LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Estimation of Threshold Distributions for Market Participation

Mattia Guerini [a,b,c]
Patrick Musso [a]
Lionel Nesta [a,b,d]

[a] Université Côte d'Azur, CNRS, GREDEG.
[b] OFCE, SciencesPo.
[c] Institute of Economics, Scuola Superiore Sant'Anna.
[d] SKEMA Business School.

2020/09 **May 2020**
ISSN(ONLINE) 2284-0400

Estimation of Threshold Distributions for Market Participation

Mattia Guerini[a,b,c] Patrick Musso[a] Lionel Nesta[a,b,d]

[a] *Université Côte d'Azur, CNRS, GREDEG*
[b] *OFCE, SciencesPo*
[c] *Sant'Anna School of Advanced Studies*
[d] *SKEMA Business School*

May 4, 2020

Abstract

We develop a new method to estimate the parameters of threshold distributions for market participation based upon an agent-specific attribute and its decision outcome. This method requires few behavioral assumptions, is not data demanding, and can adapt to various parametric distributions. Monte Carlo simulations show that the algorithm successfully recovers three different parametric distributions and is resilient to assumption violations. An application to export decisions by French firms shows that threshold distributions are generally right-skewed. We then reveal the asymmetric effects of past policies over different quantiles of the threshold distributions.

Keywords: *Parametric Distributions of Thresholds, Maximum Likelihood Estimation, Fixed Costs, Export Decision.*

JEL Codes: *C40, D01, F14.*

✉ Mattia Guerini – mattia.guerini@univ-cotedazur.fr.
✉ Patrick Musso – patrick.musso@univ-cotedazur.fr.
✉ Lionel Nesta – lionel.nesta@univ-cotedazur.fr.

Mattia Guerini has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 799412 (ACEPOL). This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'avenir " program (reference : ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD). The usual disclaimers apply.

1 Introduction

Most economic activities entail sunk and fixed costs stemming from the existence of technical, regulatory and information barriers. Such barriers hinder the participation of agents in market activities and bear serious welfare implications. From a static viewpoint, these may artificially protect incumbents over potential entrants, thereby augmenting firm market power, increasing prices and/or lowering quantities. From a dynamic viewpoint, participation costs may allow incumbents to embark on uncertain and risky investments with potentially substantial sunk costs, such as research and development activities. In all circumstances, knowledge of the costs associated with market participation has immediate policy implications.

Whether agents decide to participate in a specific market – be it the product, labor, technology or financial market – depends on their ability to cover such costs. Past contributions have attempted to estimate the costs of market participation. Das et al. (2007) develop a structural model that allows the empirical estimations of sunk and fixed costs of exporting for Colombian manufacturing firms. They find that sunk entry costs into export markets amount to, on average, $400 thousand, while the fixed costs appear to be negligible. Using a dynamic model of optimal stock market participation, Khorunzhina (2013) estimates that stock market participation costs for consumers amount to 4-6% of labor income. More recently, Fan and Xiao (2015) estimate that entry costs in the US local telephone industry reach $6.5 million.

In this paper, we focus on market participation thresholds, not on participation costs. Thresholds can be defined as unobservable barriers that condition the market participation of an economic actor. An agent will participate in the market only if some of her attribute – say productivity – exceeds the required threshold value of productivity. Conversely if an agent's attribute falls short of the threshold, she will choose not to participate in the market – or to exit it. Hence, it is the gap between the agent-specific attribute and the threshold value that ultimately dictates the decision.

Thresholds are thus well suited to represent, in a concise fashion, the arbitrage underlying market participation for one chief reason. Because they are different from costs, thresholds can be expressed in various dimensions and can be easily adapted to different settings.[1] Possible examples include, *inter alia*, the presence of a minimum efficiency for a firm to participate in foreign markets (Eaton et al., 2011); the existence of a minimum efficient plant size to enter an

[1]The difference between thresholds and costs is that the former are measured over a conditioning attribute domain and the latter in monetary terms.

industry (Lyons, 1980); a required level of absorptive capacity for a firm to efficiently assimilate new technologies (Cohen and Levinthal, 1989); the availability of sufficient collateral for a bank to grant a loan (Jiménez et al., 2006); and the presence of a wage offer above the reservation wage for a worker to accept it (DellaVigna et al., 2017). In all these instances, thresholds represent a minimum value above (below) which an economic agent decides (not) to participate in a given market.

The major problem is that thresholds are empirically unobservable to social scientists or policy makers. When an economic agent truly makes her participation decisions according to a threshold problem, the external observer can only witness the decision outcome itself, and certainly not the threshold *per se*. Most often, thresholds appear as a theoretical parameter that firms must overcome to participate in market activities (e.g., Melitz and Ottaviano, 2008, for export markets). In some instances, specific surveys may explicitly focus on particular thresholds. In their analysis of wage formation and unemployment, Brown and Taylor (2013) exploit information on individual-specific reservation wages obtained from the British Household Panel Survey (BHPS), a nationally representative random sample survey of more than 5000 private households to elicit participation thresholds. However because of the setup costs for data generation, existence of such information remains necessarily limited in size and scope.

This paper presents a new method to estimate the parameters of an underlying distribution of thresholds that hinder market participation. Key to our approach is the intuition that such thresholds are agent-specific rather than common to all or a group of agents. The immediate consequence is that rather than estimating one threshold only, we estimate the parameters underlying threshold distributions. Specifically, we develop a parametric method where the observation of (i) the agents' decision outcomes and (ii) some individual characteristics of the decision makers fostering market participation are sufficient information for recovering the statistical properties of the underlying threshold distribution. The distinctive feature of our method lies in the absence of strong requirements: it needs few behavioral assumptions, it is not data demanding, and it can adapt to various parametric distributions, institutional contexts and, more important, markets.

Knowledge of threshold distributions is important for policy makers. Imagine two distributions with similar average values; however, one is symmetric, while the other exhibits right skewness. In the symmetric distribution, all agents cluster around the mean with a given standard deviation. In the asymmetric distribution, most agents cluster around the mode, but some agents face particularly high thresholds inhibiting market participation. The policy responses to these two distributions may vary substantially. In the former case, nondiscriminatory/general policies designed to affect the common factors of all individual thresholds should presumably be favored. In the latter case, discriminatory/targeted policies focusing on specific agents might instead be privileged. Therefore, decisions on whether to implement discriminatory or nondiscriminatory policies should be grounded on a knowledge of the entire shape of the threshold distribution that goes beyond the first two moments.

Our contribution is fourfold. First, we develop a parametric maximum likelihood estimation (MLE) method to reveal the underlying parameters characterizing threshold distributions. Conditional on the agent's decision outcome and a critical variable called the θ-attribute, we derive a likelihood function that allows the recovery of the concealed threshold distribution. Importantly, we make use of Vuong (1989) procedure to select the most qualified density among the various candidate distribution laws. Second, we use stochastic Monte Carlo simulations to study the reliability of our approach when two important underlying assumptions – one distributional and one behavioral – are violated, and we broadly define the boundaries of its application. Third, we provide a primer empirical application to the problem of export thresholds. The existing literature on international trade generally assumes the existence of fixed/sunk costs associated with export activities. We detect and estimate threshold distributions at the sectoral level documenting significant right-skewness and leptokurtosis within most industries. Fourth, we employ year by year estimates to investigate the possible effects of policy shocks on different quantiles of threshold distributions. Overall, our results indicate that accounting for agents' heterogeneity and for higher order moments allows one to gain valuable information regarding the discriminatory aspect of policies.

This paper is structured as follows. Section 2 formally describes the economic problem under consideration and the tool employed to solve it. Section 3 presents the rationale for Monte Carlo exercises and their results. An empirical application of our strategy to the export decision problem of French firms is presented in Section 4, together with an exercise designed to explain participation rates through the moments of the threshold distributions. Section 5 concludes the paper.

2 Econometric Strategy

2.1 The intuition

We consider a series of agents $i = 1, \ldots, N$ making an economic decision, the outcome of which can be encoded as a binary variable $\chi_i \in \{0, 1\}$ representing their market participation. Each agent is characterized by a specific attribute θ_i that affects the decision outcome. This θ-attribute can be considered a single characteristic or a combination of several distinct features that ease or hinder the realization of a positive outcome $\chi_i = 1$. We assume that an agent decides to participate only when the θ-attribute is sufficiently large that it exceeds threshold c_i. Conversely, if an agent's attribute falls short of the threshold, the agent will choose not to participate in the market. Thresholds can thus be defined as unobservable barriers that dictate the market participation of an economic agent. What we call the perfect sorting hypothesis can be formalized as follows:

$$\begin{cases} \chi_i = 1 & \text{if} \quad \theta_i \geq c_i \\ \chi_i = 0 & \text{if} \quad \theta_i < c_i \end{cases} \tag{1}$$

The theoretical economic literature typically assumes homogeneous thresholds across agents – i.e., $c_i \sim \delta$, with δ representing the Dirac delta distribution (see Pissarides, 1974; McDonald and Siegel, 1986; Dixit, 1989, as early developers of such an approach).[2] Figure 1 describes this particular case by assuming an agent-specific, normally distributed θ_i and a unique threshold c. The shaded area highlights the part of the distribution where all the agents withdraw from the market ($\chi_i = 0$) over the θ-attribute domain ($\theta_i < c$). This representation implies a complete separation of agents, where only those whose θ_i values exceed threshold c participate in the market.

[Figure 1 about here.]

From an empirical perspective, the issue is that the social scientist observes the decision outcome χ_i and the agent's θ-attribute. However, in most situations, the threshold variable c_i is the private information of the decision maker and is thus unobservable to the external observer. This is why the empirical literature has instead focused on the probability of an agent's market

[2] To the best of our knowledge, Cogan (1981) is the sole work to have estimated heterogeneous thresholds in the labor market by means of a 2-step strategy relying on a structural equations model.

participation, conditional on her θ-attribute (e.g., Kau and Hill, 1972; Wei and Timmermans, 2008, among a large series of papers). This type of exercise is technically straightforward using, for example, a simple probit model. This approach matches the fraction of actors with a positive (negative) decision outcome with the θ-attribute and produces the probability of participation conditional on the θ-attribute: $P(\chi = 1|\theta_i)$. The caveat is that such a modeling approach cannot reveal the threshold distribution *per se*.

Furthermore, evidence in several empirical fields – in particular those on the efficiency of exporters (Bernard and Jensen, 2004) and the efficiency of labor market bargaining processes (Alogoskoufis and Manning, 1991) – generally contradicts the simple implication that nonparticipating agents locate in the left tail of the θ distribution, whereas participating agents locate in the right tail. The rule appears to be that some agents well-endowed in their θ-attribute do not participate in a given market, and conversely, some poorly endowed agents nevertheless choose to participate. In other words, the supports of the two populations significantly overlap. To account for this persistent overlap, one must relax the assumption of a unique threshold and favor the converse assumption that thresholds are, instead, agent-specific, as is done, for example, in the theoretical work of Mayer et al. (2014). This makes it possible that well-endowed agents choose to withdraw from a market, given that their specific threshold c_i exceeds their own θ-attribute or, conversely, that poorly endowed agents choose to participate.

Assuming away the homogeneity of thresholds represents an opportunity to shift the traditional perspective of Figure 1. Our perspective inverts the representation proposed by Figure 1 of the θ distribution and a unique threshold c. Figure 2 now takes the observed, agent-specific θ-attribute as given and locates it in an unknown distribution of agent-specific thresholds c_i.[3] The left panel represents the case where an agent participates in a given market. Given θ_i, observing a positive decision outcome $\chi_i = 1$ implies that agent-specific threshold c_i is located somewhere to the left of θ_i, as indicated by the shaded area. The right panel represents the converse: given the attribute θ_i, observing a negative decision outcome $\chi_i = 0$ implies that agent-specific threshold c_i is located to the right of θ_i, as indicated by the shaded area.

[Figure 2 about here.]

Our framework does not change the fundamental mechanism at work, since Equation 1 holds. However, it allows the implementation of a new strategy that can estimate threshold

[3] In Figure 2, we arbitrarily assume normally distributed thresholds.

distributions by means of maximum likelihood estimation, where the parameters of interest are those that define the whole threshold distribution.

2.2 The formal model

We define the probability of an agent participating in the market as the probability that the unobserved, agent-specific threshold c_i is lower than the agent-specific attribute θ_i:

$$p(\chi = 1|\theta_i) = F(\theta_i; \mathbf{\Omega}), \tag{2}$$

where F represents the cumulative density function of the probability distribution f and $\mathbf{\Omega}$ is the vector of distribution-specific parameters to be estimated. In turn, the probability that the unobserved, agent-specific threshold c_i is higher than the agent-specific attribute θ_i is:

$$p(\chi = 0|\theta_i) = 1 - F(\theta_i; \mathbf{\Omega}). \tag{3}$$

The likelihood function $L(\mathbf{\Omega})$ then takes the generic form:

$$L(\mathbf{\Omega}; \theta_i, \chi_i) = \prod_{i=1}^{N} [F(\theta_i; \mathbf{\Omega})]^{\chi_i} \times [1 - F(\theta_i; \mathbf{\Omega})]^{1-\chi_i} \tag{4}$$

The log-likelihood ℓ function reads:

$$\ell(\mathbf{\Omega}; \theta_i, \chi_i) = \sum_{i=1}^{N} \chi_i \log [F(\theta_i; \mathbf{\Omega})] + (1 - \chi_i) \log [1 - F(\theta_i; \mathbf{\Omega})] \tag{5}$$

The decision for the social scientist is to choose a given parametric density function $f(\mathbf{\Omega})$, where $\mathbf{\Omega}$ is the vector of parameters characterizing the distribution f. Given f, the objective function is that of estimating $\mathbf{\Omega}$ such that $\hat{\mathbf{\Omega}} = \arg\max_{\mathbf{\Omega}} \hat{L}(\mathbf{\Omega}; \chi_i, \theta_i)$.

We have two remarks. First, an important ingredient of our framework is the monotonicity of the relationship between the probability of participating in a given market and the θ-attribute. If monotonicity is not empirically verified, then our behavioral assumption formalized in Equation 1 does not hold, and the corresponding likelihood function $L(\mathbf{\Omega})$ will prove difficult to converge. Besides, the case in which the θ-attribute is a limiting rather than an enhancing factor can easily be envisaged. The formal model simply becomes the complement of Equation 1.[4]

[4]This yields following log-likelihood function:

Second, in the case in which the cumulative distribution function is Gaussian, Equation 4 resembles a traditional probit model. However, it differs from it in two important aspects. On the one hand, in a probit model, one assumes specific values for $\mathbf{\Omega}$ by setting $\mu = 0$ and $\sigma = 1$, whereas in our case, our aim is to estimate $\mathbf{\Omega}$. On the other hand, while in our framework, the domain of the support is fully observed with the vector of agent-specific attribute θ_i, in a probit model, the support of the distribution is an unobserved domain. Given a vector of explanatory variables $\mathbf{Z}$ and a decision variable y_i, the objective function of the probit model is then to choose a vector $\boldsymbol{\beta}$ such that $\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \hat{L}(\boldsymbol{\beta}; \mathbf{Z}, y_i | \mu = 0, \sigma = 1)$.[5]

Taking stock of the above, our framework relies on the following core assumptions:

- Assumption **A1**. Agents are heterogeneous in their θ-attribute θ_i.
- Assumption **A2**. Thresholds are agent-specific (c_i) and follow a density distribution f with unknown vector of parameters $\mathbf{\Omega}$: $\boldsymbol{C} \sim F(\mathbf{\Omega})$.
- Assumption **A3**. The relationship between the probability of participating in a market and the θ-attribute is either monotonically increasing $p(\chi = 1|\theta_i) > p(\chi = 1|\theta_j), \forall \theta_i > \theta_j$, as in the case of an enhancing θ-attribute, or monotonically decreasing $p(\chi = 1|\theta_i) < p(\chi = 1|\theta_j), \forall \theta_i > \theta_j$, as in the case of a limiting θ-attribute.
- Assumption **A4**. Agents decide to participate if and only if their specific attribute θ_i exceeds (is below) threshold c_i (perfect sorting hypothesis).

Provided that Assumptions **A1**-**A4** hold, the observation of vector $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_i, \ldots, \chi_N)$ stacking all decision outcomes χ_i and vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i, \ldots, \theta_N)$ of agent-specific attributes θ_i is sufficient information to estimate the vector of parameters $\mathbf{\Omega}$ defining the underlying threshold distribution F. Hence the distinctive feature of our method resides in the absence

$$\ell(\mathbf{\Omega}) = \sum_{i=1}^{N} (1 - \chi_i) \log\left[F(\theta_i; \mathbf{\Omega})\right] + \chi_i \log\left[1 - F(\theta_i; \mathbf{\Omega})\right]$$

[5] In fact, the probit model is often presented as being derived from an underlying latent variable model similar to Equation 1. Using our notation and modelling strategy in the context of a latent variable, the decision outcome χ_i is set to unity if threshold $c_i^* < 0$, and to zero otherwise. This implies that agent i participates only if the threshold is negative. Variable c_i^* is unobserved but defined as a linear function of – in our case – the θ-attribute such that $c_i^* = \beta_0 + \beta_1 \theta_i + z_i$. If z_i is modelled as a standard normal, then the model reads: $p(\chi = 1|\theta_i) = p(c_i^* < 0|\theta) = p(z_i < -(\beta_0 + \beta_1 \theta_i))$. If $z \sim \mathcal{N}(0,1)$, then $p(z_i < -(\beta_0 + \beta_1 \theta_i)) = F_z(-(\beta_0 + \beta_1 \theta_i))$, where F_z is the standard normal cumulative function, and $p(\chi = 1|\theta_i) = F(\theta_i | \mu = 0, \sigma = 1)$. The corresponding likelihood function is equivalent to our Equation 4, the only difference being in the parameter vector $\mathbf{\Omega}$. The direct implication of this is that our framework, when using the normal distribution function as our prior density, and the probit model yield exactly the same log-likelihood value. This is not surprising given that we exploit the same information in the data.

of strong requirements: it needs few behavioral assumptions, it is not data demanding, and it can adapt to various parametric distributions.

