(\underline{p}) = 0.$$

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

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

plus the value matching conditions:

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F = \frac{M}{r}(-f + D\phi_j^{\varepsilon-1}[\bar{p}e^{\bar{\mu}} + (1-\bar{p})e^{\underline{\mu}}]) - F, \quad v(\underline{p}) = 0,$$

and smooth pasting conditions:

$$v'(\bar{p}) = \tilde{V}'(\bar{p}) = \frac{M}{r}D\phi_j^{\varepsilon-1}(e^{\bar{\mu}} - e^{\underline{\mu}}), \quad v'(\underline{p}) = 0.$$

The solution is graphically depicted in Figure 4. Since both z and v are strictly increasing, we know that $n(p) = z(rv(p))$ is strictly increasing in p: as beliefs p increase, the optimal number of consumers increases.[6]

[6]Since the value of experimentation $v(p)$ should be convex, we require $c'(z(rv(p))) > D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]$,

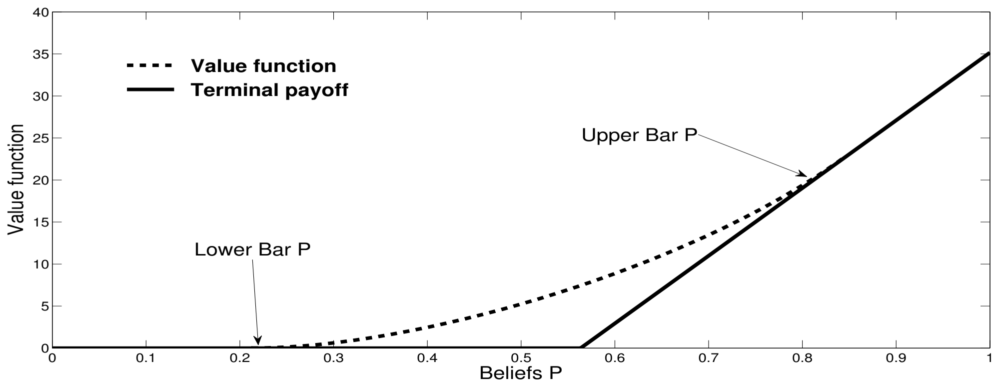

Figure 4 – The solid line shows the ultimate payoff function, $K(p) \equiv \max\{\tilde{V}(p) - F, 0\}$, and the dashed curve shows the value function. The value Lower Bar P$\equiv \underline{p}$ determines the cutoff for quitting, and the value Upper Bar P$\equiv \bar{p}$ determines the cutoff for entering the market (paying the sunk cost F).

2.5. Comparative Statics Predictions

Here we only state the propositions and provide the proofs in Appendix 5.5. These are analogous to the comparative statics predictions of Moscarini and Smith (2001). These results will be needed in later sections.

Proposition 1. *As any of real income y, aggregate price index P, firm productivity ϕ_j increases or wages w decrease, the optimal number of consumers in the experimentation stage, $n(p)$, shifts up, and the thresholds for quitting and for entering the market at full scale decrease, if $n(p)$ is sufficiently bounded away from the full scale M.*

Proposition 2. *As sunk entry cost F rises, so that the final payoff to entry falls, the optimal number of consumers in the experimentation stage, $n(p)$, shifts down, and the thresholds for quitting and for entering the market at full scale increase.*

Proposition 3. *As the testing cost function $c(n)$ grows more convex and initially (for n close to 0) weakly higher and steeper, the optimal number of consumers in the experimentation stage, $n(p)$, shifts down, the threshold for quitting increases, and the threshold for entering the market at full scale decreases.*

for all p. If $c'(z(0)) > D\phi_j^{\varepsilon-1}e^{\bar{\mu}} \geq D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]$, then, since $v(p) \geq 0$ and $c'' > 0$, $c'(z(rv(p))) > D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]$, for all p. Moreover, this assumption (along with the assumption we stated earlier, $c(0) > 0$) is sufficient for us to be able to directly apply all the proofs of existence and uniqueness of a solution in Moscarini and Smith (2001). We discuss this point in greater detail in Appendix 5.4.

2.6. Export Participation Condition

We consider only a partial equilibrium in this model, taking the domestic side of the economy, as well as the aggregate expenditures and the aggregate price index in the foreign market, 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 initial 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 $p_0 = Prob(\mu = \bar{\mu})$. As was stated in Proposition 1, as productivity ϕ rises, both the threshold for switching to full-scale exports, $\bar{p}$, and the threshold for quitting the market, $\underline{p}$, fall. 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 $\underline{\phi}$.

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

This export participation condition is 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 (e.g. Melitz (2003)).

Similarly, it is not possible to tell whether a firm will forgo 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 (i.e. ϕ, such that $\frac{M}{r}(-f + D\phi^{\varepsilon-1}[p_0 e^{\bar{\mu}} + (1-p_0)e^{\underline{\mu}}]) = F$) will choose to experiment first, since its $\bar{p}$ will be above p_0 (and its $\underline{p}$ will be below p_0). This can be seen in Figure 5. Even though its expected profits from entry exceed F, it chooses to experiment first because of the information value it gains.

The productivity of the highest productivity experimenter is given by[7]

$$\bar{\phi} = \min\{\phi_1, \phi_2\}, \qquad \text{where} \qquad \frac{M}{r}(-f + D\phi_1^{\varepsilon-1} e^{\underline{\mu}}) = F, \quad \bar{p}(\phi_2) = p_0.$$

We show the two firms - the lowest productivity exporter, and the highest productivity experimenter in Figure 6. All firms with productivity $\phi \in [\underline{\phi}, \bar{\phi}]$ will experiment.

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 Melitz-type model: $\underline{\phi} < \tilde{\phi} < \bar{\phi}$, where $\tilde{\phi}$ is the cutoff for exporting in the standard framework (obtained by equating

[7] As we keep increasing productivity, one of the two things can happen, depending on the parameter values. 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 $\frac{M}{r}(-f + D\phi_1^{\varepsilon-1} e^{\underline{\mu}}) = F$), 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, enter right away. This is depicted in Figure 5.

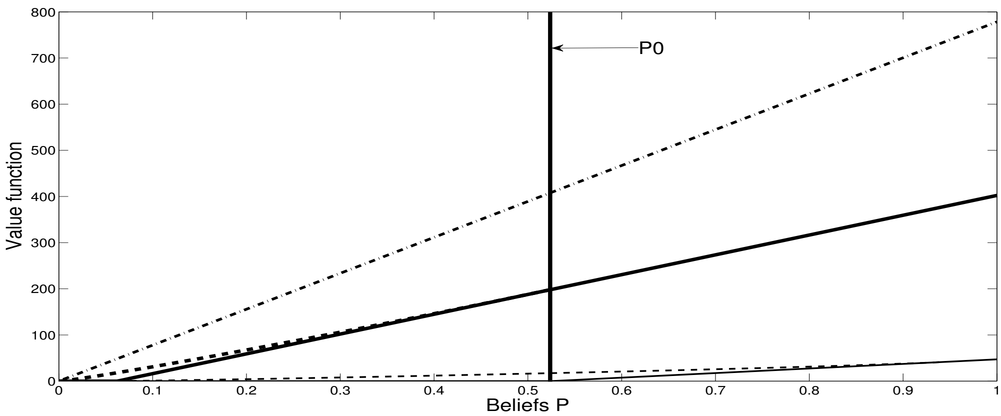

Figure 5 – The thin solid line (lowest on the graph) shows the terminal payoff function, and the thin dashed curve - the value function, for the exporter with ϕ, such that $\frac{M}{r}(-f+D\phi^{\varepsilon-1}[p_0e^{\bar{\mu}}+(1-p_0)e^{\underline{\mu}}])=F$. The thick solid line shows the terminal payoff function, and the thick dashed curve - the value function, for the exporter with productivity ϕ_2. The dashed-dotted line (the highest on the graph) shows the terminal payoff function for the exporter with productivity ϕ_1.

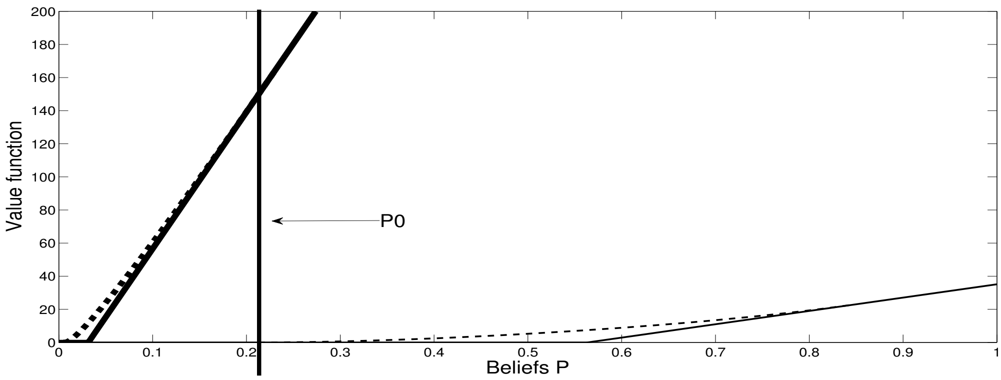

Figure 6 – The thin dashed curve shows the value function, and the thin solid line - the terminal payoff function, for the lowest productivity exporter. The thick dashed curve shows the value function, and the thick solid line - the terminal payoff function, for the highest productivity experimenter.

the expected discounted lifetime profits from selling to the entire market with the sunk cost F). Moreover, in Appendix 5.6 we show, using Proposition 2, that the estimate of F in a Melitz-type model will be lower than the estimate of F in our model, if one fits both models to the same exporter dataset. This is very intuitive. Consider the least productive exporter observed. In a Melitz-type setting, this producer is believed to have found the sunk cost low enough to export. In our setting, this firm may not have incurred the sunk entry cost yet, and may instead be experimenting. If the experimentation model is a better depiction of reality, and the objective is correct identification of the sunk entry cost, with the purpose of making policy recommendations or advising new exporters, fitting the wrong model can therefore provide underestimated values of F.

3. STRUCTURAL ESTIMATION

The model laid out in previous sections is concerned with the optimal behavior of a firm facing demand uncertainty in a new market and choosing its exit and full-scale entry policy, as well as experimentation intensity - the number of consumers that it accesses to learn about demand. One way to interpret our model is viewing each country within a certain geographical region as a consumer, and the entire region as the market to be tested. Suppose we divide the world into regions in such a way that the structure of demand uncertainty is the same in all countries within the same region. That is, the unknown demand parameter is common across these countries, and the observation error has the same distribution in all the countries. In that case, the behavior of a firm that faces a high sunk cost of entering the region and can learn about the unknown demand parameter through its sales in individual countries can be predicted by our model in a straightforward manner.

3.1. Description of the Data

The two main data sources are the exports database, collected by the French Customs, and the French Annual Business Surveys for the manufacturing sector, provided by the French Ministry of Industry. The exports database provides records of French firms' exports - quantities in tons and export values in euros - aggregated by firm, year, destination country and 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 French Customs) over the 1995-2005 period. The annual business surveys contain information on firms that have more than 20 employees, and the variables listed are the address, the identification number of the firm, total sales, number of employees, wages, capital stock and intermediate inputs use. Therefore, we are able to obtain measures of productivity for the firms with employment above 20, even though these measures are not product-specific.

We consider only consumer non-durable products,[8] and focus on 4-digit-code products, rather than 8-digit-code products. This allows us to avoid the issue of the interaction in terms of learning across individual 8-digit-code products within the same 4-digit-code category. We aggregate the quantities exported and total value of exports (in euros) of all 8-digit products within a given 4-digit-code category. Unit values of 4-digit-code products are then obtained as total values exported divided by quantities exported.

[8] According to the classification by Broad Economic Categories of the United Nations Statistics Division.

We form 17 geographical regions out of all countries in the world. We do so because it is much more plausible that demand structure is the same across destinations within a certain region in the world, than across all countries in the world. The categorization we make in this paper is very crude and was based for the most part on geographical proximity (Table 3 in the Appendix).

We define a pair as a combination of a region and a 4-digit-code product. The estimation is then carried out for each pair individually. This allows us to estimate separately the features of export behavior and demand in each 4-digit-code category and region. We focus on new exporters, which are defined as firms that start exporting to the given pair 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'. We calculate the number of countries within the given region that a firm exported to in any given year to measure the experimentation intensity (or consumer sample size), which is a central variable of the model. To make sure that we do not consider insignificant export destinations, where French firms export negligible quantities, we eliminate the countries where the total export volume by French firms over 1995-2005 is less than 5 per cent of the entire export volume by French firms to the region over 1995-2005 (in the given 4-digit-code category).

3.2. Estimation Procedure

We employ Bayesian techniques which are very suitable for situations with many unknowns. The estimation proceeds in two steps, producing estimates of two sets of parameters - Θ_1 and Θ_2. First, we estimate the demand structure - parameters $\bar{\mu}, \underline{\mu}, \sigma_x, p_0$, which we jointly denote as Θ_1 - and beliefs p, based on the time series of export quantities to individual countries. As an intermediate step, we produce a measure of firm profitability that relies directly on the estimated productivity of firms (TFP). Second, given the estimates of $\Theta_1 \equiv \{\bar{\mu}, \underline{\mu}, \sigma_x, p_0\}$, and beliefs p, we look for cost parameters (coefficients of $c(n)$ and F) and full scale M, jointly forming a set Θ_2, that provide the best fit between the optimal number of export destinations, denoted by n^*, and that observed, n. Given the profitability of a firm, and for fixed parameter values, the model predicts the optimal decision rule (experimentation versus full-scale exports versus exit), and the optimal number of destinations($n(p)$ in the experimentation stage, and full-scale export size M in the post-entry stage) of a new exporter. We can therefore calculate the likelihood of the data for any values of parameters Θ_2, relying on the difference between the predicted and the observed time series of n for each firm and year, for fixed values of Θ_1 and beliefs, estimated in Step 1.[9]

[9] In Step 1, we estimate $\Theta_1 \equiv \{\bar{\mu}, \underline{\mu}, \sigma_x, p_0\}$ and beliefs p, conditional on observed export quantities per destination country. However, the observed behavior of firms in terms of the number of export destinations also provides information for the posterior of Θ_1. We did consider this effect, and found that the posterior of Θ_1, conditional on just destination-specific quantities, is tight enough that the information about the number of export destinations does not change it much. Hence, we locate the median values of the posterior distributions of Θ_1, obtained in Step 1, fix these as the values of $\bar{\mu}, \underline{\mu}, \sigma_x, p_0$, calculate the beliefs of all firms, based on these parameters, and proceed to estimate the cost parameters in Step 2.

3.2.1. Step 1: Estimating the demand parameters $\Theta_1 \equiv \{\bar{\mu}, \underline{\mu}, \sigma_x, p_0\}$ and beliefs p.

Fix a pair of a 4-digit-code product and region. We adapt the model to discrete time data. As is shown in Appendix 5.1, using discrete time approximation, the observed value of log-quantity sold in destination country k at time t by firm j is

$$\widetilde{\ln q_{jt}^k} = \ln y_t^k - \varepsilon \ln h_{jt}^k + (\varepsilon - 1) \ln P_t^k + \alpha_{jt}^k.$$

where $j = 1, ..., J$ denotes firms, $t = 1, ..., \bar{T} = 11$ denotes years, $k = 1, ..., K$ denotes countries within the region, and

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

The α_{jt}^k in the above equation have a mean that varies from one firm to another (each firm has a μ_j that is either $\bar{\mu}$ or $\underline{\mu}$). Therefore, we cannot treat these as a usual error term, and just run a regression of $\widetilde{\ln q_{jt}^k}$ on the other variables. Instead, first we 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 h_{jt}^k, 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 these using firm productivities. We calculated TFP a la Ackerberg, Caves and Frazer (2006), to address the issue of endogeneity of production inputs in the production function and to avoid the collinearity issue arising in Olley and Pakes (1996) or Levinsohn and Petrin (2003). For more details on the estimation of TFP, please, see Appendix 5.8. The regression used to estimate the elasticities is:

$$\widetilde{\ln q_{jt}^k} = -\varepsilon \ln u_{jt}^k + d_t^k + d_j + e_{jt}^k,$$

where u_{jt}^k are the unit values and are instrumented for with firm's TFP, d_t^k are destination-time fixed effects, and d_j are firm fixed effects. For some of the pairs, we obtained values of ε that are smaller than 1. We then do not consider these particular pairs in further analysis, since we assume that $\varepsilon > 1$ in the model. We have 921 pairs that we can work with, as a result.

