

## Can you jump this high? Quantifying barriers to market participation<sup>☆</sup>

Mattia Guerini<sup>a,b,c,\*</sup>, Patrick Musso<sup>b</sup>, Lionel Nesta<sup>b,d,e</sup>

<sup>a</sup> Università di Brescia, Italy

<sup>b</sup> Université Côte d'Azur, CNRS, GREDEG, France

<sup>c</sup> Sant'Anna School of Advanced Studies, Italy

<sup>d</sup> OFCE, SciencesPo, France

<sup>e</sup> SKEMA Business School, France



### ARTICLE INFO

#### JEL classification:

C40

D01

F14

#### Keywords:

Parametric distributions of thresholds

Maximum likelihood estimation

Fixed costs

Export decision

### ABSTRACT

We develop a new method to estimate the parameters of threshold distributions for market participation based on 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. 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.

### 1. Introduction

This paper proposes a method to estimate barriers to market participation. 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 importantly, markets. As such, our method offers a simple and intuitive solution to measuring barriers that inhibit market participation.

Previous works have attempted to measure such barriers, focusing essentially on costs. Das et al. (2007) developed a structural model that allows the empirical estimations of sunk and fixed costs of exporting for Colombian manufacturing firms. They found that sunk entry costs into export markets amount to, on average, \$400,000, while the fixed costs appeared to be negligible. Using a dynamic model of optimal stock market participation, Khorunzhina (2013) estimated that stock market participation costs for consumers amount to 4–6% of labor income. More recently, Fan and Xiao (2015) estimated that entry costs in the U.S.

local telephone industry reach \$6.5 million.

Barriers, however, are more general than costs, and stem from the existence of technical, regulatory, and information barriers. 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 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, barriers represent a minimum value above (below) which an economic agent decides (not) to participate in a given market.

In this paper, we focus on thresholds, defined as unobservable firm-specific attributes that condition the market participation of an economic actor. While barriers pertain uniformly to all economic agents, thresholds relate to an agent-specific characteristic conditioning market participation. More precisely, an agent will participate in the mar-

<sup>☆</sup> Mattia Guerini has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-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 authors wish to thank the anonymous referee and the associate editor for their helpful comments and suggestions. The usual disclaimers apply.

\* Corresponding author. Università di Brescia, Dipartimento di Economia e Management, Via San Faustino 74/b, 25122 Brescia, Italy.

E-mail addresses: [mattia.guerini@unibs.it](mailto:mattia.guerini@unibs.it) (M. Guerini), [patrick.musso@univ-cotedazur.fr](mailto:patrick.musso@univ-cotedazur.fr) (P. Musso), [lionel.nesta@univ-cotedazur.fr](mailto:lionel.nesta@univ-cotedazur.fr) (L. Nesta).

ket only if some of his/her attributes – say productivity – exceeds the required threshold value of productivity. Conversely, if an agent's attribute falls short of the threshold, he/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.

The major problem is that thresholds are empirically unobservable to social scientists or policymakers. In some instances, specific surveys may explicitly focus on particular thresholds. In their analysis of wage formation and unemployment, [Brown and Taylor \(2013\)](#) exploited 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 scant. An alternative is to model thresholds as a theoretical parameter that firms must overcome to participate in market activities. For example, [Eaton et al. \(2011\)](#) modeled export costs with a series of behavioral assumptions and of data requirement which are needed in order to estimate the associated structural parameters. Although analytically rigorous, such models remain hard to replicate, first and foremost because they cannot be easily adapted to other contexts.

In this paper, 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. Our method is extremely flexible to a variety of contexts and data. As such, it should be viewed as a reduced form of a more complex web of forces determining market participation that although embodying costs, may include unobserved agent-specific characteristics that may increase or decrease the required level of endowments in order to enter a given market. Thus, thresholds are more general than mere costs, and two agents coping with the same entry costs may cope with different thresholds for market entry.

Knowledge of threshold distributions is also important for policy assessment. In fact, a policy fostering market entry may affect participation either for all potential entrants or for specific agents, and as such, it may represent a symmetric or conversely an asymmetric shock. Symmetric shocks are those policies that affect the common determinants of thresholds, shifting the whole distribution on its support, but keeping constant the relative position of agents. Asymmetric shocks are shocks that affect specific categories of agents within the distribution, thereby affecting higher moments of the distribution. To be able to qualify whether a given policy shock has constituted a symmetric or asymmetric shock is important for a more precise and nuanced policy assessment. It also helps depict unintended consequences of given policies in terms of market participation by agents.

Our contribution is four-fold. 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  $\theta$ -attribute, we derive a likelihood function that allows the recovery of the concealed threshold distribution. Importantly, we make use of [Vuong \(1989\)](#)'s 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 also 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 two structural shocks in the European Union on different quantiles of threshold distributions. We also show

that the establishment of the single market and the introduction of the euro translated into significantly lower thresholds for the majority of firms. However, while the first shock has been symmetric, the second one has been asymmetric, excluding a minor share of manufacturing firms from international trade.

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

## 2. Econometric strategy

### 2.1. The intuition

We consider a series of agents  $i = 1, \dots, N$  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  $\theta_i$  that affects the decision outcome. This  $\theta$ -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  $\theta$ -attribute is sufficiently large and 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:

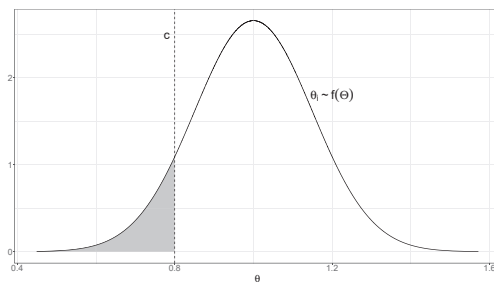
$$\begin{cases} \chi_i = 1 & \text{if } \theta_i \geq c_i \\ \chi_i = 0 & \text{if } \theta_i < c_i \end{cases} \quad (1)$$

The theoretical economic literature typically assumes homogeneous thresholds across agents, i.e.,  $c_i \sim \delta$ , with  $\delta$  representing the Dirac delta distribution (see [Pissarides, 1974](#); [McDonald and Siegel, 1986](#); [Dixit, 1989](#), as early developers of such an approach).<sup>1</sup> Fig. 1 describes this particular case by assuming an agent-specific, normally distributed  $\theta_i$  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  $\theta$ -attribute domain ( $\theta_i < c$ ). This representation implies a complete separation of agents, where only those whose  $\theta_i$  values exceed threshold  $c$  participate in the market.

From an empirical perspective, the issue is that the social scientist observes the decision outcome  $\chi_i$  and the agent's  $\theta$ -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 participation, conditional on his/her  $\theta$ -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 also matches the fraction of actors with a positive (negative) decision outcome with the  $\theta$ -attribute and produces the probability of participation conditional on the  $\theta$ -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 non-participating agents are located in the left tail of the  $\theta$  distribution, whereas participating agents are located in the right tail. The rule

<sup>1</sup> 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.



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

Fig. 1. Example of a  $\theta$ -attribute distribution with a common threshold  $c$  for all the individuals.

appears to be that some agents well-endowed in their  $\theta$ -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  $\theta$ -attribute or, conversely, that poorly endowed agents choose to participate.

An important element in our framework is to assume heterogeneity in thresholds. Because they extend well beyond costs, thresholds must reflect agent-specific characteristics that cannot be easily quantified. For example, in the labor market, two individuals applying for the same job with similar diplomas may cope with different thresholds, depending on their gender, ethnicity, or geographical location. In a similar fashion, two potential entrepreneurs with an identical project, hence identical entry costs, may cope with different likelihood of entering the market, depending on their own access to external financial resources, that is, on personal networks, family background, or any other unobservable factors affecting one's ability to access financial resources. Thus heterogeneity in thresholds is the rule, not the exception.<sup>2</sup>

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

Our framework does not change the fundamental mechanism at work, as Equation (1) holds. However, it allows the implementation of a new strategy that can estimate threshold distributions by means of maximum likelihood estimation, where the parameters of interest are those that define the whole threshold distribution.

<sup>2</sup> More formally, one may think of thresholds as the aggregation of different components such that:  $c_{i,t,s,r} = c + c_t + c_s + c_r + c_i$ , where  $c$  represents a common component pertaining to all agents,  $c_t$ ,  $c_s$  and  $c_r$  are specific components linked to time, sectoral and regional characteristics, while  $c_i$  is a very fine-grained component linked to all the possible agent-specific characteristics. For the sake of simplicity in the paper, we equate  $c_i$  with  $c_{i,t,s,r}$ .

<sup>3</sup> In Fig. 2, we arbitrarily assume normally distributed thresholds.

## 2.2. The 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  $\theta_i$ :

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

where  $F$  represents the cumulative density function of the probability distribution  $f$  and  $\Omega$  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  $\theta_i$  is:

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

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

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

The log-likelihood  $\ell$  function reads:

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

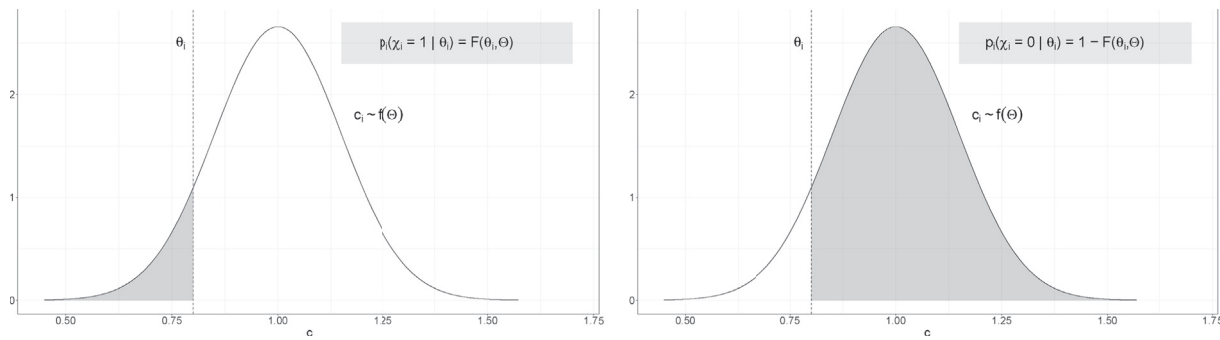
The decision for the social scientist is to choose a given parametric density function  $f(\Omega)$ , where  $\Omega$  is the vector of parameters characterizing the distribution  $f$ . Given  $f$ , the objective function is that of estimating  $\Omega$  such that  $\hat{\Omega} = \arg \max_{\Omega}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

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  $\theta$ -attribute. If monotonicity is not empirically verified, then our behavioral assumption formalized in Equation (1) does not hold, and the corresponding likelihood function  $L(\Omega)$  will prove difficult to converge.<sup>4</sup>

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  $\Omega$  by setting  $\mu = 0$  and  $\sigma = 1$ , whereas in our case, our aim is to estimate  $\Omega$ . On the other hand, while in our framework, the domain of the support is fully observed with the vector of agent-specific attribute  $\theta_i$ , in a probit model, the support of the distribution is an unobserved domain. Given a vector of explanatory variables  $Z$  and a decision variable  $y_i$ , the objective function of the probit model is then to choose a vector  $\beta$  such that  $\hat{\beta} = \arg \max_{\beta} \hat{L}(\beta; Z, y_i | \mu = 0, \sigma = 1)$ .