2.3 Model selection

The critical choice concerns the distribution of thresholds, as the scientist may choose among a large series of data generating processes. In the absence of prior information about the true distribution of thresholds, we rely on Vuong's test (Vuong, 1989) for the selection of non-nested models. The attractive features of Vuong's test in our framework is twofold: (i) it does not require preexisting knowledge of the true density; (ii) it is directional, allowing one to arbitrate between any pair of assumed density functions.

The starting point is to select, among two candidate densities f_p and f_q, the model that is closest to the true, unknown, density f^0. The statistic tests the null hypothesis H_0 that the two models (f_p and f_q) are equally close to the true data generating process, against the alternative that one model is closer. We write:

$$H_0 : \mathbb{E}^0\left[\log \frac{f_p(\chi_i|\theta_i, \mathbf{\Omega}^*_{f_p})}{f_q(\chi_i|\theta_i, \mathbf{\Omega}^*_{f_q})}\right] = 0, \tag{6}$$

where $\mathbb{E}^0$ is the expectation indicator, and $\mathbf{\Omega}^*_f$ is the candidate, pseudo-value of the true vector $\mathbf{\Omega}$. Interestingly, Equation 6 does not necessitate knowledge of the true the true density f^0, but it provides information about the best model between the alternatives f_p and f_q. Under the null hypothesis, the distributions F_p and F_q are equivalent ($F_p \sim F_q$). The two directional hypotheses read:

$$H_{F_p} : \mathbb{E}^0\left[\log \frac{f_p(\chi_i|\theta_i, \mathbf{\Omega}^*_{f_p})}{f_q(\chi_i|\theta_i, \mathbf{\Omega}^*_{f_q})}\right] > 0, \tag{7}$$

meaning that F_p is a better fit than F_q ($F_1 \succ F_q$), and:

$$H_{F_q} : \mathbb{E}^0\left[\log \frac{f_p(\chi_i|\theta_i, \mathbf{\Omega}^*_{f_p})}{f_q(\chi_i|\theta_i, \mathbf{\Omega}^*_{f_q})}\right] < 0, \tag{8}$$

meaning that F_q is a better fit than F_p ($F_p \prec F_q$). Vuong (1989) shows that the indicator $\mathbb{E}^0$ can be estimated by the Likelihood Ratio statistic such that:

$$\log LR(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) = \ell_p(\hat{\mathbf{\Omega}}_{f_p}) - \ell_q(\hat{\mathbf{\Omega}}_{f_q}) = \sum_{i=1}^{N} \big(\ell_{p,i}(\chi_i|\theta_i, \hat{\mathbf{\Omega}}_{f_p}) - \ell_{q,i}(\chi_i|\theta_i, \hat{\mathbf{\Omega}}_{f_q})\big) = \sum_{i=1}^{N} d\ell_i, \quad (9)$$

where $\ell_{p,i}(\chi_i|\theta_i, \hat{\mathbf{\Omega}}_{f_p})$ (resp. $\ell_{q,i}(\chi_i|\theta_i, \hat{\mathbf{\Omega}}_{f_q})$) is observation i's contribution to the log likelihood ℓ_p (resp. l_q) using density f_p (resp. f_q). The ratio $d\ell_i$ simply represents the difference in the log-contributions of the i^{th} observation. In addition, Vuong (1989) suggests to account for differences in the number of parameters in the two models as the in the Akaike Information Criterion such that:

$$\log \tilde{LR}(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) = \log LR(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) - (k_p - k_q)\frac{\log N}{2} \quad (10)$$

where k_p and k_q represent the number of parameters in density functions f_p and f_q, respectively. Given the above setting, Vuong's z statistic reads:

$$\text{Vuong's } z = (\sigma_{dl}\sqrt{N})^{-1} \log \tilde{LR}(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) \quad (11)$$

where σ_{dl} is the standard deviation of dl. Vuong test statistic is asymptotically normally distributed by the central limit theorem. In other words, cumulative function F_p is preferred over cumulative density function F_q if Vuong's z exceed the $(1-\alpha)^{th}$ percentile of the standard normal distribution. Setting a 5% significance level, the corresponding z statistic in a bilateral test is $|z| >= 1.96$.

A last attractive feature of Vuong's test for the selection of non-nested models is that the ranking between any pair of models is transitive. This implies that if F_p is preferred over F_q, and F_q is preferred over F_r, then F_p is preferred over F_r.[6] This is relevant in that our method can envisage a large number of density functions and then recover a complete rank order between the competing models. If N_f densities are being tested, $N_f \times (N_f - 1)/2$ pairwise comparisons will allow one to recover a complete rank order of preferences across the competing density functions.

[6] This is true for what concerns the comparisons of l scores. In few cases, the transitivity may be affected by the denominator of Vuong's z, that is, the standard deviation of individual log difference σ_{dl}.

3 Monte Carlo Simulations

3.1 Monte Carlo settings

The first choice concerns the candidate parametric densities to fit to the data. The number of candidate distribution being virtually infinite, we arbitrarily choose parametric densities with two parameters only. Remember, however, that our framework can easily adapt to parametric density functions which include a higher number of parameters.

We simulate the heterogeneous threshold distribution F as extracted from three distribution densities: the normal distribution $\mathcal{N}$, the gamma distribution Γ, and the beta distribution $\mathcal{B}$: $F \in \{\mathcal{N}; \Gamma; \mathcal{B}\}$. The choice of the normal, gamma and beta distributions for thresholds is motivated by the fact that they allow us to compare a symmetric distribution in the case of the normal and asymmetric distributions of thresholds in the case of the gamma and the beta. In addition, the gamma, being very flexible, envisages various distributional shapes that may prove empirically relevant in the presence of right-skewness. The choice of the gamma also implies that we constrain the support of θ-attributes to be strictly positive: $\boldsymbol{\theta} \in \mathbb{R}^{+}$. The beta distribution is by far the most flexible, as it encapsulates all sorts of distribution shapes, ranging from left-skewed, symmetric or right-skewed distribution. The inclusion of the beta distribution implies that the support lies over the 0-1 segment: $\boldsymbol{\theta} \in (0, 1)$. This is extremely binding, because it implies that the cumulative distribution function be unity when θ exceeds one.

Following the description in Appendix A, we fix the number of agents to $N = 50,000$, the number of Monte Carlo simulations to $M = 1,000$, and impose a normally distributed θ-attribute $\boldsymbol{\theta} \sim \mathcal{N}(.5, .15)$.[7] In our simulation, threshold C is random variable drawn from: (i) a normal distribution with $\mu = .7$ and $\sigma = .2$, i.e. $\boldsymbol{C} \sim \mathcal{N}(.7, .2)$; (ii) a right-skewed gamma distribution with shape and scale parameters $\alpha_\Gamma = .1.5$ and $\beta_\Gamma = .5$, i.e. $\boldsymbol{C} \sim \Gamma(1.5, .5)$; (iii) a left-skewed beta distribution with shape-one and shape-two parameters $\alpha_\mathcal{B} = 5$ and $\beta_\mathcal{B} = 2$, i.e. $\boldsymbol{C} \sim \mathcal{B}(5, 2)$. For all three threshold distributions, given the vector of θ-attribute, we computed the vector of decision outcomes $\boldsymbol{\chi}$ according to Equation 1.[8]

In what follows, we conduct several Monte Carlo simulation exercises to investigate whether the estimation of Equation 5 holds when Assumptions **A1**-**A4** hold. We then explore the robust-

[7] We verified that our results are qualitatively robust to alternative distributions of θ_i. We tested $\Theta \sim \mathcal{U}(min, max)$, $\Theta \sim \mathcal{B}(\alpha, \beta)$, and $\Theta \sim \mathcal{P}(min, \alpha)$ where $\mathcal{U}, \mathcal{B}, \mathcal{P}$ represent the uniform, beta and Pareto type-II distributions, respectively. The results are available from the authors upon request.

[8] Details on the functional forms of the three distributions are available in Appendix B.

ness of the estimator under violations of certain assumptions.

3.2 Baseline results

We begin with a perfect scenario, where all Assumptions **A1**-**A4** hold. Using only limited information $\boldsymbol{\theta}$ and $\boldsymbol{\chi}$, we then apply our MLE algorithm to estimate the vector of parameters $\boldsymbol{\Omega} = \{\mu, \sigma\}$ for the Gaussian case, $\boldsymbol{\Omega} = \{\alpha_\Gamma, \beta_\Gamma\}$ for the gamma case and $\boldsymbol{\Omega} = \{\alpha_\mathcal{B}, \beta_\mathcal{B}\}$ for the beta case.[9] The distributions of the estimated parameters across $M = 1000$ Monte Carlo simulations are displayed in the different panels of Figures 3 for the normal, the gamma and the beta scenarios, respectively.

[Figure 3 about here.]

Figure 3 shows that, on average, our estimation strategy accurately estimates the true parameters. In fact, a simple t-test never rejects the null hypothesis of equality of the estimated parameters with the target, true, parameter. However, the MLE approach provides evidence that our framework can sometimes be off target, i.e., it over- or underestimates the true parameters in a magnitude reaching approximately 10% of the true parameter, especially for the gamma and beta cases. A closer inspection of the results suggests that this is due to a strong – respectively negative and positive for the gamma and the beta distributions – correlation between the parameters α and β. As shown in Figure 4, there is a tight negative relationship linking the estimates of the two parameters in the gamma case (left panel) and for the beta case (right panel). Thus in the gamma case, when one of the two parameters is underestimated with respect to the true value, the other compensates and becomes overestimated. In the beta case the compensation effect goes instead in the same direction.

[Figure 4 about here.]

This compensation mechanism is a positive feature of the two asymmetric distributions. The depicted correlation between α and β simply signifies that there are several parameter combinations that allow us to unravel the density function f sufficiently close to the true one.

In Section 3.3 we move towards imperfect scenarios, testing the robustness of the estimation when at least one of the assumptions fails. We focus on Assumptions **A2** and **A4**, which may prove hard to meet.

[9]Note that estimations of α and β for the gamma and the beta distributions allow one to analytically retrieve the first four moments.

3.3 Violations of assumptions

3.3.1 Violation of A2

Assumption **A2** concerns the functional form f of the threshold distribution. Since thresholds are unobservable, we need a prior concerning the density function. A *specification* error occurs when a functional form f assumed by the scientist is different from the *true* one. Without a strong prior, any probability distribution is eligible. Given this uncertainty, understanding the consequences of a violated Assumption **A2** is of crucial importance. Our intuition is that the probability density function f must be sufficiently flexible to encompass a variety of shapes, so that it can adapt from case to case.

We set $N = 50,000$ agents and run $M = 1,000$ Monte Carlo simulations, comparing three alternatives for the likelihood function where the cumulative distribution function F may be either a normal, a gamma or a beta distribution $F = \{\mathcal{N}(\mu, \sigma), \Gamma(\alpha_\Gamma, \beta_\Gamma), \mathcal{B}(\alpha_\mathcal{B}, \beta_\mathcal{B})\}$. This represents our prior about the threshold distribution. In turn, the true distribution of thresholds $\boldsymbol{C}$ may alternatively follow a normal, a gamma or a beta distribution. For the gamma case, we consider a parametric configuration giving rise to a right-skewness with a fat right tail. For the beta distribution instead, we choose a left-skewed distribution. We set the θ-attribute such that it follows a normal distribution: $\boldsymbol{\theta} \sim \mathcal{N}(.5, .0225)$. The set of parameters for threshold distributions is the following: (i) $\boldsymbol{C} \sim \mathcal{N}(.7, .04)$; (ii) $\boldsymbol{C} \sim \Gamma(1.5, .5)$; and (iii) $\boldsymbol{C} \sim \mathcal{B}(5, 2)$.[10] Combining all the possible choices for the prior F and the true distribution of $\boldsymbol{C}$, we obtain nine different cases. In three of them, Assumption **A2** is satisfied and give rise to the situations observed in the previous section. In six of them, Assumption **A2** is violated.

To evaluate the performance of our framework under a violation of Assumption **A2**, we are restricted to compare the moments estimated by our densities, since one cannot estimate the parameters of the normal distribution (μ and σ) using a gamma or an beta prior. Instead, we compute the root mean squared errors (RMSEs) of the estimated first four moments (the mean μ; the variance σ^2; the skewness sk; and the excess kurtosis k) and compare the estimated values with their true counterparts. We do this for all the scenarios combining our prior, assumed, density for the MLE exercise and the true densities. The results displayed in Table 1 thoroughly corroborate the results of the previous section. Absent a misspecification, the true moments are

[10] The guidelines for the algorithm followed for this Monte Carlo exercise are presented in the Appendix A. Values for the variance σ^2 in the normal distribution stem from setting σ to .15 for θ and .2 for the threshold normal distributions.

consistently recovered. However, when the false prior is used, a significant error is always present.

The first three rows of the bottom panel display the performance of the three priors ($F = \{\mathcal{N}(\mu, \sigma), \Gamma(\alpha_\Gamma, \beta_\Gamma), \mathcal{B}(\alpha_\mathcal{B}, \beta_\mathcal{B})\}$) when the normal density is true. Consistently with Section 3.2, it shows that the normal prior performs well at recovering the first two moments of the true normal, with errors amounting to only .3% and 2.3% of the true mean and variance, respectively. Under violation of Assumption **A2**, the beta prior (with 2% and 24% errors for mean and variance, respectively) outperforms the gamma prior (with errors amounting to 5% and 53%, respectively). Our interpretation is that the beta prior outperforms the gamma one due to the impossibility of the latter to recover symmetric distributions.

We then investigate the case where thresholds are gamma distributed. Not surprisingly, the gamma prior properly recovers the first four moments. The largest RMSE concerns the second moment and reaches 5.2% of the true variance. Instead, both the normal and beta priors deteriorate when applied to a gamma distribution. This observation spans over all investigated moments: errors amounting to 20% and 23% of the true mean, and errors for all higher order moments exceed 100% of the true moments. This is due to two conditions. First, it is impossible for the normal prior to recover asymmetric and fat tailed distributions. Second, it is impossible for a beta prior to recover distribution outside the 0-1 support. The final three rows presents the results under a true beta distribution of thresholds. Our previous observation also prevail in this case. The beta prior under a true beta distribution performs well, whereas the normal and gamma priors fail in accurately estimating the first four moments of the distribution.

[Table 1 about here.]

The conclusion of the simulation exercise is clear: without proper knowledge about the true threshold distribution, misspecification can generate off-target predictions, leading to a wrong inference about participation threshold distributions. In this context, Vuong (1989) likelihood ratio test for model selection is fully advocated in order to recover a complete ranking of the candidate densities.

Table 2 presents the relative frequencies of each ranking of distributions, according to the Vuong (1989) test using the simulations used above. Similarly to the sensitivity (detecting true positives) and specificity (detecting true negatives) of a test, this exercise investigates whether Vuong's procedure correctly discriminates between the three alternative densities. Table 2 reports the results. We observe that the test is successful at pointing to the correct true density, with the exception of only two cases in the $3,000$ simulation runs. Besides, the ranking of the two remaining alternatives are, in the vast majority of cases, consistent with the percentage errors reported in Table 1. When the true distribution of thresholds is normal, the order of preference is $\mathcal{N} \succ \mathcal{B} \succ \Gamma$ in the majority of cases; when the thresholds are gamma distributed, then $\Gamma \succ \mathcal{N} \succ \mathcal{B}$ is the typical ranking; and when the true density of thresholds is the beta distribution, then $\mathcal{B} \succ \mathcal{N} \succ \Gamma$ is the emerging ranking. In all instances, the sensitivity and specificity of Vuong's test hold under violations of Assumption **A2**.

[Table 2 about here.]

Confident that: (i) in the absence of a specification error, one can use the MLE to consistently recover the true parameters, and (ii) Vuong's test for model selection successfully discriminates between competing prior densities, our procedure allows one to estimate participation thresholds in various empirical setups.

3.3.2 Violation of A4

Assumption **A4** concerns the decision process of economic agents and relates to the information set available either to the entrepreneur or to the external investigator. We examine two potential sources that may affect the robustness of the estimation framework.

The first source of imperfect sorting arises when economic agents have limited information about either their own threshold c_i or their own θ-attribute.[11] The immediate consequence is that agents base their decision on an erroneous inference, violating the perfect sorting hypothesis. We define a *sorting error* as a situation in which agent i, characterized by $\theta_i > c_i$ (resp. $\theta_i < c_i$), selects the negative (resp. positive) outcome χ_i due to agent i having an imperfect evaluation of her threshold c_i. With respect to Figure 2, a *sorting error* implies that by observing $\chi_i = 0$ (resp. $\chi_i = 1$), we wrongly assign a threshold c_i to the right (resp. to the left) of the observed θ_i. To generate a *sorting error* in our Monte Carlo exercise, we modify the problem in Equation 1 as follows:

$$\begin{cases} \chi_i = 1 \quad \text{if} \quad \theta_i \geq c_i + \varepsilon_i^c \\ \chi_i = 0 \quad \text{if} \quad \theta_i < c_i + \varepsilon_i^c \end{cases} \tag{12}$$

[11] The contributions of Jovanovic (1982) and Hopenhayn (1992) on the dynamics of industries are examples of models where agents – firms – learn about their own θ-attribute and may make erroneous decisions.

where we assume that c_i represents the *true* threshold and $\varepsilon_i^c \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon)$ measures instead the erroneous information of the agent, which is summarized by σ_ε.

The second source of imperfect sorting is due to *measurement errors* of the θ-attribute by the observer. The quality of an agent's decision is not at stake but this issue may harm the estimation due to a measurement error in the support of the threshold distribution. With respect to Figure 2, a *measurement error* yields a noisy location of the θ-attributes for all agents. To simulate it, we hold the original decision problem of Equation 1 fixed, but we use the noisy measure of the θ-attribute when maximizing the likelihood, defined as $\theta_i + \varepsilon_i^\theta$, where again $\varepsilon_i^\theta \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon)$ is a mean-preserving spread.