Given the estimates of ε, we calculate the following auxiliary variable for each firm j, destination country k and time t, whenever $\widetilde{\ln q_{jt}^k}$ is observed:

$$\upsilon_{jt}^k \equiv \widetilde{\ln q_{jt}^k} + \varepsilon \ln u_{jt}^k,$$

which by assumption contains the aggregate factors, common to all firms, and the firm-destination-specific, time-varying demand shocks α_{jt}^k. We would like to filter out the aggregate factors, common to all firms and destinations, and obtain demand signals α_{jt}^k. 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 α_{jt}^k, $\bar{\mu}$, $\underline{\mu}$, σ_x, p_0 and μ_j for each j. We have

$$\begin{aligned}
\upsilon_{jt}^k &= \gamma_1 A_{t1} + \gamma_2 A_{t2} + \alpha_{jt}^k, \\
\alpha_{jt}^k &= \mu_j + \sigma_x \eta_{jt}^k, \quad \eta_{jt}^k \sim iid \;\; N(0,1), \\
A_{t1} &= \rho_1 A_{(t-1)1} + \zeta_1 + \nu_{t1}, \\
A_{t2} &= \rho_2 A_{(t-1)2} + \zeta_2 + \nu_{t2},
\end{aligned}$$

$\mu_j = \bar{\mu}$ with probability p_0, and $\underline{\mu}$ with probability $1-p_0$, $\rho_1 \in (-1,1)$, $\rho_2 \in (-1,1)$, $\zeta_1, \zeta_2 \in R$ are constants, and ν_{t1} and ν_{t2} are normal error terms with means zero and covariance matrix Ω.

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 Appendix 5.7.

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:

$$Prob_{jt}(\mu_j = \bar{\mu}) \equiv p_{jt} = \frac{\psi\left(\frac{\overline{\alpha}_{jt} - \bar{\mu}}{\frac{\sigma_x}{\sqrt{n}_{jt}}}\right) p_{j(t-1)}}{\psi\left(\frac{\overline{\alpha}_{jt} - \bar{\mu}}{\frac{\sigma_x}{\sqrt{n}_{jt}}}\right) p_{j(t-1)} + \psi\left(\frac{\overline{\alpha}_{jt} - \underline{\mu}}{\frac{\sigma_x}{\sqrt{n}_{jt}}}\right) (1 - p_{j(t-1)})},$$

where ψ is the standard normal density, $\overline{\alpha}_{jt}$ is the average of α_{jt}^k over all countries sampled, and n_{jt} is the sample size at time t, which in this application is the number of countries within a given pair that the firm exports to.

3.2.2. Intermediate step: Measure of firm heterogeneity/profitability.

To predict the behavior of a firm, we need a measure of its profitability, independent of the demand shift parameter, which is determined by aggregate demand, wages and firm productivity. We treat the 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. The profits of a firm j from exporting to a destination k (in logs) are given by

$$\begin{aligned}
\ln \pi_{jt}^k &= \ln \frac{1}{\varepsilon - 1} - \varepsilon \ln \frac{\varepsilon}{\varepsilon - 1} + \ln y_t^k - (\varepsilon - 1) \ln w_t + (\varepsilon - 1) \ln \phi_{jt} + (\varepsilon - 1) \ln P_t^k + \mu_j \\
&\equiv C + (\varepsilon - 1) \ln \phi_{jt} + A_t^k + \mu_j,
\end{aligned}$$

where C collects all the constant terms, and A_t^k collects all the aggregate terms (y_t^k, w_t, P_t^k). As mentioned before, we assume that the aggregate variables are stationary. Moreover, we set $E(A_t^k) = \bar{A}, Var(A_t^k) = \sigma_A^2$, for all k,t. We assume that $\ln\phi_{jt}$ follows an AR(1) process for every firm j: $\ln\phi_{jt} = C_\phi^j + \rho_\phi \ln\phi_{j(t-1)} + e_{jt}^\phi$, where $\rho_\phi \in (-1,1)$, $e_{jt}^\phi \sim N(0,\sigma_\phi)$, and C_ϕ^j 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, for a fixed μ_j,

$$E(\pi_j) = \exp\left[E(\ln\pi_j) + \frac{Var(\ln\pi_j)}{2}\right] =$$
$$= e^{\mu_j}\exp\left[(\varepsilon-1)\frac{C_\phi^j}{(1-\rho_\phi)}\right]\exp\left[(\varepsilon-1)^2\frac{\sigma_\phi^2}{2(1-(\rho_\phi)^2)}\right]\exp\left[\frac{\sigma_A^2}{2}\right]\tilde{C},$$

where $\tilde{C} \equiv \exp[C+\bar{A}]$. We take the time series of ϕ for all firms, and evaluate ρ_ϕ, σ_ϕ, and C_ϕ^j for each firm j. To measure heterogeneity between firms, it suffices to focus on $\exp[\frac{C_\phi^j}{(1-\rho_\phi)}]$, which we will call as smoothed average productivity of firm j and denote by $\tilde{\phi}_j$. We show in Appendix 5.9 that normalizing the profits by any constant results in the estimates of cost variables that are scaled down by the same constant. Therefore, we can scale down expected profits π_j, given a fixed μ_j:

$$\hat{\pi}_j \equiv \frac{E(\pi_j)}{\exp\left[(\varepsilon-1)^2\frac{\sigma_\phi^2}{2(1-(\rho_\phi)^2)}\right]\exp\left[\frac{\sigma_A^2}{2}\right]\tilde{C}m} = e^{\mu_j}\frac{\exp\left[(\varepsilon-1)\frac{C_\phi^j}{(1-\rho_\phi)}\right]}{m} \equiv e^{\mu_j}\tilde{\pi}_j, \qquad (4)$$

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

3.2.3. *Step 2: Estimating the cost parameters and full scale M, given estimates of $\bar{\mu}, \underline{\mu}, \sigma_x, p_0$ and beliefs p.*

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

$$c(n) = g_0 + g_1 n + g_2 n^2 + g_3 n^3 + g_4 n^4 + g_5 n^5.$$

From now on, denote the set of cost parameters $\{g_0, g_1, g_2, g_3, g_4, g_5\}$ by G.

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 number of export destinations. We substitute the variables $D\phi_j^{\varepsilon-1}$ with the profitability measure $\tilde{\pi}_j$ from (4).

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.[10] 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 $S_t = 2$. We assume that in the full-scale export phase firms exit only due to exogenous shocks and normalize $f = 0$ (we show in Appendix 5.10 that any $f > 0$ and F are equivalent to $f' = 0$ and $F' = F + \frac{fM}{r}$).

Relying on the BVP, and given the profitability of a firm, $\tilde{\pi}_j$, the cost parameters, the full scale M and the beliefs of the firm, one can calculate the thresholds $\bar{p}$ and $\underline{p}$, the state and optimal number of destinations for this firm in every year. In particular, we can predict which new exporters are experimenters - these are firms with $\tilde{\pi}_j$, s.t.

$$\frac{M}{r}(\tilde{\pi}_j e^{\underline{\mu}} - f) < F, \quad \bar{p}(\tilde{\pi}_j) \geq p_0 \geq \underline{p}(\tilde{\pi}_j).$$

Given that a firm chooses to experiment first, we can predict its state and export size at every $t = 1, ..., \bar{T} = 11$: with probability δ (an exogenous death rate, introduced here for the first time), the firm exits in any given period, but otherwise,

$$\begin{aligned}
t = 1 \Longrightarrow S_{jt} = 0, &\quad n^*_{jt} = z(rv(p_{jt})|\tilde{\pi}_j);\\
t > 1, S_{j(t-1)} = 0, p_{jt} \in (\bar{p}_j, \underline{p}_j) \Longrightarrow S_{jt} = 0, &\quad n^*_{jt} = z(rv(p_{jt})|\tilde{\pi}_j);\\
t > 1, S_{j(t-1)} = 0, p_{jt} \geq \bar{p}_j, \Longrightarrow S_{jt} = 1, &\quad n^*_{jt} = M;\\
t > 1, S_{j(t-1)} = 1 \Longrightarrow S_{jt} = 1, &\quad n^*_{jt} = M;\\
t > 1, S_{j(t-1)} = 0, p_{jt} \leq \underline{p}_j \Longrightarrow S_{jt} = 2, &\quad n^*_{jt} = 0;\\
t > 1, S_{j(t-1)} = 2 \Longrightarrow S_{jt} = 2, &\quad n^*_{jt} = 0.
\end{aligned}$$

Given that a new exporter chooses to enter at full scale right away ($\frac{M}{r}(\tilde{\pi}_j e^{\underline{\mu}} - f) \geq F$ or $\bar{p}(\tilde{\pi}_j) \leq p_0$), we can also predict its behavior: with probability δ it exits in any given period, and if it

[10] In the data, we assume that a firm that did not export a given 4-digit-code product to any country within the region in some year and in all years after that, up to 2005, is in the non-exporting state. However, when we observe a firm not exporting to any destination in some year, followed by exports to at least one destination within the given region in any year afterwards, the non-exporting year is treated as just a temporary zero in the time series of the number of destinations of this firm. That is, we simply set $n_{jt} = 0$ for that year for that firm.

does not, it exports to M countries in the region. Thus, the model gives us predictions with respect to the optimal export size of firms, as measured by the number of export destinations within the region, and optimal exit behavior.

We assume that the total discount rate r is equal to the sum of the exogenous firm death rate, δ, and a pure discount rate, say, dr. We fix the value of dr to 0.05. We choose the value of 0.05, since that is the average long-term interest rate in France over 1995-2005, and it is also the value of discount rate suggested by the Guide to Cost Benefit Analysis of Investment Projects (2008) of the European Commission. Thus, we need to estimate δ only, rather than both δ and r.

There are some additional parameters that we introduce to accommodate the data better. We let there be a discrepancy between the observed n_{jt} and the predicted n^*_{jt}:

$$n_{jt} = n^*_{jt} + e^N_{jt},$$

where e^N_{jt} is a mean zero normal error term, with variance σ^2_N, which we will also estimate. We view the error term as the discrepancy caused by managerial error, delays and disruptions in international delivery, and considerations affecting the choice of the number of destinations that are outside the model. Additionally, the model predicts that in the experimentation stage, the firm quits as soon as its beliefs p_{jt} fall below some level $\underline{p}_j$. We cannot ensure that this holds true for all j,t, and therefore introduce $\widetilde{\underline{p}}_j$:

$$\widetilde{\underline{p}}_j = \underline{p}_j - e_{pd},$$

where $\underline{p}_j$ is the threshold for quitting as predicted by the model, and $\widetilde{\underline{p}}_j$ is the actual threshold for quitting, which differs from that predicted by some error e_{pd}. The variable e_{pd} has an exponential distribution with mean $m_p > 0$. So we allow for the actual quitting threshold for any firm to be lower than the one predicted by the model.

To summarise, we need to estimate the following parameters: $\Theta_2 \equiv \{G, F, M, \delta, \sigma_N, m_p\}$. We do this as follows:

- Fix values of M, δ, $r \equiv \delta + dr$, σ_N, and m_p. Given these values, carry out the Metropolis-Hastings step (which essentially lets you pick the values with the highest likelihood, given flat priors), to produce new estimates for the cost parameters G and F.
- For given values of G and F, update the values of M, δ, r, σ_N and m_p, that is, generate draws from their new posteriors, conditional on the values of G and F.
- Iterate on the previous two steps until convergence.

In Appendix 5.11, we explain in greater detail how we set the initial values of M, δ, σ_N, and m_p, how we calculate the likelihood, and how we carry out the two steps above.

3.3. General Results

We run a regression of the number of export destinations within a region on the beliefs of firms, to see if the latter have a positive effect on the former, as predicted by the model. Note that beliefs were estimated in Step 1, without relying on the information on the number of export destinations. We define a variable ES to be the number of countries within a given region (for a given 4-digit-code product) that the firm exports to until it exits the region entirely, if at all. We let it take the value 0 in the year when the firm quits. Thus, we are capturing both expansions and contractions, as well as absolute exits. One can run a regression of ES on estimated beliefs, controlling for aggregate demand, firm productivity and firm fixed effects, to evaluate the effect of beliefs on firm behavior. Since the beliefs are not observed, but instead have been estimated, we do the following: we generate thirty draws of the vector of beliefs from the posterior obtained in the first step of the estimation (estimating demand parameters), for each firm in every year. We then generate a new dataset, where the observed independent variables are matched with the vectors of beliefs, and stacked together, so that we have 30 times the original number of observations. That is, initially we have a regression equation for the export size ES as a function of beliefs P

$$ES_{jt}^{k} = \beta_0 + \beta_1 P_{jt}^{k} + \beta_2 \ln \phi_{jt}^{k} + d_t^k + d_j^k + \varepsilon_{jt}^{k},$$

where k denotes a pair - combination of 4-digit-code product and region, j denotes firms exporting the given 4-digit product to the given region, and t - years, ϕ_{jt}^{k} denotes productivity of firm j exporting to pair k, d_t^k are pair-year fixed effects, d_j^k are pair-firm fixed effects and ε_{jt}^{k} is the error term.

Table 1 – Regression of the number of export destinations on beliefs, controlling for productivity and aggregate demand.

	Number of destinations
Beliefs P	0.239 (18.27)**
Log-TFP	0.02 (2.03)*
Constant	0.313 (42.61)**
Pair-year fixed effects	YES
Pair-firm fixed effects	YES
Observations	6612082
R-squared	0.01

Note: absolute values of t-statistics in parentheses;
*significant at 5%; **significant at 1%.

After we generate 30 draws of beliefs P for each firm within a given pair, we have

$$ES_{jt}^{k} = \beta_0 + \beta_1 P_{jt}^{sk} + \beta_2 \ln \phi_{jt}^{k} + d_t^k + d_j^k + \varepsilon_{jt}^{sk},$$

where s denotes the draw of beliefs P. Note that the error term has the form

$$\varepsilon_{jt}^{sk} = \bar{\varepsilon}_{jt}^{k} + \tilde{\varepsilon}_{jt}^{sk},$$

since the beliefs P_{jt}^{sk} come from the same distribution for every j,t,k. Hence, we calculate clustered standard errors in this regression, where each combination of pair-firm and year forms a cluster. The results are presented in Table 1. As is shown in the table, beliefs have a positive, highly significant, effect on export size, as measured by number of destinations. This suggests that beliefs have the effect predicted by the model.

3.4. Estimates for Two Pairs

As an illustration we present the results of the structural estimation for two pairs: exports of the 4-digit code 4202 (trunks, suitcases, etc.) to the Middle East and exports of the 4-digit code 6403 (footwear) to East Asia.[11] In Figures 7 and 8, we show how the average number of destinations accessed within each pair evolves with the date since first exports in the region. On the same graphs, we plot the evolution of the average number of export destinations (within the pair) of old exporters, where the horizontal axis now measures chronological time (year since 1995, 1995 being 1). One can see that the behavior of old exporters in East Asia and Middle East is much more stable than that of new exporters, and is close to the estimated full scale, on average. As an additional check, we plot the aggregate demand in each pair in Figure 9. Aggregate demand is measured as the weighted average over all countries within the region of quantity imported from the entire world of the 4-digit-code product. We see that in both East Asia and Middle East the aggregate demand has multiple ups and downs, and does not exhibit steady upward movement over time, so that the gradual expansion of new exporters cannot be explained away by an aggregate expansion.