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

<sup>4</sup> Besides, the case in which the  $\theta$ -attribute is a limiting rather than an enhancing factor can easily be envisaged. The formal model simply becomes the complement of Equation (1). This yields following log-likelihood function:  $\ell(\Omega) = \sum_{i=1}^N (1 - \chi_i) \log [F(\theta_i; \Omega)] + \chi_i \log [1 - F(\theta_i; \Omega)]$



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

Fig. 2. Example of threshold distribution with observed  $\theta$ -attribute.

- Assumption A1. Agents are heterogeneous in their  $\theta$ -attribute  $\theta_i$ .
- Assumption A2. Thresholds are agent-specific ( $c_i$ ) and follow a density distribution  $f$  with unknown vector of parameters  $\Omega$ :  $C \sim F(\Omega)$ .
- Assumption A3. The relationship between the probability of participating in a market and the  $\theta$ -attribute is either monotonically increasing  $p(\chi = 1|\theta_{i,t_1}) > p(\chi = 1|\theta_{i,t_0}), \forall \theta_{i,t_1} > \theta_{i,t_0}$ , as in the case of an enhancing  $\theta$ -attribute, or monotonically decreasing  $p(\chi = 1|\theta_{i,t_1}) < p(\chi = 1|\theta_{i,t_0}), \forall \theta_{i,t_1} > \theta_{i,t_0}$ , as in the case of a limiting  $\theta$ -attribute. Importantly, subscripts  $t_0$  and  $t_1$  refer to different values of the  $\theta$ -attribute that pertain to the same economic agent  $i$ .
- Assumption A4. Agents decide to participate if and only if their specific attribute  $\theta_i$  exceeds (is below) threshold  $c_i$  (perfect sorting hypothesis).

Provided that Assumptions A1-A4 hold, the observation of vector  $\chi = (\chi_1, \dots, \chi_N)$  stacking all decision outcomes  $\chi_i$  and vector  $\theta = (\theta_1, \dots, \theta_N)$  of agent-specific attributes  $\theta_i$  is sufficient information to estimate the vector of parameters  $\Omega$  defining the underlying threshold distribution  $F$ . Hence the distinctive feature of our method resides in the absence of strong requirements: it needs few behavioral assumptions, it is not data demanding, and it can adapt to various parametric distributions.

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 (see Appendix A). The attractive features of Vuong’s test in our framework is two-fold: (i) it does not require preexisting knowledge of the true density and (ii) it is directional, allowing one to arbitrate between any pair of assumed density functions.

A first possible application is to examine the performance of the method on simulated data. Appendix B conducts several Monte Carlo simulations and investigates the estimation of Equation (5) when Assumptions A1-A4 hold. It then explores the robustness of the estimator under violations of Assumptions A2 and A4. In what follows, we apply our framework on empirical data, investigating the case of firm export decisions in French manufacturing.

### 3. Empirical application to international trade

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).<sup>5</sup>

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

#### 3.1. Choice of support

We use a panel database of French manufacturing firms covering the period 1990–2007 and 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 sector (excluding food and beverages), and the data provide information about their income statements and balance sheets. The complete dataset gathers the financial statements of 43,000 companies, yielding 350,000 firm-year observations. We also use information on sales, exports, value added, the wage bill, the number of employees and hours worked, capital stock, investments, and intermediate inputs as the main variables used to compute the firm-specific  $\theta$ -attribute.

We have two eligible measures for  $\theta$ , namely apparent labor productivity (ALP) or total factor productivity (TFP), as these two measures are used interchangeably in the empirical literature.<sup>6</sup> We choose to use TFP as the preferred  $\theta$ -attribute. This choice is motivated by the fact that TFP is the one that is most tightly related with the efficiency term used in the theoretical literature (e.g. Melitz, 2003). Although TFP is more prone to measurement errors, it accounts for more inputs and firm characteristics than mere ALP.<sup>7</sup> The fact that TFP is prone to mis-

<sup>5</sup> 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).

<sup>6</sup> Appendix C provides the details of the industry-wide deflators used and the computations that yield labor and total factor productivity.

<sup>7</sup> 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, Wooldridge (2009), Akerberg et al. (2015), and De Loecker and Goldberg (2014) for a thorough discussion on these issues. Appendix D provides all of the results using ALP.



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

| Industry Name                      | # Obs.  | PR   | TFFP  |
|------------------------------------|---------|------|-------|
| <i>Market participation</i>        |         |      |       |
| All Manufacturing                  | 339,088 | .740 | .042  |
| Automobile                         | 9276    | .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                    | 9067    | .914 | .030  |
| Printing and Publishing            | 29,307  | .612 | .046  |
| Textile                            | 21,659  | .800 | .075  |
| Transportation Machinery           | 5125    | .795 | .057  |
| Wood and Paper                     | 22,062  | .692 | -.014 |
| <i>Market entry</i>                |         |      |       |
| All Manufacturing                  | 74,832  | .231 | .017  |
| Automobile                         | 1635    | .252 | .006  |
| Chemicals                          | 5026    | .266 | .008  |
| Clothing and Footwear              | 7358    | .197 | .055  |
| Electric and Electronic Components | 2858    | .246 | .035  |
| Electric and Electronic Equipment  | 3775    | .218 | .018  |
| House Equipment and Furnishings    | 3488    | .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            | 9735    | .218 | .020  |
| Textile                            | 3687    | .234 | .037  |
| Transportation Machinery           | 886     | .234 | .003  |
| Wood and Paper                     | 5750    | .203 | -.019 |
| <i>Market remaining</i>            |         |      |       |
| All Manufacturing                  | 219,747 | .929 | .044  |
| Automobile                         | 6577    | .947 | .021  |
| Chemicals                          | 26,278  | .956 | .029  |
| Clothing and Footwear              | 15,974  | .913 | .129  |
| Electric and Electronic Components | 9762    | .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                    | 7334    | .975 | .062  |
| Printing and Publishing            | 15,490  | .870 | .044  |
| Textile                            | 15,233  | .947 | .071  |
| Transportation Machinery           | 3570    | .947 | .073  |
| Wood and Paper                     | 13,476  | .920 | .001  |

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

measurement is also a test for the robustness of our framework.

The top panel of Table 1 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 1 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 productivity 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*.

### 3.2. Export participation thresholds

We set the outcome decision variable  $\chi_i$  to unity if we observe positive exports by firm  $i$ , or 0 otherwise:  $\chi = (\chi_1, \dots, \chi_N)$ . In our framework, total factor productivity represents the  $\theta$ -attribute:  $\theta =$

$(\theta_1, \dots, \theta_N)$ . Given vectors  $\chi$  and  $\theta$ , we estimate Equation (5) for all the manufacturing firms. Because the support must be strictly positive for the gamma density and below unity for the beta density, we transform  $\theta$  (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  $\theta'$  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  $\Omega$  for each sector:

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

where subscript  $s$  stands for sector  $s$ . We have no prior information 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 Table 2 for the estimated first two moments ( $\mu$  and  $\sigma^2$ ). For the gamma and beta densities, we also report the estimated median to provide initial insights into the presence of skewness. We have three major observations.

First, 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.

Second, 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  $\theta$ -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.

Third, 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.

### 3.3. Threshold distributions for entry into and remaining in export markets

Our method also applies to decisions about market entry and remaining, conditional on the availability of a time dimension in the data. We now exploit it 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

**Table 2**  
Maximum likelihood estimation of participation threshold distributions, using Total Factor Productivity as the  $\theta$ -attribute.

| Industry Name                      | $\mathcal{N}_\mu$ | $\mathcal{N}_{\sigma^2}$ | $\Gamma_\mu$ | $\Gamma_{\sigma^2}$ | $\Gamma_{p50}$ | $\mathcal{B}_\mu$ | $\mathcal{B}_{\sigma^2}$ | $\mathcal{B}_{p50}$ |
|------------------------------------|-------------------|--------------------------|--------------|---------------------|----------------|-------------------|--------------------------|---------------------|
| <i>Market participation</i>        |                   |                          |              |                     |                |                   |                          |                     |
| 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                     | -                 | -                        | -            | -                   | -              | -                 | -                        | -                   |
| <i>Market entry</i>                |                   |                          |              |                     |                |                   |                          |                     |
| 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                     | -                 | -                        | -            | -                   | -              | -                 | -                        | -                   |
| <i>Market remaining</i>            |                   |                          |              |                     |                |                   |                          |                     |
| 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;  $\Gamma_\mu$ : estimated mean of the gamma distribution;  $\Gamma_{\sigma^2}$ : estimated variance of the gamma distribution;  $\Gamma_{p50}$ : 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; and  $\mathcal{B}_{p50}$ : estimated median of the beta distribution.

as those that decide to start exporting or remain exporters at time  $t$ . Such a distinction between entry and exit produces sharp differences in rates, where the entry rate drops to 23% whereas the remaining rate reaches 94%.<sup>8</sup>

We first focus on the results for threshold distributions for entry into export markets. Table 1 shows that the share of firms entering into export markets reaches 23% for all manufacturing. Thus, novel entry into export market reflects a fiercer selection process. The estimation results are reported in the middle panel of Table 2. 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 borne. Hence, one should expect entry thresholds to significantly increase, notably above the maximum value of the  $\theta$ -attribute. However, 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  $\theta$  support.

This is confirmed when looking at the estimated mean values for the normal and the gamma entry threshold distributions. When using the normal prior, 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 (-0.039% for the former, -0.019% and for the latter), implying that Equation (1), on average, does not hold for these sectors. There is less convergence with the gamma density, although

<sup>8</sup> Whether we focus on remaining or exit thresholds is essentially a semantic matter. By setting  $\chi_i = 1$  if the firm exits export markets or 0 otherwise would become an exit decision instead of a remaining one.

the algorithm seem successful in industries where the normal density fails to converge.

Table 1 shows instead that the share of firms remaining in export markets reaches almost 93% for all manufacturing. Exit from export market is, therefore, a relatively rare phenomenon, and that the associated threshold distribution must allow for the vast majority of firm to remain in such market. The estimation results are reported in the bottom panel of Table 2. Looking at *All Manufacturing*, the average threshold for remaining in export markets is substantially lower than that for market entry. Also, 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. This 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 found that fixed costs borne in 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 in export markets is far less challenging than market entry into export markets.

### 3.4. 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, but 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. Table 3 displays all pairwise Vuong's  $z$ :  $z_{N,\Gamma}$  comparing the normal and the gamma density,  $z_{N,B}$  comparing the normal and the beta density, and  $z_{\Gamma,B}$  comparing the gamma and the beta density. The last two columns of Table 3 provide the overall ranking of the 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 threshold exceeds six and the median thresholds locates at about three, whereas the  $\theta$  support ranges from zero to unity, this means that a large number of these firms are excluded from international trade.