Both the sorting and measurement errors may occur simultaneously within an empirical exercise.[12] For all these scenarios exploring the robustness of the method under the violation of Assumption **A4**, we instead set Assumption **A2** to be valid: no specification error can therefore arise. Combining the possibility of *imperfect sorting* (IS) or/and *imperfect measurement* (IM) give rise to three possible scenarios (labeled as IS-PM, PS-IM and IS-IM). We also discriminated for different sizes of the errors, characterized by σ_ε which lies in 5%, 10% and 15% of the original variance of the θ-attribute (σ_θ). This exercise is performed for all the three (i.e. normal, gamma and beta) threshold densities.[13]

Each scenario gives rise to 1,000 Monte Carlo simulation, allowing us to compute the RMSE for the estimated first four moments. We express each RMSE relatively to the true moments, i.e. as percentage errors, in Table 3. We observe that, as long as the informational error is small or medium sized (i.e. 5% or 10%) the moments are recovered without the generation of large errors for all the three densities. The largest error is a 8.5% deviation from the true value in the estimate of the kurtosis for the beta case in the PS-IM scenario. When the size of the shock reaches its 15%, the central moment is typically precisely estimated, however, for the IS-IM scenario both the gamma and the beta distribution commit errors a sizeable 10% for the variance and the kurtosis.

[Table 3 about here.]

In general, we observe that the normal is more resilient to shocks, the gamma follows and

[12] Details of the algorithmic guidelines are presented in Appendix A.

[13] Unreported results show that for the gamma and beta distributions, a pattern similar to Figure 4 on the correlation between estimates of α and β is found in each of the three scenarios. A downward bias in the estimates of α_Γ is compensated by an upward bias in the estimates of β_Γ for the gamma case; and downward bias in the estimates of $\alpha_\mathcal{B}$ is compensated by a downward bias in the estimates of $\beta_\mathcal{B}$ for the beta case.

eventually might inconsistently estimate the variance. The beta distribution seems instead the least resilient to violations of **A4**, possibly leading to errors of higher magnitude when both the θ-attribute is badly measured and a significant fraction of the firms do not take decisions according to Equation 1.

3.3.3 Joint violation of A2 and A4

We have observed in Section 3.3.1 that, in the absence of measurement and sorting errors, the true density function is correctly recovered and its true parameters consistently estimated. Section 3.3.2 has documented that, with the correct prior density, sorting and measurement errors mildly impact the estimation precision. We now turn to investigating the effect of a joint violation of Assumptions **A2** and **A4**. This amounts to questioning the capacity of Vuong's test to correctly identify the true density in the presence of measurement and sorting errors. This is an empirically relevant issue, as in most cases: (i) a researcher does not have any prior knowledge about the true underlying density function (hence Assumption **A2** is violated); (ii) some firms do not behave according to Equation 1 and/or the θ-attribute is measured only imperfectly (hence also Assumption **A4** is violated).

We perform 1,000 Monte Carlo simulations combining the settings performed in the two previous subsections. In particular, we use the three threshold distributions – namely the normal, the gamma, and the beta – with the joint presence of imperfect sorting and imperfect measurement (the IS-IM scenario). For each simulation run, we confront the predicted density as identified by Vuong's procedure for model selection with the true density. The results are presented in Table 4. They suggest that the Vuong's test is successful at identifying the proper density, even in presence of a joint violation of the Assumption **A2** and Assumption **A4**. If the true underlying density is normal, the test always excludes the gamma and beta alternatives. Similarly, if thresholds are distributed gamma, the test predicts the correct density in virtually all cases (99.9% of the performed simulations).

The only issue concerns the beta case. In 32% of the simulation runs, Vuong z statistics points to a normal density, whereas the true density is the beta. The excess presence of false normal in lieu of the beta one casts doubts on the reliability of Vuong's test in the presence of sorting and measurement errors. However, by truncating the normal distribution over the $(0, 1)$ support, we notice that the pattern of the estimated normal distribution has a shape similar to the one of the true beta distribution. More generally, over the $(0; 1)$ support, truncation of the normal distribution yields distribution similar extremely close to the beta case. We therefore conclude that the prediction of a gamma or a beta distribution according to the Vuong's test is reliable. When estimating a normal distribution instead, there remains a substantial degree of uncertainty about the true underlying density function.

[Table 4 about here.]

Collecting all results obtained by means of Monte Carlo simulations, we conclude that our estimation strategy is robust under Assumptions **A1**-**A4**. Violations of Assumption **A2** might lead to severe errors in the predicted moments, but Vuong's test for model selection allows one to correctly select among the set of competing densities. When Assumption **A4** does not hold, both imperfect sorting and the imperfect measurement have similar effects on the robustness of the framework. We found that estimation errors reflect the magnitude of the shock, i.e. they increase with the shock size and that, in general, the estimation of the symmetric normal distribution seems more resilient than the asymmetric ones. Applications to situations where Assumptions **A2** and **A4** are violated leads to satisfactory results, although caution must be taken when the normal law is identified as the correct density.

4 Empirical Application to International Trade

We apply our framework to the case of firm export decisions. Following the seminal contributions by Melitz (2003) and Melitz and Ottaviano (2008), the recent international trade literature has modeled the export decision as being conditional on a unique export productivity threshold called the export productivity cutoff. Only the most productive firms, which have a productivity level that exceeds the homogeneous threshold, enter foreign markets. The assumption of a unique threshold is extremely restrictive and is at odds with robust empirical evidence about the coexistence of high-efficiency firms that do not export and inefficient firms that export (Bernard and Jensen, 2004; Eaton et al., 2011; Impullitti et al., 2013).[14]

Based on Assumptions **A1**-**A4**, we estimate the export threshold distribution for French manufacturing firms using firm-specific productive efficiency as the θ-attribute. Our framework allows us to reconcile the appearing empirical paradox with the theory.

[14]The authors of the theoretical literature also recognize this limitation, but for reasons of analytical tractability, they cannot leave this assumption aside. Only recent versions of these models overcome this issue by accounting for product variety and a heterogeneous product mix. Thresholds are equal within varieties, but product mixes are firm specific and, in turn, generate firm-specific productivity cutoffs (Mayer et al., 2014).

4.1 Choice of support

We use a panel database of French manufacturing firms covering the period 1990–2007 found in the annual survey of companies (*Enquête annuelle d'entreprises*) led by the statistical department of the French Ministry of Industry. The survey covers all firms with at least 20 employees in the manufacturing sectors (excluding food and beverages), and the data provide information about their income statements and balance sheets. The complete dataset gathers the financial statements of 43 thousand companies, yielding 350 thousand firm-year observations. We use information on sales, exports, value added, the wage bill, the number of employees and hours worked, capital stock, investment, and intermediate inputs as the main variables used to compute the firm-specific θ-attribute.

We have two eligible measures for $\boldsymbol{\theta}$, namely apparent labor productivity (ALP) or total factor productivity (TFP), as these two measures are used interchangeably in the empirical literature.[15] The choice between the two can, in principle, be based according to which variable is most tightly associated with export decision. Based on footnote 5, one could compare the log-likelihood value stemming from the use of rival variables as the support and simply choose the one yielding the highest likelihood. Alternatively, one could tailor a statistical procedure where a preliminary step would embed the choice of the most appropriate support. One can argue, instead, that this choice be theoretically grounded, and not exclusively data-driven, or say, information-content driven. Here, we decide to use TFP as θ-attribute. Although it is prone to some measurement error, TFP accounts for more inputs and firm characteristics than mere ALP.[16] The fact that TFP is prone to mismeasurement is also a test for the robustness of our framework.

The top panel of Table 5 provides the preliminary descriptive statistics for market participation. In our sample, the export participation rate reaches 74%. This is a relatively large participation rate, which is due to the fact that our dataset comprises larger firms, which are more likely to export vis-à-vis smaller firms. Table 5 also displays the export premium, that is, the productivity differential between exporters and non-exporters. Consistent with the economic literature, exporters are on average more productive than non-exporters, with a productiv-

[15] Appendix C provides the details of the industry-wide deflators used and the computations that yield labor and total factor productivity.

[16] There are various reasons underlying mismeasurement in TFP. We mainly think of the potential misspecification of the production function, measurement errors in capital stocks, or assumptions on the endogeneity of production factors. See, among others, Atkinson and Stiglitz (1969), Wooldridge (2009), Ackerberg et al. (2015) and De Loecker and Goldberg (2014) for a thorough discussion on these issues. Appendix D however, provides all the robustness results with the apparent labor productivity.

ity differential of 4.2% for all manufacturing. Although we observe significant cross-industry variation, all sectors display a positive productivity premium, with the exception of *Wood and Paper*.

[Table 5 about here.]

4.2 Export participation thresholds

We set the outcome decision variable χ_i to unity if we observe positive exports by firm i, 0 otherwise: $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_i, \ldots, \chi_N)$. In our framework, total factor productivity represents the θ-attribute: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i, \ldots, \theta_N)$. Given vectors $\boldsymbol{\chi}$ and $\boldsymbol{\theta}$, we estimate Equation 5 for all manufacturing firms. Because the support must be strictly positive for the gamma density and below unity for the beta density, we transform θ (whether ALP or TFP) such that $\theta_i \in (0;1)$, as follows:

$$\tilde{\theta}_i = \left(\frac{\theta'_i - \min \theta'}{\max \theta' - \min \theta'} \right),$$

where θ' represents the labor productivity measure net of sector-year fixed effects, $\theta'_i = \theta_i - \bar{\theta}_{sy} + \bar{\theta}$, and where subscript sy indicates the sector $\times$ year identifier. We also estimate $\boldsymbol{\Omega}$ for each sector:

$$\hat{\boldsymbol{\Omega}}_s = \arg\max_{\boldsymbol{\Omega}_s} \hat{L}(\boldsymbol{\Omega}_s; \chi_i, \theta_i) \quad \forall s \in S$$

where subscript s stands for sector s. We have no prior about the functional form of the density distribution for export thresholds. Therefore, we perform the estimation exercise using the normal, the gamma and the beta densities. The results are reported in the top panel of Tables 6 for the goodness of fit and 7 for the estimated first two moments (μ and σ^2). For the gamma and beta densities, we also report the estimated median to provide initial insights into the presence of skewness. We have four major observations.

First, the algorithm converges rapidly with an average of 5 iterations when assuming normal and beta densities. The number of iterations increases substantially when we assume gamma distributed thresholds, reaching 27 iterations for *Metallurgy, Iron and Steel*, *versus* 5 and 6 iterations for normal and beta densities. This is evidence that assuming gamma distributed thresholds introduces some computational complexity in the search grid. We find one instance of no convergence of the algorithm, regardless of the prior about the functional form of the distribution (see Table 6) for *Wood and Paper*, where all three densities fail to converge. Our interpretation is that this is due to a specification or measurement error. As a matter of fact, Table 5 shows that the export premium is negative (-1.4%) for this industry. This implies that, on average, Equation 1 does not hold, which may stem either from the absence of monotonicity between the θ-attribute and the probability of exporting (Assumption **A3**) or from measurement errors by the entrepreneur or the social scientist (Assumption **A4**). Although the gamma density needs a larger number of iteration, it also yields the largest likelihood in the majority of industries (8 out of 13), whereas the normal outperforms other densities in the remaining 5 industries. Hence the number of iterations is not a reliable proxy for the goodness of fit.

[Table 6 about here.]

[Table 7 about here.]

Second, a clear pattern emerges in the mean and variance of the three densities. The normal density systematically estimates the lowest mean, whereas the gamma estimated mean is systematically the largest. Conversely, the normal density yields the largest variance (with the exception of *Metallurgy, Iron and Steel*, whereas the beta distribution produces the lowest one. Hence, the choice of the underlying density is a choice which predetermines the ultimate distribution shape. This reinforces the need for Vuong's procedure for model selection.

Third, most estimated mean export thresholds lie within the $(0; 1)$ interval. This is especially true when we assume gamma or beta distributed thresholds. This is consistent with the idea that the participation rates are generally high, exceeding 70% in most industries. When we focus on the assumption of normally distributed thresholds, we also observe negative mean values when we impose normally distributed thresholds on the data, for *Automobile*, *Chemicals*, *Metallurgy, Iron and Steel* and *Pharmaceuticals*. Although this is at odds with the positive support for the θ-attribute, it reflects an interesting feature of the normal law. In fact, a negative mean implies that the shape of the distribution is truncated normal on $\mathbb{R}+$, allowing for the presence of right skewness and fat right tails in threshold distributions.

Fourth, the gamma and beta densities discard the possibility of normally distributed thresholds. In fact, we observe positive skewness in most, if not all sectors, including *All Manufacturing*, and the estimated median is significantly below the estimated average. *Metallurgy, Iron* *and Steel* and *Pharmaceuticals* stand out with median values which are extremely low. These two sectors are precisely those with the lowest (positive) productivity premium (*Metallurgy, Iron and Steel*) and highest export participation rates (*Pharmaceuticals*). Looking at the parameter estimates for these two sectors suggests that searching for alternative densities may be advocated.

Altogether, we find heterogeneity in two dimensions. We find cross-density heterogeneity and cross-sectoral heterogeneity in the shapes of the threshold distributions, their mean values, their variances, their medians and their (unreported) higher moments.

4.3 Threshold distributions for entry and remaining into export markets

Our method also applies to decisions about market entry and exit, conditional on the availability of a time dimension in the data. We now exploit this time dimension and condition the decision on the export status observed the preceding year. We define (i) the pool of potential entrants into export markets as the firms that do not export at time $t-1$ and (ii) the pool of potential remaining firms as those that already exported at time $t-1$. We then define actual entering and actual remaining firms those that decide to start exporting or remain exporters at time t. We now estimate entry and remaining threshold distributions.[17]

We first focus on results for threshold distributions for entry into export markets. Table 5 shows that the share of firms entering into export markets reaches 23% for all manufacturing. This suggests that, although export market participation is pervasive in the data, entry into export market reflects a fiercer selection process. The estimation results are reported in the middle panel of Tables 6 and 7. The most immediate observation is the poor performance of the beta prior but this is not surprising. Entry thresholds are presumably higher than mere market participation as it focuses on pure entry, incorporating sunk entry costs which are otherwise not born. Hence one should expect entry thresholds to significantly increase, notably above the maximum value of the θ-attribute. Remember, however, that the beta distribution imposes an upper limit for the support at unity. This imposes that the cumulative distribution of beta be unity when $\theta = 1$. In practice, this is very unlikely to hold. On the contrary, when focusing on entry, we should expect the threshold distribution to go well beyond the θ support.

This is confirmed when looking at the estimated mean values for the normal and the gamma

[17]Whether we focus on remaining or exit thresholds is essentially a matter of semantic. By setting $\chi_i = 1$ if the firm exits export markets and 0 otherwise would become an exit decision. In this case, we would need to adapt Equation 4 as presented in footnote 4. The likelihood function is ultimately the same maximization problem, yielding an identical vector of parameter $\hat{\mathbf{\Omega}}$.

entry threshold distributions. When using the normal prior instead, we find convergence for all industries with the exception of *Clothing and Footwear*, *Pharmaceuticals* and *Wood and Paper*. This is not surprising for *Pharmaceuticals* and *Wood and Paper*, due to their negative export premia ($-.039\%$ for the former, $-.019\%$ and for the latter), implying that Equation 1, on average, does not hold for these two sectors. There is less convergence with the gamma density, although the algorithm seem successful in industries where the normal density fails to converge.

We now turn to threshold distributions for remaining in export markets. Table 5 shows that the share of firms remaining in export markets reaches almost 93% for all manufacturing. This suggests that exit from export market is a relatively rare phenomenon, and that the associated threshold distribution must allow for the vast majority of firm to remain in export market, once overcome the entry hurdle. The estimation results are reported in the bottom panel of Tables 6 and 7. Looking at *All Manufacturing*, the average threshold for remaining in export markets is substantially lower than that for market entry. This result is valid across all sectors and irrespective of the prior MLE densities. Since entry costs are essentially sunk, one would expect higher thresholds for entering firms than for remaining firms. Previous exporters have already borne any sunk entry costs and only have to cope with fixed and variable costs. It suggests that whereas the entry thresholds are substantial, the remaining thresholds are of a much lower magnitude. This is consistent with Das et al. (2007), who find that fixed costs borne each period are negligible, whereas sunk entry costs are of a considerable magnitude.

The normal density yields systematically negative mean values for thresholds for market remaining. In fact, this reflects the flexibility of the normal distribution whose support spans over the entire spectrum for real numbers. This search for the maximum likelihood produces a distribution that easily accommodates for a truncation at zeros. In all instances, all densities estimate low mean thresholds with a low variance. This suggests that for the vast majority of companies, remaining into export markets is far less challenging than market entry into export markets.

4.4 Vuong's test for model selection of export markets

The key question is to discriminate between the candidate densities. We perform Vuong's model selection procedure testing the best fit among the three selected densities. Any pairwise comparison can be interpreted as evidence of a specification error when we find evidence that a tested density provides less information than a rival density. Vuong's test for model selection provides us with a tool to detect specification error, it cannot confirm that the selected model is indeed the optimal fit for our data. However, the sign and magnitude of Vuong's z allow for a direct interpretation of the better model, among the finite set of alternatives, on the basis of the following hypotheses testing:

$$\begin{aligned} F_p \sim F_q & \quad \text{if} \quad |z_{F_p,F_q}| & < +1.96 \\ F_p \succ F_q & \quad \text{if} \quad z_{F_p,F_q} & \geq +1.96 \\ F_q \succ F_p & \quad \text{if} \quad z_{F_p,F_q} & \leq -1.96 \end{aligned}$$

where F represent the cumulation distribution function, $p, q \in \{\mathcal{N}, \Gamma, B\}$, and $p \neq q$. Table 8 displays all pairwise Vuong's z: $z_{\mathcal{N},\Gamma}$ comparing the normal and the gamma density, $z_{\mathcal{N},\mathcal{B}}$ comparing the normal and the beta density, and $z_{\Gamma,\mathcal{B}}$ comparing the gamma and the beta density. The last two columns of Table 8 provide the overall ranking of densities and the conclusion of model selection, where the proper density is displayed with the associated vector of parameter estimates.