Table 2 – Estimates of the parameters for two pairs.

Parameter	ε	m	p_0	$\bar{\mu}$	χ	M	F	$\frac{F}{p_0 e^{\bar{\mu}} + (1-p_0) e^{\underline{\mu}}}$	r	σ_N
East Asia	3.14	7761	0.2175	3.5	1.65	2	70	8.75	0.19	0.91
Middle East	1.64	14.71	0.1183	2.64	1.84	3	59.77	23.52	0.241	1.09

[11] The choise of the two pairs is casual; the exercise can be replicated easily for all products region exisitng in the dataset. The 4-digit code 4202 encompasses 'trunks, suitcases, vanity cases, executive-cases, briefcases, school satchels, spectacle cases, binocular cases, camera cases, musical instrument cases, gun cases, holsters and similar containers; travelling-bags, insulated food or beverages bags, toilet bags, rucksacks, handbags, shopping-bags, wallets, purses, map-cases, cigarette-cases, tobacco-pouches, tool bags, sports bags, bottle-cases, jewellery boxes, powder boxes, cutlery cases and similar containers, of leather or of composition leather, of sheeting of plastics, of textile materials, of vulcanised fibre or of paperboard, or wholly or mainly covered with such materials or with paper'. The 4-digit code 6403 encompasses 'footwear with outer soles of rubber, plastics, leather or composition leather and uppers of leather'.

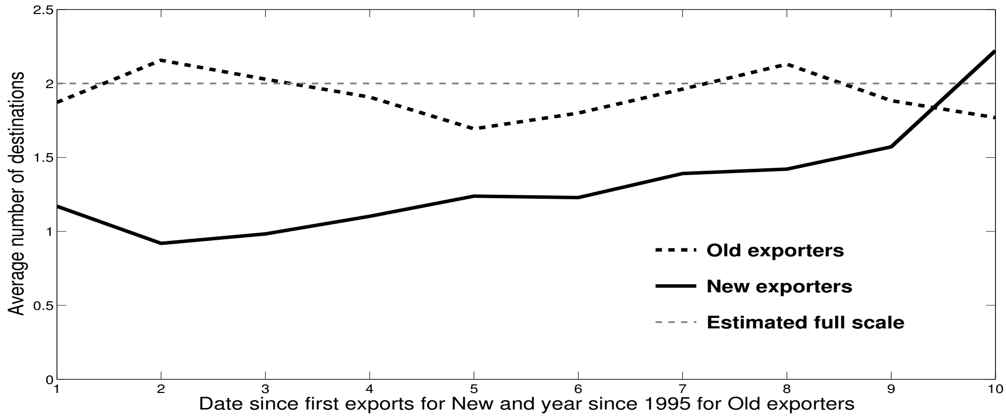

Figure 7 – Average size evolution for East Asia, Footwear

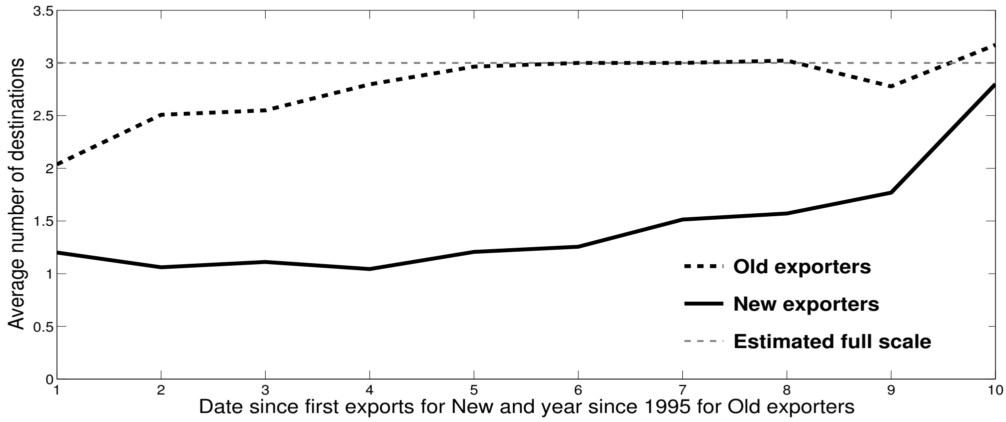

Figure 8 – Average size evolution for Middle East, Trunks, suitcases, etc.

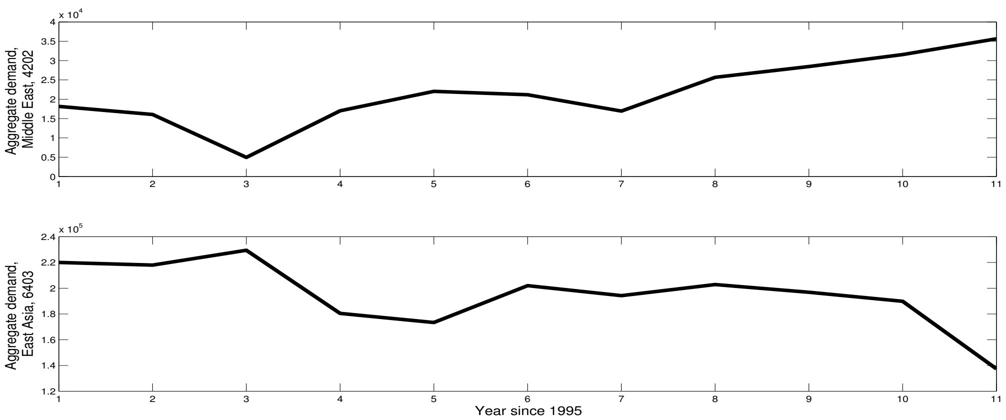

Figure 9 – Aggregate demand evolution between 1995-2005 in the two pairs

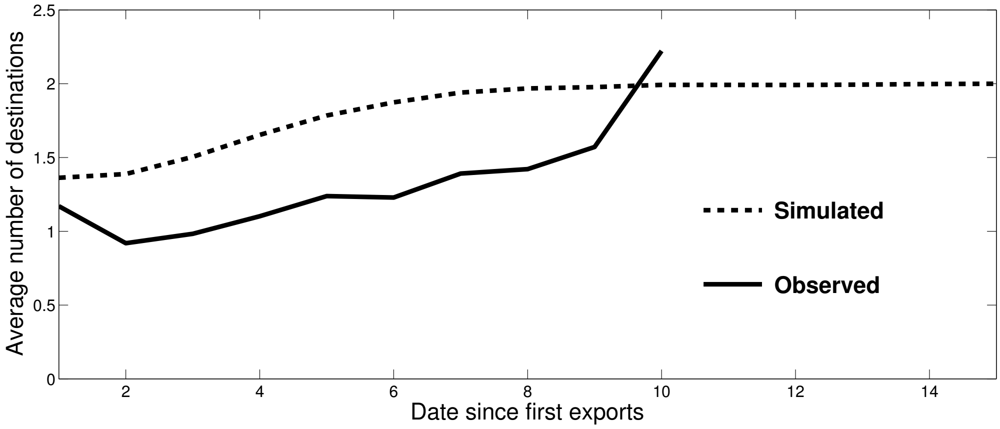

Figure 10 – Simulated average size, over 10000 exporters, for East Asia, Footwear

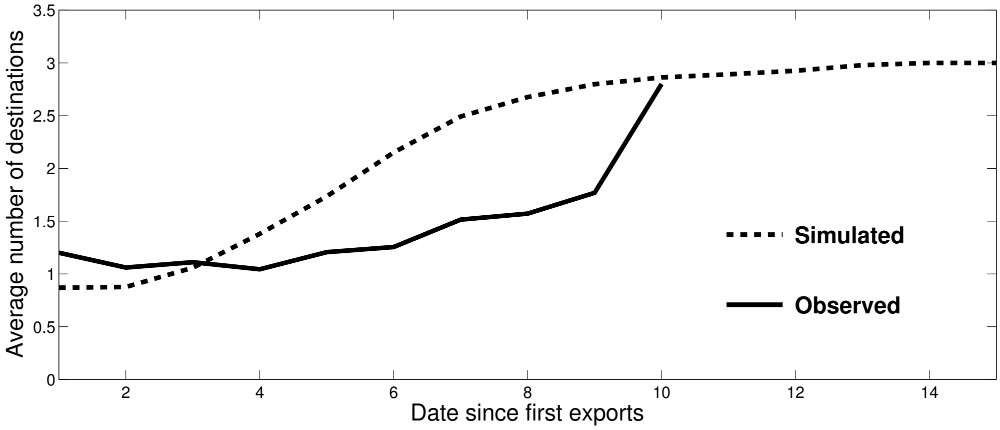

Figure 11 – Simulated average size, over 10000 exporters, for Middle East, Trunks, suitcases, etc.

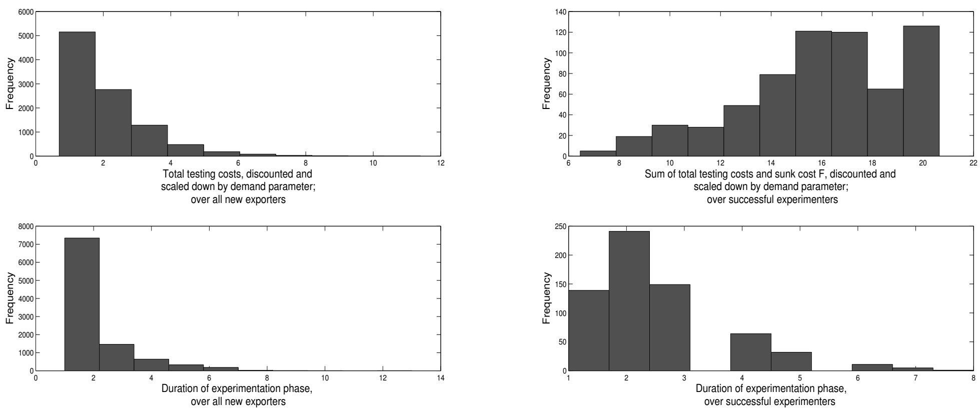

Figure 12 – Histograms of new exporters' duration and total cost of experimentation.

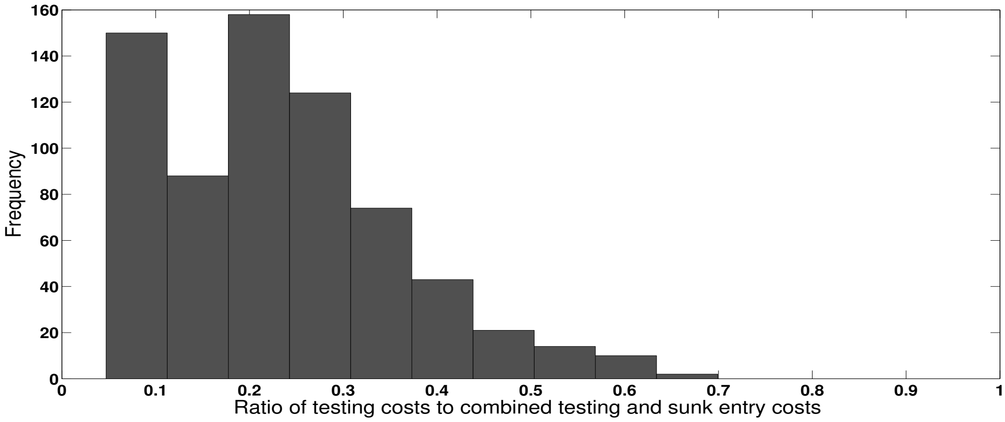

Figure 13 – Histogram of the ratio of testing costs to combined testing and sunk entry costs, incurred by successful experimenters, over 10000 exporters, for Middle East, Trunks, suitcases, etc.

In Table 2 we summarize the estimates of the main parameters (means of/draws from their posterior distributions). One can see that the estimated value of the signal-to-noise ratio χ is lower in East Asia than in the Middle East. The signal-to-noise ratio affects the speed of learning and convergence of the average export size to the full scale, and we would expect that this would result in faster learning and faster convergence in the Middle East. We simulate the behavior of new exporters in both pairs. We start with a fixed number of new exporters (10000), 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 $n(p)$. In this way we produce a smoothed average size evolution, without the sharp turns observed in our (much smaller) dataset. In Figures 10 and 11, we observe that the simulated average number of destinations in East Asia and the Middle East approach the estimated full scale by dates 14 and 15, respectively. That is, the average converges to the full scale faster in East Asia than in the Middle East. The effect of the higher signal-to-noise ratio is cancelled out, and dominated, by the effect of higher initial beliefs p_0 and higher $\bar{\mu}$ in East Asia: firms experiment more intensively and learn faster due to higher expected terminal payoff. We also use the simulated time series of export size in the two regions as an indication of in-sample goodness-of-fit. Notice that we estimated the model using the firm-level behavior, and did not pick the parameters to match the aggregate time series. That the simulated average number of destinations follows the observed to a certain degree tells us that we can numerically evaluate policy experiments using our estimates.

In Table 2, we show the estimated sunk cost estimates, as well as the scaled down estimates $(F/[p_0e^{\bar{\mu}} + (1-p_0)e^{\underline{\mu}}])$. We scale the costs down by the demand portion that does not appear in our measure of profitability, $\tilde{\pi}$, but still should be taken into account - since total expected profits of a firm will be given by the product of $\tilde{\pi}$ and $E(e^{\mu})$. The raw estimate of F is higher in East Asia, but when scaled down by the expected value of the demand shift parameter, it is much lower there. This is again due to higher $\bar{\mu}$ and p_0 in East Asia.

We can also study the total cost of experimentation and the duration of experimentation phase of new exporters. We sum the experimentation costs, discounted and scaled down by $E(e^{\mu})$, over all periods for each new exporter in Middle East, and plot the histogram of these in Figure 12, upper left corner. In the upper right corner, we plot the total costs of entering the new market (sum of testing costs and sunk cost F, discounted and scaled down by $E(e^{\mu})$) of successful experimenters (those who switch to the full scale at some point). We also plot the histogram of the duration of the experimentation phase, unconditional, as well as conditional on being a successful experimenter, in the left and right lower corners, respectively. The mean experimentation cost in the Middle East is 5.22 (in units of median profitability, since we scaled down firm profitability and therefore costs by the median profitability in the pair to obtain $\tilde{\pi}$ in Section 3.2.2). The mean duration of experimentation phase for successful experimenters is 2.52, and for unsuccessful experimenters - 2. We see that in our model both the optimal experimentation duration and the total cost of entering a new market (total experimentation cost plus the sunk cost of entry for successful experimenters) are firm-specific and random. In Figure 13, we show the histogram of the ratio of testing costs to combined testing and sunk entry costs of successful experimenters in the simulated sample (Middle East, Trunks, suitcases, etc.). This ratio ranges between 0.05 and 0.7, with a mean of 0.23. Thus, on average, about one fourth of the costs incurred in entering a new market by experimenters is accounted for by testing costs (in this particular example). Taking the ratio of total testing costs (sum over all new exporters in the simulated sample) to total combined testing and sunk entry costs gives us the result of 0.61. That is, about two thirds of all costs incurred by new exporters in the process of accessing new markets (successfully or not) is allocated to experimentation efforts.