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 also 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, or remaining). More generally, caution is needed when the diagnosis is dominance of the normal over the gamma.

Fig. 3 plots the estimated density functions for *All Manufacturing* and for three selected 2-digit 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 examples 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  $\theta$ -attribute, corroborating that for the vast majority of already exporting firms, remaining thresholds are virtually nil. The threshold distributions for *Printing and Publishing* display various densities.

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

$$\hat{\Omega}_{st} = \arg \max_{\Omega_{st}} \hat{L}(\Omega_{st}; \chi_t, \theta_t) \quad \forall s \in S \text{ and } t \in T$$

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 4 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; and (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 4 illustrates the various rankings and arbitrage in the better fit. Of all the three candidate densities, the gamma distribution dominates in all types of market decisions. However, there is a need for alternative densities. We see, for example, that the beta distribution represents a better fit in a sizable number of occurrences. Table D.4 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  $\theta$ -attribute, the normal density (left-truncated at 0) is the dominating density. Table D.4 also implies that all three densities are relevant when using the alternative support.

**Table 3**  
Vuong's  $z$  test for model selection using Total Factor Productivity as the  $\theta$ -attribute.

| Industry Name                      | $z_{N,\Gamma}$ | $z_{N,B}$ | $z_{\Gamma,B}$ | Ranking                                 | Conclusion  |
|------------------------------------|----------------|-----------|----------------|---|---|
| <i>Market participation</i>        |                |           |                |   |   |
| All Manufacturing                  | -371.9         | 1601.1    | 66.5           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(8, .5)$  |
| Automobile                         | -146.3         | 1001.6    | 246.9          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(4, .9)$  |
| Chemicals                          | -397.5         | 608.2     | 352.3          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(3, .8)$  |
| Clothing and Footwear              | -49            | 127.1     | 69.6           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(4.2, .1)$                                      |
| Electric and Electronic Components | 75.4           | 372.3     | -8.9           | $\mathcal{N} > \mathcal{B} > \Gamma$    | $C \sim \mathcal{N}(2, .4)^a$                                 |
| Electric and Electronic Equipment  | 84.7           | 132.1     | -111.8         | $\mathcal{N} > \mathcal{B} > \Gamma$    | $C \sim \mathcal{N}(3, .4)^a$                                 |
| House Equipment and Furnishings    | -8.4           | 524.4     | 15.3           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(9, .3)$  |
| Machinery and Mechanical Equipment | 14.7           | 1146.0    | 132.1          | $\mathcal{N} > \Gamma > \mathcal{B}$    | $C \sim \mathcal{N}(3, .4)$                                   |
| Metallurgy, Iron and Steel         | -1031.4        | 1551.8    | 704.4          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(1, 12.4)$                                      |
| Pharmaceuticals                    | -307.5         | 284.2     | 165.6          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(1, 1.4)$                                       |
| Printing and Publishing            | 26.6           | 30.5      | 63.7           | $\mathcal{N} > \Gamma > \mathcal{B}$    | $C \sim \mathcal{N}(4, .6)$                                   |
| Textile                            | -193.2         | 23.3      | 14.8           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(1.3, .3)$                                      |
| Transportation Machinery           | -.6            | 179.5     | 42.4           | $\mathcal{N} \sim \Gamma > \mathcal{B}$ | $C \sim \mathcal{N}(2, .4)$ or $C \sim \Gamma(1.1, .2)$       |
| Wood and Paper                     | -              | -         | -              | $\emptyset$                             | $\emptyset$   |
| <i>Market entry</i>                |                |           |                |   |   |
| All Manufacturing                  | -18.7          | -         | -              | $\Gamma > \mathcal{N}$                  | $C \sim \Gamma(5, 13.5)$                                      |
| 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          | $\mathcal{N} > \mathcal{B} > \Gamma$    | $C \sim \mathcal{N}(1.0, .7)^a$                               |
| Electric and Electronic Equipment  | 133.6          | -         | -              | $\mathcal{N} > \Gamma$                  | $C \sim \mathcal{N}(1.5, 1.3)$                                |
| House Equipment and Furnishings    | 75.6           | -         | -              | $\mathcal{N} > \Gamma$                  | $C \sim \mathcal{N}(1.0, .8)$                                 |
| Machinery and Mechanical Equipment | -149.1         | -         | -              | $\Gamma > \mathcal{N}$                  | $C \sim \Gamma(6, 8.5)$                                       |
| Metallurgy, Iron and Steel         | -              | -         | -              | $\mathcal{N}$                           | $C \sim \mathcal{N}(3.1, 3.4)$                                |
| Pharmaceuticals                    | -              | -         | -              | $\emptyset$                             | $\emptyset$   |
| Printing and Publishing            | -287.4         | -         | -              | $\Gamma > \mathcal{N}$                  | $C \sim \Gamma(6, 9.4)$                                       |
| Textile                            | -154           | 337.9     | 171            | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(9, 2.6)$                                       |
| Transportation Machinery           | -              | -         | -              | $\mathcal{N}$                           | $C \sim \mathcal{N}(6.0, 7.6)$                                |
| Wood and Paper                     | -              | -         | -              | $\emptyset$                             | $\emptyset$   |
| <i>Market remaining</i>            |                |           |                |   |   |
| All Manufacturing                  | -402.3         | 868.1     | 467.7          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(3, .4)$  |
| Automobile                         | -223.7         | 517.3     | 196.2          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(1, .8)$  |
| Chemicals                          | -106.7         | 236       | 94.6           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(2, .6)$  |
| Clothing and Footwear              | -71.7          | 68.5      | 48.6           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(1.8, .1)$                                      |
| Electric and Electronic Components | 9.7            | -.4       | -13.7          | $\mathcal{N} \sim \mathcal{B} > \Gamma$ | $C \sim \mathcal{N}(-.3, .5)$ or $C \sim \mathcal{B}(0, 1.4)$ |
| Electric and Electronic Equipment  | 77             | 193.7     | -84.1          | $\mathcal{N} > \mathcal{B} > \Gamma$    | $C \sim \mathcal{N}(-.1, .4)^a$                               |
| House Equipment and Furnishings    | -157.4         | 372.2     | 15.6           | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(3, .4)$  |
| Machinery and Mechanical Equipment | 228.6          | 832.6     | 22.4           | $\mathcal{N} > \Gamma > \mathcal{B}$    | $C \sim \mathcal{N}(-.0, .4)$                                 |
| Metallurgy, Iron and Steel         | -94.9          | 1005.0    | 562.4          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(0, 4.7)$                                       |
| Pharmaceuticals                    | -225.2         | 236.3     | 125.3          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(2, .4)$  |
| Printing and Publishing            | 92.5           | 68.2      | -4.3           | $\mathcal{N} > \mathcal{B} > \Gamma$    | $C \sim \mathcal{N}(-.3, .7)^a$                               |
| Textile                            | -175.9         | 435.2     | 161.8          | $\Gamma > \mathcal{N} > \mathcal{B}$    | $C \sim \Gamma(4, .3)$  |
| Transportation Machinery           | 19.6           | -30       | -11.5          | $\mathcal{B} > \mathcal{N} > \Gamma$    | $C \sim \mathcal{B}(3, 2.0)$                                  |
| Wood and Paper                     | -              | -993.3    | -              | $\mathcal{B} > \mathcal{N}$             | $C \sim \mathcal{B}(0, .0)$                                   |

$z_{N,\Gamma}: H_{N \sim \Gamma} : |z| < +1.96 H_{N > \Gamma} : z \geq +1.96 H_{\Gamma > N} : z \leq -1.96$ .

$z_{N,B}: H_{N \sim B} : |z| < +1.96 H_{N > B} : z \geq +1.96 H_{B > N} : z \leq -1.96$ .

$z_{\Gamma,B}: H_{\Gamma \sim B} : |z| < +1.96 H_{\Gamma > B} : z \geq +1.96 H_{B > \Gamma} : z \leq -1.96$ .

The [a] symbol indicates that caution is needed in the dominance of  $\mathcal{N}$  over  $\mathcal{B}$ , as revealed by the Monte Carlo results presented in Section B.3.3.

### 3.5. Threshold dynamics and globalization

We use our method to evaluate the impact of structural shocks on export thresholds between 1990 and 2007. This period is characterized by major shocks, intended to reduce export barriers: the establishment of the single market in 1993; the birth of the euro in 1999; and the implementation of the single currency for all transactions in 2002.<sup>9</sup> Much has been written about the pro-competitive consequences of European integration (e.g., [Bouhol, 2009](#), among a large series of contributions) or globalization ([De Loecker and Goldberg, 2014](#)) on markups, but evidence of its effect on export thresholds is yet to emerge. Two contrasting effects are expected. Because European integration is

