The added value of management skill in the explanation of the distribution of firm size

Luis Medrano-Adán · Vicente Salas-Fumás

Abstract The stochastic explanations of skewed distribution of firm size question the value added of management theories that trace differences in market share and profits of firms to strategy and organization choices, which are in turn linked to better entrepreneurial skills. This paper explains the distribution of firm size as the market equilibrium outcome of individual occupational choices of working as entrepreneurs or employees. The distribution of size and profit of firms in the equilibrium is directly related to the distribution of skills within the group of individuals who, in the same equilibrium, choose to work as entrepreneurs, restoring the importance of management choice in the performance of firms. The paper highlights the importance of having a theoretical model as guidance in the interpretation of regularities observed in the distribution of firm size, and the ability to distinguish between "resembling" and "true" power laws of such distributions.

L. Medrano-Adán
Centro Universitario de la Defensa, Zaragoza, Spain
e-mail: lmedrano@unizar.es

V. Salas-Fumás (✉)
Faculty of Economics, University of Zaragoza, C. Gran Vía n° 4, 50005 Zaragoza, Spain
e-mail: vsalas@unizar.es

1 Introduction

Observed empirical regularities in the distribution of organization variables challenge the theories of individual and group behavior that must explain the "stylized facts." This paper is concerned with the theoretical understanding—from individual occupational choices of working as employees or as entrepreneur-managers—of the long-term, empirically established regularity that the distribution of firm size (DFS from now on) follows a power law (Simon, 1955; Axtell, 2001). Existing explanations of why the size of the firm and other organization variables are distributed as a power law appeal to stochastic, multiplicative, or additive cumulative shocks to the value of the reference variable (Andriani and McKelvey, 2009; Gabaix, 2016).[1] This paper takes a different stance and proposes an explanation of the distribution of firm size that follows Axtell's (2001:1820) request for a "microeconomic model in which individual agents interact to form productive teams (that can explain) … the undisputed empirical evidence that the size of firms in the US follows a Zipf distribution" (a special case of power law distribution). To this end, we propose an explanation of the DFS with entrepreneurial-managerial skill as a main determinant.

---

[1] The earlier explanation is from Gibrat's (1931) law of independence between growth rates of firms and their respective initial absolute size. Ijiri and Simon (1967) combine proportionate growth rates (Gibrat's law) with a minimum scale in the entry of new firms to show that the distribution of firm size will converge to a Pareto distribution (power law distribution).

🙋 Springer

The proposed microeconomic explanation of the DFS draws on occupational choice theory (Lucas, 1978; Rosen, 1982), where management teams compete for control of resources to produce goods and services until the market equilibrium is reached. Individuals in the economy differ in attributes important for income-maximizing occupational choices that in this paper are summarized in the construct "entrepreneurial skills" (Lucas, 1978; Jovanovic, 1994). The way that the entrepreneurial skill determines the income from working as employee (salary) or as entrepreneur-manager (profit), implies that in the market equilibrium, individuals in the upper tail of the distribution of skill choose to work as entrepreneur-managers and the rest as employees. The profit-maximizing inputs and output quantities of each entrepreneur-firm depend on the skill of the respective entrepreneur-manager. A correspondence is then established between the distribution of skills in the population, the truncated upper tail (from the skill level of the less skilled entrepreneur) of this distribution that corresponds to the distribution of skills for the sub-group of individuals who work as entrepreneurs, and the distribution of the profit-maximizing values of the outcome variables of firms such as sales, employees, capital, and profits - each a function of the respective entrepreneur's skill. From this correspondence, under the assumptions of the model, the DFS will be a power law only when the distribution of skill in the population is also a power law.

The purely stochastic or "chance" explanation of the distribution of variables, such as growth rate and size of the firm, that has dominated the literature, particularly that around the explanation of the power law distribution attributed to the variable size of the firm (Andriani and McKelvey, 2009), and that Axtell (2006) qualifies "more as fables of firm growth than as credible explanations," question the "value added" of theories of Management. Such theories trace the differences in performance among firms to better strategy and organization choices linked to better managerial skills (Geroski 2000; Denrell et al. 2014; Knudsen et al. 2017). The representation of the firm in occupational choice models explicitly incorporates the input of the entrepreneur (quality of strategic decisions and skill-weighted monitoring time) as determining the organization's production output and performance (profit), in combination with direct labor and capital. Consequently, in the explanation of the heterogeneity observed in the size and performance of firms from occupational

choice models, entrepreneur-managers' skills matter, and the value added of management theory is restored. In this respect, the entrepreneur in occupational choice models is characterized as a director of resources (Coase, 1937; Penrose, 1959), rather than as an individual involved in the introduction of innovations and new ventures (Schumpeter, 1934).[2]