3.4.1. Temporary currency appreciation in a single country

Suppose that the currency of only one country within a region appreciates relatively to the French currency by 67 percent. Assume that France is a small exporter in this market. In that case, the aggregate variables (in particular, the aggregate price index) will not be affected by the behavior of French exporters, and we can focus only on the direct effect of lower exchange rate (French versus foreign currency) on the profits of French exporters. To study the effect of the currency appreciation on French exporters' profits, we introduce a trade barrier term τ_k which is a product of e_k, the exchange rate between the French currency and the foreign country k currency, and b_k, all other trade costs, so that demand in country k is given by $q(h^*_{jk}) = (h^*_{jk})^{-\varepsilon_k} D_k = (\tau_k h_j)^{-\varepsilon_k} D_k = (e_k b_k h_j)^{-\varepsilon_k} D_k$ where D_k incorporates all the aggregate demand variables in market k, h_j is the price of the product in France in French currency, and $h^*_{jk} = h_j e_k b_k$ is the price of the product in the foreign country in the foreign currency. We have the following expression for the profits earned by firm j in the foreign market k (disregarding the time dimension and the firm-specific demand portion e^{μ_j})

$$\pi_{jk} = h_j q(h^*_{jk}) - q(h^*_{jk})mc_j = q(h^*_{jk})(h_j - mc_j) = (e_k b_k h_j)^{-\varepsilon_k} D_k(h_j - mc_j),$$

where mc_j is the marginal cost of production. When the exchange rate falls from $3e_k$ to e_k, resulting in a 67 % appreciation of the foreign currency, profits of French exporters increase. Assume that the lower exchange rate applies only for one year. This requires that we assume a hybrid model with discrete time in the stage 0 (the year of lower exchange rate), and continuous time afterwards (the stages 1 and 2 of the model we studied in the theoretical part). We ignore here the effect this temporary exchange rate appreciation will have on the exporters already present in the region. Any firm that did not choose to export to the region before the appreciation will either continue not exporting there, or will export only to the country whose currency gained in value in the year when that happens, to benefit from the temporarily higher profits. This would require that the profits from exporting to that destination, net of exporting costs, are positive:

$$E_0(e^{\mu}\tilde{\pi}'_k) - c(1) = E_0(e^{\mu}\tilde{\pi}_k 3^{\varepsilon_k}) - c(1) > 0,$$

where $e^{\mu}\tilde{\pi}'_k$ are the profits from exporting to the destination k whose currency appreciated in the year of this event, $e^{\mu}\tilde{\pi}_k$ are the profits from exporting to destination k under old conditions, E_0 is the expected value given beliefs p_0, and $c(1)$ is the testing cost of accessing one destination, which is the exporting cost, since the firm does not invest in full-scale entry to export to a single destination for one year. To locate all new exporters, we therefore find all firms that do not export under the original setting, but for whom $E_0(e^{\mu}\tilde{\pi}'_k) - c(1) > 0$. In Appendix 5.12, we show how we predict the new exporting threshold on productivity and the number of new exporters in response to higher profits. We count 37 and 29 such new exporters in East Asia and Middle East, respectively.

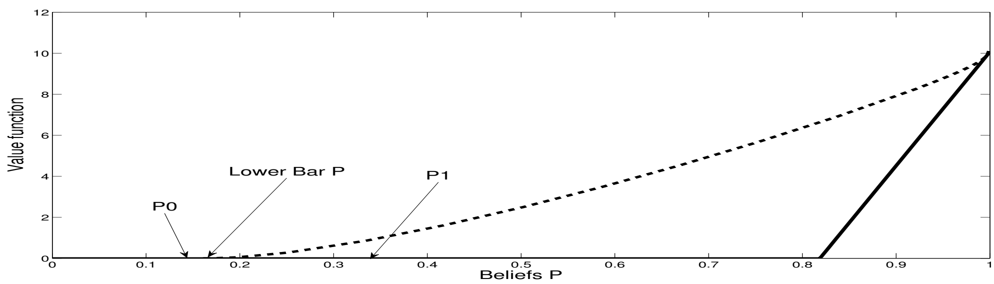

Figure 14 – A firm for whom initial beliefs p_0 are too low to export to the region, but a change in beliefs to $p_1 > p_0$ is sufficient to start experimenting in the region.

In Figure 14, we show the solution to the BVP from the theoretical part for a firm whose productivity is lower than that of exporters already present in the market, so that $\underline{p} > p_0$. After exporting to one destination and updating its beliefs to some $p_1 > \underline{p} > p_0$, it decides to continue

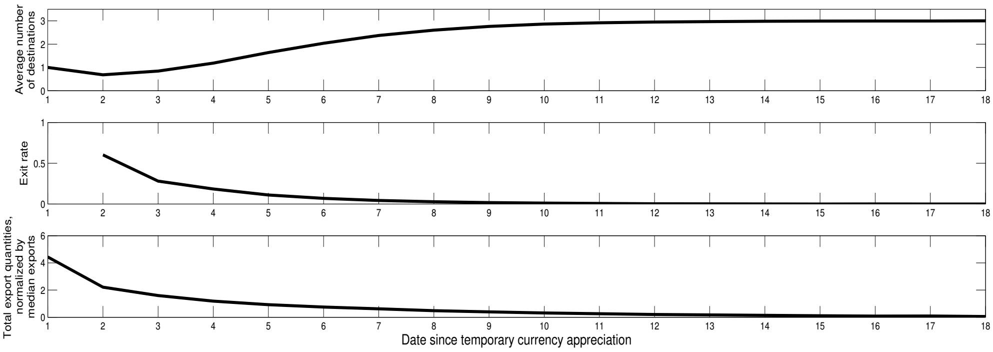

Figure 15 – Response to a 67 % temporary currency appreciation in a single country within Middle East, exports of Trunks, suitcases, etc. First panel: average number of destinations, second panel: exit rate, third panel: total quantity exported, scaled down by the median quantity exported.

exporting in the region even after the exchange rate reverts to its original level. We simulate the behavior of new exporters 10000 times, and take an average of the resulting time series (averaging over all 10000 samples, for each date since first exports), to obtain a smoothed out pattern. We plot the response of the new exporters in the Middle East in Figure 15, and of the new exporters in East Asia - in Figure 17 in the Appendix. The new exporters all start with 1 destination in the year of lower exchange rate, and later either quit or decide to export/experiment in the region further. It takes 18 and 20 years for the average number of destinations of these exporters (conditional on survival) to reach the full scale in the Middle East and East Asia, respectively. In both pairs around 40 percent of those firms that exported in the lower exchange rate period stay afterwards, and out of these around 30 percent exit in the third year. Alongside the evolution of the average number of destinations, we show the total export quantities over all new exporters, scaled down by the median quantity exported (median of $\exp\left[\varepsilon_k \frac{C_\phi^j}{(1-\rho_\phi)}\right] y \left(\frac{\varepsilon_k}{\varepsilon_k-1} w\right)^{-\varepsilon_k} P^{\varepsilon_k-1}$), analogously to the way we scaled down firm profits earlier.We observe non-negligible export volumes resulting from a temporary impulse in the form of a transient currency appreciation in a single destination.

3.4.2. Response to a decrease in the testing costs

Suppose testing costs shift down in such a way that for the new testing cost schedule $\hat{c}(0) \equiv \hat{g}_0 = g_0 = c(0)$, and $\hat{c}(n) = 0.9c(n), \forall n > 0$. According to Proposition 4, the optimal testing size will shift up for all p, for given ϕ, the threshold for quitting will decrease, and the threshold for entering at full scale will increase. We calculate the new exporting cutoff on productivity, and evaluate the number of new exporters, as before.

We want to show the non-equivalence between decreasing testing costs and the sunk entry cost. To do so, we find a new value of F that would result in the same new exporting cutoff on

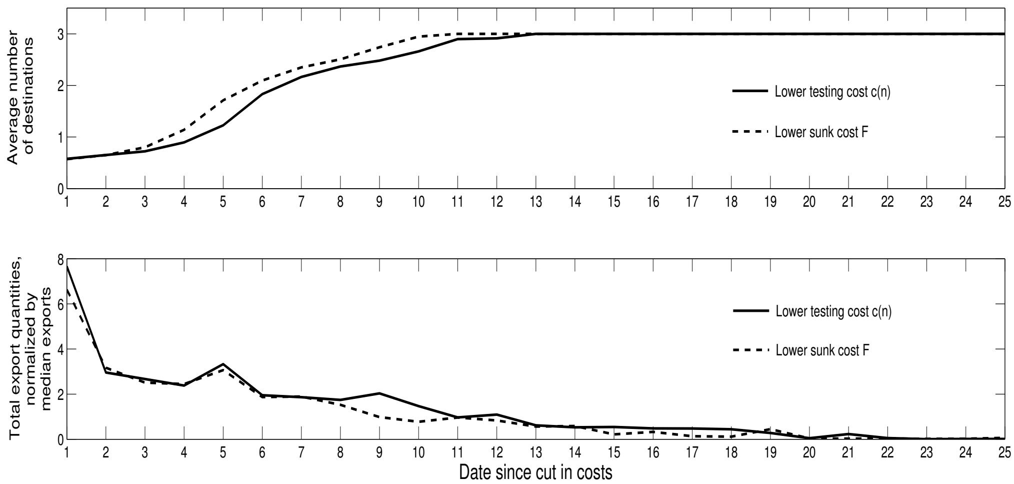

Figure 16 – Response to a 10 % decrease in the testing costs $c(n)$ and to an 'equivalent' cut in the sunk entry cost F. Exports of Trunks, suitcases, etc., by French firms to the Middle East. First panel: average number of destinations, second panel: total quantity exported, scaled down by the median quantity exported.

productivity and thus in the same number of new exporters, given original testing costs. By Proposition 2, for any ϕ the thresholds on entering at full scale and quitting will decrease. The simulated evolution of new exporters in the Middle East and East Asia are depicted in Figures 16 and 18, respectively. As can be seen in the first panel, the average export size converges to full scale faster when the sunk entry cost is cut, since the threshold on beliefs for entering at full scale is lower with lower F. However, the total quantities exported are in general higher with lower testing costs then with lower F. With lower testing costs the threshold on beliefs for entering at full scale increases, while with lower sunk cost this threshold decreases. Firms will keep experimenting for some high values of beliefs under lower testing costs, but will decide to switch to full scale for some relatively low values of beliefs under lower sunk cost. This leads to a higher ratio of high-demand firms among those that enter the market after experimenting, when testing costs are lower. For example, for the Middle East, the ratio of low-demand firms ($\mu_j = \underline{\mu}$) among all 'successful' experimenters is 0.0015 for lower $c(n)$, compared with 0.0091 for lower F, and, as a result, the total quantity exported over the first 25 years is 33.8 for lower $c(n)$, compared with 29.2 for lower F. Thus, lower testing costs contribute to better selection of high-demand firms into full-scale exporting, which in turn leads to higher total export volumes. The average cost (present value of the sum of all experimentation costs and sunk entry cost in the event of full-scale entry) per new exporter over 10000 simulations is 5.2185 for lower $c(n)$ and 5.2179 for lower F. While expected average cost of experimentation/entry is the same in the two scenarios, lowering testing costs results in better screening of firms and higher export volumes.

4. Conclusion

We present a model with demand uncertainty and sunk costs of entry. The firm is able to learn more about the demand for its product 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 demand shift parameter. Thus, the decision for new exporters is not 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 and the market. For policy purposes, any exogenous shock prompting entry of new exporters will not have an immediate impact. Instead, a transition period will be observed, and the dynamics and duration of this period can be predicted using the model. We apply this framework to the problem of a firm that wishes to export to a certain geographical region, with uncertain demand and a high sunk entry cost, and can learn more about demand there by accessing a few individual countries in the region first. This allows us to structurally estimate the model, using French firm-destination-product-level export data between 1995-2005. As an illustration we show estimates of all parameters of the model for exports by French firms of two casual pairs of product-region: trunks, suitcases, etc., to the Middle East and of footwear to East Asia. Several simulation exercises are carried out. We consider the effects on French exporters of a temporary currency appreciation in a single country within a region and of lower testing costs. The model generates interesting new dynamics. In the case of a temporary currency appreciation in a single destination the new information obtained in one country allows the firm to update beliefs about the entire region and possibly start exporting to other countries within the region. Lower testing costs are shown to be non-equivalent to lower sunk entry cost, since lower testing costs lead to better selection of high-demand producers into exporting, and higher total export volumes. Thus, policies aimed at increasing export volumes by domestic firms should focus on cutting testing costs, rather than the sunk entry cost, all else equal. This model can be applied to multiple other settings, such as the introduction by firms of new products in domestic markets, and we plan to carry out this application using marketing data. Another extension is considering this model in a general equilibrium setting, where firm exporting behavior affects the aggregate variables in the market. We will consider this in future work.

REFERENCES

ACKERBERG, D., CAVES, K. AND FRAZER, G., 2006. Structural Identification of Production Functions. *Econometrica* R&R.

ALBORNOZ, F., PARDO, H. F. C., CORCOS, G. AND ORNELAS, E., 2012. Sequential Exporting. *Journal of International Economics* 88, 17-31.

ARKOLAKIS, C., 2010. Market Penetration Costs and the New Consumers Margin in International Trade. *Journal of Political Economy* 118(6), 1151-1199.

BERNARD, A. B., JENSEN, J. B., REDDING, S.J. AND SCHOTT, P.K., 2007. Firms in International Trade. *Journal of Economic Perspectives* 21(3), 105–130.

BOLTON, P., HARRIS, C., 1999. Strategic Experimentation. *Econometrica* 67(2), 349-374.

DAS, S., ROBERTS, M. J. AND TYBOUT, J. R., 2007. Market Entry Costs, Producer Heterogeneity, and Export Dynamics. *Econometrica* 75(3), 837-873.

DE LOECKER, J., 2011. Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica* 79(5), 1407-1451.

EATON, J., ESLAVA, M., KRIZAN, C.J., KUGLER, M. AND TYBOUT, J., 2007. Export Dynamics in Colombia: Firm-Level Evidence. *Working Paper, Banco de la Republica.*

EATON, J., ESLAVA, M., KRIZAN, C.J., KUGLER, M. AND TYBOUT, J., 2008. A Search and Learning Model of Export Dynamics. *Working Paper.*

EUROPEAN COMMISSION, 2008. Guide to Cost Benefit Analysis of Investment Projects.

FREUND, C., PIEROLA, M. D., 2010. Export Entrepreneurs: Evidence from Peru. *Policy Research Working Paper Series 5407, The World Bank.*

HITSCH, G. J., 2006. An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty. *Marketing Science* 25(1), 25-50.

HORSTMANN, I. J., MARKUSEN, J. R., 1996. Exploring New Markets: Direct Investment, Contractual Relations and the Multinational Enterprise. *International Economic Review* 37(1), 1-19.

KLOMPMAKER, J. E., HUGHES, G. D. AND HALEY, R. I., 1976. Test Marketing in New Product Development. *Harvard Business Review.*

LEVINSOHN, J., MELITZ, M., 2002. Productivity in a Differentiated Products Market Equilibrium. *Harvard University mimeo.*

LEVINSOHN, J., PETRIN, A., 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies* 70(2), 317-342.

MELITZ, M., 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* 71(6), 1695-1725.

MOSCARINI, G., SMITH. L., 1998. Wald Revisited: The Optimal Level of Experimentation. *Working Paper.*

MOSCARINI, G., SMITH. L., 2001. The Optimal Level of Experimentation. *Econometrica* 69(6), 1629-1644.

NGUYEN, D. X. (2012). Demand Uncertainty: Exporting Delays and Exporting Failures. *Journal of International Economics* **86(2)** 2012.

OLLEY, S., PAKES, A., 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64(6), 1263-1295.

RAUCH, J. E., WATSON, J., 2003. Starting Small in an Unfamiliar Environment. *International Journal of Industrial Organization* 21(7), 1021-1042.