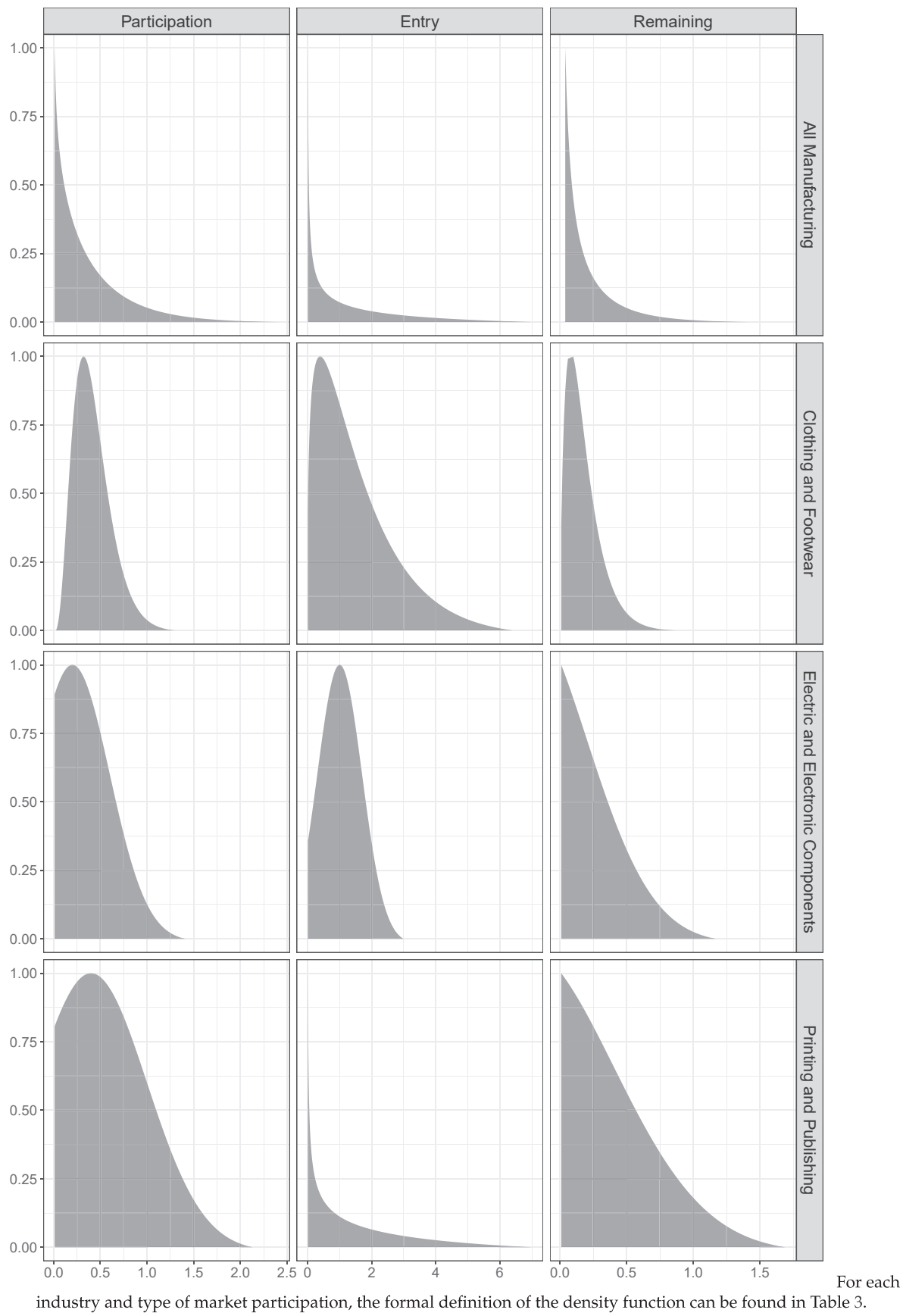
about decreasing trade barriers among member countries, one should expect a significant decrease in thresholds overall. However, European integration also fosters competition for firms across member states, raising the productivity threshold for potential exporters. How does this translate into changes in the threshold distribution? In fact, although we do expect a reduction in the first moment of the threshold distribution, we remain agnostic about the effect of the aforementioned shocks on higher moments of the threshold distribution. In particular, we explore whether these shocks have been of the symmetric or asymmetric type by estimating the threshold distribution for market participation as follows:

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

We do not report the results of our estimations, but consistent with prior findings, the gamma density is the best fit according the Vuong's test. [Fig. 4](#) displays the dynamics of the first four moments of export

<sup>9</sup> More remotely, this period also witnessed the entry of China, India, and more generally the BRICS countries as major players on the international trade scene.





**Fig. 3.** Threshold distributions for all manufacturing and three selected sectors.

**Table 4**  
Occurrence of diagnosis, according to the type of market participation.

| Ranking                                    | Participation | Entry | Remaining |
|--|---------------|-------|-----------|
| <i>No density fit</i>                      |               |       |           |
| $\emptyset$                                | 22            | 67    | 19        |
| <i>Unique density fit</i>                  |               |       |           |
| $\mathcal{N}$                              | 1             | 20    | 0         |
| $\Gamma$                                   | 0             | 20    | 1         |
| $\mathcal{B}$                              | 1             | 10    | 0         |
| <i>Two density fits</i>                    |               |       |           |
| $\mathcal{N} > \Gamma$                     | 0             | 6     | 0         |
| $\mathcal{N} > \mathcal{B}^a$              | 3             | 2     | 2         |
| $\Gamma > \mathcal{B}$                     | 0             | 23    | 0         |
| $\Gamma > \mathcal{N}$                     | 0             | 22    | 0         |
| $\mathcal{B} > \mathcal{N}$                | 3             | 0     | 5         |
| $\mathcal{B} > \Gamma$                     | 0             | 12    | 0         |
| $\Gamma \sim \mathcal{B}$                  | 0             | 2     | 0         |
| <i>Tree density fits</i>                   |               |       |           |
| $\mathcal{N} > \Gamma > \mathcal{B}$       | 25            | 3     | 16        |
| $\mathcal{N} > \mathcal{B} > \Gamma^a$     | 44            | 7     | 32        |
| $\mathcal{N} \sim \Gamma > \mathcal{B}$    | 5             | 0     | 1         |
| $\mathcal{N} \sim \Gamma \sim \mathcal{B}$ | 2             | 2     | 1         |
| $\mathcal{N} \sim \mathcal{B} > \Gamma$    | 0             | 1     | 0         |
| $\Gamma > \mathcal{N} > \mathcal{B}$       | 98            | 13    | 78        |
| $\Gamma > \mathcal{B} > \mathcal{N}$       | 11            | 7     | 14        |
| $\Gamma \sim \mathcal{B} > \mathcal{N}$    | 2             | 0     | 0         |
| $\mathcal{B} > \mathcal{N} > \Gamma$       | 14            | 2     | 48        |
| $\mathcal{B} > \Gamma > \mathcal{N}$       | 3             | 2     | 3         |
| $\mathcal{B} > \mathcal{N} \sim \Gamma$    | 0             | 0     | 1         |
| <i>Overall dominance</i>                   |               |       |           |
| $\mathcal{N}$                              | 80            | 47    | 51        |
| $\Gamma$                                   | 111           | 87    | 94        |
| $\mathcal{B}$                              | 21            | 26    | 57        |
| <i>Total</i>                               |               |       |           |
|  | 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 B.3.3.

thresholds. As expected, we observe an overall downward trend in thresholds by 15%. A closer look at the evolution of the mean suggests 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 also concomitant to significant decreases in the mean value of export thresholds. This may be due to the 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.

Looking at higher moments of 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 symmetric shock, homogeneous to all manufacturing firms. Higher moments of the distribution exhibit an important upward trend from 1999 onwards, that is, after the birth of the euro. The synchronized increase in the variance, skewness, and kurtosis implies that, oppositely, the euro constituted an asymmetric shock

to French firms. We argue that many firms have benefited from the introduction of the euro through the direct effect of the reduction in information costs. Thus, the mass of the threshold’s distribution moved toward the left as barriers have diminished. However, for a number of firms, the indirect effect due to an increase in competition with foreign firms belonging to the European Monetary Union, has outplayed the benefits, leading to an increase in thresholds located in the last decile of the distribution. This interpretation is consistent with the observed dynamics of the extreme deciles displayed in Fig. 4.

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 symmetric and homogeneous shock to all firms, the latter constituted an asymmetric shock, definitely excluding a minor share of large manufacturing firms from international trade.

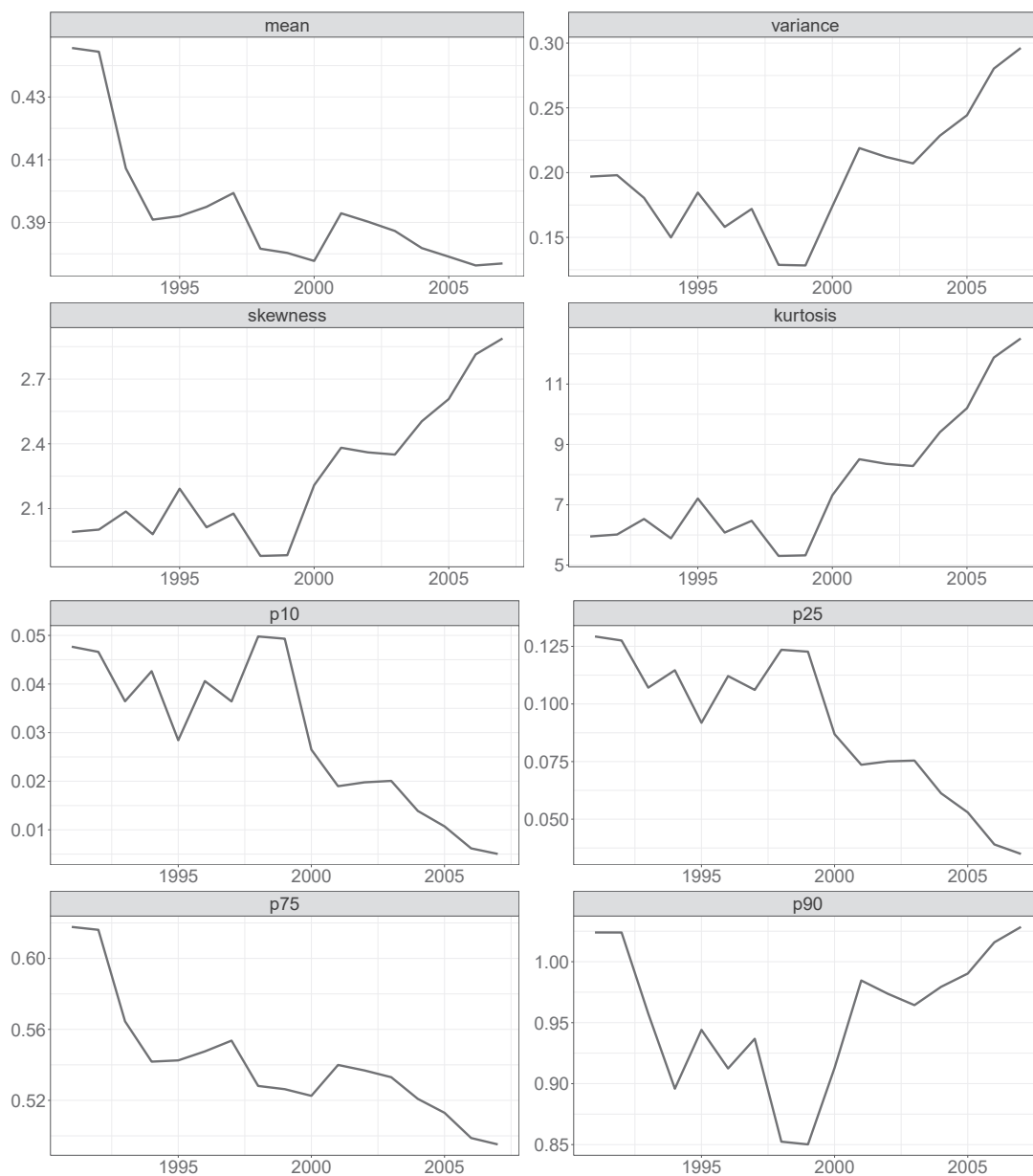
#### 4. Conclusion

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 also 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 a hypothesis testing procedure, based on Vuong (1989), to discriminate between a set of competing priors.

We found that heterogeneity is relevant across several domains: (i) within-sector estimates, suggesting that participation thresholds are characterized by right-skewness and leptokurtosis; (ii) between-sector estimates, conditional upon participating, entering, or remaining in a market, suggesting that gamma, normal, or beta densities are all empirically relevant; and (iii) across-year estimates, suggesting that the establishment of the common market and the introduction of the euro have decreased average thresholds. However, due to increased foreign competition, the common currency has led to the exclusion of a minor share of firms from international trade. Overall, our results indicate that accounting for heterogeneity in thresholds allows one to gain new insights on how policies affect barriers to market participation.