In the empirical section, we use Axtell's (2001) data about the size of firms, from the US census, to test some predictions from the theory, and to compare these predictions with the results from assuming that the firm size data follows a power low (Axtell's assumption). The research interest is similar to that of Joo et al. (2017) on whether the taxonomy of non-normal distributions that fits individual output variables includes distributions other than the power law. The difference, in our paper, is that, first, we connect the empirical results with the predictions from a model of individual and collective behavior; in particular, we derive the DFS in the market equilibrium assuming that the distribution of skill in the population is lognormal and then fit the distribution to the actual data. And second, it is possible to relate the parameters of the distribution of outcome variables with the parameters of the theoretical model.

This paper contributes to the literature that aims to explain the causal processes underlying observed patterns of heterogeneity across firms, particularly in their size and profits, with new views from occupational choice models that complement the existing explanations, mainly coming from complexity science (Andriani and McKelvey, 2009). The repetition of the same pattern of a highly skewed-to-the-right distribution of the values of organization and entrepreneurial variables has led to the conclusion that the description of a social world is far from "normality." The paper then complements the existing explanations of the social world of power laws, with an explanation of the heterogeneity in size and performance of firms from a model of rational individual behavior and team production that responds to the request of Axtell (2001), and restores the value of

---

[2] Coase (1937) defines the entrepreneur as "the person or persons that in a competitive system take the place of the price system in the direction of resources." Penrose (1959) explicitly refers to the managers' limited working time as determining the costs of growth and limits to the size of the firm; in organizational choice models, the number of entrepreneurs and the size of the respective production team depend on the organizational size diseconomies, similar to Penrose's limits to the growth of the firm.

management and entrepreneurial inputs as determinants of size and performance of firms.

Entrepreneurship research, see Crawford et al. (2015), Aguinis et al. (2018), and Joo et al. (2017), has established empirical links between the distribution of entrepreneurial inputs (one of these being entrepreneurial skill) and the distribution of entrepreneurial outcomes (one of these being the size of the firm). The occupational choice model presented in this paper complements this literature with a market equilibrium distribution of firm size (outcome) that is the result of a transformation of the distribution of skill (input). Specifically, for the standard formulation of the production function of firms, the DFS will be a convex transformation of the left-truncated distribution of entrepreneurial skills in the population, and the probability density function of the DFS will generally be strictly (strictly) decreasing and convex. The visual examination of the function will resemble that of a power law, but the sufficient conditions for the size distribution of firms being a strict power law are more restrictive, as this paper shows. The lower bound and the shape of the DFS will depend, among other things, on parameters of the distribution of skills in the population, and on parameters of the production and organization technologies.[3]

Although the occupational choice theory is well established and has been influential in the literature (Lucas' 1975 paper has more than 4000 citations in Google), its practical relevance is conditioned by the difficulty of directly measuring the entrepreneurial skill of individuals. The use of human capital data, i.e., education, training, and experience of entrepreneur-managers (Storey, 1994; Roper, 1998; Ferrante, 2005) as proxies of entrepreneurial skill, leaves out many personal attributes that differentiate entrepreneurs from salaried employees. This difficulty may explain why the predictions from the theory have been tested in indirect ways. For example, the positive association between size of the firm and compensation of managers has been interpreted as evidence that the higher productivity of the management team (more volume of managed resources) is rewarded with higher compensation (Rosen, 1981, 1982; Gabaix and Landier, 2008).

Ferrante (2005) finds that the time entrepreneurs dedicate to entrepreneurial tasks increases with their skills, which is interpreted as evidence that the return from the time dedicated to entrepreneurial tasks increases with skill. Vendrell-Herrero et al. (2014) find a positive correlation between total factor productivity, their proxy for entrepreneurial talent, and the rate of return in a sample of technological firms. Although there are ways to overcome the difficulty of directly measuring entrepreneurial skills, more needs to be done for occupational choice theory becoming more relevant for policy and management.