The most immediate observation is that the gamma density outperforms the normal and the beta densities for *All Manufacturing* and for market participation, market entry and market remaining. This is evidence of the presence of right-skewness and leptokurtosis in export thresholds. For market participation and remaining, this implies that most firms cope relatively low export thresholds, whereas a minority of them cope with higher export thresholds. As for market entry, the presence of right-skewness is secondary. Bearing in mind that the average thresholds exceeds 6, that the median thresholds locates at almost 3, whereas the θ support ranges from 0 to unity, the policy recommendation is to act upon those firms located in the left part of the threshold distributions. In this case, most firms are excluded from international trade.

[Table 8 about here.]

The second observation concerns the cross-sectoral heterogeneity in the best density functions among the three alternatives. Although the gamma distribution dominates the overall industry threshold distribution, sector-specific threshold distributions do not necessarily follow a gamma density. We find evidence of beta densities for market remaining in *Transportation Machinery* and *Wood and Paper*. In many instances, the normal distribution is the advocated better density according to Vuong's z test for model selection. In *Electric and Electronic Components* and *Electric and Electronic Equipment* for example, the normal density is systematically selected as the best fit among the three, irrespective of the type of market participation (participation, entry, remaining). More generally, caution is needed when the diagnosis is dominance of the normal over the gamma

[Figure 5 about here.]

Figure 5 plots the estimated density functions for *All Manufacturing* and for three selected 2-digits industries: *Clothing and Footwear*; *Electric and Electronic Components*; and *Printing and Publishing*. It is important to notice the difference in the magnitude of the support when considering alternatively mere market participation, market entry or market remaining, as it underlines the different types of costs to be borne for participation in general, entry, and remaining. The top two panels provide examples of the variety of possible shapes for threshold distributions that stem from a gamma density: whether the mode is located at the minimum and higher values, whether there exist fat tails, etc. The three distributions for *Electric and Electronic Components* are instead example of normal distributions truncated at zero. For market remaining, we observe that the mode of the distribution is located at the left of the minimum θ-attribute, corroborating that for the vast majority of already exporting firms, remaining thresholds are virtually nil. The threshold distributions for *Printing and Publishing* displays various densities.

[Table 9 about here.]

Last, we estimate the vector of parameters for the normal, gamma and the beta densities at the industry×year level such that:

$$\hat{\boldsymbol{\Omega}}_{st} = \underset{\boldsymbol{\Omega}_{st}}{\arg\max}\, \hat{L}(\boldsymbol{\Omega}_{st}; \chi_i, \theta_i) \quad \forall s \in S \text{ and } t \in T$$

where subscripts s and t stand for sector s at time t. This amounts to running 13 industries $\times$ 18 years Vuong's procedures for model selection for market participation, yielding various ranking in densities. We proceed similarly for market entry and market remaining. Accounting for entry or remaining imply the loss of the first year of observation due to the use of a lagged year in identifying firm market entry and/or remaining. Table 9 presents the various rankings obtained and the associated count.

Of the 234 estimations performed for market participation (221 for market remaining), convergence is achieved in 204 (respectively 194) cases for all three candidate densities. In 22 (respectively 19) cases only, none of the candidate densities succeed in estimating the densities. This is in contrast with market entry, where no convergence is achieved for 67 of the 221 industry-year estimations, whereas all three candidate densities can be estimated in only 37 cases. The lack of convergence for market entry may stem from: (i) a violation of Assumption **A2** on the proper density and the need for alternative densities with possibly more parameters; (ii) a violation of Assumption **A3** on the monotonicity of the relationship between the support and the probability of export; (iii) a strong violation of Assumption **A4** on perfect sorting, stemming from either wrong decisions by firms or measurement errors in the support. In fact, entry into export markets involves a host of factors that may stem well beyond mere productivity. This suggests that a possible development of our framework is to consider more than one support to accurately estimate the thresholds hindering entry decisions by agents.

Table 9 illustrates the various rankings and arbitrage in the better fit. Of all the three candidate densities, the gamma distribution dominates in all type of market decision. However, there is a need for alternative densities. We see, for example, that the beta distribution represents a better fit in a seizable number of occurrences. Table D.5 of Appendix D shows that one shall not conclude that the gamma density is the best prior, irrespective of the support. In fact, it reveals that when using labor productivity as the θ-attribute, the normal density (left-truncated at 0) is the dominating density. Table D.5 also implies that all three densities are relevant when using the alternative support.

4.5 Threshold dynamics and globalization

We use our method to evaluate the impact of policy shocks on export thresholds between 1990 and 2007. This period is characterized by major structural shocks, all intended to reduce export barriers: the establishment of the single market in 1993; the birth of the euro in 1999; the implementation of the single currency for all transactions in 2002; and the entry of China, India, and more generally the BRICS countries as major players on the international trade scene. Much has been written about the pro-competitive consequences of European integration (e.g., Boulhol, 2009, amongst a large series of contributions) or globalization (De Loecker and Goldberg, 2014) on markups, but evidence of decreased export thresholds has yet to emerge. One should expect a significant decrease in thresholds overall. However, we remain agnostic about the effect of the aforementioned shocks on higher moments of the distribution. We explore this issue by estimating the threshold distribution for market participation as follows:

$$\hat{\mathbf{\Omega}}_t = \arg\max_{\mathbf{\Omega}_t} \hat{L}(\mathbf{\Omega}_t; \chi_i, \theta_i) \quad \forall t \in T = (1990, \ldots, t, \ldots, 2007)$$

We do not report the results of our estimations, but consistently with prior findings, the gamma density is the best fit according the Vuong's test. Figure 6 displays the dynamics of the first four moments of export thresholds. As expected, we observe an overall downward trend in thresholds by 15%. A closer look at the evolution of the mean suggest that the establishment of the single market in 1993 was a major step towards lower export thresholds (-8%). The introduction of the euro as a common currency for all transactions is concomitant to significant decreases in the mean value of export thresholds. This may be due to increased competition by rival eurozone firms or the simultaneous arrival of major players outside Europe, such as China joining the World Trade Organization in 2001, in the export markets, both of which would exclude the least efficient firms from export markets.

[Figure 6 about here.]

Looking at higher moments of the distribution, we observe that the establishment of the single market and the introduction of the euro represent shocks of different natures. The significant decrease in the mean in 1993 leaves all higher moments nearly unchanged. By securing the free movement of goods, services, capital and persons and by removing customs barriers between member states, the establishment of the single market represented a shock homogeneous to all manufacturing firms. Higher moments of the distribution exhibit an important upward trend from 1999 onwards, that is, at the birth of the euro. The synchronized increase in the variance, skewness and kurtosis implies that the euro constituted an asymmetric shock to French firms: it created threshold distributions with more extreme values (higher kurtosis) located in the right tail of the distribution (positive skewness). Regarding the dynamics of the last decile in Figure 6, we immediately observe that the birth of the euro in 1999 is associated with higher thresholds values; conversely, the lower percentiles of the distributions enjoy persistently decreasing thresholds.

Overall, the establishment of the single market and the introduction of the euro – together with the arrival of major countries in export markets – represented major shocks for French manufacturing firms. These shocks translated into significantly lower thresholds for the majority of firms. Whereas the former represented a homogeneous shock to all firms, the latter constituted an asymmetric shock, definitely excluding a minor share of large manufacturing firms from international trade.

5 Conclusion

This paper has developed a new method to estimate the parameters of the threshold distribution for market participation, requiring few working assumptions. Stochastic Monte Carlo simulations have shown that our method is resilient to specification, sorting and measurement errors. We have applied our method to unravel the productivity threshold distribution for export markets for French manufacturing firms. In most cases, the likelihood function needs few iterations to converge, except when some underlying assumption is not empirically supported. We have implemented an hypothesis testing procedure based upon Vuong (1989) to discriminate between a set of competing priors. We find that heterogeneity is relevant across several domains: (i) within-sector estimates suggest that participation thresholds are characterized by right-skewness and leptokurtosis; (ii) between-sector estimates, conditional upon participating, entering or remaining in a market, suggest that gamma, normal or beta densities are all empirically relevant; (iii) across-year estimates, suggest that European integration policies and access to foreign markets have decreased average thresholds, at the cost of an increased heterogeneity, captured by an increase in higher order moments and on the top decile thresholds. Overall, our results indicate that accounting for heterogeneity in thresholds allows one to gain new insights on policy effectiveness and in particular about the perils of basing policies solely on centrality statistics.

The distinctive feature of our method is its flexibility: it needs few behavioral assumptions, is not data demanding, and can adapt to various parametric distributions. Two column vectors – one about the decision outcome and one about the θ-attribute supporting the decision – suffice to recover the distribution parameters and to test for alternative priors. The critical assumption is to choose the parametric form of the unobserved thresholds. In this paper, we experimented with the normal, gamma and beta distributions. Whether simulated or empirical, our findings unambiguously support the idea that, in the absence of a strong prior about the distribution of thresholds, it is important to use an hypothesis testing procedure to discriminate between the set of priors.

Our method can be applied to various issues within the realm of economics. Beyond export productivity threshold distributions, our method could be used to unravel distributions of reservation wages, skill requirements, and financial constraints: all agent-specific characteristics that support or limit the decision to participate in a market. All that is required is to gather the two vectors of agent-specific decision outcomes and θ-attributes. The existence of dynamic information on whether agents enter or exit a market can also be exploited to reveal distributions of entry and remaining thresholds. Moreover, this method may also be of interest outside the realm of economics, although the range of relevance is difficult to forecast.

We intend to extend this work in three directions. The most immediate extension is to condition the vector of parameters on agent-specific characteristics. The gain is twofold. First, one can analyze the determinants affecting the parameters of the threshold distribution. Second, one could then compute agent-specific thresholds. In other words, beyond the estimation of distribution shapes, one could locate each agent within the distribution with a certain degree of confidence. Such an exercise would be particularly useful for policy purposes. In the case of export thresholds, policy markers could target export subsidies to a set of identified firms that combine a high level of productivity but a high export threshold; alternatively, policies could be designed to maximize participation, conditional on a given budget. Second, by broadening the meaning of the θ-attribute toward a multivariate context, we could envisage narrowing the definition of a market to a more fine-grained scale. One could, for example, distinguish among destinations for export markets or among industries for labor markets. In all instances, the road ahead is to extend the univariate case presented in this paper to multivariate (normal, gamma and beta) distributions while accounting for the correlation between markets. Third, we intend to identify the underlying relationship between thresholds and costs. While the first can be applied to more domains, the second can provide a direct monetary evaluation of the problem under investigation for policy makers. Increases in thresholds imply higher barriers to market participation. Although the support of the thresholds is expressed in a linear scale, the underlying series of investments and efforts allowing agents to climb the threshold ladder is likely to involve highly convex costs.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Alogoskoufis, G. and A. Manning (1991). Tests of alternative wage employment bargaining models with an application to the UK aggregate labour market. *European Economic Review 35*(1), 23 – 37.

Atkinson, A. B. and J. E. Stiglitz (1969). A new view of technological change. *The Economic Journal 79*(315), 573–578.

Bernard, A. B. and J. B. Jensen (2004). Why some firms export. *The Review of Economics and Statistics 86*(2), 561–569.

Boulhol, H. (2009, April). The convergence of price-cost margins. *Open Economies Review 19*(2), 221-240.

Brown, S. and K. Taylor (2013). Reservation wages, expected wages and unemployment. *Economics Letters 119*(3), 276 – 279.

Caves, D. W., L. R. Christensen, and W. E. Diewert (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal 92*(365), 73–86.

Cogan, J. F. (1981). Fixed costs and labor supply. *Econometrica 49*(4), 945–963.

Cohen, W. M. and D. A. Levinthal (1989). Innovation and learning: The two faces of r&d. *The Economic Journal 99*(397), 569–596.

Das, S., M. J. Roberts, and J. R. Tybout (2007). Market entry costs, producer heterogeneity, and export dynamics. *Econometrica 75*(3), 837–873.

De Loecker, J. and P. Goldberg (2014). Firm performance in a global market. *Annual Review of Economics 6*, 201–227.

DellaVigna, S., A. Lindner, B. Reizer, and J. F. Schmieder (2017, 05). Reference-Dependent Job Search: Evidence from Hungary*. *The Quarterly Journal of Economics 132*(4), 1969–2018.

Dixit, A. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy 97*(3), 620–638.

Eaton, J., S. Kortum, and F. Kramarz (2011, September). An anatomy of international trade: Evidence from french firms. *Econometrica 79*, 1453–1498.

Fan, Y. and M. Xiao (2015). Competition and subsidies in the deregulated us local telephone industry. *The RAND Journal of Economics 46*(4), 751–776.

Good, D. H., M. I. Nadiri, and R. Sickles (1997). Handbook of applied econometrics: Microeconometrics. Chapter Index Number and Factor Demand Approaches to the Estimation of Productivity. Blackwell, Oxford.

Hall, R. E. (1988). The relation between price and marginal cost in u.s. industry. *Journal of Political Economy 96*(5), 921–947.

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

Impullitti, G., A. A. Irarrazabal, and L. D. Opromolla (2013). A theory of entry into and exit from export markets. *Journal of International Economics 90*(1), 75 – 90.

Jiménez, G., V. Salas, and J. Saurina (2006). Determinants of collateral. *Journal of Financial Economics 81*(2), 255 – 281.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica 50*(3), 649–670.

Kau, P. and L. Hill (1972). A threshold model of purchasing decisions. *Journal of Marketing Research 9*(3), 264–270.

Khorunzhina, N. (2013). Structural estimation of stock market participation costs. *Journal of Economic Dynamics and Control 37*(12), 2928 – 2942.

Lyons, B. (1980). A new measure of minimum efficient plant size in uk manufacturing industry. *Economica 47*(185), 19–34.

Mayer, T., M. J. Melitz, and G. I. P. Ottaviano (2014, February). Market size, competition, and the product mix of exporters. *American Economic Review 104*(2), 495–536.

McDonald, R. and D. Siegel (1986, 11). The Value of Waiting to Invest*. *The Quarterly Journal of Economics 101*(4), 707–727.

Melitz, M. J. (2003, November). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*, 1695–1725.

Melitz, M. J. and G. I. P. Ottaviano (2008, 01). Market size, trade, and productivity. *The Review of Economic Studies 75*(1), 295–316.

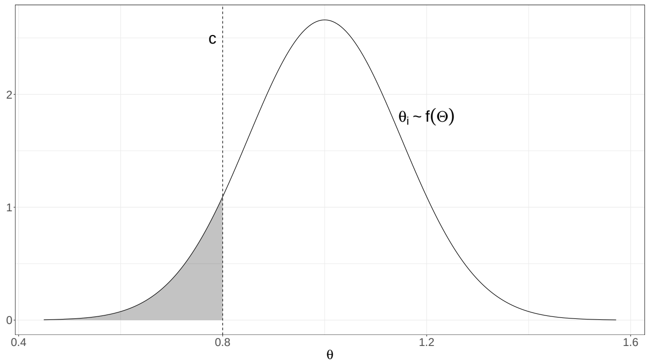
Pissarides, C. A. (1974). Risk, job search, and income distribution. *Journal of Political Economy 82*(6), 1255–1267.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*(2), 307–333.

Wei, Z. and H. Timmermans (2008). Cut-off models for the 'go-home' decision of pedestrians in shopping streets. *Environment and Planning B: Planning and Design 35*(2), 248–260.