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  $\theta$ -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 two-fold. 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, policymakers 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



The moments and percentiles of the threshold distribution are based on the estimated gamma parameters for all manufacturing using total factor productivity as the  $\theta$ -attribute, net of industry  $\times$  year effects.  $p_{10}$ ,  $p_{25}$ ,  $p_{75}$  and  $p_{90}$  stand for the 1<sup>st</sup> decile, the 1<sup>st</sup> quartile, the 3<sup>rd</sup> quartile and the 9<sup>th</sup> decile of the threshold distribution, respectively.

**Fig. 4.** The dynamics of threshold distributions in French manufacturing: 1990–2007.

to maximize participation, conditional on a given budget. Second, by broadening the meaning of the  $\theta$ -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 policymakers. 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.

### Appendix A. Model selection

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:

$$H_0 : \mathbb{E}^0 \left[ \log \frac{f_p(\chi_i|\theta_i, \mathbf{\Omega}_{f_p}^*)}{f_q(\chi_i|\theta_i, \mathbf{\Omega}_{f_q}^*)} \right] = 0, \tag{A.1}$$

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 A.1 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:

$$H_{F_p} : \mathbb{E}^0 \left[ \log \frac{f_p(\chi_i|\theta_i, \mathbf{\Omega}_{f_p}^*)}{f_q(\chi_i|\theta_i, \mathbf{\Omega}_{f_q}^*)} \right] > 0, \tag{A.2}$$

meaning that  $F_p$  is a better fit than  $F_q$  ( $F_p > F_q$ ), and:

$$H_{F_q} : \mathbb{E}^0 \left[ \log \frac{f_p(\chi_i|\theta_i, \mathbf{\Omega}_{f_p}^*)}{f_q(\chi_i|\theta_i, \mathbf{\Omega}_{f_q}^*)} \right] < 0, \tag{A.3}$$

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

$$\log LR(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) = \ell_p(\hat{\mathbf{\Omega}}_{f_p}) - \ell_q(\hat{\mathbf{\Omega}}_{f_q}) = \sum_{i=1}^N (\ell_{p,i}(\chi_i|\theta_i, \hat{\mathbf{\Omega}}_{f_p}) - \ell_{q,i}(\chi_i|\theta_i, \hat{\mathbf{\Omega}}_{f_q})) = \sum_{i=1}^N d\ell_i, \tag{A.4}$$

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  $\ell_p$  (resp.  $\ell_q$ ) 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 \(1989\)](#) suggested to account for differences in the number of parameters in the two models as the in the Akaike Information Criterion such that:

$$\log \widetilde{LR}(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) = \log LR(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) - (k_p - k_q) \frac{\log N}{2} \tag{A.5}$$

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:

$$\text{Vuong}^s z = (\sigma_{dl} \sqrt{N})^{-1} \log \widetilde{LR}(\hat{\mathbf{\Omega}}_{f_p}, \hat{\mathbf{\Omega}}_{f_q}) \tag{A.6}$$

where  $\sigma_{dl}$  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  $F_r$ .<sup>10</sup> 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.

### Appendix B. Monte Carlo simulations

#### B.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  $\Gamma$ , 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  $\theta$ -attributes to be strictly positive:  $\theta \in \mathbb{R}^+$ . 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:  $\theta \in (0, 1)$ . This is extremely binding, because it implies that the cumulative distribution function be unity when  $\theta$  exceeds one.

Following the description in [Appendix E](#), 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  $\theta$ -attribute  $\theta \sim \mathcal{N}(.5, .15)$ .<sup>11</sup> In our simulation, threshold  $C$  is random variable drawn from: (i) a normal distribution

<sup>10</sup> 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  $\sigma_{dl}$ .

<sup>11</sup> We verified that our results are qualitatively robust to alternative distributions of  $\theta_i$ . 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.

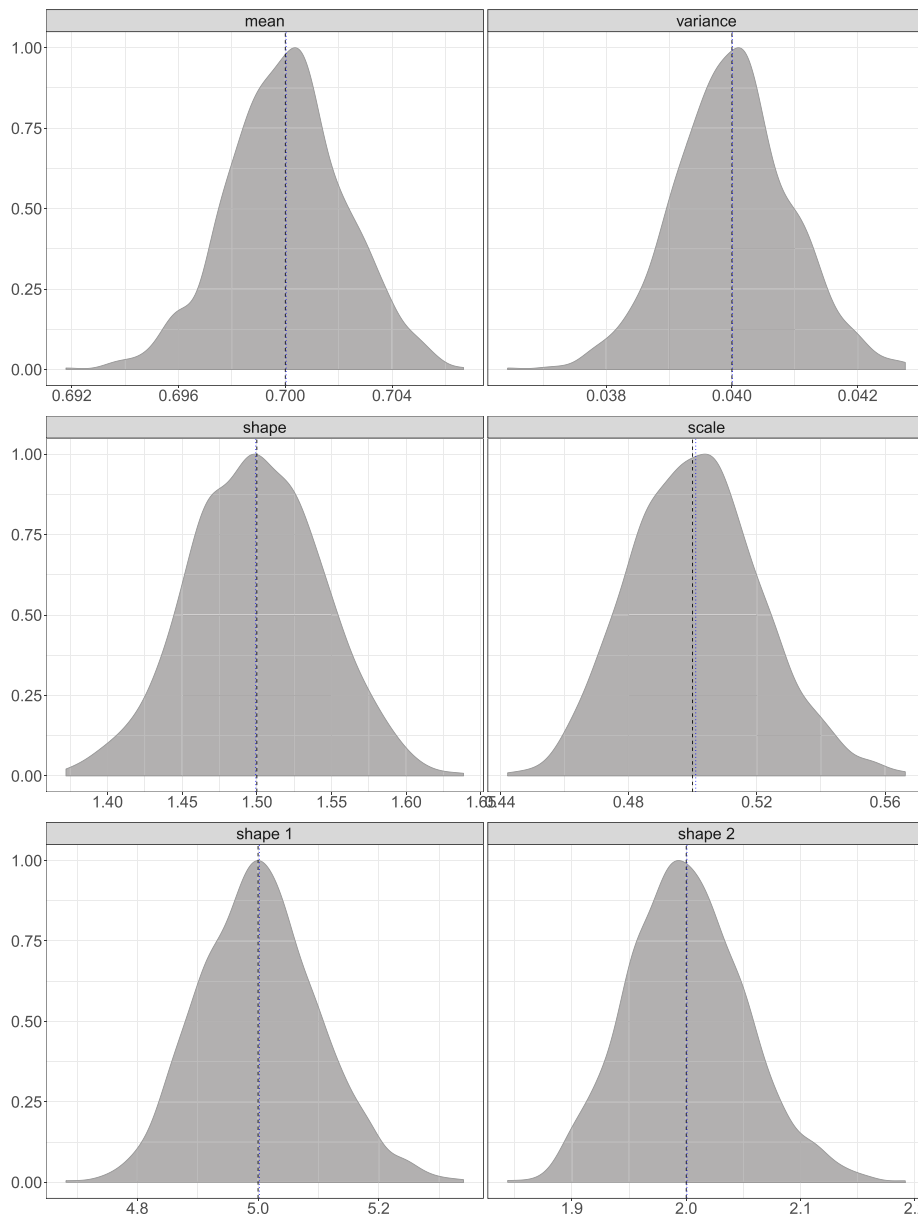


with  $\mu = 0.7$  and  $\sigma = 0.2$ , i.e.,  $C \sim \mathcal{N}(.7, .2)$ ; (ii) a right-skewed gamma distribution with shape and scale parameters  $\alpha_\Gamma = .15$  and  $\beta_\Gamma = 0.5$ , i.e.,  $C \sim \Gamma(1.5, 0.5)$ ; and (iii) a left-skewed beta distribution with shape-one and shape-two parameters  $\alpha_B = 5$  and  $\beta_B = 2$ , i.e.,  $C \sim \mathcal{B}(5, 2)$ . For all three threshold distributions, given the vector of  $\theta$ -attribute, we computed the vector of decision outcomes  $\chi$  according to Equation (1).<sup>12</sup>

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 robustness of the estimator under violations of certain assumptions.

**B.2. Baseline results**

We begin with a perfect scenario, where all Assumptions A1-A4 hold. Using only limited information  $\theta$  and  $\chi$ , we then apply our MLE algorithm to estimate the vector of parameters  $\Omega = \{\mu, \sigma\}$  for the Gaussian case,  $\Omega = \{\alpha_\Gamma, \beta_\Gamma\}$  for the gamma case, and  $\Omega = \{\alpha_B, \beta_B\}$  for the beta case.<sup>13</sup> The distributions of the estimated parameters across  $M = 1000$  Monte Carlo simulations are displayed in the different panels of Fig. B.1 for the normal, the gamma, and the beta scenarios, respectively.



$M = 1,000$  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.

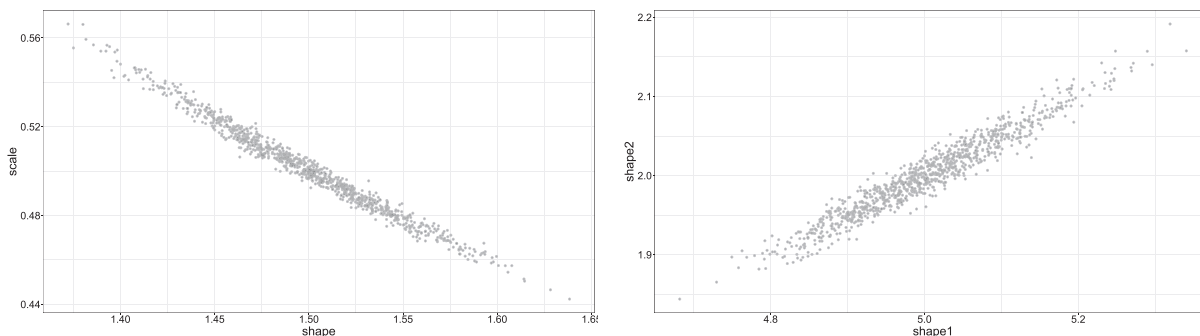
For all simulations, we set the  $\theta$ -attribute such that it follows a normal distribution:  $\theta \sim \mathcal{N}(.5, .0225)$ . Set of parameters for threshold distributions: (i)  $C \sim \mathcal{N}(.7, .004)$ ; (ii)  $C \sim \Gamma(1.5, .5)$ ; and (iii)  $C \sim \mathcal{B}(5, 2)$ .

**Fig. B.1** Distribution of Monte Carlo estimates of the parameters  $\Omega$  for three different threshold distributions.

<sup>12</sup> Details on the functional forms of the three distributions are available in Appendix F.

<sup>13</sup> Note that estimations of  $\alpha$  and  $\beta$  for the gamma and the beta distributions allow one to analytically retrieve the first four moments.

Fig. B.1 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  $\alpha$  and  $\beta$ . As shown in Fig. B.2, there is a tight negative relationship linking the estimates of the two parameters in the gamma case (left panel) and in 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.