The rest of the paper is organized as follows. Section 1 presents a brief introduction to power laws. Section 2 introduces the basic theory of the size distribution of firms from occupational choices. Section 3 extends the results to outside occupational choice models. Section 4 re-examines the empirical tests of whether the size distribution of firms in the USA, with census data, follows a Zipf distribution. The Conclusions summarize the main findings of the paper and draw some theoretical and empirical implications as well as extensions of future research.

2 Power laws

The term *power law* is used in statistics, the natural sciences, and the social sciences to refer to relationships between two quantities or variables that can be mathematically described by power functions. Thus, a power law defines a functional relationship between two variables, y and x, such that the value of y is equal or proportional to a value of x risen to a power parameter. More formally, a power law is usually represented by a function of the form $y = c\ x^{-a}$ with c, $a > 0$. In statistics, the distribution of a variable x is said to follow a power law if and only if its probability density function $f(x)$ (or its survival function) is a power function.

A distinctive feature of the power function is that the elasticity of y to changes in x is constant and equal to a for all values of x (a log-linear relationship of variables y and x). Power functions have the property that, if $g(x)$ and $h(x)$ are power functions, then $g'(x)$ and $h(g(x))$ are also power functions. Mathematically, if $g(x) = c_1 x^{a_1}$ and $h(x) = c_2 x^{a_2}$, then $g'(x) = c_1\ a_1 x^{a_1-1}$, $h'(x) = c_2\ a_2 x^{a_2-1}$, and $h(g(x)) = c_2 c_1^{a_2} x^{a_1 a_2}$.

---

[3] Although not considered in this study, empirical evidence (Davis and Henrekson 1999; Henrekson and Johansson 1999) shows that the distribution of firm size will also depend on market frictions and the institutional environment (taxes, employment laws, regulation of financial markets, and size of the public sector).

The only random variables whose probability density functions are power functions in a strict sense are the Pareto distribution and the Zeta (also known as Zipf distribution), the former being a continuous version of the latter.[4] Their probability density functions are given by:

$$\text{Pareto distribution}: \ f(x) = \alpha e_m^{\alpha} x^{-1-\alpha} \ \ (\alpha > 0, x \geq e_m > 0, x \in \mathbb{R})$$

$$\text{Zeta distribution} \ \ \ p(x) = \frac{x^{-1-\rho}}{\zeta(1+\rho)} \ \ \ \ (\rho > 0, x \geq 1, x \in \mathbb{N})$$

where $\zeta(x)$ is the Riemann zeta function.

2.1 Power laws of individual and organizational variables

Most of the explanations of why entrepreneurial and organizational variables, including firm size, follow a true or apparent power law include some stochastic element. A common reference, see Crawford et al. (2015) and Crawford (2018), is Andriani and McKelvey (2009) that explains power law distributions as an outcome consistent with the predictions from complexity science, with extensions to theories of organizational change and development that are "scale-free," i.e., there is one primary driver that explains the relationship between inputs and outputs at multiple levels, regardless the scale at which the phenomenon is measured.[5] These authors view organization development as the outcome of *independent-multiplicative* causal elements that operate as the system complexity increases, or as the result of random events that are *interdependent*, *interactive*, or both. In these situations, "Pareto distributions dominate because the positive feedback (and other processes) leading to extreme effects occur more

frequently than "normal" Gaussian-based statistics lead us to expect" (p. 1055).

Gabaix (2016) considers two mechanisms that generate power laws in the distribution of firm size: random growth, and "transfer" of that power from another variable via matching and optimization. The random growth model that generates power laws of firm sizes is the *proportional* random growth of Gibrat (1931) complemented with other assumptions, for example a minimum firm size required for survival (Ijiri and Simon 1967). The random growth rates that interact multiplicatively generate bell-shaped but skewed distributions (lognormal distribution). With a lower bound of the size variable the bell shape is lost, and the distribution converges to the Pareto one. The "scale free" condition in these papers is that of constant returns to scale in the production technology, so that unit production cost is constant for all values of size (i.e., differences in size do not become differences in unit costs).

The matching and optimization explanation of the power law of firm size draws from Rosen's (1981) "economics of superstars" that provides a market explanation of the positive correlation between compensation of managers and size of the respective firm (Gabaix and Landier, 2008). Firms with different sizes compete for hiring managers of different talent; the outcome of the competition process assigns the most talented person to the largest firm and so on. The distribution of managers' talent is not known but, under certain assumptions, given the ordering of talent, the approximate difference of talent between two adjacent managers varies like a power law of their rank. The matching and optimization processes imply that the power law distribution of the rank of managerial talent is "transferred" to the distribution of firm size, and to the distribution of managers' compensation.