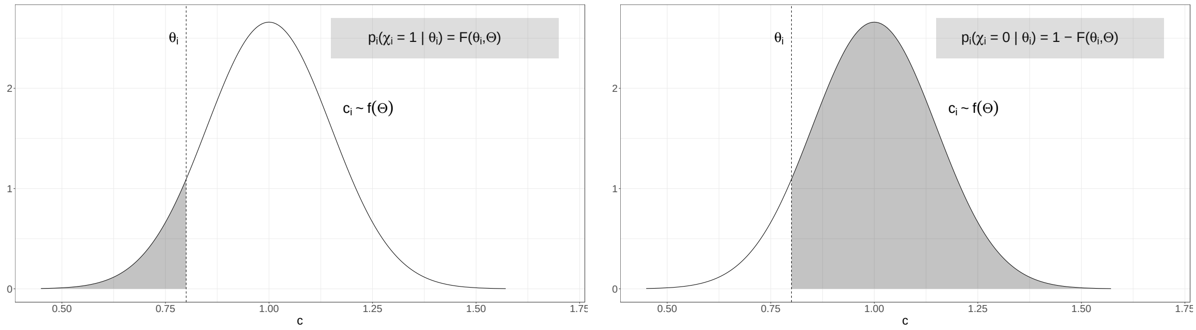
Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

Figure 1: *Example of a θ-attribute distribution with a common threshold c for all the individuals*



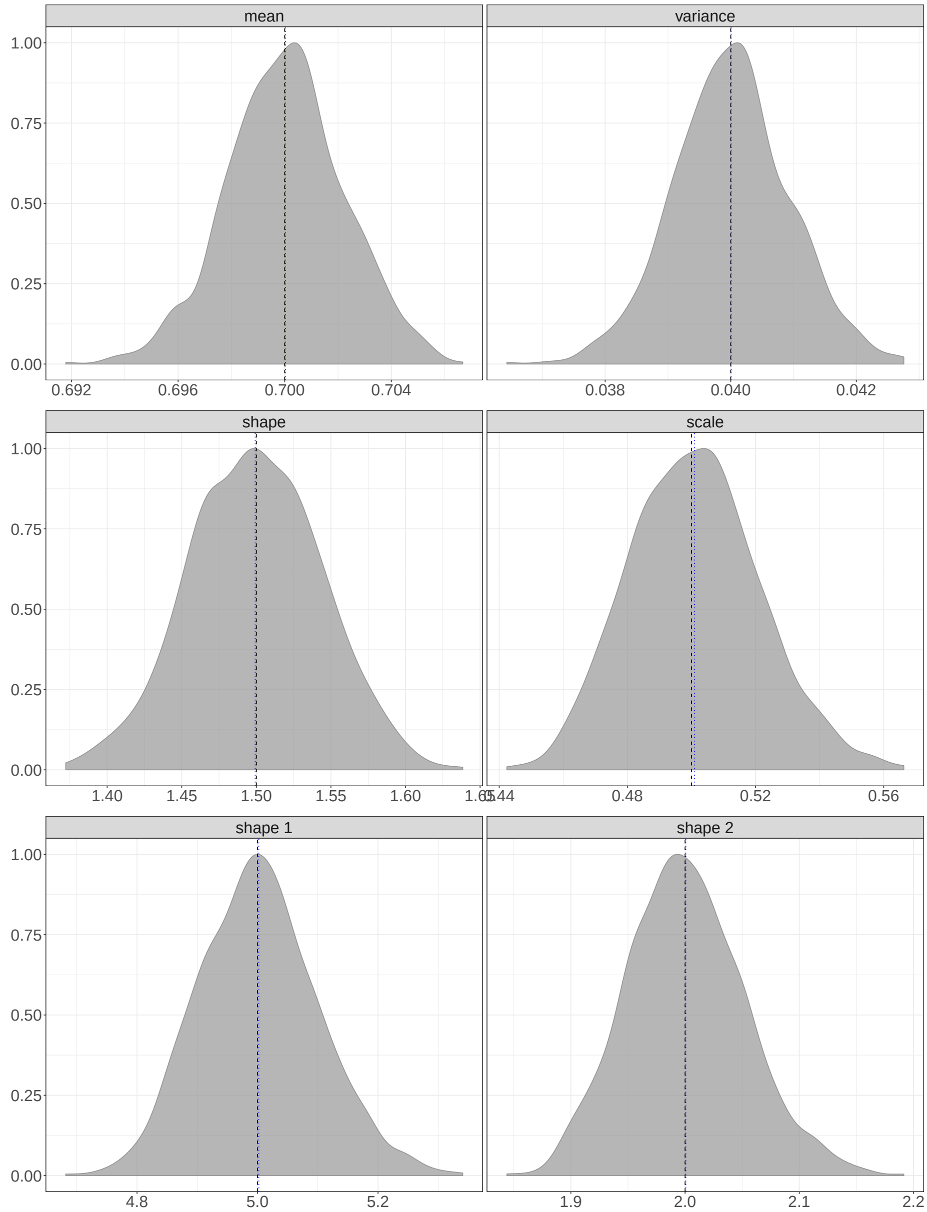
Agents deciding to participate in a market are those whose θ-attribute values exceed threshold c. Conversely, agents deciding not to participate in a market are those whose θ-attribute lies to the left of threshold c, as indicated by the shaded area.
In this example, we assume a normally distributed θ-attribute.

Figure 2: *Example of threshold distribution with observed θ-attribute*



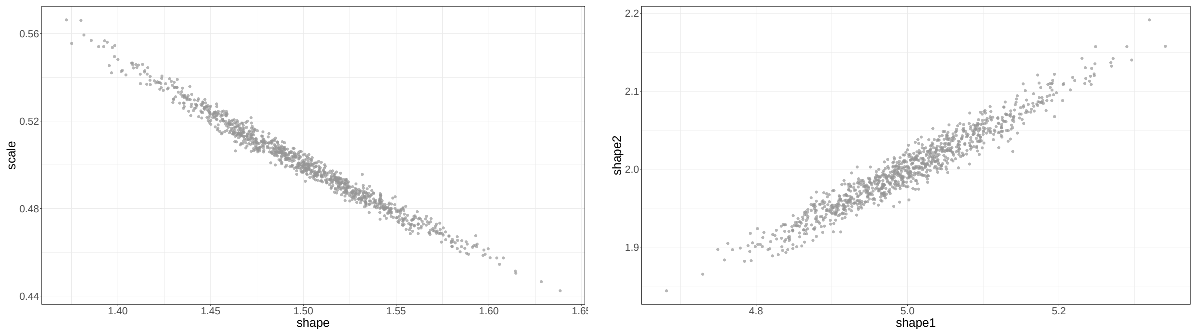
Left panel: Given the θ-attribute θ_i, observing a positive decision outcome $\chi_i = 1$ implies that agent-specific threshold c_i is located to the left of θ_i, as indicated by the shaded area. *Right panel*: Given the θ-attribute θ_i, observing a negative decision outcome $\chi_i = 0$ implies that agent-specific threshold c_i is located to the right of θ_i, as indicated by the shaded area.
In this example, we assume normally distributed thresholds.

Figure 3: *Distribution of Monte Carlo estimates of the parameters* $\mathbf{\Omega}$ *for three different threshold distributions.*



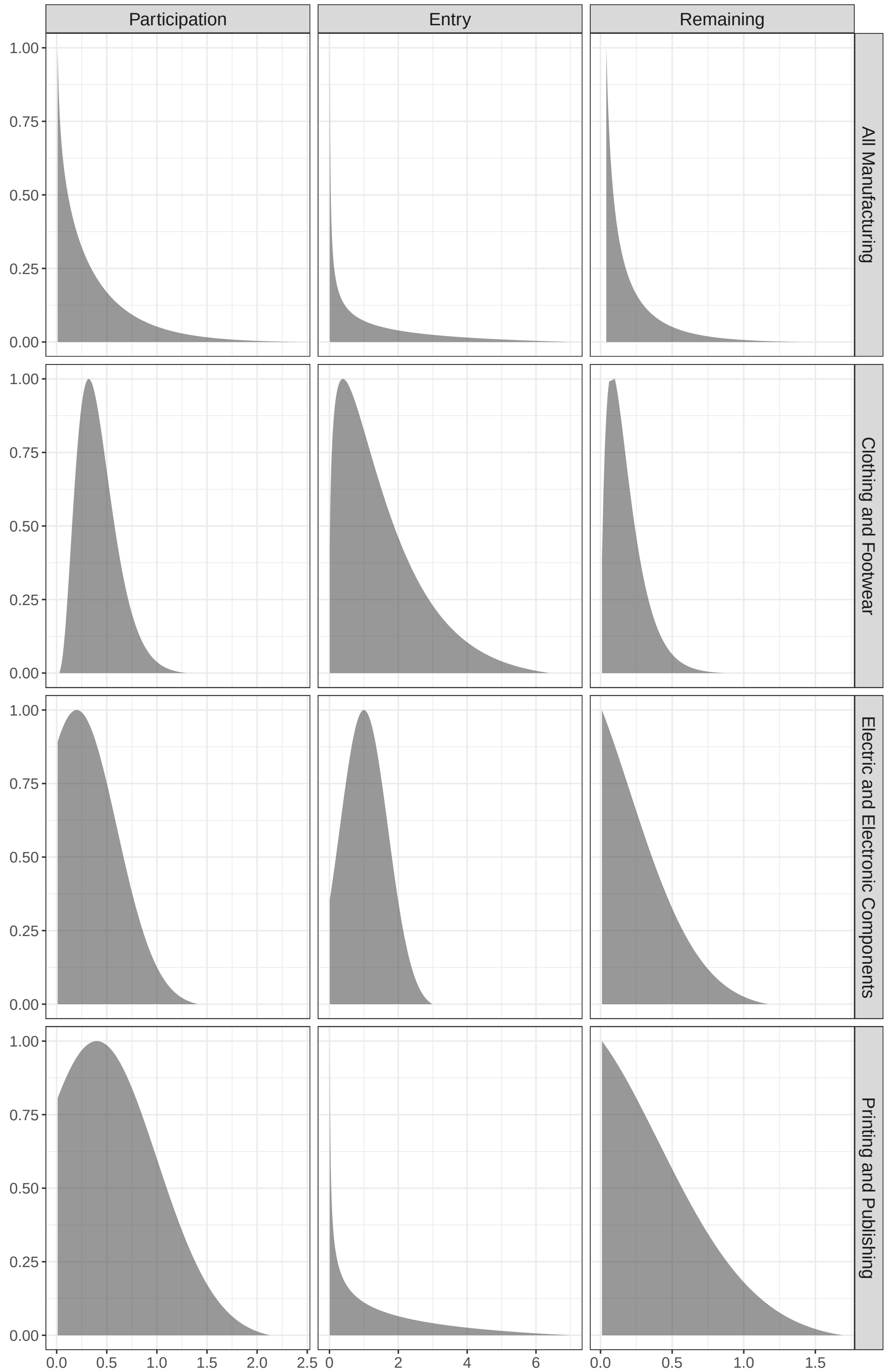
$M = 1,000$ Monte Carlo simulations with $N = 50,000$ agents. The true values are represented by a black vertical line with short dashes. The average estimate is represented by the blue vertical line with dots. Top panels: estimates of normal distributed thresholds. Middle panel: estimates of gamma distributed thresholds. Bottom panels: estimates of beta distributed thresholds. For all simulations, we set the θ-attribute such that it follows a normal distribution: $\boldsymbol{\theta} \sim \mathcal{N}(.5, .0225)$. Set of parameters for threshold distributions: (i) $\boldsymbol{C} \sim \mathcal{N}(.7, .004)$; (ii) $\boldsymbol{C} \sim \Gamma(1.5, .5)$; and (iii) $\boldsymbol{C} \sim \mathcal{B}(5, 2)$.

Figure 4: *Scatter plot of the parameters' estimates for the gamma and beta scenarios*



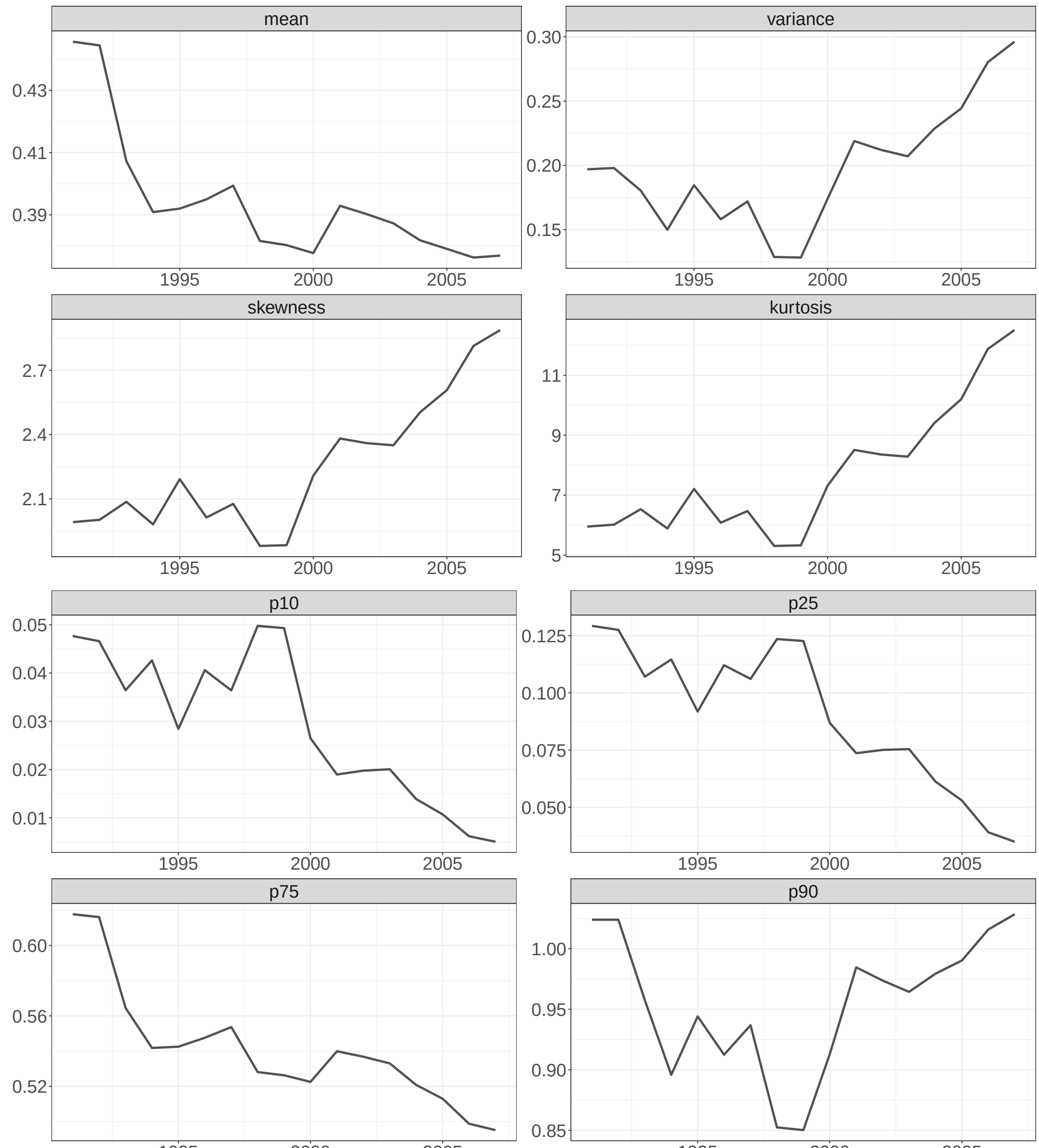
$M = 1,000$ Monte Carlo simulations with $N = 50,000$ agents. Left panel: estimates of shape and scale parameters of gamma distributed thresholds. Right panel: estimates of shape 1 and shape 2 parameters of beta distributed thresholds.

Figure 5: *Threshold distributions for all manufacturing and three selected sectors.*



For each industry and type of market participation, the formal definition of the density function can be found in Table 8.

Figure 6: *The dynamics of threshold distributions in French manufacturing: 1990-2007*



The moments and percentiles of the threshold distribution are based on the estimated gamma parameters for all manufacturing using total factor productivity as the θ-attribute, net of industry×year effects. $p10$, $p25$, $p75$ and $p90$ stand for the 1^{st} decile, the 1^{st} quartile, the 3^{rd} quartile and the 9^{th} decile of the threshold distribution, respectively.

Table 1: *Mean squared errors of the estimated first four moments over the different scenarios.*

Assumed c	True c	A2	μ	σ^2	sk	k
			True Values			
	$\mathcal{N}$		0.700	0.040	0.000	0.000
	Γ		0.750	0.375	1.633	4.000
	$\mathcal{B}$		0.714	0.026	0.596	0.120
			RMSE as Share of True Values			
$\mathcal{N}$	$\mathcal{N}$	✓	0.003	0.023	-	-
Γ	$\mathcal{N}$	✗	0.049	0.533	-	-
$\mathcal{B}$	$\mathcal{N}$	✗	0.022	0.242	-	-
$\mathcal{N}$	Γ	✗	0.203	0.452	1.000	1.000
Γ	Γ	✓	0.012	0.052	0.015	0.029
$\mathcal{B}$	Γ	✗	0.243	0.731	1.152	1.321
$\mathcal{N}$	$\mathcal{B}$	✗	0.104	1.508	1.000	1.000
Γ	$\mathcal{B}$	✗	0.074	1.158	2.025	5.675
$\mathcal{B}$	$\mathcal{B}$	✓	0.002	0.016	0.015	0.064

$M = 1,000$ Monte Carlo simulations with $N = 50,000$ agents. ✓: Assumption A2 holds. ✗: Assumption A2 is violated. sk: Skewness. ku: Kurtosis. For all simulations, we set the θ-attribute such that it follows a normal distribution: $\boldsymbol{\theta} \sim \mathcal{N}(.5, .0225)$. Set of parameters for threshold distributions: (i) $\boldsymbol{C} \sim \mathcal{N}(.7, .004)$; (ii) $\boldsymbol{C} \sim \Gamma(1.5, .5)$; and (iii) $\boldsymbol{C} \sim \mathcal{B}(5, 2)$.

Table 2: *Vuong's ranking of distributions, by type of true threshold distribution*

Vuong diagnosis	True c: $\mathcal{N}$	True c: Γ	True c: $\mathcal{B}$
$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	0.961		
$\mathcal{N} \succ \Gamma \sim \mathcal{B}$	0.025		
$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	0.011		
$\mathcal{N} \sim \mathcal{B} \succ \Gamma$	0.001		0.001
$\Gamma \succ \mathcal{N} \sim \mathcal{B}$		0.026	
$\Gamma \succ \mathcal{N} \succ \mathcal{B}$		0.707	
$\Gamma \succ \mathcal{B} \succ \mathcal{N}$		0.267	
$\mathcal{B} \succ \mathcal{N} \succ \Gamma$	0.002		0.867
$\mathcal{B} \succ \Gamma \succ \mathcal{N}$			0.131
$\mathcal{B} \succ \Gamma \sim \mathcal{N}$			0.001

$M = 1,000$ Monte Carlo simulations with $N = 50,000$ agents for each true threshold distribution. Figures represent the share of simulation representing Vuong's diagnosis appearing as row heads. For example: $\mathcal{N} \succ \mathcal{B} \succ \Gamma$ must be read as "The normal distribution is preferred over the beta distribution which is preferred over the gamma distribution.