$M = 1,000$  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.

Fig. B.2 Scatter plot of the parameters estimates for the gamma and beta scenarios.

This compensation mechanism is a positive feature of the two asymmetric distributions. The depicted correlation between  $\alpha$  and  $\beta$  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 B.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.

### B.3. Violations of assumptions

#### B.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, gamma, or beta distribution  $F = \{\mathcal{N}(\mu, \sigma), \Gamma(\alpha_\Gamma, \beta_\Gamma), \mathcal{B}(\alpha_\beta, \beta_\beta)\}$ . This represents our prior about the threshold distribution. In turn, the true distribution of thresholds  $C$  may alternatively follow a normal, gamma or 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 instead choose a left-skewed distribution. We set the  $\theta$ -attribute such that it follows a normal distribution:  $\theta \sim \mathcal{N}(.5, .0225)$ . The set of parameters for threshold distributions is the following: (i)  $C \sim \mathcal{N}(.7, .04)$ ; (ii)  $C \sim \Gamma(1.5, 0.5)$ ; and (iii)  $C \sim \mathcal{B}(5, 2)$ .<sup>14</sup> Combining all the possible choices for the prior  $F$  and the true distribution of  $C$ , we obtain nine different cases. In three of them, Assumption A2 is satisfied and gives 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 ( $\mu$  and  $\sigma$ ) using a gamma or an beta prior. Instead, we compute the root mean squared errors (RMSEs) of the estimated first four moments (the mean  $\mu$ ; the variance  $\sigma^2$ ; the skewness  $sk$ ; and the excess kurtosis  $k$ ) and compare the estimated values with their true counterparts. We do this for all of the scenarios combining our prior, assumed, density for the MLE exercise and the true densities. The results displayed in Table B.1 thoroughly corroborate the results of the previous section. Absent a misspecification, the true moments are consistently recovered. However, when the false prior is used, a significant error is always present.

<sup>14</sup> The guidelines for the algorithm followed for this Monte Carlo exercise are presented in Appendix E. The values for the variance  $\sigma^2$  in the normal distribution stem from setting  $\sigma$  to 0.15 for  $\theta$  and 0.2 for the threshold normal distributions.

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

| Assumed $c$                         | True $c$      | A2 | $\mu$ | $\sigma^2$ | sk    | k     |
|-------------------------------------|---------------|----|-------|------------|-------|-------|
| <i>True Values</i>                  |               |    |       |            |       |       |
|                                     | $\mathcal{N}$ |    | 0.700 | 0.040      | 0.000 | 0.000 |
|                                     | $\Gamma$      |    | 0.750 | 0.375      | 1.633 | 4.000 |
|                                     | $\mathcal{B}$ |    | 0.714 | 0.026      | 0.596 | 0.120 |
| <i>RMSE as Share of True Values</i> |               |    |       |            |       |       |
| $\mathcal{N}$                       | $\mathcal{N}$ | ✓  | 0.003 | 0.023      | –     | –     |
| $\Gamma$                            | $\mathcal{N}$ | ×  | 0.049 | 0.533      | –     | –     |
| $\mathcal{B}$                       | $\mathcal{N}$ | ×  | 0.022 | 0.242      | –     | –     |
| $\mathcal{N}$                       | $\Gamma$      | ×  | 0.203 | 0.452      | 1.000 | 1.000 |
| $\Gamma$                            | $\Gamma$      | ✓  | 0.012 | 0.052      | 0.015 | 0.029 |
| $\mathcal{B}$                       | $\Gamma$      | ×  | 0.243 | 0.731      | 1.152 | 1.321 |
| $\mathcal{N}$                       | $\mathcal{B}$ | ×  | 0.104 | 1.508      | 1.000 | 1.000 |
| $\Gamma$                            | $\mathcal{B}$ | ×  | 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 of the simulations, we set the  $\theta$ -attribute such that it follows a normal distribution:  $\theta \sim \mathcal{N}(.5, .0225)$ . The set of parameters for the threshold distributions: (i)  $C \sim \mathcal{N}(.7, .004)$ ; (ii)  $C \sim \Gamma(1.5, 0.5)$ ; and (iii)  $C \sim \mathcal{B}(5, 2)$ .

The first three rows of the bottom panel display the performance of the three priors ( $F = \{\mathcal{N}(\mu, \sigma), \Gamma(\alpha_\Gamma, \beta_\Gamma), \mathcal{B}(\alpha_B, \beta_B)\}$ ) when the normal density is true. Consistently with Section B.2, it shows that the normal prior performs well at recovering the first two moments of the true normal, with errors amounting to only 0.3% 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 all of the 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 present the results under a true beta distribution of thresholds. Our previous observation also prevails 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.

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’s likelihood ratio test for model selection is fully advocated in order to recover a complete ranking of the candidate densities.

Table B.2 presents the relative frequencies of each ranking of distributions, according to the Vuong’s test using the simulations used above. Similarly to the sensitivity (detecting true positives) and specificity (detecting true negatives) of the test, this exercise investigates whether Vuong’s procedure correctly discriminates between the three alternative densities. Table B.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 B.1. When the true distribution of thresholds is normal, the order of preference is  $\mathcal{N} > \mathcal{B} > \Gamma$  in the majority of cases; when the thresholds are gamma distributed, then  $\Gamma > \mathcal{N} > \mathcal{B}$  is the typical ranking; and when the true density of thresholds is the beta distribution, then  $\mathcal{B} > \mathcal{N} > \Gamma$  is the emerging ranking. In all instances, the sensitivity and specificity of Vuong’s test hold under violations of Assumption A2.

**Table B.2**  
Vuong’s ranking of distributions, by type of true threshold distribution.

| Vuong’s diagnosis                       | True $c$ : $\mathcal{N}$ | True $c$ : $\Gamma$ | True $c$ : $\mathcal{B}$ |
|---|--------------------------|---------------------|--------------------------|
| $\mathcal{N} > \mathcal{B} > \Gamma$    | 0.961                    |                     |                          |
| $\mathcal{N} > \Gamma > \mathcal{B}$    | 0.025                    |                     |                          |
| $\mathcal{N} > \Gamma > \mathcal{B}$    | 0.011                    |                     |                          |
| $\mathcal{N} \sim \mathcal{B} > \Gamma$ | 0.001                    |                     | 0.001                    |
| $\Gamma > \mathcal{N} \sim \mathcal{B}$ |                          | 0.026               |                          |
| $\Gamma > \mathcal{N} > \mathcal{B}$    |                          | 0.707               |                          |
| $\Gamma > \mathcal{B} > \mathcal{N}$    |                          | 0.267               |                          |
| $\mathcal{B} > \mathcal{N} > \Gamma$    | 0.002                    |                     | 0.867                    |
| $\mathcal{B} > \Gamma > \mathcal{N}$    |                          |                     | 0.131                    |
| $\mathcal{B} > \Gamma \sim \mathcal{N}$ |                          |                     | 0.001                    |

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

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.

**B.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  $\theta$ -attribute.<sup>15</sup> 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  $\chi_i$  due to agent  $i$  having an imperfect evaluation of her threshold  $c_i$ . With respect to Fig. 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  $\theta_i$ . To generate a *sorting error* in our Monte Carlo exercise, we modify the problem in Equation (1) as follows:

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

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  $\sigma_\varepsilon$ .

The second source of imperfect sorting is due to *measurement errors* of the  $\theta$ -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 Fig. 2, a *measurement error* yields a noisy location of the  $\theta$ -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  $\theta$ -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.<sup>16</sup> For all of 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 discriminate for different sizes of the errors, characterized by  $\sigma_\varepsilon$  which lies in 5%, 10% and 15% of the original variance of the  $\theta$ -attribute ( $\sigma_\theta$ ). This exercise is performed for all the three (i.e., normal, gamma, and beta) threshold densities.<sup>17</sup>

Each scenario gives rise to 1000 Monte Carlo simulations, 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 B.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 an 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.

<sup>15</sup> The contributions of Jovanovic (1982) and Hopenhayn (1992) on the dynamics of industries are examples of models where agents – firms – learn about their own  $\theta$ -attribute and may make erroneous decisions.

<sup>16</sup> Details of the algorithmic guidelines are presented in Appendix E.

<sup>17</sup> The unreported results show that for the gamma and beta distributions, a pattern similar to Fig. B.2 on the correlation between estimates of  $\alpha$  and  $\beta$  is found in each of the three scenarios. A downward bias in the estimates of  $\alpha_\Gamma$  is compensated by an upward bias in the estimates of  $\beta_\Gamma$  for the gamma case, while a downward bias in the estimates of  $\alpha_B$  is compensated by a downward bias in the estimates of  $\beta_B$  for the beta case.



**Table B.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 | $\sigma_\epsilon = 5\%$ | $\sigma_\epsilon = 10\%$ | $\sigma_\epsilon = 15\%$ |
|---------------------------------|----|----|-------------------------|--------------------------|--------------------------|
| <i>N</i> distributed thresholds |    |    |                         |                          |                          |
| $\mu$                           | ✓  |    | 0.003                   | 0.003                    | 0.003                    |
| $\sigma^2$                      | ✓  |    | 0.023                   | 0.024                    | 0.026                    |
| sk                              | ✓  |    | –                       | –                        | –                        |
| ku                              | ✓  |    | –                       | –                        | –                        |
| $\mu$                           |    | ✓  | 0.003                   | 0.003                    | 0.003                    |
| $\sigma^2$                      |    | ✓  | 0.023                   | 0.023                    | 0.023                    |
| sk                              |    | ✓  | –                       | –                        | –                        |
| ku                              |    | ✓  | –                       | –                        | –                        |
| $\mu$                           | ✓  | ✓  | 0.003                   | 0.003                    | 0.003                    |
| $\sigma^2$                      | ✓  | ✓  | 0.023                   | 0.024                    | 0.026                    |
| sk                              | ✓  | ✓  | –                       | –                        | –                        |
| ku                              | ✓  | ✓  | –                       | –                        | –                        |
| $\Gamma$ distributed thresholds |    |    |                         |                          |                          |
| $\mu$                           | ✓  |    | 0.012                   | 0.012                    | 0.012                    |
| $\sigma^2$                      | ✓  |    | 0.052                   | 0.052                    | 0.052                    |
| sk                              | ✓  |    | 0.015                   | 0.015                    | 0.015                    |
| ku                              | ✓  |    | 0.030                   | 0.030                    | 0.030                    |
| $\mu$                           |    | ✓  | 0.012                   | 0.014                    | 0.020                    |
| $\sigma^2$                      |    | ✓  | 0.053                   | 0.065                    | 0.098                    |
| sk                              |    | ✓  | 0.015                   | 0.018                    | 0.027                    |
| ku                              |    | ✓  | 0.030                   | 0.037                    | 0.056                    |
| $\mu$                           | ✓  | ✓  | 0.012                   | 0.014                    | 0.021                    |
| $\sigma^2$                      | ✓  | ✓  | 0.054                   | 0.066                    | 0.101                    |
| sk                              | ✓  | ✓  | 0.015                   | 0.018                    | 0.028                    |
| ku                              | ✓  | ✓  | 0.031                   | 0.037                    | 0.056                    |
| <i>B</i> distributed thresholds |    |    |                         |                          |                          |
| $\mu$                           | ✓  |    | 0.002                   | 0.002                    | 0.002                    |
| $\sigma^2$                      | ✓  |    | 0.016                   | 0.017                    | 0.020                    |
| sk                              | ✓  |    | 0.015                   | 0.015                    | 0.015                    |
| ku                              | ✓  |    | 0.064                   | 0.063                    | 0.068                    |
| $\mu$                           |    | ✓  | 0.002                   | 0.003                    | 0.006                    |
| $\sigma^2$                      |    | ✓  | 0.017                   | 0.025                    | 0.047                    |
| sk                              |    | ✓  | 0.016                   | 0.025                    | 0.045                    |
| ku                              |    | ✓  | 0.066                   | 0.085                    | 0.137                    |
| $\mu$                           | ✓  | ✓  | 0.002                   | 0.003                    | 0.005                    |
| $\sigma^2$                      | ✓  | ✓  | 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  $\theta$ -attribute such that it follows a normal distribution:  $\theta \sim \mathcal{N}(5, .0225)$ . The set of parameters for threshold distributions: (i)  $C \sim \mathcal{N}(7, .004)$ ; (ii)  $C \sim \Gamma(1.5, 0.5)$ ; and (iii)  $C \sim \mathcal{B}(5, 2)$ . The true moment values are displayed in Table B.1.