RUHL, K. J., WILLIS, J. L., 2009. New Exporter Dynamics. *Working Paper, NYU Stern School of Business.*

SEGURA-CAYUELA, R., VILARRUBIA, J. M., 2008. Uncertainty and Entry into Export Markets. *Banco de Espana Working Papers.*

5. APPENDIX

5.1. 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 k:

$$X_{jt}^k \equiv \int_0^t \ln\left(\frac{q_{js}^k}{y(h_j)^{-\varepsilon}P^{\varepsilon-1}}\right)ds + \sigma_x W_{jt}^k = \int_0^t \mu_j ds + \sigma_x W_{jt}^k,$$

so that

$$dX_{jt}^k = \mu_j dt + \sigma_x dW_{jt}^k,$$

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

$$\Delta X_{jt}^k \equiv \ln\left(\frac{q_{jt}^k}{y(h_j)^{-\varepsilon}P^{\varepsilon-1}}\right)\Delta t + \sigma_x \Delta W_{jt}^k,$$

and with $\Delta t = 1$, denoting $\Delta X_{jt}^k \equiv \alpha_{jt}^k$, and $\Delta W_{jt}^k \equiv \eta_{jt}^k \sim N(0,1)$, from the properties of the Wiener process, we get

$$\alpha_{jt}^k = \ln q_{jt}^k - \ln y + \varepsilon \ln h_j - (\varepsilon - 1)\ln P + \sigma_x \eta_{jt}^k = \mu_j + \sigma_x \eta_{jt}^k,$$

i.e. $\alpha_{jt}^k \sim N(\mu_j, \sigma_x^2)$. Denote

$$\widetilde{\ln q_{jt}^k} \equiv \ln q_{jt}^k + \sigma_x \eta_{jt}^k.$$

The firm observing α_{jt}^k is equivalent to the firm observing

$$\widetilde{\ln q_{jt}^k} \equiv \ln q_{jt}^k + \sigma_x \eta_{jt}^k = \alpha_{jt}^k + \ln y - \varepsilon \ln h_j + (\varepsilon - 1)\ln P,$$

where $\ln y, \varepsilon, \ln h_j, \ln P$ are known/observed. 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.

5.2. Deriving the evolution of beliefs

Here we derive the evolution of beliefs of the firm, given the stochastic process for the firm's observations. This proof follows closely the proof in Bolton and Harris (1999), Lemma 1. Firm j samples n observations, indexed by k, each of which follows the independent process

$$dX_{jt}^k = \mu_j dt + \sigma_x dW_{jt}^k,$$

where W_{jt}^k is a Wiener process. Denoting by $p(t)$ the probability that $\mu_j = \bar{\mu}$, given all information up to time t (p_{jt} in the main text) and applying Bayes' rule,

$$\begin{aligned}
p(t+dt) &= \frac{Prob([dX_{jt}^k]_{k=1}^n, \mu_j = \bar{\mu})}{Prob([dX_{jt}^k]_{k=1}^n)} \\
&= \frac{Prob([dX_{jt}^k]_{k=1}^n | \mu_j = \bar{\mu}) Prob(\mu_j = \bar{\mu})}{Prob([dX_{jt}^k]_{k=1}^n | \mu_j = \bar{\mu}) Prob(\mu_j = \bar{\mu}) + Prob([dX_{jt}^k]_{k=1}^n | \mu_j = \underline{\mu}) Prob(\mu_j = \underline{\mu})} \\
&= \frac{p(t)\tilde{F}(\bar{\mu})}{p(t)\tilde{F}(\bar{\mu}) + (1-p(t))\tilde{F}(\underline{\mu})},
\end{aligned}$$

where

$$\begin{aligned}
\tilde{F}(\mu) &\equiv Prob([dX_{jt}^k]_{k=1}^n | \mu_j = \mu) \\
&= Prob\left(\left[dW_{jt}^k = \frac{dX_{jt}^k - \mu_j dt}{\sigma_x}\right]_{k=1}^n | \mu_j = \mu\right) \\
&= \prod_{k=1}^n \frac{1}{\sqrt{(2\pi dt)}} \exp\left[-\frac{1}{2dt}\left(\frac{dX_{jt}^k - \mu dt}{\sigma_x}\right)^2\right] \\
&= (2\pi dt)^{-\frac{n}{2}} \exp\left[-\frac{1}{2dt}\sum_{k=1}^n \left(\frac{dX_{jt}^k - \mu dt}{\sigma_x}\right)^2\right],
\end{aligned}$$

since $dW_{jt}^k \sim N(0, dt)$.

Drop the j and t indices in what follows (with the exception of μ_j, which still denotes the μ that firm j holds), to simplify notation. The change in beliefs is

$$dp = p(t+dt) - p(t) = \frac{p(1-p)(\tilde{F}(\bar{\mu}) - \tilde{F}(\underline{\mu}))}{p\tilde{F}(\bar{\mu}) + (1-p)\tilde{F}(\underline{\mu})},$$

and since

$$\begin{aligned}
\tilde{F}(\mu) &= (2\pi dt)^{-\frac{n}{2}} \exp\left[-\frac{1}{2dt\sigma_x^2}\sum_{k=1}^n \left((dX_k)^2 - 2dX_k\mu dt + \mu^2(dt)^2\right)\right] \\
&= (2\pi dt)^{-\frac{n}{2}} \exp\left[\sum_{k=1}^n \left(-\frac{(dX_k)^2}{2dt\sigma_x^2} + \frac{dX_k\mu}{\sigma_x^2} - \frac{\mu^2 dt}{2\sigma_x^2}\right)\right],
\end{aligned}$$

we have

$$dp = \frac{p(1-p)(2\pi dt)^{-\frac{n}{2}} \exp\left[\sum_{k=1}^{n}(-\frac{(dX_k)^2}{2dt\sigma_x^2})\right]\left(\hat{F}(\bar{\mu}) - \hat{F}(\underline{\mu})\right)}{(2\pi dt)^{-\frac{n}{2}} \exp\left[\sum_{k=1}^{n}(-\frac{(dX_k)^2}{2dt\sigma_x^2})\right]\left(p\hat{F}(\bar{\mu}) + (1-p)\hat{F}(\underline{\mu})\right)}$$
$$= \frac{p(1-p)(\hat{F}(\bar{\mu}) - \hat{F}(\underline{\mu}))}{p\hat{F}(\bar{\mu}) + (1-p)\hat{F}(\underline{\mu})},$$

where

$$\hat{F}(\mu) \equiv \exp\left[\sum_{k=1}^{n}\left(\frac{dX_k\mu}{\sigma_x^2} - \frac{\mu^2 dt}{2\sigma_x^2}\right)\right].$$

Under the Maclaurin expansion of the exponent,

$$\hat{F}(\mu) = 1 + \sum_{k=1}^{n}\left(\frac{dX_k\mu}{\sigma_x^2} - \frac{\mu^2 dt}{2\sigma_x^2}\right) + \frac{1}{2}\left[\sum_{k=1}^{n}(\frac{dX_k\mu}{\sigma_x^2} - \frac{\mu^2 dt}{2\sigma_x^2})\right]^2 + \ldots$$
$$= 1 + \sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2} - \sum_{k=1}^{n}\frac{\mu^2 dt}{2\sigma_x^2} + \frac{1}{2}\left[\sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2}\right]^2 - \sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2}\sum_{k=1}^{n}\frac{\mu^2 dt}{2\sigma_x^2}$$
$$+ \frac{1}{2}\left[\sum_{k=1}^{n}\frac{\mu^2 dt}{2\sigma_x^2}\right]^2 + \ldots$$
$$= 1 + \sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2} - \sum_{k=1}^{n}\frac{\mu^2 dt}{2\sigma_x^2} + \frac{1}{2}\sum_{k=1}^{n}\left[\frac{dX_k\mu}{\sigma_x^2}\right]^2 + \sum_{k=1}^{n}\sum_{i<k}\frac{dX_k\mu}{\sigma_x^2}\frac{dX_i\mu}{\sigma_x^2}$$
$$- \sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2}n\frac{\mu^2 dt}{2\sigma_x^2} + \frac{1}{2}n^2\left[\frac{\mu^2 dt}{2\sigma_x^2}\right]^2 + \ldots$$
$$= 1 + \sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2} - \sum_{k=1}^{n}\frac{\mu^2 dt}{2\sigma_x^2} + \sum_{k=1}^{n}\frac{\mu^2 dt}{2\sigma_x^2}$$
$$= 1 + \sum_{k=1}^{n}\frac{dX_k\mu}{\sigma_x^2},$$

since we ignore everywhere terms of order $dt^{\frac{3}{2}}$ and higher, and

$$(dX_k)^2 = \mu^2(dt)^2 - 2\mu dt dW_k\sigma_x + \sigma_x^2(dW_k)^2 = \sigma_x^2 dt,$$
$$dX_k dX_i = \mu^2(dt)^2 + \mu dt\sigma_x dW_k + \mu dt\sigma_x dW_i + \sigma_x^2 dW_k dW_i = 0,$$

since dW_k and dW_i are independent.

We get

$$dp = \frac{p(1-p)\sum_{k=1}^{n} dX_k \frac{\bar{\mu}-\underline{\mu}}{\sigma_x^2}}{1+\frac{p\bar{\mu}+(1-p)\underline{\mu}}{\sigma_x^2}\sum_{k=1}^{n} dX_k} \equiv \frac{p(1-p)\sum_{k=1}^{n} \frac{dX_k}{\sigma_x}\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}}{1+\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}},$$

where $\tilde{m}(p) \equiv \frac{p\bar{\mu}+(1-p)\underline{\mu}}{\sigma_x}$.

Since

$$\left(1+\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right)\left(1-\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right) = 1-(\tilde{m}(p))^2\frac{1}{\sigma_x^2}\left(\sum_{k=1}^{n}(dX_k)^2+\sum_{k=1}^{n}\sum_{i<k} dX_k dX_i\right) = 1-(\tilde{m}(p))^2 ndt,$$

and

$$\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\left(1+\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right)\left(1-\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right) = \sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\left(1-(\tilde{m}(p))^2 ndt\right)$$
$$= \sum_{k=1}^{n}\frac{dX_k}{\sigma_x} - (\tilde{m}(p))^2 n\frac{1}{\sigma_x}\sum_{k=1}^{n}\left(\mu(dt)^2+\sigma_x dt dW_k\right) = \sum_{k=1}^{n}\frac{dX_k}{\sigma_x},$$

we can write

$$\frac{\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}}{1+\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}} = \frac{\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\left(1+\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right)\left(1-\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right)}{1+\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}} = \sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\left(1-\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right),$$

and

$$\begin{aligned} dp &= p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\left(1-\tilde{m}(p)\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}\right) \\ &= p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\left(\sum_{k=1}^{n}\frac{dX_k}{\sigma_x}-\sum_{k=1}^{n}\tilde{m}(p)dt\right) \\ &= p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\frac{n}{\sigma_x}\left(\frac{\sum_{k=1}^{n}dX_k}{n}-(p\bar{\mu}+(1-p)\underline{\mu})dt\right) \\ &= p(1-p)\frac{\bar{\mu}-\underline{\mu}}{\sigma_x}\sqrt{n}d\tilde{W}, \end{aligned}$$

where

$$d\tilde{W} \equiv \frac{\sqrt{n}}{\sigma_x}\left(\frac{\sum_{k=1}^{n}dX_k}{n}-(p\bar{\mu}+(1-p)\underline{\mu})dt\right) = \frac{\sqrt{n}}{\sigma_x}\left(\overline{dX}-(p\bar{\mu}+(1-p)\underline{\mu})dt\right).$$

The sum of several independent Wiener processes is also a Wiener process. Since

$$dX_k = \mu_j dt + \sigma_x dW_k,$$

we can express

$$d\tilde{W} \equiv m_{\tilde{W}}dt + \sigma_{\tilde{W}}dW_t,$$

and find the drift and variance in the following way.

First, the drift term is given simply by the sum of all the drift terms in $d\tilde{W}$:

$$m_{\tilde{W}} = \frac{\sqrt{n}}{\sigma_x}\left(\sum_{k=1}^{n}\frac{\mu_j dt}{n}-(p\bar{\mu}+(1-p)\underline{\mu})dt\right) = \frac{\sqrt{n}}{\sigma_x}\left(\mu_j-(p\bar{\mu}+(1-p)\underline{\mu})\right)dt,$$

and the variance as

$$\sigma_{\tilde{W}} = \frac{n}{\sigma_x^2}\frac{nVar(dW_t)}{n^2} = \frac{n}{\sigma_x^2}\frac{n\sigma_x^2 dt}{n^2} = dt.$$

Thus,

$$\begin{gathered} d\tilde{W}_t = \frac{\sqrt{n}}{\sigma_x}\left(\mu_j-(p_t\bar{\mu}+(1-p_t)\underline{\mu})\right)dt + dW_t, \\ E(\mu_j-(p_t\bar{\mu}+(1-p_t)\underline{\mu})|I_t) = 0, \end{gathered}$$

$\tilde{W}$ has zero mean, conditional on information up to time t, i.e. $\tilde{W}$ is an observation-adapted Wiener innovation process.

5.3. Solution of the problem of Stage 2

In stage 2, the firm may still be learning about demand and updating its beliefs p_{jt}. Hence, the value function of the firm j is given by:

$$V(p_{jt}) = \sup_{<n_{js}>} E\left[\int_t^\infty \left(-fn_{js} + n_{js}D\phi_j^{\varepsilon-1}[p_{js}e^{\bar{\mu}} + (1-p_{js})e^{\underline{\mu}}]\right) e^{-r(s-t)}ds|p_{jt}\right],$$

subject to (1),(2). Note that we do not take into account the possibility that a firm that exited the market may choose to re-enter the market in the future, since in the model the aggregate variables and firm productivity are constant over time. When firm j is not selling in the market, it 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.

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

$$rv(p) = \max_{0\le n\le M}\left[n(-f + D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]) + n\frac{1}{2}(p(1-p)\chi)^2 v''(p)\right],$$

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

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

$$rv(p) = M\left(-f + D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]\right) + M\frac{1}{2}(p(1-p)\chi)^2 v''(p).$$

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:

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

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, so 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. The threshold value of p_{jt} below which the firm would quit the market (sell to 0 consumers) is obtained by setting the $v(p)$ above to 0 and is given by:

$$\underline{\underline{p_j}} = \max\left\{ \frac{\frac{f}{D\phi_j^{\varepsilon-1}} - e^{\underline{\mu}}}{e^{\bar{\mu}} - e^{\underline{\mu}}}, 0 \right\}.$$

5.4. Comparing the problem of Stage 1 with that of Moscarini and Smith (2001)

The problem formulated in Moscarini and Smith (2001) concerns the choice between two terminal actions, A and B, that have different payoffs in two possible states of the world ($\theta = H, L$): if p denotes the probability of high state $\theta = H$,

$$\pi_A(p) = p\pi_A^H + (1-p)\pi_A^L,$$
$$\pi_B(p) = p\pi_B^H + (1-p)\pi_B^L,$$

where $\pi_B^H > \pi_A^H$, $\pi_B^H > \pi_B^L$, and $\pi_B^L < \pi_A^L$.

The cost of experimentation is $c(n)$, twice differentiable on $(0,\infty)$, increasing and strictly convex on $[0,\infty)$. The main results in the paper require that either $c(0) > 0$ or $\pi_A(p) > 0$. Since in our case $\pi_A(p) \equiv 0$ (action A corresponds to quitting, and action B corresponds to entering at full scale in our case), assume that $c(0) > 0$. They also assume that $\lim_{n\to\infty} g(n) > r\max_{\theta,a}\pi_a^\theta$.

There is an observation process $[\bar{x}_t]$, which is a Brownian motion with constant uncertain drift μ^θ in state θ, $\mu_H = -\mu_L = \mu > 0$. More precisely, $\bar{x}_t$ follows

$$d\bar{x}_t^\theta = \mu^\theta dt + \frac{\sigma}{\sqrt{n_t}} dW_t,$$

in state θ, where W_t is a Wiener process, and n_t is the level or intensity of experimentation. Denoting by p_t the probability that $\theta = H$, given all information up to time t,

$$p_t = p_0 + \int_0^t p_s(1-p_s)\chi\sqrt{n_s}d\bar{W}_s,$$

where $d\bar{W}_s \equiv \frac{\sqrt{n_s}}{\sigma}(d\bar{x}_s - [p_s\mu + (1-p_s)(-\mu)]ds)$, the observation-adapted Wiener innovation process, and $\chi \equiv \frac{\mu-(-\mu)}{\sigma}$. Then the problem of the agent can be expressed as

$$V(p_0) = \sup_{T,<n_t>} E\left[\int_0^T -c(n)e^{-rt}dt + e^{-rT}\pi\left(p_0 + \int_0^T p_t(1-p_t)\chi\sqrt{n_t}d\bar{W}_t\right)|p_0\right],$$

i.e. the agent has to set the stopping time T, and the experimentation schedule $< n_t >, t \in [0,T]$.