Table 3: *Root Mean Squared Errors (RMSE) as a share of the true value for the estimated first four moments over the different scenarios*

MSE	IS	IM	$\sigma_\varepsilon = 5\%$	$\sigma_\varepsilon = 10\%$	$\sigma_\varepsilon = 15\%$
					$\mathcal{N}$ distributed thresholds
μ	✓		0.003	0.003	0.003
σ^2	✓		0.023	0.024	0.026
sk	✓		-	-	-
ku	✓		-	-	-
μ		✓	0.003	0.003	0.003
σ^2		✓	0.023	0.023	0.023
sk		✓	-	-	-
ku		✓	-	-	-
μ	✓	✓	0.003	0.003	0.003
σ^2	✓	✓	0.023	0.024	0.026
sk	✓	✓	-	-	-
ku	✓	✓	-	-	-
					Γ distributed thresholds
μ	✓		0.012	0.012	0.012
σ^2	✓		0.052	0.052	0.052
sk	✓		0.015	0.015	0.015
ku	✓		0.030	0.030	0.030
μ		✓	0.012	0.014	0.020
σ^2		✓	0.053	0.065	0.098
sk		✓	0.015	0.018	0.027
ku		✓	0.030	0.037	0.056
μ	✓	✓	0.012	0.014	0.021
σ^2	✓	✓	0.054	0.066	0.101
sk	✓	✓	0.015	0.018	0.028
ku	✓	✓	0.031	0.037	0.056
					$\mathcal{B}$ distributed thresholds
μ	✓		0.002	0.002	0.002
σ^2	✓		0.016	0.017	0.020
sk	✓		0.015	0.015	0.015
ku	✓		0.064	0.063	0.068
μ		✓	0.002	0.003	0.006
σ^2		✓	0.017	0.025	0.047
sk		✓	0.016	0.025	0.045
ku		✓	0.066	0.085	0.137
μ	✓	✓	0.002	0.003	0.005
σ^2	✓	✓	0.017	0.029	0.058
sk	✓	✓	0.016	0.025	0.046
ku	✓	✓	0.066	0.077	0.111

$M = 1,000$ Monte Carlo simulations with $N = 50,000$ agents. ✓: Assumption in the specific column is violated. sk: Skewness. ku: Kurtosis. For all simulations, we set the θ-attribute such that it follows a normal distribution: $\boldsymbol{\theta} \sim \mathcal{N}(.5, .0225)$. Set of parameters for threshold distributions: (i) $\boldsymbol{C} \sim \mathcal{N}(.7, .004)$; (ii) $\boldsymbol{C} \sim \Gamma(1.5, .5)$; and (iii) $\boldsymbol{C} \sim \mathcal{B}(5, 2)$. The true moment values are displayed in Table 1.

Table 4: *Vuong's ranking of distributions, by type of true threshold distribution*

Vuong diagnosis	True c: $\mathcal{N}$	True c: Γ	True c: $\mathcal{B}$
$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	0.990		0.326
$\mathcal{N} \succ \Gamma \sim \mathcal{B}$	0.008		
$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	0.002		
$\mathcal{N} \sim \mathcal{B} \succ \Gamma$			0.015
$\Gamma \succ \mathcal{N} \sim \mathcal{B}$		0.016	
$\Gamma \succ \mathcal{N} \succ \mathcal{B}$		0.110	
$\Gamma \succ \mathcal{B} \succ \mathcal{N}$		0.873	
$\mathcal{B} \succ \mathcal{N} \succ \Gamma$			0.550
$\mathcal{B} \succ \Gamma \succ \mathcal{N}$		0.001	0.109

$M = 1,000$ Monte Carlo simulations with $N = 50,000$ agents for each true threshold distribution under the presence of imperfect sorting and imperfect measurement (IS-IM) with a 5% shock size. Figures represent the share of simulation representing Vuong's diagnosis appearing as row heads. For example: $\mathcal{N} \succ \mathcal{B} \succ \Gamma$ must be read as "The normal distribution is preferred over the beta distribution which is preferred over the gamma distribution.

Table 5: *Participation rates and productivity premium, by industry*

Industry Name	♯ Obs.	PR	TFPP
	Market participation		
All Manufacturing	339,088	.740	.042
Automobile	9,276	.798	.022
Chemicals	35,360	.836	.025
Clothing and Footwear	27,632	.675	.136
Electric and Electronic Components	14,421	.773	.059
Electric and Electronic Equipment	18,522	.752	.063
House Equipment and Furnishings	23,707	.822	.049
Machinery and Mechanical Equipment	62,094	.702	.043
Metallurgy, Iron and Steel	60,856	.728	.008
Pharmaceuticals	9,067	.914	.030
Printing and Publishing	29,307	.612	.046
Textile	21,659	.800	.075
Transportation Machinery	5,125	.795	.057
Wood and Paper	22,062	.692	-.014
	Market entry		
All Manufacturing	74,832	.231	.017
Automobile	1,635	.252	.006
Chemicals	5,026	.266	.008
Clothing and Footwear	7,358	.197	.055
Electric and Electronic Components	2,858	.246	.035
Electric and Electronic Equipment	3,775	.218	.018
House Equipment and Furnishings	3,488	.280	.029
Machinery and Mechanical Equipment	15,612	.236	.014
Metallurgy, Iron and Steel	14,362	.231	.006
Pharmaceuticals	660	.312	-.039
Printing and Publishing	9,735	.218	.020
Textile	3,687	.234	.037
Transportation Machinery	886	.234	.003
Wood and Paper	5,750	.203	-.019
	Market remaining		
All Manufacturing	219,747	.929	.044
Automobile	6,577	.947	.021
Chemicals	26,278	.956	.029
Clothing and Footwear	15,974	.913	.129
Electric and Electronic Components	9,762	.943	.056
Electric and Electronic Equipment	11,918	.942	.074
House Equipment and Furnishings	17,018	.948	.048
Machinery and Mechanical Equipment	38,160	.912	.048
Metallurgy, Iron and Steel	38,957	.926	.010
Pharmaceuticals	7,334	.975	.062
Printing and Publishing	15,490	.870	.044
Textile	15,233	.947	.071
Transportation Machinery	3,570	.947	.073
Wood and Paper	13,476	.920	.001

PR: participation rate. TFPP: Total Factor Productivity Premium.

Table 6: *Maximum likelihood scores and number of iterations, by industry*

Industry Name	$l_{\mathcal{N}}(\mu,\sigma)$	$l_{\Gamma}(\alpha,\beta)$	$l_{\mathcal{B}}(\alpha,\beta)$	♯ite.$\mathcal{N}$	♯ite.Γ	♯ite.$\mathcal{B}$
			Market participation			
All Manufacturing	-191,631.4	-191,567.3	-191,782.3	4	7	5
Automobile	-4,640.3	-4,639.1	-4,642.3	4	9	7
Chemicals	-15,669.5	-15,663.3	-15,672.1	6	9	6
Clothing and Footwear	-15,806.7	-15,791.5	-15,895.9	6	12	7
Electric and Electronic Components	-7,541.9	-7,550.2	-7,549.3	4	9	4
Electric and Electronic Equipment	-10,084.2	-10,115.2	-10,095.5	4	9	4
House Equipment and Furnishings	-10,886.4	-10,882.5	-10,897.3	5	7	6
Machinery and Mechanical Equipment	-37,089.7	-37,100.5	-37,117.1	4	9	3
Metallurgy, Iron and Steel	-35,626.5	-35,615.8	-35,629.7	5	27	6
Pharmaceuticals	-2,639.4	-2,634.0	-2,642.3	6	11	8
Printing and Publishing	-19,319.5	-19,320.6	-19,326.8	5	12	5
Textile	-10,488.5	-10,475.7	-10,509	4	7	3
Transportation Machinery	-2,540.8	-2,540.8	-2,542.7	3	10	5
Wood and Paper	-	-	-	-	-	-
			Market entry			
All Manufacturing	-40,329.1	-40,325.9	-	4	18	-
Automobile	-922.6	-	-	4	-	-
Chemicals	-2,908.5	-	-	4	-	-
Clothing and Footwear	-	-3,598.5	-3,598.5	-	11	4
Electric and Electronic Components	-1,581.9	-1,583.4	-1,583.1	19	6	6
Electric and Electronic Equipment	-1,974.8	-1,975.3	-	6	22	-
House Equipment and Furnishings	-2,053.9	-2,054.6	-	4	11	-
Machinery and Mechanical Equipment	-8,510.4	-8,509.9	-	8	16	-
Metallurgy, Iron and Steel	-7,753	-	-	3	-	-
Pharmaceuticals	-	-	-	-	-	-
Printing and Publishing	-5,092.6	-5,090.6	-	5	22	-
Textile	-1,991.1	-1,989.7	-1,992.7	9	8	5
Transportation Machinery	-481.6	-	-	5	-	-
Wood and Paper	-	-	-	-	-	-
			Market remaining			
All Manufacturing	-55,426.8	-55,410.1	-55,466.6	5	6	12
Automobile	-1,359.1	-1,358.3	-1,359.4	5	13	12
Chemicals	-4,686.9	-4,686.5	-4,687.2	5	13	11
Clothing and Footwear	-4,403.4	-4,398.2	-4,427.9	4	9	5
Electric and Electronic Components	-2,086.7	-2,086.9	-2,086.7	4	7	9
Electric and Electronic Equipment	-2,565	-2,571.7	-2,567.7	4	8	8
House Equipment and Furnishings	-3,446.9	-3,445.1	-3,449.4	4	6	8
Machinery and Mechanical Equipment	-11,113.9	-11,118.9	-11,119.8	4	12	7
Metallurgy, Iron and Steel	-10,308.2	-10,305.3	-10,309.2	7	24	13
Pharmaceuticals	-852.3	-848.7	-855.5	4	7	9
Printing and Publishing	-5,919.8	-5,920.7	-5,920.6	4	8	9
Textile	-3,070.6	-3,068.4	-3,073.8	5	6	6
Transportation Machinery	-717.3	-718.2	-717.2	4	7	7
Wood and Paper	-3,746.7	-	-3,746.7	9	-	12

θ-attribute: Total Factor Productivity. l: log likelihood value. ♯ite: Number of iterations.

Table 7: *Maximum likelihood estimation of participation threshold distributions, using Total Factor Productivity as the θ-attribute*

Industry Name	$\mathcal{N}_{\mu}$	$\mathcal{N}_{\sigma^2}$	Γ_{μ}	Γ_{σ^2}	Γ_{p50}	$\mathcal{B}_{\mu}$	$\mathcal{B}_{\sigma^2}$	$\mathcal{B}_{p50}$
	Market participation							
All Manufacturing	.195	.284	.397	.189	.254	.315	.100	.198
Automobile	-.079	.550	.336	.291	.118	.238	.103	.050
Chemicals	-.212	.587	.276	.230	.079	.198	.090	.023
Clothing and Footwear	.419	.060	.455	.050	.419	.426	.049	.412
Electric and Electronic Components	.229	.168	.359	.119	.256	.307	.077	.227
Electric and Electronic Equipment	.279	.142	.382	.116	.287	.334	.074	.272
House Equipment and Furnishings	.105	.216	.302	.105	.197	.247	.075	.131
Machinery and Mechanical Equipment	.332	.152	.436	.120	.348	.375	.080	.325
Metallurgy, Iron and Steel	-.995	6.451	1.463	18.098	.022	.279	.174	.000
Pharmaceuticals	-1.253	1.711	.157	.215	.002	.094	.063	.000
Printing and Publishing	.370	.363	.582	.387	.381	.425	.126	.359
Textile	.194	.161	.334	.087	.252	.281	.072	.194
Transportation Machinery	.185	.181	.337	.101	.244	.284	.075	.194
Wood and Paper	-	-	-	-	-	-	-	-
	Market entry							
All Manufacturing	1.581	2.048	6.516	87.751	2.878	-	-	-
Automobile	2.354	7.426	-	-	-	-	-	-
Chemicals	2.022	5.666	-	-	-	-	-	-
Clothing and Footwear	-	-	1.665	2.133	1.263	.771	.090	.934
Electric and Electronic Components	1.018	.526	2.073	4.856	1.364	.741	.110	.930
Electric and Electronic Equipment	1.532	1.692	6.988	99.457	3.128	-	-	-
House Equipment and Furnishings	1.002	.673	2.017	5.164	1.254	-	-	-
Machinery and Mechanical Equipment	1.456	1.665	4.701	4.062	2.322	-	-	-
Metallurgy, Iron and Steel	3.060	11.710	-	-	-	-	-	-
Pharmaceuticals	-	-	-	-	-	-	-	-
Printing and Publishing	1.605	1.918	5.241	49.044	2.620	-	-	-
Textile	1.140	.748	2.220	5.665	1.449	.755	.116	.965
Transportation Machinery	6.046	57.867	-	-	-	-	-	-
Wood and Paper	-	-	-	-	-	-	-	-
	Market remaining							
All Manufacturing	-.335	.355	.148	.065	.043	.109	.045	.004
Automobile	-.873	.769	.100	.077	.002	.071	.037	.000
Chemicals	-.719	.543	.089	.052	.004	.066	.031	.000
Clothing and Footwear	.126	.094	.256	.037	.210	.206	.042	.136
Electric and Electronic Components	-.265	.259	.132	.049	.041	.105	.037	.009
Electric and Electronic Equipment	-.084	.157	.158	.042	.083	.131	.036	.038
House Equipment and Furnishings	-.357	.305	.125	.046	.036	.091	.035	.003
Machinery and Mechanical Equipment	-.021	.173	.210	.054	.133	.168	.047	.067
Metallurgy, Iron and Steel	-2.765	5.249	.210	.989	.000	.079	.060	.000
Pharmaceuticals	-.898	.537	.063	.026	.003	.036	.019	.000
Printing and Publishing	-.287	.546	.227	.161	.061	.165	.077	.010
Textile	-.263	.243	.136	.041	.056	.100	.035	.007
Transportation Machinery	-.087	.148	.164	.035	.100	.129	.034	.042
Wood and Paper	-3.187	476.9	-	-	-	.080	.072	.000

$\mathcal{N}_{\mu}$: Estimated mean of the normal distribution; $\mathcal{N}_{\sigma^2}$: estimated variance of the normal distribution; Γ_{μ}: estimated mean of the gamma distribution; Γ_{σ^2} estimated variance of the gamma distribution; Γ_{p50}: estimated median of the gamma distribution; $\mathcal{B}_{\mu}$: estimate mean of the beta distribution; $\mathcal{B}_{\sigma^2}$: estimated variance of the beta distribution; $\mathcal{B}_{p50}$: estimated median of the beta distribution.

Table 8: *Vuong's z test for model selection using Total Factor Productivity as the θ-attribute*

Industry Name	$z_{\mathcal{N},\Gamma}$	$z_{\mathcal{N},\mathcal{B}}$	$z_{\Gamma,\mathcal{B}}$	Ranking	Conclusion
				Market participation	
All Manufacturing	-371.9	1,601.1	66.5	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.8, .5)$
Automobile	-146.3	1,001.6	246.9	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.4, .9)$
Chemicals	-397.5	608.2	352.3	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.3, .8)$
Clothing and Footwear	-49	127.1	69.6	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(4.2, .1)$
Electric and Electronic Components	75.4	372.3	-8.9	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.2, .4)$[a]
Electric and Electronic Equipment	84.7	132.1	-111.8	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.3, .4)$[a]
House Equipment and Furnishings	-8.4	524.4	15.3	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.9, .3)$
Machinery and Mechanical Equipment	14.7	1,146.0	132.1	$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	$C \sim \mathcal{N}(.3, .4)$
Metallurgy, Iron and Steel	-1,031.4	1,551.8	704.4	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.1, 12.4)$
Pharmaceuticals	-307.5	284.2	165.6	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.1, 1.4)$
Printing and Publishing	26.6	30.5	63.7	$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	$C \sim \mathcal{N}(.4, .6)$
Textile	-193.2	23.3	14.8	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(1.3, .3)$
Transportation Machinery	-.6	179.5	42.4	$\mathcal{N} \sim \Gamma \succ \mathcal{B}$	$C \sim \mathcal{N}(.2, .4)$ or $C \sim \Gamma(1.1, .2)$
Wood and Paper	-	-	-	$\emptyset$	$\emptyset$
				Market entry	
All Manufacturing	-18.7	-	-	$\Gamma \succ \mathcal{N}$	$C \sim \Gamma(.5, 13.5)$
Automobile	-	-	-	$\mathcal{N}$	$C \sim \mathcal{N}(2.4, 2.7)$
Chemicals	-	-	-	$\mathcal{N}$	$C \sim \mathcal{N}(2.0, 2.4)$
Clothing and Footwear	-	-	.3	$\Gamma \sim \mathcal{B}$	$C \sim \Gamma(1.3, 1.3)$ or $C \sim \mathcal{B}(.7, .2)$
Electric and Electronic Components	33.3	98.8	-23.6	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(1.0, .7)$[a]
Electric and Electronic Equipment	133.6	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(1.5, 1.3)$
House Equipment and Furnishings	75.6	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(1.0, .8)$
Machinery and Mechanical Equipment	-149.1	-	-	$\Gamma \succ \mathcal{N}$	$C \sim \Gamma(.6, 8.5)$
Metallurgy, Iron and Steel	-	-	-	$\mathcal{N}$	$C \sim \mathcal{N}(3.1, 3.4)$
Pharmaceuticals	-	-	-	$\emptyset$	$\emptyset$
Printing and Publishing	-287.4	-	-	$\Gamma \succ \mathcal{N}$	$C \sim \Gamma(.6, 9.4)$
Textile	-154	337.9	171	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.9, 2.6)$
Transportation Machinery	-	-	-	$\mathcal{N}$	$C \sim \mathcal{N}(6.0, 7.6)$
Wood and Paper	-	-	-	$\emptyset$	$\emptyset$
				Market remaining	
All Manufacturing	-402.3	868.1	467.7	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.3, .4)$
Automobile	-223.7	517.3	196.2	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.1, .8)$
Chemicals	-106.7	236	94.6	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.2, .6)$
Clothing and Footwear	-71.7	68.5	48.6	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(1.8, .1)$
Electric and Electronic Components	9.7	-.4	-13.7	$\mathcal{N} \sim \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(-.3, .5)$ or $C \sim \mathcal{B}(.0, 1.4)$
Electric and Electronic Equipment	77	193.7	-84.1	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(-.1, .4)$[a]
House Equipment and Furnishings	-157.4	372.2	15.6	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.3, .4)$
Machinery and Mechanical Equipment	228.6	832.6	22.4	$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	$C \sim \mathcal{N}(-.0, .4)$
Metallurgy, Iron and Steel	-94.9	1,005.0	562.4	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.0, 4.7)$
Pharmaceuticals	-225.2	236.3	125.3	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.2, .4)$
Printing and Publishing	92.5	68.2	-4.3	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(-.3, .7)$[a]
Textile	-175.9	435.2	161.8	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.4, .3)$
Transportation Machinery	19.6	-30	-11.5	$\mathcal{B} \succ \mathcal{N} \succ \Gamma$	$C \sim \mathcal{B}(.3, 2.0)$
Wood and Paper	-	-993.3	-	$\mathcal{B} \succ \mathcal{N}$	$C \sim \mathcal{B}(.0, .0)$

$z_{\mathcal{N},\Gamma}$: $H_{\mathcal{N}\sim\Gamma} : |z| < +1.96$ $H_{\mathcal{N}\succ\Gamma} : z \geq +1.96$ $H_{\Gamma\succ\mathcal{N}} : z \leq -1.96$.
$z_{\mathcal{N},\mathcal{B}}$: $H_{\mathcal{N}\sim\mathcal{B}} : |z| < +1.96$ $H_{\mathcal{N}\succ\mathcal{B}} : z \geq +1.96$ $H_{\mathcal{B}\succ\mathcal{N}} : z \leq -1.96$.
$z_{\Gamma,\mathcal{B}}$: $H_{\Gamma\sim\mathcal{B}} : |z| < +1.96$ $H_{\Gamma\succ\mathcal{B}} : z \geq +1.96$ $H_{\mathcal{B}\succ\Gamma} : z \leq -1.96$.
The $[a]$ symbol indicates that caution is needed in the dominance of $\mathcal{N}$ over $\mathcal{B}$, as revealed by Monte Carlo results presented in Section 3.3.3.