In general, we observe that the normal is more resilient to shocks, while the gamma follows and 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  $\theta$ -attribute is badly measured and a significant fraction of the firms do not take decisions according to Equation (1).

**B.3.3. Joint violation of A2 and A4**

We observe in Section B.3.1 that, in the absence of measurement and sorting errors, the true density function is correctly recovered and its true parameters are consistently estimated. Section B.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) and (ii) some firms do not behave according to Equation (1) and/or the  $\theta$ -attribute is measured only imperfectly (hence, Assumption A4 is also violated).

We perform 1000 Monte Carlo simulations combining the settings performed in the two previous subsections. In particular, we use the three threshold distributions, namely, the normal, gamma, and 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 B.4. They suggest that the Vuong’s test is successful at identifying the proper density, even in the presence of a joint violation of Assumption A2 and Assumption A4. If the true underlying density is normal, the test always excludes the gamma and beta alternatives. Similarly, if the thresholds are distributed gamma, the test predicts the correct density in virtually all of the 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 a similar distribution that is extremely close. 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 B.4**  
Vuong’s ranking of distributions, by type of true threshold distribution.

| Vuong diagnosis                         | True c: $\mathcal{N}$ | True c: $\Gamma$ | True c: $\mathcal{B}$ |
|---|-----------------------|------------------|-----------------------|
| $\mathcal{N} > \mathcal{B} > \Gamma$    | 0.990                 |                  | 0.326                 |
| $\mathcal{N} > \Gamma \sim \mathcal{B}$ | 0.008                 |                  |                       |
| $\mathcal{N} > \Gamma > \mathcal{B}$    | 0.002                 |                  |                       |
| $\mathcal{N} \sim \mathcal{B} > \Gamma$ |                       |                  | 0.015                 |
| $\Gamma > \mathcal{N} \sim \mathcal{B}$ |                       | 0.016            |                       |
| $\Gamma > \mathcal{N} > \mathcal{B}$    |                       | 0.110            |                       |
| $\Gamma > \mathcal{B} > \mathcal{N}$    |                       | 0.873            |                       |
| $\mathcal{B} > \mathcal{N} > \Gamma$    |                       |                  | 0.550                 |
| $\mathcal{B} > \Gamma > \mathcal{N}$    |                       | 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:  $\mathcal{N} > \mathcal{B} > \Gamma$  must be read as “The normal distribution is preferred over the beta distribution which is preferred over the gamma distribution.”

After collecting all of the results 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.

**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 the workers taken from other firms) by the average hours worked. The annual series for hours worked are available at the 2-digit industry level and provided by the GGDC Groningen Growth Development Center). This choice has been made because there are no data on hours worked in the EAE (*Enquête annuelle d’entreprises*) survey. Also note also that a large decline in the hours worked occurred between 1999 and 2000 because of the specific “French 35 h policy” (on average, the 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  $\delta_t$  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 the 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,  $CT_{kt} = w_{kt}L_{kt} + c_{It}K_{kt} + m_{It}M_{kt}$  represent 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 the 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 the 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)’s 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<sup>18</sup>

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

where  $\tau_t$  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 depreciation 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 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:

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

where  $Y_{it}$  denotes real gross output produced by firm  $i$  at time  $t$  using the set of  $n$  inputs  $X_{nit}$ , where input  $X$  is alternatively capital stocks ( $K$ ), labor in terms of the hours worked ( $L$ ), and intermediate inputs ( $M$ ).  $S_{nit}$  is the cost share of input  $X_{nit}$  in the total cost. Subscripts  $\tau$  and  $n$  are indexes for time and inputs, respectively. Symbols with an 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 Equation (C.5) implies that reference 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 the 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 of the results on participation, entry, and remaining threshold distributions using labor productivity as the  $\theta$ -attribute. We stack all of the tables without commenting further.

<sup>18</sup> In this equation, we abstract from tax credit allowance.

Table D.1

Participation rates and labor productivity premium, by industry

| Industry Name                      | # Obs.  | PR   | ALPP  |
|------------------------------------|---------|------|-------|
| <i>Market participation</i>        |         |      |       |
| All Manufacturing                  | 337,275 | .738 | .171  |
| Automobile                         | 9213    | .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                    | 8677    | .915 | .170  |
| Printing and Publishing            | 29,001  | .608 | .145  |
| Textile                            | 21,418  | .796 | .256  |
| Transportation Machinery           | 5076    | .794 | .221  |
| Wood and Paper                     | 21,946  | .691 | .168  |
| <i>Market entry</i>                |         |      |       |
| All Manufacturing                  | 74,832  | .231 | .017  |
| Automobile                         | 1635    | .252 | .006  |
| Chemicals                          | 5026    | .266 | .008  |
| Clothing and Footwear              | 7358    | .197 | .055  |
| Electric and Electronic Components | 2858    | .246 | .035  |
| Electric and Electronic Equipment  | 3775    | .218 | .018  |
| House Equipment and Furnishings    | 3488    | .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            | 9735    | .218 | .020  |
| Textile                            | 3687    | .234 | .037  |
| Transportation Machinery           | 886     | .234 | .003  |
| Wood and Paper                     | 5750    | .203 | -.019 |
| <i>Market remaining</i>            |         |      |       |
| All Manufacturing                  | 218,102 | .929 | .157  |
| Automobile                         | 6517    | .946 | .122  |
| Chemicals                          | 25,926  | .956 | .193  |
| Clothing and Footwear              | 15,708  | .911 | .373  |
| Electric and Electronic Components | 9759    | .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                    | 7015    | .974 | .241  |
| Printing and Publishing            | 15,241  | .869 | .137  |
| Textile                            | 14,981  | .946 | .214  |
| Transportation Machinery           | 3535    | .946 | .256  |
| Wood and Paper                     | 13,391  | .920 | .161  |

PR: participation rate. ALPP: Labor Productivity Premium.



**Table D.2**  
Maximum likelihood estimation of participation threshold distributions.

| Industry Name                      | $\mathcal{N}_\mu$ | $\mathcal{N}_{\sigma^2}$ | $\Gamma_\mu$ | $\Gamma_{\sigma^2}$ | $\Gamma_{p50}$ | $\mathcal{B}_\mu$ | $\mathcal{B}_{\sigma^2}$ | $\mathcal{B}_{p50}$ |
|------------------------------------|-------------------|--------------------------|--------------|---------------------|----------------|-------------------|--------------------------|---------------------|
| <i>Market participation</i>        |                   |                          |              |                     |                |                   |                          |                     |
| 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                |
| <i>Market entry</i>                |                   |                          |              |                     |                |                   |                          |                     |
| 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         | 1522.1              | 5.438          | .754              | .150                     | .999                |
| Wood and Paper                     | –                 | –                        | 2.960        | 11.9                | 1.777          | .759              | .109                     | .956                |
| <i>Market remaining</i>            |                   |                          |              |                     |                |                   |                          |                     |
| 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;  $\Gamma_\mu$ : estimated mean of the gamma distribution;  $\Gamma_{\sigma^2}$ : estimated variance of the gamma distribution;  $\Gamma_{p50}$ : 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; and  $\mathcal{B}_{p50}$ : estimated median of the beta distribution.