The matching explanation of the power law distribution of firm size is somehow related to the matching between entrepreneurial skills and firm sizes in occupational choice models, although this relation has not yet been formally established. The model presented in the following section summarizes the heterogeneity of the working population in the construct "entrepreneurial skills." The heterogeneity of skills in the population will explain the occupational choice of working as salaried employee or as entrepreneur-manager, as well as the differences in the number of employees working under

---

[4] In a less strict sense, other distributions, which are not represented by pure power functions, are also referred to as power laws, like the "power law with exponential cut-off" or the piecewise functions consisting of two or more power functions. In this respect, Aguinis et al. (2018) use the term power law to refer to those heavy-tailed distributions where output is clearly dominated by a small group of elites and most individuals in the distribution are far to the left of the mean". This definition of power law can accommodate distributions other than the Pareto and the Zeta distributions, the only two technically acceptable power laws. Then, the fact that a distribution resembles a power law does not necessarily imply that it is strictly a power law. In this paper, the power law is restricted to Pareto and Zeta distributions.

[5] See also McKelvey (2004) on complexity science and entrepreneurship.

the direction of each entrepreneur (different firm size).

The condensation of the heterogeneity in the working population into the single attribute of entrepreneurial skill responds to tractability reasons. Entrepreneurship is a multidimensional phenomenon and the list of variables that have been used to characterize entrepreneurs in the working population is rather long (Shepherd et al., 2019). Conceptually at least, the occupational choice theory and the predictions from the market equilibrium could be extended substituting the single construct of entrepreneurial skill, for a vector z of attributes (skills, risk attitude, preference for independent work, empathy...), with joint cumulative distribution functions in the population, $H(z)$. The individuals who choose to work as entrepreneurs in the market equilibrium will occupy a region R of the space of attributes determined from supply and demand conditions, similar to what happens with the single construct of skill. The outcome variables from the market equilibrium (production inputs labor and capital, output, profits) would then depend on the values of the variables in z, and the cumulative distribution functions of those outcome variables would continue being a function (transformation) of the joint distribution function H with the values of the attributes z restricted to values in the region R.

3 Occupational choices explanation of firm size

The main premise of occupational choice models is that the distribution of firm size is the equilibrium outcome in a market where individuals with different skills compete for the control of productive resources, labor, and capital. This section presents a summary of the basic results of occupational choice models (Lucas 1978; Rosen 1982; Jovanovic 1994) on the determinants of the distribution of firm size in the economy.

3.1 Market equilibrium and distribution of firm sizes

Consider a production technology that uses two inputs[6] to produce a quantity of output, labor services supplied by employees, and the skill-weighted time of the entrepreneur who manages the production process. The skill

of the entrepreneur is fixed, and employees can be hired in the market at a given salary. Output produced is sold in the market at a normalized price of one. The working time is normalized to one unit. The entrepreneur earns the profit of the firm and employees receive a market-determined salary.

Let $Q = g(e, L)$ be the production technology, where Q is the total output produced; e is a number that represents the skill of the entrepreneur; L is the number of employees; and $g()$ is the production function, increasing in e and L, and concave in L. The production function is formally derived from the aggregation of outputs from individual job, each occupied by an employee monitored by a single entrepreneur of given skill, as in Rosen (1982). The entrepreneur decides how to allocate the skill-weighted limited working time among the employees-jobs; in the optimal solution the output in each job depends on the entrepreneur's monitoring time and of the direct working time of the employee. The aggregate output of the entrepreneur-team of employees is further leveraged by the quality of the strategic decisions of the entrepreneur that contribute equally to the productivity of all employees. The monitoring function of the entrepreneur-manager takes place under decreasing returns to scale (which capture the organizational size diseconomies that limit the growth of firms in Penrose's (1959) theory of the growth of the firm). The quality of strategic decisions is a "public good" that affects positively the output of all jobs equally. The contribution to the total productivity of the group from the quality of the decisions of the entrepreneur increases with the level of skill. Occupational choice models modify the neoclassical production function to incorporate the contribution to the total output (from quality of decisions and from monitoring of their implementation) of the skill-weighted working time of the entrepreneur who manages the production team.