Table 9: *Occurrence of diagnosis, according to the type of market participation*

Ranking	Participation	Entry	Remaining
	No density fit		
$\emptyset$	22	67	19
	Unique density fit		
$\mathcal{N}$	1	20	0
Γ	0	20	1
$\mathcal{B}$	1	10	0
	Two density fits		
$\mathcal{N} \succ \Gamma$	0	6	0
$\mathcal{N} \succ \mathcal{B}^a$	3	2	2
$\Gamma \succ \mathcal{B}$	0	23	0
$\Gamma \succ \mathcal{N}$	0	22	0
$\mathcal{B} \succ \mathcal{N}$	3	0	5
$\mathcal{B} \succ \Gamma$	0	12	0
$\Gamma \sim \mathcal{B}$	0	2	0
	Tree density fits		
$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	25	3	16
$\mathcal{N} \succ \mathcal{B} \succ \Gamma^a$	44	7	32
$\mathcal{N} \sim \Gamma \succ \mathcal{B}$	5	0	1
$\mathcal{N} \sim \Gamma \sim \mathcal{B}$	2	2	1
$\mathcal{N} \sim \mathcal{B} \succ \Gamma$	0	1	0
$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	98	13	78
$\Gamma \succ \mathcal{B} \succ \mathcal{N}$	11	7	14
$\Gamma \sim \mathcal{B} \succ \mathcal{N}$	2	0	0
$\mathcal{B} \succ \mathcal{N} \succ \Gamma$	14	2	48
$\mathcal{B} \succ \Gamma \succ \mathcal{N}$	3	2	3
$\mathcal{B} \succ \mathcal{N} \sim \Gamma$	0	0	1
	Overall dominance		
$\mathcal{N}$	80	47	51
Γ	111	87	94
$\mathcal{B}$	21	26	57
	Total		
	234	221	221

Figures represent counts of estimated densities for market participation, market entry or market remaining. The overall number of estimated densities for market participation is 13 industries observed for 18 years, yielding 234 trials of density estimation. Accounting for entry or remaining imply the loss of the first year of observation due to the use of a lagged year in identifying firm market entry and/or remaining.
The [a] symbol indicates that caution is needed in the dominance of $\mathcal{N}$ over $\mathcal{B}$, as revealed by Monte Carlo results presented in Section 3.3.3.

Appendix A. Algorithms for the Monte Carlo simulation exercises

Baseline Monte Carlo Settings

The Monte Carlo simulations are carried out as follows:

1. fix a sufficiently large number of agents N;
2. simulate the *true* θ-attribute data $\boldsymbol{\theta}^T$ from a known distribution g;
3. simulate the *true* threshold data $\boldsymbol{C}^T$ from a known distribution f;
4. let the agents compute their individual decision outcomes $\boldsymbol{\chi}$ according to Equation 1;
5. using the information available to the social researcher (i.e., $\boldsymbol{\theta}$ and $\boldsymbol{\chi}$), estimate via maximum likelihood the parameters $\hat{\Omega}$ that characterize the threshold distribution f;
6. repeat steps 2 to 5 a sufficient number of times M; and
7. use the M estimates $\hat{\boldsymbol{\Omega}}$ to evaluate the goodness of the estimation.

Monte Carlo Settings - Testing Assumption A2

The Monte Carlo simulations are carried out as follows:

1. fix a sufficiently large number of agents N;
2. fix a set of probability density functions $\mathcal{F} = f_1, f_2, \ldots, f_K$ to be used for the generation of the true thresholds and as priors (i.e. $\hat{\mathcal{F}}$) for the estimation of the threshold distribution parameters;
3. simulate the *true* θ-attribute data $\boldsymbol{\theta}^T$ from a known distribution g;
4. simulate the *true* threshold data $\boldsymbol{C}^T$ from the distribution f_k;
5. let the agents compute their individual decision outcomes $\boldsymbol{\chi}$ according to Equation 1;
6. using the information available to the social researcher (i.e., $\boldsymbol{\theta}$ and $\boldsymbol{\chi}$), estimate via maximum likelihood the parameters $\hat{\Omega}$ that characterize all the threshold distributions priors $\hat{\mathcal{F}}$;
7. compute all the pairwise Vuong z statistics;
8. repeat steps 3 to 7 a sufficient number of times M;
9. use all the M estimates $\hat{\boldsymbol{\Omega}}$ to evaluate the goodness of the estimation;
10. repeat steps 3 to 9 K times, each time using as true distribution of threshold a new the density function belonging to the set $\mathcal{F}$, as defined at step 2;
11. for each of the true f_k evaluate and compare the estimation errors generated by all the priors in $\hat{\mathcal{F}}$;
12. for each of the true f_k evaluate and compare the Vuong tests, to verify if the correct prior density $\hat{f}_k$ has been preferred to the alternative priors in the set $\hat{\mathcal{F}}$.

Monte Carlo Settings - Testing Assumption A4

The Monte Carlo simulations are carried out as follows:

1. fix a sufficiently large number of agents N;
2. fix a vector of noise $\boldsymbol{\sigma} = \sigma_1, \sigma_2, \ldots, \sigma_K$;
3. simulate the *true* θ-attribute data $\boldsymbol{\theta}^T$ from a known distribution g;
   - generate also the noisy θ-attribute data $\boldsymbol{\theta}^{\varepsilon} = \boldsymbol{\theta}^T + \boldsymbol{\varepsilon}^{\theta}$;
4. simulate the *true* threshold data $\boldsymbol{C}^T$ from the known distribution f_k;
   - generate also the noisy threshold data $\boldsymbol{C}^{\varepsilon} = \boldsymbol{C}^T + \boldsymbol{\varepsilon}^{c}$;
5. let the agents compute their individual decision outcomes $\boldsymbol{\chi}$ according to Equation 12;
6. using the information available to the social researcher, estimate via maximum likelihood the parameters $\hat{\Omega}$ that characterize the threshold distribution f_k;
7. repeat steps 3 to 6 a sufficient number of times M;
8. use the M estimates $\hat{\boldsymbol{\Omega}}$ to evaluate the goodness of the estimation;
9. repeat steps 3 to 8 for all the noise levels $\boldsymbol{\sigma}$, as defined at step 2; and
10. evaluate and compare the goodness of all the values of $\boldsymbol{\sigma}$.

Appendix B. Distributions

B.1 Univariate normal distribution

In the case of a normal distribution, the probability density function is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{B.1}$$

where μ and σ represent the average and the standard deviation, respectively. This distribution is typically denoted as $x \sim \mathcal{N}(\mu, \sigma^2)$. Integrating over the interval $(-\infty, \bar{x}]$ yields the cumulative density function:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} exp \int_{-\infty}^{\bar{x}} \left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx \tag{B.2}$$

B.2 Univariate gamma distribution

In the case of a gamma distribution, the probability density function is defined as:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} exp\left(\frac{-x}{\beta}\right) \tag{B.3}$$

where $\alpha > 0$ and $\beta > 0$ represent the shape and scale parameters, respectively, and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} exp(-t)dt$ is the gamma function. The mean and variance of this distribution are a combination of shape and scale parameters and read as $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$, respectively. This distribution is typically denoted as $x \sim \Gamma(\alpha, \beta)$. Integrating over the interval $(0, \bar{x}]$ yields the cumulative density function:

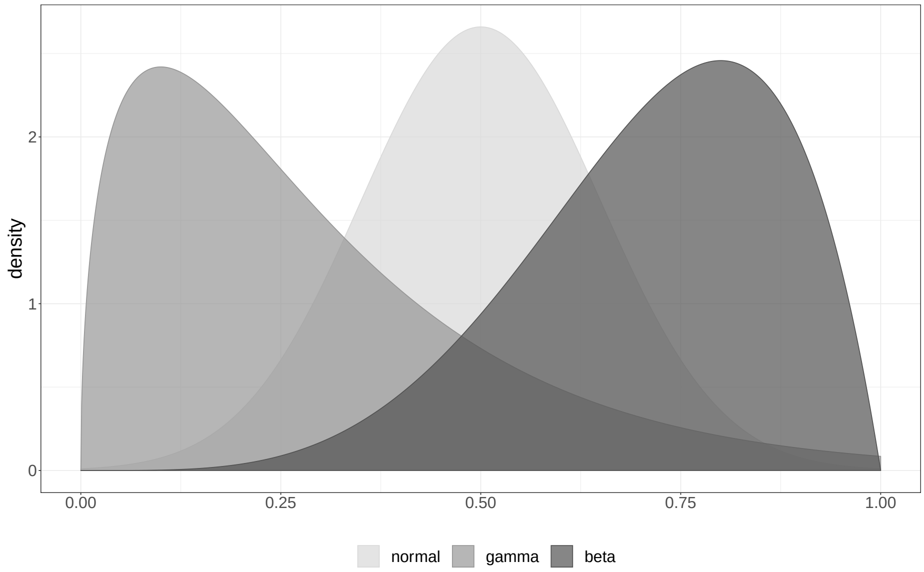
$$F(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\bar{x}} x^{\alpha-1} exp\left(\frac{-x}{\beta}\right) dx \tag{B.4}$$

B.3 Univariate beta distribution

For a beta distribution, the probability density function is defined as:

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{B.5}$$

Figure B.1: *Examples of normal, gamma and beta distributions.*



Figures are drawn with parametrization $\mathcal{N}(0.5, 0.0225)$ for the normal, $\Gamma(1.5, .2)$ for the gamma and $B(5, 2)$ for the beta.

where $\alpha > 0$ and $\beta > 0$ represent the shape$_1$ and shape$_2$ parameters, and $B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ is the beta function, a normalization constant derived from the composition of gamma functions and that ensures the total probability is one. The mean and variance of this distribution are a combination of shape$_1$ and shape$_2$ parameters and read as $\mu = \frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \alpha\beta^2$, respectively. This distribution is typically denoted as $x \sim \mathcal{B}(\alpha, \beta)$. Integrating over the interval $(0, 1)$ yields the cumulative density function:

$$F(x) = I_x(\alpha, \beta) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt} \tag{B.6}$$

where $I_x(\alpha, \beta)$ is the regularized incomplete beta function, which is the ratio between the incomplete beta function (integral between 0 and x), while the denominator the complete beta function (integral between 0 and 1).

Examples of the three distributions are reported in Figure B.1.

Appendix C. Productivity measures

All nominal output and input variables are available at the firm level. Industry-level data are used for price indexes, hours worked and depreciation rates.

Output

Gross output deflated using sectoral price indexes published by INSEE (French System of National Accounts).

Labor

Labor input is obtained by multiplying the number of effective workers (i.e., the number of employees plus the number of outsourced workers minus workers taken from other firms) by average hours worked. The annual series for hours worked are available at the 2-digit industry level and provided by *GGDC Groningen Growth Development Center*). This choice has been made because there are no data on hours worked in the EAE survey. Note also that a large decline in hours worked occurred between 1999 and 2000 because of the specific "French 35 hours policy" (on average, hours worked fell from 38.39 in 1999 to 36.87 in 2000).

Capital input

Capital stocks are computed from the investment and book value of tangible assets following the traditional perpetual inventory method (PIM):

$$K_t = (1 - \delta_{t-1})\, K_{t-1} + I_t \tag{C.1}$$

where δ_t is the depreciation rate and I_t is real investment (deflated nominal investment). Both investment price indexes and depreciation rates are available at the 2-digit industrial classification from INSEE data series.

Intermediate inputs

Intermediate inputs are defined as purchases of materials and merchandise, transport and traveling, and miscellaneous expenses. They are deflated using sectoral price indexes for intermediate inputs published by INSEE (French System of National Accounts).

Input cost shares

With w, c and m denoting the wage rate, user cost of capital and price index for intermediate inputs, respectively, $CT_{kt} = w_{kt}L_{kt} + c_{It}K_{kt} + m_{It}M_{kt}$ represents the total cost of production of firm k at time t. Labor, capital and intermediate input cost shares are then respectively given by

$$s_{Lkt} = \frac{w_{kt}L_{kt}}{CT_{kt}} \; ; \; s_{Kkt} = \frac{c_{It}K_{kt}}{CT_{kt}} \; ; \; s_{Mkt} = \frac{m_{It}M_{kt}}{CT_{kt}} \tag{C.2}$$

To compute the labor cost share, we rely on the variable "labor compensation" in the EAE survey. This value includes total wages paid as salaries plus income tax withholding and is used to approximate the theoretical variable $w_{kt}L_{kt}$. To compute the intermediate input cost share, we use variables on intermediate goods consumption in the EAE survey and the price index for intermediate inputs in industry I provided by INSEE.

We compute the user cost of capital by using the Hall (1988) methodology where the user cost of capital (i.e., the rental price of capital) in the presence of a proportional tax on business income and of a fiscal depreciation formula is given by[18]

$$c_{It} = (r_t + \delta_{It} - \pi_t^e) \left(\frac{1 - \tau_t z_I}{1 - \tau_t} \right) p_{IKt} \tag{C.3}$$

where τ_t is the business income tax in period t and Z_I denotes the present value of the depreciation deduction on one nominal unit investment in industry I. A complex depreciation formula can be employed for tax purposes in France. To simplify, we choose to rely on the usual following depreciation formula

$$z_I = \sum_{t=1}^{n} \frac{(1 - \bar{\delta}_I)^{t-1}\delta}{(1 + \bar{r})^{t-1}}$$

where $\bar{\delta}_I$ is a mean of the industrial deprecation rates for the period 1984-2002 and $\bar{r}$ is the mean nominal interest rate over the period 1990-2002.

We measure firm productive efficiency by means of two complementary indicators, namely apparent labor productivity (ALP) and total factor productivity (TFP). *Labor productivity* is defined as the log-ratio of real value added on labor (hours worked):

$$\ln LP_{it} = \ln \left(\frac{V_{it}}{L_{it}} \right) \tag{C.4}$$

where V_{it} denotes the value added of the firm deflated by sectoral price indexes published

[18]In this equation, we abstract from tax credit allowance.

by INSEE (French System of National Accounts). Next, we compute *total factor productivity* by using the so-called *multilateral productivity index* first introduced by Caves et al. (1982) and extended by Good et al. (1997). This methodology consists of computing the TFP index for firm i at time t as follows:

$$\ln TFP_{it} = \ln\ Y_{it} - \overline{\ln\ Y_t} + \sum_{\tau=2}^{t} \left(\overline{\ln\ Y_\tau} - \overline{\ln\ Y_{\tau-1}}\right) - \left[\begin{array}{l} \sum_{n=1}^{N} \frac{1}{2} \left(S_{nit} + \overline{S_{nt}}\right) \left(\ln\ X_{nit} - \overline{\ln\ X_{nt}}\right) \\ + \sum_{\tau=2}^{t} \sum_{n=1}^{N} \frac{1}{2} \left(\overline{S_{n\tau}} + \overline{S_{n\tau-1}}\right) \left(\overline{\ln\ X_{n\tau}} - \overline{\ln\ X_{n\tau-1}}\right) \end{array} \right] \tag{C.5}$$

where Y_{it} denotes real gross output produced by firm i at time t using the set of n inputs X_{nit} and input X is alternatively capital stocks (K), labor in terms of hours worked (L) and intermediate inputs (M). S_{nit} is the cost share of input X_{nit} in the total cost. Subscripts τ and n are indexes for time and inputs, respectively. Symbols with upper bar correspond to measures for the reference point (the hypothetical firm), computed as the means of the corresponding firm-level variables, over all firms in year t. Note that Eq.(C.5) implies that references points $\overline{\ln Y}$ and $\overline{\ln X}$ are the geometric means of the firm's output quantities and input quantities, respectively, whereas the cost share of inputs of the representative firms $\overline{S}$ is computed as the arithmetic mean of the cost share of all firms in the dataset.

This methodology is particularly well suited for comparisons within firm-level panel data sets across industries because it guarantees the transitivity of any comparison between two firm-year observations by expressing each firm's input and output as deviations from a single reference point.

Appendix D. Robustness checks: application to labor productivity

This appendix presents robustness checks for all results on participation, entry and remaining threshold distributions using labor productivity as the θ-attribute. We stack all tables without commenting further.