**Table D.3**  
 Vuong's  $z$  test for model selection using Labor Productivity as the  $\theta$ -attribute.

| Industry Name                      | $z_{N,\Gamma}$ | $z_{N,B}$ | $z_{\Gamma,B}$ | Ranking                    | Conclusion                       |
|------------------------------------|----------------|-----------|----------------|----------------------------|----------------------------------|
| <i>Market participation</i>        |                |           |                |                            |                                  |
| All Manufacturing                  | 712.4          | 1294.7    | -523.3         | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.3, .3)^a$   |
| Automobile                         | 115            | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(.2, .3)$     |
| Chemicals                          | 51.2           | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.0, .5)$    |
| Clothing and Footwear              | -1.4           | 113.6     | 44.3           | $\Gamma > \mathcal{N} > B$ | $C \sim \Gamma(6.6, 1)$          |
| Electric and Electronic Components | 13.8           | 227.2     | -82.3          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.3, .2)^a$   |
| Electric and Electronic Equipment  | 42.7           | 67.8      | -78.9          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.3, .3)^a$   |
| House Equipment and Furnishings    | 252.2          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(.1, .3)$     |
| Machinery and Mechanical Equipment | 591.6          | 1,39.1    | -643.8         | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.3, .3)^a$   |
| Metallurgy, Iron and Steel         | 275.8          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(.1, .6)$     |
| Pharmaceuticals                    | -196.4         | -         | -              | $\Gamma > \mathcal{N}$     | $C \sim \Gamma(2, .9)$           |
| Printing and Publishing            | 37.8           | 51.6      | -432.9         | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.4, .5)^a$   |
| Textile                            | 219.9          | 506.6     | -76.9          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.2, .3)^a$   |
| Transportation Machinery           | 91.6           | 428.3     | -49.2          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.2, .3)^a$   |
| Wood and Paper                     | 409.6          | 64.8      | -586.1         | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.3, .3)^a$   |
| <i>Market entry</i>                |                |           |                |                            |                                  |
| All Manufacturing                  | -              | -         | -197.1         | $B > \Gamma$               | $C \sim B(.5, .2)$               |
| Automobile                         | 14.9           | 366.5     | -179.4         | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(1.1, 1.0)^a$ |
| Chemicals                          | -              | -         | -              | $\emptyset$                | $\emptyset$                      |
| Clothing and Footwear              | -              | -         | 72.6           | $\Gamma > B$               | $C \sim \Gamma(2.7, .3)$         |
| Electric and Electronic Components | 76.6           | 198.3     | -41            | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.8, .5)^a$   |
| Electric and Electronic Equipment  | -              | -         | -28.8          | $B > \Gamma$               | $C \sim B(.5, .2)$               |
| House Equipment and Furnishings    | 81.5           | 29.3      | -176.6         | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.8, .6)^a$   |
| Machinery and Mechanical Equipment | -              | -         | -176           | $B > \Gamma$               | $C \sim B(.5, .2)$               |
| Metallurgy, Iron and Steel         | -154.4         | -93.2     | 315.3          | $\Gamma > B > \mathcal{N}$ | $C \sim \Gamma(5, 13.1)$         |
| Pharmaceuticals                    | -              | -         | -              | $\emptyset$                | $\emptyset$                      |
| Printing and Publishing            | -              | -         | -165.4         | $B > \Gamma$               | $C \sim B(.4, 1)$                |
| Textile                            | 10.8           | 297.6     | -78.2          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(.8, .5)^a$   |
| Transportation Machinery           | -67            | 92.8      | 121.1          | $\Gamma > \mathcal{N} > B$ | $C \sim \Gamma(3, 7.5)$          |
| Wood and Paper                     | -              | -         | -303.9         | $B > \Gamma$               | $C \sim B(.5, .2)$               |
| <i>Market remaining</i>            |                |           |                |                            |                                  |
| All Manufacturing                  | 1036.7         | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.2, .5)$    |
| Automobile                         | 29.9           | 525.6     | 52.8           | $\mathcal{N} > \Gamma > B$ | $C \sim \mathcal{N}(-.4, .5)$    |
| Chemicals                          | 214.4          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.3, .5)$    |
| Clothing and Footwear              | -4.3           | 352.2     | 89.9           | $\Gamma > \mathcal{N} > B$ | $C \sim \Gamma(2.0, 1)$          |
| Electric and Electronic Components | 245.2          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.1, .3)$    |
| Electric and Electronic Equipment  | 179            | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.0, .3)$    |
| House Equipment and Furnishings    | 118.1          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.3, .5)$    |
| Machinery and Mechanical Equipment | 636.9          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.1, .4)$    |
| Metallurgy, Iron and Steel         | 376.4          | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.6, .8)$    |
| Pharmaceuticals                    | -10.7          | -         | -              | $\Gamma > \mathcal{N}$     | $C \sim \Gamma(1, .5)$           |
| Printing and Publishing            | 443.2          | 224.6     | -42.2          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(-.2, .6)^a$  |
| Textile                            | 244.9          | 191.4     | -54.8          | $\mathcal{N} > B > \Gamma$ | $C \sim \mathcal{N}(-.2, .4)^a$  |
| Transportation Machinery           | 23.1           | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.0, .3)$    |
| Wood and Paper                     | 44.4           | -         | -              | $\mathcal{N} > \Gamma$     | $C \sim \mathcal{N}(-.1, .4)$    |

$z_{N,\Gamma}: H_{N \sim \Gamma} : |z| < +1.96$   $H_{N > \Gamma} : z \geq +1.96$   $H_{\Gamma > N} : z \leq -1.96$ .

$z_{N,B}: H_{N \sim B} : |z| < +1.96$   $H_{N > B} : z \geq +1.96$   $H_{B > N} : z \leq -1.96$ .

$z_{\Gamma,B}: H_{\Gamma \sim B} : |z| < +1.96$   $H_{\Gamma > B} : z \geq +1.96$   $H_{B > \Gamma} : z \leq -1.96$ .

The [a] symbol indicates that caution is needed in the dominance of  $\mathcal{N}$  over  $B$ , as revealed by the Monte Carlo results presented in Section B.3.3.

**Table D.4**  
Occurrence of diagnosis, according to the type of market participation.

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

The figures represent the 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 the Monte Carlo results presented in Section B.3.3.

**Appendix E. 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*  $\theta$ -attribute data  $\theta^T$  from a known distribution  $g$ ;
3. Simulate the *true* threshold data  $C^T$  from a known distribution  $f$ ;
4. Let the agents compute their individual decision outcomes  $\chi$  according to Equation (1);
5. Using the information available to the social researcher (i.e.,  $\theta$  and  $\chi$ ), 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  $\hat{\Omega}$  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  $\mathcal{F} = f_1, f_2, \dots, f_K$  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*  $\theta$ -attribute data  $\theta^T$  from a known distribution  $g$ ;
4. Simulate the *true* threshold data  $C^T$  from the distribution  $f_k$ ;
5. Let the agents compute their individual decision outcomes  $\chi$  according to Equation (1);
6. Using the information available to the social researcher (i.e.,  $\theta$  and  $\chi$ ), estimate via maximum likelihood the parameters  $\hat{\Omega}$  that characterize all the threshold distributions priors  $\hat{\mathcal{F}}$ ;
7. Compute all of the pairwise Vuong's  $z$  statistics;
8. Repeat steps 3 to 7 a sufficient number of times  $M$ ;
9. Use all of the  $M$  estimates  $\hat{\Omega}$  to evaluate the goodness of the estimation;
10. Repeat steps 3–9  $K$  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 of the priors in  $\hat{\mathcal{F}}$ ;
12. For each of the true  $f_k$  evaluate and compare the Vuong's 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  $\sigma = \sigma_1, \sigma_2, \dots, \sigma_K$ ;
3. Simulate the *true*  $\theta$ -attribute data  $\theta^T$  from a known distribution  $g$ ;
  - Also generate also the noisy  $\theta$ -attribute data  $\theta^\varepsilon = \theta^T + \varepsilon^\theta$ ;
4. Simulate the *true* threshold data  $C^T$  from the known distribution  $f_k$ ;
  - Also generate also the noisy threshold data  $C^\varepsilon = C^T + \varepsilon^c$ ;
5. Let the agents compute their individual decision outcomes  $\chi$  according to Equation B.1;
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  $\hat{\Omega}$  to evaluate the goodness of the estimation;
9. Repeat steps 3 to 8 for all of the noise levels  $\sigma$ , as defined at step 2; and
10. Evaluate and compare the goodness of all the values of  $\sigma$ .

Appendix F. Distributions

F.1. Univariate normal distribution

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

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

where  $\mu$  and  $\sigma$  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:

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

F.2. Univariate gamma distribution

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

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

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:

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

F.3. Univariate beta distribution

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

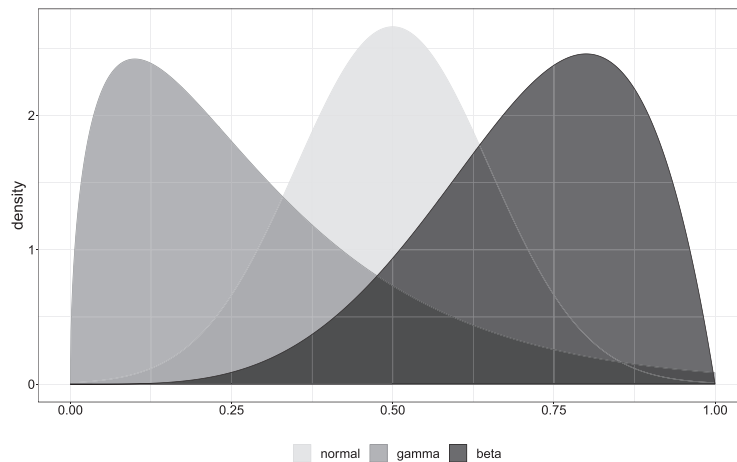
$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \tag{F.5}$$

where  $\alpha > 0$  and  $\beta > 0$  represent the shape<sub>1</sub> and shape<sub>2</sub> parameters, and  $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$  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<sub>1</sub> and shape<sub>2</sub> parameters and read as  $\mu = \frac{\alpha}{\alpha+\beta}$  and  $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2}$ , respectively. This distribution is typically denoted as  $x \sim \mathcal{B}(\alpha, \beta)$ . Integrating over the interval  $(0, 1)$  yields the cumulative density function:

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

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 is the complete beta function (integral between 0 and 1).

Examples of the three distributions are reported in Fig. F.1.



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.

Fig. F.1 Examples of normal, gamma and beta distributions.

References

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

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

Bernard, A.B., Jensen, J.B., 2004. Why some firms export. *Rev. Econ. Stat.* 86 (2), 561–569.

Boulhol, H., 2009. The convergence of price-cost margins. *Open Econ. Rev.* 19 (2), 221–240.

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

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

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

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

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

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

- DellaVigna, S., Lindner, A., Reizer, B., Schmieder, J.F., 2017. Reference-dependent job search: evidence from Hungary\*. *Q. J. Econ.* 132 (4), 1969–2018.
- Dixit, A., 1989. Entry and exit decisions under uncertainty. *J. Polit. Econ.* 97 (3), 620–638.
- Eaton, J., Kortum, S., Kramarz, F., 2011. An anatomy of international trade: evidence from French firms. *Econometrica* 79, 1453–1498.
- Fan, Y., Xiao, M., 2015. Competition and subsidies in the deregulated us local telephone industry. *Rand J. Econ.* 46 (4), 751–776.
- Good, D.H., Nadiri, M.I., Sickles, R., 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. *J. Polit. Econ.* 96 (5), 921–947.
- Hopenhayn, H., 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60 (5), 1127–1150.
- Impullitti, G., Irarrazabal, A.A., Oromolla, L.D., 2013. A theory of entry into and exit from export markets. *J. Int. Econ.* 90 (1), 75–90.
- Jiménez, G., Salas, V., Saurina, J., 2006. Determinants of collateral. *J. Financ. Econ.* 81 (2), 255–281.
- Jovanovic, B., 1982. Selection and the evolution of industry. *Econometrica* 50 (3), 649–670.
- Kau, P., Hill, L., 1972. A threshold model of purchasing decisions. *J. Market. Res.* 9 (3), 264–270.
- Khorunzhina, N., 2013. Structural estimation of stock market participation costs. *J. Econ. Dynam. Contr.* 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., Melitz, M.J., Ottaviano, G.I.P., 2014. Market size, competition, and the product mix of exporters. *Am. Econ. Rev.* 104 (2), 495–536.
- McDonald, R., Siegel, D., 1986. The value of waiting to invest\*. *Q. J. Econ.* 101 (4), 707–727.
- Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71, 1695–1725.
- Melitz, M.J., Ottaviano, G.I.P., 2008. Market size, trade, and productivity. *Rev. Econ. Stud.* 75 (1), 295–316.
- Pissarides, C.A., 1974. Risk, job search, and income distribution. *J. Polit. Econ.* 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., Timmermans, H., 2008. Cut-off models for the 'go-home' decision of pedestrians in shopping streets. *Environ. Plann. Plann. Des.* 35 (2), 248–260.
- Wooldridge, J.M., 2009. On estimating firm-level production functions using proxy variables to control for unobservables. *Econ. Lett.* 104 (3), 112–114.