The Hamilton-Jacobi-Bellman equation for the control problem is

$$rv(p) = \sup_{n \geq 0} \left(-c(n) + n \frac{p^2(1-p)^2\chi^2}{2} v''(p) \right),$$

subject to the value matching conditions:

$$v(\underline{p}) = \underline{p}\pi_A^H + (1-\underline{p})\pi_A^L, \quad v(\bar{p}) = \bar{p}\pi_B^H + (1-\bar{p})\pi_B^L.$$

The 'generalized Stefan' ODE problem for the optimal stopping problem given the control policy $n(p)$ is

$$rv(p) = -c(n(p)) + n(p)\frac{p^2(1-p)^2\chi^2}{2} v''(p),$$

or substituting $c'(n) = \frac{p^2(1-p)^2\chi^2}{2} v''(p)$ (from the FOC),

$$v''(p) = \frac{c'(z(rv(p)))}{\frac{p^2(1-p)^2\chi^2}{2}},$$

where $z \equiv g^{-1}$, and $g(n) \equiv nc'(n) - c(n)$, plus the value-matching conditions

$$v(\underline{p}) = \underline{p}\pi_A^H + (1-\underline{p})\pi_A^L, \quad v(\bar{p}) = \bar{p}\pi_B^H + (1-\bar{p})\pi_B^L,$$

and the smooth pasting conditions

$$v'(\underline{p}) = \pi_A^H - \pi_A^L, \quad v'(\bar{p}) = \pi_B^H - \pi_B^L.$$

Compare this problem with the one we have in our case:

$$v''(p) = \frac{c'(z(rv(p))) - (\tilde{D}_1 p + \tilde{D}_2(1-p))}{\frac{p^2(1-p)^2\chi^2}{2}}$$

where $z \equiv g^{-1}$ as before, $\chi \equiv \frac{\bar{\mu} - \underline{\mu}}{\sigma_x}$,

$$\tilde{D}_1 \equiv D\phi_j^{\varepsilon-1} e^{\bar{\mu}}, \quad \tilde{D}_2 \equiv D\phi_j^{\varepsilon-1} e^{\underline{\mu}},$$

plus the value-matching conditions

$$v(\underline{p}) = \underline{p}\tilde{\pi}_A^H + (1-\underline{p})\tilde{\pi}_A^L, \quad v(\bar{p}) = \bar{p}\tilde{\pi}_B^H + (1-\bar{p})\tilde{\pi}_B^L,$$

and the smooth pasting conditions

$$v'(\underline{p}) = \tilde{\pi}_A^H - \tilde{\pi}_A^L, \quad v'(\bar{p}) = \tilde{\pi}_B^H - \tilde{\pi}_B^L,$$

where

$$\tilde{\pi}_B^H \equiv \frac{M}{r}(-f + \tilde{D}_1) - F, \quad \tilde{\pi}_B^L \equiv \frac{M}{r}(-f + \tilde{D}_2) - F, \quad \tilde{\pi}_A^H \equiv 0, \quad \tilde{\pi}_A^L \equiv 0.$$

Notice that the two problems are extremely similar, with one difference being the additional linear term $(\tilde{D}_1 p + \tilde{D}_2(1-p))$ appearing in the nominator of the differential equation in our case. The presence of this term does not affect the derivation of the main results in Moscarini and Smith (2001), in particular those of the existence and uniqueness of a solution to the problem of the firm. The only way in which it does affect the problem, is that now we have to make sure that the nominator $c'(z(rv(p))) - (\tilde{D}_1 p + \tilde{D}_2(1-p))$ is positive for all p, since we need $v''(p) > 0$ for all $p \in [\underline{p}, \bar{p}]$. In the main text, we made the assumption

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

which implies that

$$c'(z(rv(p))) > D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}], \quad \forall p.$$

This is a sufficient condition for $v''(p) > 0$. However, even some $c(n)$ that do not satisfy this condition may still allow for the existence and uniqueness of a solution, if $c(n)$ is such that

$$c'(z(rv(p))) > D\phi_j^{\varepsilon-1}[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}], \quad \forall p \in [\underline{p}, \bar{p}].$$

Another condition that we have to impose is $c(0) > 0$, since in our problem $\tilde{\pi}_A^H \equiv 0$, $\tilde{\pi}_A^L \equiv 0$. Otherwise, as pointed out in Moscarini and Smith (2001), Appendix.B.4, boundary problems may arise: as p approaches $\underline{p}$ from above, $n(p)$ goes to 0, and $\underline{p}$ may not be reached in finite time with positive probability.

Given these two assumptions we impose ($c(0) > 0$, and $c'(z(0)) > D\phi_j^{\varepsilon-1}e^{\bar{\mu}}$), all the main proofs of Moscarini and Smith (2001) apply directly to our problem. The only results that need modification are the comparative statics results. Below we first show how introducing the profits in the experimentation phase (the linear term in the ODE nominator), affects the value function and the optimal experimentation behavior, and then re-derive the comparative statics results for our case.

5.4.1. The effect of including profits in the experimentation phase

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, and if it is 1, they do so fully.

$$V(p_t) = \sup_{T,<n_s>} E\left[e^{-rT}\pi(p_T) + \int_t^T \left(-c(n_s) + \Delta n_s A(p_s e^{\bar{\mu}} + (1-p_s)e^{\underline{\mu}})\right) e^{-r(s-t)} ds | p_t\right],$$

where $A \equiv y\frac{1}{\varepsilon-1}[\frac{\varepsilon}{\varepsilon-1}]^{-\varepsilon}[\frac{\phi_j}{w}]^{\varepsilon-1}P^{\varepsilon-1}$.

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

$$V_\Delta(p_t) = E\left[\int_0^{T^*} n_s^* A(p_s e^{\bar{\mu}} + (1-p_s)e^{\underline{\mu}})e^{-r(s-t)} ds | p_t\right] > 0,$$

where T^* and $n_s^*, s \in [t, T^*]$ are the optimal values of the stopping time and experimentation intensity, respectively. Hence, the value function increases for all p_t (shifts up) when Δ goes from 0 to 1. This implies, by Proposition 1 (Monotonicity in values) in Moscarini and Smith (2001), that in the case with testing profits in the value function, compared with the case without, $n(p_t) = z(rv(p_t))$ goes up for all p_t. Also, since the terminal payoff does not change in any way, and $v(p)$ is increasing and convex, and given the value matching and smooth pasting conditions, the threshold $\bar{p}$ goes up and the threshold $\underline{p}$ 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.

5.5. Comparative statics results

Proposition 1. *As any of real income y, aggregate price index P, firm productivity ϕ_j increases or wages w decrease, the optimal number of consumers in the experimentation stage, $n(p)$, shifts up, and the thresholds for quitting and for entering the market at full scale decrease, if $n(p)$ is sufficiently bounded away from the full scale M.* **Proof.** Consider the comparative statics with respect to $A \equiv y\frac{1}{\varepsilon-1}[\frac{\varepsilon}{\varepsilon-1}]^{-\varepsilon}[\frac{\phi_j}{w}]^{\varepsilon-1}P^{\varepsilon-1}$, which will effectively give us the comparative statics with respect to ϕ, P, w and y. The derivative of the value function

$$V(p_t) = \sup_{T,<n_s>} E\left[e^{-rT}\pi(p_T) + \int_t^T \left(-c(n_s) + n_s A(p_s e^{\bar{\mu}} + (1-p_s)e^{\underline{\mu}})\right) e^{-r(s-t)} ds | p_t\right],$$

with respect to A, using the envelope theorem, is

$$\begin{aligned} V_A(p_t) &= E\left[\int_t^{T^*} n_s^*(p_s e^{\bar{\mu}} + (1-p_s)e^{\underline{\mu}})e^{-r(s-t)}ds|p_t\right] \\ &\quad + Prob(p_{T^*} = \bar{p}|p_t, n^*, T^*)E[e^{-rT^*}|p_T = \bar{p}, p_t]\frac{M}{r}(\bar{p}e^{\bar{\mu}} + (1-\bar{p})e^{\underline{\mu}})) > 0, \end{aligned}$$

since $\pi(p_T) \equiv \max\left[0, \frac{M}{r}(p_T(e^{\bar{\mu}}A - f) + (1-p_T)(e^{\underline{\mu}}A - f)) - F\right]$.

Hence, we already know that the value function increases for all p_t as A rises, and since $n(p_t)$ is increasing in $v(p_t)$, 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. The value function of the problem for $A = A_1$, evaluated at $\underline{p}_0$ is, by linear approximation,

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

since $V(\underline{p}_0|A_0) = 0$, by value matching, and $V_A(p) > 0$. Hence, we get $\underline{p}_1 < \underline{p}_0$, since we require $V(\underline{p}_1|A_1) = 0$, and $V(p)$ is increasing in p . That is, the threshold for quitting goes down as A increases.

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

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

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

$$\begin{aligned} V_{pA}(p) &= V_{Ap}(p) \\ &= E_t\left[\int_t^{T^*} \left(n_s^*(e^{\bar{\mu}} - e^{\underline{\mu}}) + \frac{dn_s^*}{dp_t}(p_s e^{\bar{\mu}} + (1-p_s)e^{\underline{\mu}})\right)e^{-r(s-t)}ds\right] \\ &\quad + E_t\left[n_{T^*}^*(p_{T^*}e^{\bar{\mu}} + (1-p_{T^*})e^{\underline{\mu}})e^{-rT^*}\right]\frac{dE_t(T^*)}{dp_t} \\ &\quad + \frac{dProb(p_T = \bar{p}|p_t)}{dp_t}E_t[e^{-rT^*}|p_T = \bar{p}]\frac{M}{r}(\bar{p}e^{\bar{\mu}} + (1-\bar{p})e^{\underline{\mu}}) \\ &\quad + Prob(p_T = \bar{p}|p_t)\frac{dE_t[e^{-rT^*}|p_T = \bar{p}]}{dp_t}\frac{M}{r}(\bar{p}e^{\bar{\mu}} + (1-\bar{p})e^{\underline{\mu}}), \end{aligned}$$

where $E_t(x)$ now denotes $E(x|p_t)$. For p_t small enough and close to 0, $\frac{dE_t(T^*)}{dp_t} > 0$. This follows from the behavior of the expected remaining time until stopping, $E[T|p_t]$, which according to Moscarini and Smith (1998), obeys the boundary conditions $\tau(\underline{p}) = \tau(\bar{p}) = 0$, and the ODE

$$-1 = 0 + \frac{p^2(1-p)^2\chi^2}{2} n(p)\tau''(p),$$

i.e. $\tau''(p) < 0$, and $\tau(p)$ is hill-shaped. For any p_t, we have

$$\frac{dProb(p_T = \bar{p}|p_t)}{dp_t} > 0,$$

and

$$\frac{dE_t[e^{-rT^*}|p_T = \bar{p}]}{dp_t} > 0.$$

So even if p_t is close to 1, so that $\frac{dE_t(T^*)}{dp_t} < 0$, if the last two terms in the expression for the cross-derivative dominate the second term ($E_t[n^*_{T^*}(p_{T^*}e^{\bar{\mu}} + (1-p_{T^*})e^{\underline{\mu}})e^{-rT^*}]\frac{dE_t(T^*)}{dp_t}$) in absolute value, then the cross-derivative is still positive. In particular, if the term $\frac{M}{r}(\bar{p}e^{\bar{\mu}} + (1-\bar{p})e^{\underline{\mu}})$ is large enough compared to $E_t[n^*_{T^*}(p_{T^*}e^{\bar{\mu}} + (1-p_{T^*})e^{\underline{\mu}})e^{-rT^*}]$, then we can obtain a positive cross-derivative. That is, if the optimal experimentation intensity n^* is sufficiently bounded away from M, so that $n^*_{T^*}$ is relatively small compared to the entire market size M, then the threshold for entering the market, $\bar{p}$, falls as A increases. This is intuitive: suppose the opposite holds, and the optimal testing intensity is large, relative to the full scale. 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 will dominate the pure testing costs ($c(n)$). Hence, as A rises, the threshold for entering will in fact rise, rather than fall. ■

Proposition 2. *As sunk entry cost F rises, so that the final payoff to entry falls, the optimal number of consumers in the experimentation stage, $n(p)$, shifts down, and the thresholds for quitting and for entering the market at full scale increase.*

Proof. The derivative of the value function with respect to F, using the envelope theorem, is

$$V_F(p_t) = -Prob(p_{T^*} = \bar{p}|p_t, n^*, T^*)E_t[e^{-rT^*}|p_T = \bar{p}] < 0,$$

since $\pi(p_T) \equiv \max\left[0, \frac{M}{r}(A(p_T e^{\bar{\mu}} + (1-p_T)e^{\underline{\mu}}) - f) - F\right]$.

Hence, the value function shifts down as F increases. Since $n(p) \equiv z(rv(p))$, and z is increasing, we know that $n(p)$ also shifts down. Consider now two values of F, $F_1 > F_0$, 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. The value function of the problem for $F = F_1$, evaluated at $\underline{p}_0$ is, by linear approximation,

$$V(\underline{p}_0|F_1) \simeq V(\underline{p}_0|F_0) + V_F(\underline{p}_0|F_0)(F_1 - F_0) < 0,$$

since $V(\underline{p}_0|F_0) = 0$, by value matching, and $V_F(p) < 0$. Hence, we get $\underline{p}_1 > \underline{p}_0$, since we require $V(\underline{p}_1|F_1) = 0$, and $V(p)$ is increasing in p . That is, the threshold for quitting increases as F increases.

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

$$\begin{aligned} V_p(\bar{p}_0|F_1) &\simeq V_p(\bar{p}_0|F_0) + V_{pF}(\bar{p}_0|F_0)(F_1 - F_0) \\ &= M\frac{A(e^{\bar{\mu}} - e^{\underline{\mu}})}{r} + V_{Fp}(\bar{p}_0|F_0)(F_1 - F_0) \\ &= M\frac{A(e^{\bar{\mu}} - e^{\underline{\mu}})}{r} - \frac{dProb(p_{T^*} = \bar{p}|p_t, n^*, T^*)}{dp_t} E_t[e^{-rT^*}|p_T = \bar{p}] - \\ &- Prob(p_{T^*} = \bar{p}|p_t, n^*, T^*) \frac{dE_t[e^{-rT^*}|p_T = \bar{p}]}{dp_t}. \end{aligned}$$

We have

$$\frac{dProb(p_{T^*} = \bar{p}|p_t, n^*, T^*)}{dp_t} > 0,$$

and

$$\frac{dE_t[e^{-rT^*}|p_T = \bar{p}, p_t]}{dp_t} > 0.$$

Hence,

$$V_p(\bar{p}_0|F_1) < M\frac{A(e^{\bar{\mu}} - e^{\underline{\mu}})}{r},$$

and since we need $V_p(\bar{p}_1|F_1) = M\frac{A(e^{\bar{\mu}} - e^{\underline{\mu}})}{r}$, and $V(p)$ is convex, we know that $\bar{p}_1 > \bar{p}_0$, i.e. the threshold for entering the market at full scale, $\bar{p}$, increases as F increases. ■

Proposition 3. *As the testing cost function $c(n)$ grows more convex and initially (for n close to 0) weakly higher and steeper (i.e. $c(n)$ is replaced by $\hat{c}(n)$, where $\hat{c}(0) = c(0)$, $\hat{c}'(0) \geq c'(0)$, and $\hat{c}''(n) \geq c''(n)$ for all n), the optimal number of consumers in the experimentation stage, $n(p)$, shifts down, the threshold for quitting increases, and the threshold for entering the market at full scale decreases.*

Proof. The proof of Moscarini and Smith (2001), Proposition 5 (c), applies directly. ■

5.6. Showing that an estimate of the sunk entry cost F is higher when fitting the experimentation model than when fitting an alternative model to the same dataset

In a given exporter dataset, locate the lowest productivity exporter. Denote this firm's productivity by $\underline{\phi}$. In a model where there is no experimentation and all exporters sell at full scale right away, the following condition should hold:

$$\frac{M}{r}\left(-f+D\underline{\phi}^{\varepsilon-1}[p_0e^{\bar{\mu}}+(1-p_0)e^{\underline{\mu}}]\right)-F=0,$$

so that the estimate of F would be $\hat{F}_{alt}=\frac{M}{r}(-f+D\underline{\phi}^{\varepsilon-1}[p_0e^{\bar{\mu}}+(1-p_0)e^{\underline{\mu}}])$. In the experimentation model, from Figure 5, this firm's threshold for quitting $\underline{p}(\underline{\phi})$ is lower than p_0. Since this firm is lowest productivity exporter, it should satisfy the condition

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

From Proposition 2, for fixed values of demand parameters and testing costs, the only way to do this is increase F, until $\underline{p}(\underline{\phi})$ increases sufficiently to satisfy the above condition. Thus, $\hat{F}_{ex}>\hat{F}_{alt}$.