Table D.1: *Participation rates and labor productivity premium, by industry*

Industry Name	♯ Obs.	PR	ALPP
	Market participation		
All Manufacturing	337,275	.738	.171
Automobile	9,213	.797	.154
Chemicals	34,886	.836	.161
Clothing and Footwear	27,475	.668	.422
Electric and Electronic Components	14,413	.772	.252
Electric and Electronic Equipment	19,006	.753	.217
House Equipment and Furnishings	23,526	.822	.162
Machinery and Mechanical Equipment	61,870	.701	.131
Metallurgy, Iron and Steel	60,768	.727	.080
Pharmaceuticals	8,677	.915	.170
Printing and Publishing	29,001	.608	.145
Textile	21,418	.796	.256
Transportation Machinery	5,076	.794	.221
Wood and Paper	21,946	.691	.168
	Market entry		
All Manufacturing	74,832	.231	.017
Automobile	1,635	.252	.006
Chemicals	5,026	.266	.008
Clothing and Footwear	7,358	.197	.055
Electric and Electronic Components	2,858	.246	.035
Electric and Electronic Equipment	3,775	.218	.018
House Equipment and Furnishings	3,488	.280	.029
Machinery and Mechanical Equipment	15,612	.236	.014
Metallurgy, Iron and Steel	14,362	.231	.006
Pharmaceuticals	660	.312	-.039
Printing and Publishing	9,735	.218	.020
Textile	3,687	.234	.037
Transportation Machinery	886	.234	.003
Wood and Paper	5,750	.203	-.019
	Market remaining		
All Manufacturing	218,102	.929	.157
Automobile	6,517	.946	.122
Chemicals	25,926	.956	.193
Clothing and Footwear	15,708	.911	.373
Electric and Electronic Components	9,759	.943	.220
Electric and Electronic Equipment	12,286	.943	.220
House Equipment and Furnishings	16,890	.948	.143
Machinery and Mechanical Equipment	37,983	.912	.122
Metallurgy, Iron and Steel	38,870	.925	.070
Pharmaceuticals	7,015	.974	.241
Printing and Publishing	15,241	.869	.137
Textile	14,981	.946	.214
Transportation Machinery	3,535	.946	.256
Wood and Paper	13,391	.920	.161

PR: participation rate. ALPP: Labor Productivity Premium.

Table D.2: *Maximum likelihood scores and number of iterations, by industry*

Industry Name	$l_{\mathcal{N}}(\mu,\sigma)$	$l_{\Gamma}(\alpha,\beta)$	$l_{\mathcal{B}}(\alpha,\beta)$	♯ite.$\mathcal{N}$	♯ite.Γ	♯ite.$\mathcal{B}$
	Market participation					
All Manufacturing	-187,902.0	-188,348.8	-188,093	5	12	8
Automobile	-4,523.7	-4,530.3	-	4	14	-
Chemicals	-15,244.2	-15,248.2	-	4	6	-
Clothing and Footwear	-15,086.9	-15,079.5	-15,186.7	6	12	4
Electric and Electronic Components	-7,260.6	-7,291.9	-7,275	4	12	7
Electric and Electronic Equipment	-10,073.6	-10,136.9	-10,102.6	4	9	7
House Equipment and Furnishings	-10,704.3	-10,716.2	-	4	9	-
Machinery and Mechanical Equipment	-36,738.3	-36,828.9	-36,765.6	5	8	7
Metallurgy, Iron and Steel	-35,334.9	-35,345.3	-	4	10	-
Pharmaceuticals	-2,493.6	-2,488.3	-	4	7	-
Printing and Publishing	-19,030.4	-19,071.4	-19,037.7	6	10	6
Textile	-10,271.1	-10,301.5	-10,287.1	3	9	6
Transportation Machinery	-2,474	-2,483.6	-2,479.4	3	9	7
Wood and Paper	-13,105.3	-13,157.5	-13,116.1	5	9	7
	Market entry					
All Manufacturing	-	-40,166.9	-40,162.0	-	13	7
Automobile	-923.1	-923.8	-923.5	4	15	6
Chemicals	-	-	-	-	-	-
Clothing and Footwear	-	-3,541.8	-3,549.8	-	13	8
Electric and Electronic Components	-1,569.1	-1,570.9	-1,570.4	8	6	5
Electric and Electronic Equipment	-	-2,010.8	-2,009.6	-	12	7
House Equipment and Furnishings	-2,026.7	-2,027.7	-2,026.7	9	8	6
Machinery and Mechanical Equipment	-	-8,484.5	-8,483.7	-	12	7
Metallurgy, Iron and Steel	-7,741.5	-7,740.4	-7,741.4	7	16	7
Pharmaceuticals	-	-	-	-	-	-
Printing and Publishing	-	-5,075.3	-5,073.9	-	19	6
Textile	-1,985.3	-1,988.5	-1,987.3	27	7	8
Transportation Machinery	-474.0	-473.8	-474	6	22	8
Wood and Paper	-	-2,872.4	-2,871.2	-	11	9
	Market remaining					
All Manufacturing	-54,869.3	-54,950.6	-	4	7	-
Automobile	-1,346.4	-1,346.6	-1,347.1	5	7	12
Chemicals	-4,555.4	-4,560.4	-	4	8	-
Clothing and Footwear	-4,367.9	-4,364.7	-4,381.7	4	13	7
Electric and Electronic Components	-2,048.8	-2,055.4	-	4	5	-
Electric and Electronic Equipment	-2,571.2	-2,581.8	-	4	6	-
House Equipment and Furnishings	-3,410.4	-3,411.6	-	4	5	-
Machinery and Mechanical Equipment	-11,092.9	-11,111.6	-	4	10	-
Metallurgy, Iron and Steel	-10,301.8	-10,304.5	-	4	8	-
Pharmaceuticals	-816.5	-815.3	-	4	6	-
Printing and Publishing	-5,836.5	-5,846.5	-5,837.1	5	8	9
Textile	-3,066.6	-3,071.1	-3,068.8	4	5	9
Transportation Machinery	-709.6	-714.1	-	3	5	-
Wood and Paper	-3,645.2	-3,653.6	-	3	9	-

θ-attribute: Apparent Labor Productivity. l: log likelihood value. ♯ite: Number of iterations.

Table D.3: *Maximum likelihood estimation of participation threshold distributions.*

Industry Name	$\mathcal{N}_{\mu}$	$\mathcal{N}_{\sigma^2}$	Γ_{μ}	Γ_{σ^2}	Γ_{p50}	$\mathcal{B}_{\mu}$	$\mathcal{B}_{\sigma^2}$	$\mathcal{B}_{p50}$
	Market participation							
All Manufacturing	.255	.116	.350	.087	.271	.311	.062	.255
Automobile	.192	.113	.301	.068	.229	-	-	-
Chemicals	-.001	.232	.242	.099	.125	-	-	-
Clothing and Footwear	.385	.031	.405	.025	.385	.391	.028	.381
Electric and Electronic Components	.281	.062	.332	.047	.286	.311	.041	.279
Electric and Electronic Equipment	.298	.063	.346	.053	.296	.323	.044	.292
House Equipment and Furnishings	.148	.122	.274	.065	.200	-	-	-
Machinery and Mechanical Equipment	.310	.097	.385	.085	.314	.347	.059	.308
Metallurgy, Iron and Steel	.138	.316	.380	.240	.201	-	-	-
Pharmaceuticals	-.654	.669	.130	.111	.006	-	-	-
Printing and Publishing	.353	.203	.499	.242	.350	.400	.095	.349
Textile	.230	.081	.306	.052	.252	.280	.046	.233
Transportation Machinery	.214	.096	.301	.066	.232	.273	.052	.215
Wood and Paper	.321	.094	.391	.085	.321	.354	.058	.318
	Market entry							
All Manufacturing	-	-	2.653	1.074	1.543	.732	.120	.939
Automobile	1.117	1.044	4.421	4.311	1.956	.719	.138	.960
Chemicals	-	-	-	-	-	-	-	-
Clothing and Footwear	-	-	.843	.268	.740	.668	.070	.725
Electric and Electronic Components	.784	.273	1.300	1.481	.946	.678	.101	.791
Electric and Electronic Equipment	-	-	2.599	9.076	1.568	.738	.112	.930
House Equipment and Furnishings	.782	.343	1.344	1.836	.925	.661	.110	.779
Machinery and Mechanical Equipment	-	-	2.572	9.387	1.503	.728	.120	.932
Metallurgy, Iron and Steel	1.397	1.644	6.125	8.437	2.613	.750	.135	.990
Pharmaceuticals	-	-	-	-	-	-	-	-
Printing and Publishing	-	-	5.002	47.2	2.393	.757	.125	.983
Textile	.805	.271	1.318	1.453	.974	.690	.098	.807
Transportation Machinery	1.804	3.528	21.6	1,522.1	5.438	.754	.150	.999
Wood and Paper	-	-	2.960	11.9	1.777	.759	.109	.956
	Market remaining							
All Manufacturing	-.203	.215	.136	.047	.048	-	-	-
Automobile	-.389	.289	.103	.041	.020	.080	.030	.002
Chemicals	-.340	.225	.093	.030	.021	-	-	-
Clothing and Footwear	.120	.071	.233	.027	.195	.196	.032	.144
Electric and Electronic Components	-.080	.122	.139	.031	.075	-	-	-
Electric and Electronic Equipment	-.025	.099	.151	.029	.094	-	-	-
House Equipment and Furnishings	-.312	.235	.108	.036	.030	-	-	-
Machinery and Mechanical Equipment	-.094	.179	.165	.052	.078	-	-	-
Metallurgy, Iron and Steel	-.636	.600	.120	.096	.005	-	-	-
Pharmaceuticals	-.806	.424	.047	.022	.000	-	-	-
Printing and Publishing	-.158	.326	.202	.115	.064	.163	.061	.029
Textile	-.222	.186	.118	.034	.044	.095	.028	.012
Transportation Machinery	-.043	.101	.138	.028	.079	-	-	-
Wood and Paper	-.114	.177	.155	.047	.072	-	-	-

$\mathcal{N}_{\mu}$: Estimated mean of the normal distribution; $\mathcal{N}_{\sigma^2}$: estimated variance of the normal distribution; Γ_{μ}: estimated mean of the gamma distribution; Γ_{σ^2} estimated variance of the gamma distribution; Γ_{p50}: estimated median of the gamma distribution; $\mathcal{B}_{\mu}$: estimate mean of the beta distribution; $\mathcal{B}_{\sigma^2}$: estimated variance of the beta distribution; $\mathcal{B}_{p50}$: estimated median of the beta distribution.

Table D.4: *Vuong's z test for model selection using Labor Productivity as the θ-attribute*

Industry Name	$z_{\mathcal{N},\Gamma}$	$z_{\mathcal{N},\mathcal{B}}$	$z_{\Gamma,\mathcal{B}}$	Ranking	Conclusion
	Market participation				
All Manufacturing	712.4	1,294.7	-523.3	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.3, .3)^a$
Automobile	115	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(.2, .3)$
Chemicals	51.2	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.0, .5)$
Clothing and Footwear	-1.4	113.6	44.3	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(6.6, .1)$
Electric and Electronic Components	13.8	227.2	-82.3	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.3, .2)^a$
Electric and Electronic Equipment	42.7	67.8	-78.9	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.3, .3)^a$
House Equipment and Furnishings	252.2	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(.1, .3)$
Machinery and Mechanical Equipment	591.6	1,39.1	-643.8	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.3, .3)^a$
Metallurgy, Iron and Steel	275.8	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(.1, .6)$
Pharmaceuticals	-196.4	-	-	$\Gamma \succ \mathcal{N}$	$C \sim \Gamma(.2, .9)$
Printing and Publishing	37.8	51.6	-432.9	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.4, .5)^a$
Textile	219.9	506.6	-76.9	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.2, .3)^a$
Transportation Machinery	91.6	428.3	-49.2	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.2, .3)^a$
Wood and Paper	409.6	64.8	-586.1	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.3, .3)^a$
	Market entry				
All Manufacturing	-	-	-197.1	$\mathcal{B} \succ \Gamma$	$C \sim \mathcal{B}(.5, .2)$
Automobile	14.9	366.5	-179.4	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(1.1, 1.0)^a$
Chemicals	-	-	-	$\emptyset$	$\emptyset$
Clothing and Footwear	-	-	72.6	$\Gamma \succ \mathcal{B}$	$C \sim \Gamma(2.7, .3)$
Electric and Electronic Components	76.6	198.3	-41	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.8, .5)^a$
Electric and Electronic Equipment	-	-	-28.8	$\mathcal{B} \succ \Gamma$	$C \sim \mathcal{B}(.5, .2)$
House Equipment and Furnishings	81.5	29.3	-176.6	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.8, .6)^a$
Machinery and Mechanical Equipment	-	-	-176	$\mathcal{B} \succ \Gamma$	$C \sim \mathcal{B}(.5, .2)$
Metallurgy, Iron and Steel	-154.4	-93.2	315.3	$\Gamma \succ \mathcal{B} \succ \mathcal{N}$	$C \sim \Gamma(.5, 13.1)$
Pharmaceuticals	-	-	-	$\emptyset$	$\emptyset$
Printing and Publishing	-	-	-165.4	$\mathcal{B} \succ \Gamma$	$C \sim \mathcal{B}(.4, .1)$
Textile	10.8	297.6	-78.2	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(.8, .5)^a$
Transportation Machinery	-67	92.8	121.1	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(.3, 7.5)$
Wood and Paper	-	-	-303.9	$\mathcal{B} \succ \Gamma$	$C \sim \mathcal{B}(.5, .2)$
	Market remaining				
All Manufacturing	1,036.7	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.2, .5)$
Automobile	29.9	525.6	52.8	$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	$C \sim \mathcal{N}(-.4, .5)$
Chemicals	214.4	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.3, .5)$
Clothing and Footwear	-4.3	352.2	89.9	$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	$C \sim \Gamma(2.0, .1)$
Electric and Electronic Components	245.2	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.1, .3)$
Electric and Electronic Equipment	179	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.0, .3)$
House Equipment and Furnishings	118.1	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.3, .5)$
Machinery and Mechanical Equipment	636.9	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.1, .4)$
Metallurgy, Iron and Steel	376.4	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.6, .8)$
Pharmaceuticals	-10.7	-	-	$\Gamma \succ \mathcal{N}$	$C \sim \Gamma(.1, .5)$
Printing and Publishing	443.2	224.6	-42.2	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(-.2, .6)^a$
Textile	244.9	191.4	-54.8	$\mathcal{N} \succ \mathcal{B} \succ \Gamma$	$C \sim \mathcal{N}(-.2, .4)^a$
Transportation Machinery	23.1	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.0, .3)$
Wood and Paper	44.4	-	-	$\mathcal{N} \succ \Gamma$	$C \sim \mathcal{N}(-.1, .4)$

$z_{\mathcal{N},\Gamma}$: $H_{\mathcal{N}\sim\Gamma} : |z| < +1.96$ $H_{\mathcal{N}\succ\Gamma} : z \geq +1.96$ $H_{\Gamma\succ\mathcal{N}} : z \leq -1.96$.
$z_{\mathcal{N},\mathcal{B}}$: $H_{\mathcal{N}\sim\mathcal{B}} : |z| < +1.96$ $H_{\mathcal{N}\succ\mathcal{B}} : z \geq +1.96$ $H_{\mathcal{B}\succ\mathcal{N}} : z \leq -1.96$.
$z_{\Gamma,\mathcal{B}}$: $H_{\Gamma\sim\mathcal{B}} : |z| < +1.96$ $H_{\Gamma\succ\mathcal{B}} : z \geq +1.96$ $H_{\mathcal{B}\succ\Gamma} : z \leq -1.96$.
The $[a]$ symbol indicates that caution is needed in the dominance of $\mathcal{N}$ over $\mathcal{B}$, as revealed by Monte Carlo results presented in Section 3.3.3.

Table D.5: *Occurrence of diagnosis, according to the type of market participation*

Ranking	Participation	Entry	Remaining
		No density fit	
$\emptyset$	1	38	4
		Unique density fit	
$\mathcal{B}$	0	9	0
$\mathcal{N}$	0	4	1
		Two density fits	
$\mathcal{N} \succ \Gamma$	49	0	83
$\mathcal{N} \succ \mathcal{B}^a$	0	9	0
$\mathcal{N} \sim \Gamma$	1	0	3
$\Gamma \succ \mathcal{N}$	22	0	25
$\Gamma \succ \mathcal{B}$	0	45	0
$\Gamma \sim \mathcal{B}$	0	3	0
$\mathcal{B} \succ \mathcal{N}$	0	6	0
$\mathcal{B} \succ \Gamma$	0	21	0
		Tree density fits	
$\mathcal{N} \succ \Gamma \succ \mathcal{B}$	13	4	15
$\mathcal{N} \succ \mathcal{B} \succ \Gamma^a$	101	38	45
$\mathcal{N} \sim \Gamma \sim \mathcal{B}$	7	0	0
$\mathcal{N} \sim \Gamma \succ \mathcal{B}$	4	0	0
$\Gamma \succ \mathcal{N} \succ \mathcal{B}$	28	11	28
$\Gamma \succ \mathcal{B} \succ \mathcal{N}$	1	24	6
$\mathcal{B} \succ \Gamma \succ \mathcal{N}$	1	2	0
$\mathcal{B} \succ \mathcal{N} \succ \Gamma$	6	6	11
$\mathcal{B} \succ \mathcal{N} \sim \Gamma$	0	1	0
		Overall dominance	
$\mathcal{N}$	174	55	147
Γ	52	83	59
$\mathcal{B}$	7	45	11
		Total	
	234	221	221

Figures represent counts of estimated densities for market participation, market entry or market remaining. The overall number of estimated densities for market participation is 13 industries observed for 18 years, yielding 234 trials of density estimation. Accounting for entry or remaining imply the loss of the first year of observation due to the use of a lagged year in identifying firm market entry and/or remaining.

The [a] symbol indicates that caution is needed in the dominance of $\mathcal{N}$ over $\mathcal{B}$, as revealed by Monte Carlo results presented in Section 3.3.3.