If w is the market salary for employees, the profit/income of the entrepreneur of skill e is $\Pi = g(e, L) - w L$. The profit-maximizing number of employees satisfies the condition of marginal revenue equal to marginal costs, $\frac{\partial g(e,L)}{\partial L} = w$. Solving this equation, the demand for labor as a function of the entrepreneur's skill and employees' salary is given by $L = h(e; w)$, increasing with skills e and decreasing with the salary, w. Substituting into the profit function, the maximum profit for an entrepreneur of skill e is equal to $\Pi^*(e; w) = g(e, h(e; w)) - w\ h(e; w)$.

---

[6] All the analysis can be easily extended to the case where firms also use other inputs (beyond labor) in production, for instance, capital input. See Medrano-Adan et al., 2015; 2019).

Individuals with different skills choose between working as employees and earning the market salary or working as entrepreneurs and earning a profit. The distribution of skills in the population is taken as given. Let $F(e)$ and $f(e)$ denote the cumulative distribution function (CDF) and the probability density function (PDF) of entrepreneurial skills with support $[e_m, e_M]$, where e_M may be equal (or tend) to $+\infty$.

The market equilibrium is determined by two equations, one on the level of entrepreneurial skills e^* that characterizes the individual who is indifferent between working as employee or as entrepreneur, and the other on the market salary w^* for which the supply of employees is equal to the demand:

$$\Pi^*(e^*; w^*) = w^*$$
$$\int_{e_m}^{e^*} dF(e) = \int_{e^*}^{e_M} h(e; w^*) dF(e) \qquad (1)$$

Figure 1 shows the market equilibrium; more specifically, it shows the salary of employees and profits of entrepreneurs-managers for the equilibrium salary w^*. Since the maximum profit is increasing with e, for each w there will be a skill value e^* such that an individual with this skill will be indifferent between working as employee or as entrepreneur (first equation). Individuals with skills above the threshold will earn higher income as entrepreneurs, while individuals with lower skills will earn higher income as employees (see Fig. 1). If the demand for employees is higher than the supply, the market salary will increase, reducing the demand and increasing the supply of individuals who want to work as employees. The salary will adjust until supply equals demand.

The distribution of skills in the group of entrepreneurs is just the left-truncated distribution of skills in the population, so that its support is $[e^*, e_M]$ for e_M finite, and $[e^*, +\infty]$ otherwise. When we measure firm size S by the number of employees L, then firm size is a function of the entrepreneur's skill, $S = L = h(e; w^*)$, where "e" is the skill of an entrepreneur hiring $L = h(e; w^*)$ employees. The distribution of firm size, measured by the number of employees, is just a transformation of the distribution of the entrepreneurs' skills.

Consequently, **the distribution of firm size is** just **a transformation of the left-truncated distribution of skills in the whole population**. The following claim characterizes the probability cumulative function and

the probability density function of the distribution of firm size.

Claim 1. Mathematically, the probability cumulative function and the probability density function of the distribution of entrepreneurs' skills are characterized by

$$F_{ent\ skills}(e) = \frac{F(e) - F(e^*)}{F(e_M) - F(e^*)},$$

$$f_{ent\ skills}(e) = \frac{f(e)}{F(e_M) - F(e^*)} \qquad (\forall e \geq e^*)$$

Then the probability cumulative function and the probability density function of the distribution of firm size are characterized by

$$F_S(s) = \frac{F(h^{-1}(s)) - F(e^*)}{F(e_M) - F(e^*)}, \qquad (2)$$

$$f_S(s) = \frac{f(h^{-1}(s))}{F(e_M) - F(e^*)} \frac{d\ h^{-1}(s)}{d\ s} \qquad \forall s \geq S_{\min}$$

Where $S_{\min}$ is the minimum firm size, $S_{\min} = h(e^*; w^*)$.

Proof of Claim 1: See Appendix 1.

Mathematically, the shape of the DFS is determined by: (i) the shape of the left tail of the distribution of entrepreneurial skills in the population, and (ii) the shape of the (individual) labor demand as a function of the entrepreneur's skill, which is increasing and convex with skills for all reasonable production functions.