5.7. Estimating demand parameters

We would like to evaluate the posteriors of the parameters $\Theta_1\equiv\{\bar{\mu},\underline{\mu},\sigma_x,p_0\}$, as well as $\{A_{t1},A_{t2},t=1,...,T\}$ and $\rho_1,\rho_2,\gamma_1,\gamma_2,\zeta_1,\zeta_2,\Omega$, conditional on $\{\upsilon^k_{jt}\}$. It can be seen from the equations above that the unobserved α^k_{jt} and A_{t1},A_{t2} can be identified up to a constant, and therefore we normalize $\underline{\mu}=0$. 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 $\tilde{H}\equiv\{H_j,j=1,...,J\}$. $\tilde{H}$ will be estimated along with the main unknown parameters.

We set the following priors:

- $\bar{\mu}\sim N(m_h,\sigma_h)$,
- $\frac{1}{\sigma_x}\sim Gamma(\alpha_x,\beta_x)$,
- $p_0\sim Beta(\alpha_p,\beta_p)$.
- Impose flat priors on ζ_1,ζ_2.
- Impose flat priors, bounded by -1 and 1, on ρ_1,ρ_2.
- Impose flat priors on γ_1,γ_2.
- Impose a flat prior on the coefficients of the matrix Ω, the covariance of the error terms ν_{t1},ν_{t2}.

The joint posterior of these parameters, given $\{\upsilon^k_{jt}\}$, is proportional to the product of their priors and their joint likelihood:

$$f(\bar{\mu},\sigma_x,p_0,\rho_1,\rho_2,\zeta_1,\zeta_2,\gamma_1,\gamma_2,\Omega|\upsilon^k_{jt})\infty$$
$$\infty f(\bar{\mu})f(\frac{1}{\sigma_x})f(p_0)f(\rho_1,\rho_2)f(\zeta_1,\zeta_2)f(\gamma_1,\gamma_2)f(\Omega)*$$
$$*f(\upsilon^k_{jt}|\bar{\mu},\sigma_x,A_{t1},A_{t2},\gamma_1,\gamma_2,\tilde{H})f(\tilde{H}|p_0)f(A_{t1},A_{t2}|\rho_1,\rho_2,\zeta_1,\zeta_2).$$

Since the joint posterior is quite complicated, we apply Gibbs sampling. Denote by $\tilde{\Theta}_1$ the set $\{\Theta_1,\rho_1,\rho_2,\zeta_1,\zeta_2,\gamma_1,\gamma_2,\Omega\}$, and by $\tilde{\Theta}_1^{-par}$ the set $\tilde{\Theta}_1$ without the parameter *par*.

The conditional distributions of the main parameters of interest are as follows:

- $f(\bar{\mu}|\tilde{\Theta}_1^{-\bar{\mu}},\{\upsilon^k_{jt}\},\{A_{t1},A_{t2}\},\tilde{H})$ is a normal with mean $\frac{m_h\sigma_x+N_h\bar{\alpha}_h\sigma_h}{\sigma_x+N_h\sigma_h}$ and variance $\frac{\sigma_x\sigma_h}{\sigma_x+N_h\sigma_h}$, where N_h is the number of α^k_{jt} values of all high-μ firms and $\bar{\alpha}_h\equiv\frac{\sum_{j:H_j=1}\sum_t\sum_k\alpha^k_{jt}}{N_h}$ is the average of these α^k_{jt}. Given υ^k_{jt}, γ_1,γ_2 and $A_{t1},A_{t2},t=1,..,T$, α^k_{jt} is calculated as
$$\upsilon^k_{jt}-\gamma_1A_{t1}-\gamma_2A_{t2}.$$
- $f(\frac{1}{\sigma_x}|\tilde{\Theta}_1^{-\sigma_x},\upsilon^k_{jt},\{A_{t1},A_{t2}\},\tilde{H})$ is Gamma with hyperparameters $\alpha_x+\frac{N_x}{2}$ and $\beta_x+\frac{\sum_{j,t,k}(e^k_{jt}-\bar{e})^2}{2}$, where $e^k_{jt}=\alpha^k_{jt}-\bar{\mu}$ if $H_j=1$ and $e^k_{jt}=\alpha^k_{jt}-\underline{\mu}$ if $H_j=0$, $\bar{e}$ is the average and N_x is the number of these residuals.
- $f(p_0|\tilde{\Theta}_1^{-p_0},\upsilon^k_{jt},\{A_{t1},A_{t2}\},\tilde{H})$ is Beta with hyperparameters $\alpha_p+N_{\tilde{h}}$ and $\beta_p+N-N_{\tilde{h}}$, where $N_{\tilde{h}}$ is the number of high-μ firms and N is the total number of firms.
- The conditional distribution of H_j can be written as $P(H_j=1|\tilde{\Theta}_1,\upsilon^k_{jt},\{A_{t1},A_{t2}\})$. To calculate this, we apply the discrete time Bayesian updating equation:
$$Prob(\mu_j=\bar{\mu}|\tilde{\Theta}_1,\upsilon^k_{jt},\{A_{t1},A_{t2}\})\equiv P_j=\frac{\psi\left(\frac{\bar{\alpha}_j-\bar{\mu}}{\frac{\sigma_x}{\sqrt{n_j}}}\right)p_0}{\psi\left(\frac{\bar{\alpha}_j-\bar{\mu}}{\frac{\sigma_x}{\sqrt{n_j}}}\right)p_0+\psi\left(\frac{\bar{\alpha}_j-\underline{\mu}}{\frac{\sigma_x}{\sqrt{n_j}}}\right)(1-p_0)},$$
where $\bar{\alpha}_j$ is the average over all t and k of α^k_{jt} for firm j, and n_j is the number of these draws. ψ is the standard normal density.
- To evaluate the posteriors of A_{t1},A_{t2}, conditional on the rest of the parameters and υ^k_{jt}, we apply the Kalman smoother to the system
$$\begin{aligned}\upsilon^k_{jt}&=\gamma_1A_{t1}+\gamma_2A_{t2}+\alpha^k_{jt}.\\ \alpha^k_{jt}&=\bar{\mu}+\sigma_x\eta^k_{jt},\quad\text{if}\quad H_j=1,\\ \alpha^k_{jt}&=\underline{\mu}+\sigma_x\eta^k_{jt},\quad\text{if}\quad H_j=0,\\ A_{t1}&=\rho_1A_{(t-1)1}+\zeta_1+\nu_{t1},\\ A_{t2}&=\rho_2A_{(t-1)2}+\zeta_2+\nu_{t2},\end{aligned}$$

where $[v_{t1}, v_{t2}] \sim N([0,0],\Omega)$, $\gamma_1, \gamma_2, \bar{\mu}, \underline{\mu}$, $\{H_j, j = 1,...,J\}$, $\rho_1, \rho_2, \zeta_1, \zeta_2, \Omega$ are fixed and known.

- Given A_{t1}, A_{t2}, regress A_{t1} on $A_{(t-1)1}$ to update values of ζ_1, ρ_1, and regress A_{t2} on $A_{(t-1)2}$ to update values of ζ_2, ρ_2.
- Given A_{t1}, A_{t2}, and $\zeta_1, \zeta_2, \rho_1, \rho_2$, calculate $v_{t1} = A_{t1} - \rho_1 A_{(t-1)1} - \zeta_1$, $v_{t2} = A_{t2} - \rho_2 A_{(t-1)2} - \zeta_2$, and update Ω.
- Given A_{t1}, A_{t2}, υ^k_{jt}, $\bar{\mu}, \underline{\mu}$, $\tilde{H}$, regress υ^k_{jt} on A_{t1}, A_{t2} and firm-specific constant μ, to update the values of γ_1, γ_2.

We iterate on the above steps, until convergence.

5.8. Estimating TFP (total factor productivity)

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 all 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. We borrow many steps from De Loecker (2011). Begin with the pair of production and demand equations:

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

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

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):

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

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

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

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 (2002) 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 $\tilde{\omega}$ (TFP) 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

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

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 Z the combined input $L_{jt}^{\alpha_l} M_{jt}^{\alpha_m} K_{jt}^{\alpha_k}$, and by c_Z the average cost of this input: $c_Z \equiv \frac{wL+rK+mM}{L_{jt}^{\alpha_l} M_{jt}^{\alpha_m} K_{jt}^{\alpha_k}}$, 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:

$$P^*_{jt} = \tau P_{jt} = \tau \frac{c_Z}{e^{\omega_{jt}}} \frac{\varepsilon}{\varepsilon - 1},$$

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

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

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 ($\tilde{\omega}$). 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

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

where D_{st} are industry-year dummies. 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 (2006)) critique of the Levinsohn and Petrin (2003) and Olley and Pakes (1996) 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:

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

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:

$$\begin{aligned}
\widetilde{va}_{jt} \equiv \tilde{r}_{jt} - \beta_m m_{jt} &= \beta_l l_{jt} + \beta_k k_{jt} + \psi(m_{jt}, k_{jt}) + \sum_{t=1}^{T} \beta_{st} D_{st} + \varepsilon_{jt} \\
&= \beta_l l_t(k_{jt}, \psi(m_{jt}, k_{jt})) + \beta_k k_{jt} + \psi(m_{jt}, k_{jt}) + \sum_{t=1}^{T} \beta_{st} D_{st} + \varepsilon_{jt} \\
&\equiv \Psi_t(m_{jt}, k_{jt}) + \sum_{t=1}^{T} \beta_{st} D_{st} + \varepsilon_{jt},
\end{aligned}$$

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

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

where D_t are year dummies. 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

$$\tilde{\omega}_{jt} = \widetilde{va}_{jt} - \beta_l l_{jt} - \beta_k k_{jt} - \sum_{t=1}^{T} \beta_{st} D_{st},$$

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

Next, use the moments

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

to identify the coefficients on capital and labor.

Finally, given estimates $\hat{\beta}_k, \hat{\beta}_l$, the estimate of TFP, $\widehat{\tilde{\omega}_{jt}}$, is given by

$$\widehat{\tilde{\omega}_{jt}} = \widetilde{va}_{jt} - \hat{\beta}_l l_{jt} - \hat{\beta}_k k_{jt} - \sum_{t=1}^{T} \hat{\beta}_{st} D_{st}.$$

5.9. 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:

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

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

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 and smooth pasting conditions:

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

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 is now given by

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

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 and smooth pasting conditions:

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

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.

5.10. Normalizing $f = 0$

There are two equations that allow us to estimate this parameter in the model: the boundary condition from the experimentation stage

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F = \frac{M}{r}\left(-f + \tilde{\pi}_j[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}]\right) - F,$$

and the exit threshold on beliefs from the full-scale exports stage:

$$\underline{\underline{p_j}} = \max\left\{\frac{\frac{f}{D\phi_j^{\varepsilon-1}} - e^{\underline{\mu}}}{e^{\bar{\mu}} - e^{\underline{\mu}}}, 0\right\}.$$

We assume that all exits in the full-scale exports stage happen due to exogenous shocks. Thus, we can set $\underline{\underline{p_j}} = 0$, and focus on the first equation for evaluating f. Replace F with $F' \equiv F + \frac{M}{r}f$, and f with $f' = 0$:

$$v(\bar{p}) = \tilde{V}(\bar{p}) - F' = \tilde{V}(\bar{p}) - F - \frac{M}{r}f = \frac{M}{r}\tilde{\pi}_j[pe^{\bar{\mu}} + (1-p)e^{\underline{\mu}}] - F - \frac{M}{r}f.$$

Notice that the last expression is identical to the original boundary condition. Thus, F and f are not separately identifiable: for any F and f, the same kind of behavior will be produced by $F' \equiv F + \frac{M}{r}f$ and $f' = 0$. Therefore, we normalize $f = 0$ and focus on estimating F.

5.11. Estimating $\Theta_2 \equiv \{M, G, F, \delta, \sigma_N, m_p\}$, given estimates of $\bar{\mu}, \sigma_x, p_0$ and beliefs p

In the second step of the estimation, we treat the values of $\Theta_1 \equiv \{\bar{\mu}, \underline{\mu}, \sigma_x, p_0\}$ and beliefs $p_{jt}, j = 1, ..., J, t = 1, ..., T$ as given and fixed. Here we describe how we calculate the likelihood of the data, given arbitrary values of $\Theta_2 \equiv \{G, F, M, \delta, \sigma_N, m_p\}$, and these fixed values of Θ_1, and how we use this likelihood to estimate Θ_2.

- Fix the values of δ, σ_N, m_p, M. We fix the initial value of σ_N to 1, of m_p to 0.01, of δ to an estimate of the exogenous death rate based on the subsample of old exporters alone, and of M to the average number of export destinations over all old exporters. We do so because the old exporters are assumed to have completed their experimentation phase (if any), and therefore their average export size should be a good indicator of the full scale in the market, and their average exit rate should be a good indicator of the exogenous death rate. Starting the iteration from different values of δ, σ_N, m_p, M does not affect the results.
  For fixed δ, σ_N, m_p, M, and $r \equiv \delta + dr$, $dr = 0.05$, we can calculate the likelihood of the data

for any given values of the coefficients of the cost function $c(n)$ and the sunk entry cost F as

$$
\begin{aligned}
L(n_{jt}|G,F,\sigma_N,m_p,\delta,M) &\equiv Prob(n_{jt}|G,F,\sigma_N,m_p,\delta,M) \\
&= \prod_{E^0_{jt}} \psi\left(\frac{n_{jt}-n^*(p_{jt})}{\sigma_N}\right) \prod_{E^1_{jt}} \psi\left(\frac{n_{jt}-M}{\sigma_N}\right) * \\
&\prod_{X^1_{jt}} \delta \prod_{R^1_{jt}} (1-\delta) \prod_{Q^0_{jt}} \kappa(0|m_p) * \\
&\prod_{Q^1_{jt}} \delta \prod_{R^0_{jt}} (1-\delta)\kappa(p_{jt}-\underline{p}_j|m_p) \prod_{R^2_{jt}} (1-\delta),
\end{aligned}
$$

where ψ denotes the standard normal distribution, κ denotes the exponential distribution (for the discrepancy between the predicted $\underline{p}_j$ and the actual $\underline{\tilde{p}}_j$),

- $E^0_{jt} = [j,t : S_{jt} = 0]$, that is, all j,t, where the firm exports and is in the experimentation stage,
- $E^1_{jt} = [j,t : S_{jt} = 1]$, that is, all j,t, where the firm exports and is in the full export stage,
- $X^1_{jt} = [j,t : S_{jt} = 2, S_{j(t-1)} = 1]$, 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,
- $Q^0_{jt} = [j,t : S_{jt} = 2, S_{j(t-1)} = 0, p_{jt} \leq \underline{p}_j]$, that is, all j,t, where the firm exits the market from State 0, given its beliefs are below $\underline{p}_j$, which allows us to set $\underline{\tilde{p}}_j = \underline{p}_j$, to maximize the likelihood, so that $e_{pd} = \underline{p}_j - \underline{\tilde{p}}_j = 0$,
- $Q^1_{jt} = [j,t : S_{jt} = 2, S_{j(t-1)} = 0, p_{jt} > \underline{p}_j]$, that is, all j,t, where the firm exits the market from State 0, given its beliefs are above $\underline{p}_j$, which implies that the exit was caused by an exogenous death shock,
- $R^0_{jt} = [j,t : S_{jt} = 0, S_{j(t-1)} = 0, p_{jt} \leq \underline{p}_j]$, that is, all j,t, where the firm remains in the market in State 0, given its beliefs are below $\underline{p}_j$, the probability of which happening is $(1-\delta)\kappa(\underline{\tilde{p}}_j - \underline{p}_j|m_p)$, and we set $\underline{\tilde{p}}_j = p_{jt}$, to maximize the likelihood,
- and $R^2_{jt} = [j,t : S_{jt} < 2, S_{j(t-1)} = 0, p_{jt} \geq \underline{p}_j]$, that is, all j,t, where the firm remains in the market in State 0, given its beliefs are above $\underline{p}_j$, which happens with probability $(1-\delta)$.

We set flat priors for G and F, and pick new values for these parameters using the Metropolis-Hastings step, which essentially amounts to maximising the above likelihood in our case.

- Given the latest values of G and F, we can reevaluate the posteriors of δ, σ_N, m_p, M.
  - The prior distribution of σ_N is an inverse gamma distribution with hyperparameters α_N, β_N, so that

$$
f\left(\frac{1}{\sigma_N}|\alpha_N,\beta_N\right) = \frac{\frac{1}{\sigma_N}^{\alpha_N-1} e^{-\frac{1}{\sigma_N \beta_N}}}{\Gamma(\alpha_N)\beta_N^{\alpha_N}}, \quad if \quad \sigma_N > 0,
$$

and 0 otherwise. The posterior distribution of σ_N is also an inverse gamma with hyperparameters

$$\alpha_N' = \alpha_N + \frac{nb_N}{2}, \beta_N' = \beta_N + \frac{\sum_{i=1}^{nb_N}(e_i - \bar{e})^2}{2},$$

where nb_N is the number of observations on export size (number of destinations) we have, e_i are the deviations of the observed export size n from the predicted export size n^*, and $\bar{e}$ is the mean of these e_i.

– The prior distribution of δ is a beta distribution with hyperparameters $\alpha_\delta, \beta_\delta$, so that

$$f(\delta|\alpha_\delta, \beta_\delta) = \frac{\Gamma(\alpha_\delta + \beta_\delta)}{\Gamma(\alpha_\delta)\Gamma(\beta_\delta)} \delta^{\alpha_\delta - 1}(1-\delta)^{\beta_\delta - 1}, \quad if \quad 0 < \delta < 1,$$

and 0 otherwise. The posterior of δ is also a beta distribution with hyperparameters

$$\alpha_\delta' = \alpha_\delta + \sum_{i=1}^{nb_\delta} x_i, \beta_\delta' = \beta_\delta + nb_\delta - \sum_{i=1}^{nb_\delta} x_i,$$

where x_i is an indicator variable equal to 1 if a firm exits the market due to an exogenous shock, and nb_δ is the number of observations where we can infer this indicator variable's value.

– The prior distribution of m_p is an inverse gamma with hyperparameters α_d, β_d, so that

$$f\left(\frac{1}{m_p}|\alpha_d, \beta_d\right) = \frac{(\frac{1}{m_p})^{\alpha_d - 1} e^{-\frac{1}{m_p \beta_d}}}{\Gamma(\alpha_d)\beta_d^{\alpha_d}}, \quad if \quad m_p > 0,$$

and 0 otherwise. The posterior of m_p is also an inverse gamma distribution with the hyperparameters

$$\alpha_d' = \alpha_d + nb_p, \beta_d' = \frac{\beta_d}{1 + \beta_d \sum_{i=1}^{nb_p}(\underline{p}_j - \tilde{\underline{p}}_j)},$$

where nb_p is the number of all observations where we can infer the value of $e_{pd} = \underline{p}_j - \tilde{\underline{p}}_j$, that is sets Q_{jt}^0 and R_{jt}^0 described above.

– The prior distribution of M is a normal distribution with hyperparameters m, σ_m, so that

$$f(M|m, \sigma_m) = \sqrt{(}\frac{1}{2\sigma_m \pi})e^{-\frac{1}{2\sigma_m}(M-m)^2},$$

and the posterior of M, given σ_N, and the observed export sizes in state 1 (full-scale export), $\{n_{jt}|S_{jt} = 1\}$, is a normal distribution with hyperparameters

$$m' = \frac{m\sigma_N + N_m \bar{n} \sigma_m}{\sigma_N + N_m \sigma_m},$$

$$\sigma_m' = \frac{\sigma_m \sigma_N}{\sigma_N + N_m \sigma_m},$$

where $\bar{n}$ is the average over all $\{n_{jt}|S_{jt} = 1\}$, and N_m is the number of all such observations.

- Iterate on the previous two steps until convergence.

5.12. Evaluating the number of new exporters in counterfactual exercises

Recall that we used the normalized version of profits, $\tilde{\pi}_j \equiv \frac{\exp\left[(\varepsilon-1)\frac{C_\phi^j}{(1-\rho_\phi)}\right]}{m}$, where m is the median of $\exp\left[(\varepsilon-1)\frac{C_\phi^j}{(1-\rho_\phi)}\right]$ over all exporters in the given pair, to characterize the heterogeneity of exporters in a given sector. We locate the first percentile of this $\tilde{\pi}_j$ over all exporters in a given pair (rather than the minimum, which helps us disregard outliers), and call this the exporting threshold in this pair, $\tilde{\pi}_{ex}^0$. This value also implies an exporting threshold for the smoothed average productivity, $\tilde{\phi}_{ex}^0 = \exp\left[\frac{\ln(\tilde{\pi}_{ex}^0 m)}{\varepsilon-1}\right]$. Suppose that all profits $\tilde{\pi}_j$ go up by some factor γ^{ε_k} (as they did in the currency appreciation exercise, where the decrease in the exchange rate from $3e_k$ to e_k resulted in the shift in profits by a factor of 3^{ε_k}). Denote the new profit variable by $\tilde{\pi}_j'$:

$$\tilde{\pi}_j' = \tilde{\pi}_j \gamma^{\varepsilon_k}.$$

All firms with $\tilde{\pi}_j'$ at or above the exporting threshold $\tilde{\pi}_{ex}^0$ will export now. That is, in terms of the old profit variable, all firms with $\tilde{\pi}_j$ at or above $\tilde{\pi}_{ex}^1$, where $\tilde{\pi}_{ex}^1 \equiv \frac{\tilde{\pi}_{ex}^0}{\gamma^{\varepsilon_k}}$, will now export. Thus, firms with profitability $\tilde{\pi}_j$ between $\tilde{\pi}_{ex}^1$ and $\tilde{\pi}_{ex}^0$ will be new exporters. $\tilde{\pi}_{ex}^1$ can also be used to deduce the new exporting threshold for the smoothed average productivity,

$$\tilde{\phi}_{ex}^1 = \exp\left[\frac{\ln(\tilde{\pi}_{ex}^1 m)}{\varepsilon_k - 1}\right] = \frac{\exp\left[\frac{\ln(\tilde{\pi}_{ex}^0 m)}{\varepsilon_k-1}\right]}{\exp\left[\frac{\varepsilon_k}{\varepsilon_k-1}\ln\gamma\right]} = \frac{\tilde{\phi}_{ex}^0}{\exp\left[\frac{\varepsilon_k}{\varepsilon_k-1}\ln\gamma\right]} = \frac{\tilde{\phi}_{ex}^0}{\gamma^{\frac{\varepsilon_k}{\varepsilon_k-1}}}.$$

To know how many of these new exporters will be there, we need to turn to the observed distribution of productivities in the sector concerned. 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 (in our case, codes 4202 and 6403) may come from different sectors. The exporters of trunks, suitcases, etc. belong to 19 various sectors, the largest sector being Clothing and furs (20 percent), and the exporters of footwear belong to 15 various sectors, the largest being Leather products and shoes (50 percent). We therefore treat Clothing and furs as the main sector for exporters of trunks, suitcases, etc., and Leather products and shoes - as the main sector for exporters of footwear.

First, we fit a Pareto distribution to the smoothed average productivities within the main sector:

$$\ln(1 - CF(\tilde{\phi}_j)) = \varsigma(\ln \tilde{\phi}_m - \ln \tilde{\phi}_j),$$

where $\tilde{\phi}_m$ is the minimum smoothed average productivity in the group, ς is the shape parameter, and CF is the cumulative distribution. We obtain a shape parameter of 1.03 and 1.94 for Clothing and furs and Leather products and shoes, respectively. We apply the above equation to find the cumulative distribution of both values, $\tilde{\phi}^0_{ex}$, $\tilde{\phi}^1_{ex}$, and take the difference between the two. We then multiply this fraction by the total number of firms in each sector to find the number of new exporters.

5.13. Supplementary Figures and Tables

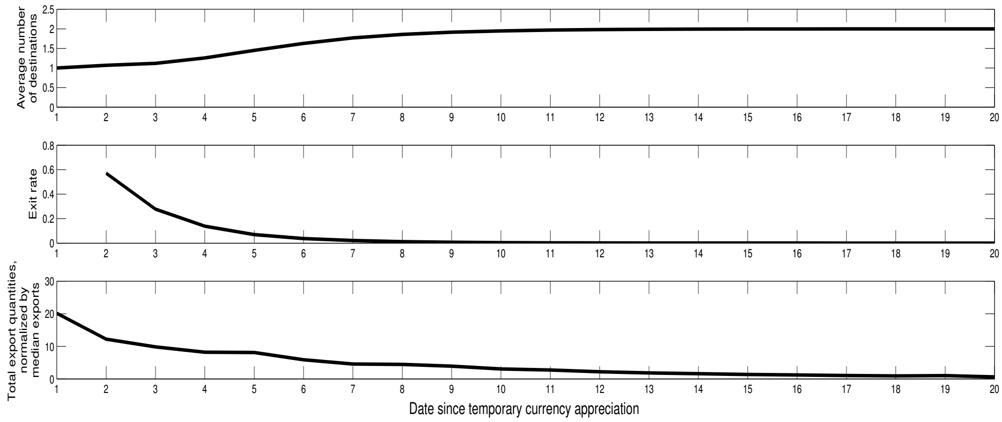

Figure 17 – Response to a 67 % temporary currency appreciation in a single country within East Asia, exports of footwear. The first panel shows the evolution of the average number of destinations (over all new exporters), the second panel - the exit rate, and the third panel - the total quantity exported, normalized by the median quantity exported.

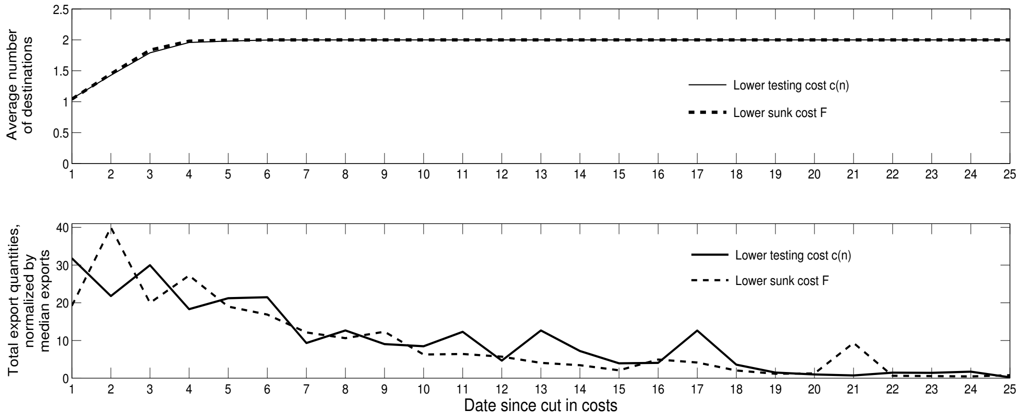

Figure 18 – Response to a 10 % decrease in the testing costs $c(n)$ and to an 'equivalent' cut in the sunk entry cost F. Exports of footwear by French firms to East Asia. The first panel shows the evolution of the average number of destinations (over all new exporters), and the second panel - the total quantity exported, normalized by the median quantity exported.

Table 3 – The 17 regions

Region	Members
Caribbean islands	Antigua&Barbuda, Anguilla, Aruba, Barbados, Bermuda, Bonaire, St. Eustatius&Saba, Bahamas, Cuba, Curacao, Dominica, Dominican Rep., Haiti, Jamaica, St. Kitts&Nevis, Virgin Isd. (U.S.), Cayman Isd., Saint Lucia, St. Martin (Dutch), Turks&Caicos Isd., Trinidad&Tobago, Grenada, St. Vincent&the Grenadines, Virgin Isd. (Brit)
Central Africa	Angola, The Democratic Rep. of Congo, Central African Rep., Congo, Cameroon, Gabon, Equatorial Guinea, Sao Tome&Principe, Chad
Eastern Asia	China, Hong Kong, Japan, Rep. of Korea, D.P. Rep. of Korea, Mongolia, Macao, Taiwan
Eastern Africa	Burundi, Djibouti, Eritrea, Ethiopia, Kenya, Comoros, Madagascar, Mauritius Malawi, Mozambique, Rwanda, Seychelles, Somalia, Tanzania, Uganda, Zambia
Eastern Europe	Albania, Bosnia and Herzegovina, Bulgaria, Czech Rep., Estonia, Croatia Hungary, Lithuania, Latvia, Montenegro, Macedonia, Poland, Romania, Serbia, Slovenia, Slovakia
European Union, the first 15 members, excluding France	Austria,Belgium, Germany, Denmark, Spain, Finland, UK, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Sweden
Latin America	Argentina, Bolivia, Brazil, Belize, Chile, Venezuela,Colombia, Costa Rica, Ecuador, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Peru, Paraguay, Suriname, El Salvador, Uruguay
Middle East	United Arab Emirates, Bahrain, Israel, Iraq, Iran, Jordan, Kuwait, Lebanon Oman, Palestinian Terr., Qatar, Saudi Arabia, Syria, Turkey, Yemen
Northern Africa	Algeria, Egypt, West. Sahara, Libya, Morocco, Sudan, Tunisia
Oceania	American Samoa, Australia, Cocos (Keeling) Isd., Cook Isd., Christmas Isd., Fiji, Micronesia, Guam, Northern Mariana Isd., Nauru, Niue, New Zealand, Papua New Guinea, Pitcairn, Palau, Solomon Isd., Tonga, Tuvalu, Vanuatu, Samoa, Marshall Isd., Kiribati, Tokelau
European countries, other than previous	Andorra, Switzerland, Cyprus, Gibraltar, Iceland, Liechtenstein, Monaco, Malta, Norway, San Marino, Vatican City State
South Asia	Afghanistan, Bangladesh, Bhutan, India, Sri Lanka, Maldives, Nepal, Pakistan
South East Asia	Brunei, Indonesia, Cambodia, Laos, Myanmar, Malaysia, Philippines, Singapore Thailand, Timor-Leste, Vietnam
South Africa	Botswana, Lesotho, Namibia, Swaziland, South Africa, Zimbabwe
Former Soviet Rep. excl. Baltic states	Armenia, Azerbaijan, Belarus, Georgia, Kyrgyzstan, Kazakhstan, Moldova, Russia,Tajikistan, Turkmenistan, Ukraine, Uzbekistan
US and Canada	Canada, Puerto Rico, US, US minor outlying islands
Western Africa	Burkina Faso, Benin, Cote D'Ivoire, Cape Verde, Ghana, Gambia, Guinea Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Sierra Leone, Senegal, Togo