3.2 Distributions of firm size for different distributions of skills

We now illustrate the implications of Claim 1 by solving for the DFS from different distributions of skills and a particular production function. In its simplest form, from the way the input of the entrepreneur enters the production function, the aggregate output from L employees-jobs managed by an entrepreneur of skill e is given by,

$$Q = \theta\ e^{(\tau+\beta)} L^{1-\beta} \qquad (3)$$

The parameter θ captures the general level of total factor productivity different from the contribution of the skill of the entrepreneur. The term e^τ ($\tau > 0$) is the contribution of the skills of the entrepreneur from the

Fig. 1 Market income and occupational choices in the market equilibrium (as a function of skills). Wages and profits (vertical axis) versus level of skill (horizontal axis). The hump-shaped curve is the probability density function of the distribution of skills, lognormal in this figure
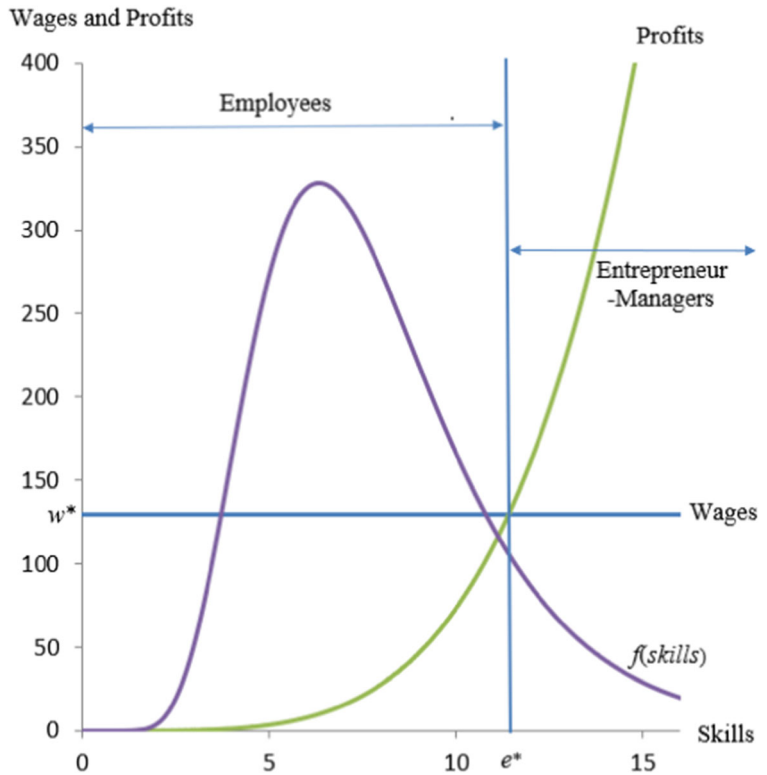
quality of the entrepreneurial and managerial decisions (public good effect). The term $e^{\beta}L^{1-\beta}$ is the contribution from the aggregation of the output at the job level, of the input of the employee and the skill-weighted individualized monitoring time of the entrepreneur. The parameter $1 > \beta > 0$ determines the degree of organizational size diseconomies from the technology used in the supervision of employees' work.

The first order conditions of profit maximization, $\frac{\partial g(e,L)}{\partial L} = w$, lead to the optimal labor demand function,

$$L = h(e; w) = \left(\frac{\theta(1-\beta)}{w}\right)^{\frac{1}{\beta}} e^{(\beta+\tau)/\beta}$$

This is a power function of the skill, $h(e; w) = a\, e^{b}$, where $a = (\theta(1-\beta)/w)^{1/\beta}$ and the power parameter is $b = (\beta + \tau)/\beta$.

The market equilibrium from occupational choices is characterized by [1], with $\Pi^*(e^*; w^*) = \theta\ e^{(\tau+\beta)}[h(e^*; w^*)]^{1-\beta} - w^* h(e^*; w^*)$. After substituting the optimal labor demand $h(e; w) = \left(\frac{\theta(1-\beta)}{w}\right)^{\frac{1}{\beta}} e^{(\beta+\tau)/\beta}$, we obtain the system of equations

$$\beta\theta^{\frac{1}{\beta}}\left(\frac{1-\beta}{w^*}\right)^{\frac{(1-\beta)}{\beta}}\left(e^*\right)^{\frac{(\beta+\tau)}{\beta}} = w^*,$$

$$\int_{e_m}^{e^*} dF(e) = \left(\frac{\theta(1-\beta)}{w^*}\right)^{\frac{1}{\beta}} \int_{e^*}^{e_M} e^{(\beta+\tau)/\beta} dF(e)$$

From the first equation, the equilibrium salary is $w^* = \theta\beta^{\beta}(1-\beta)^{(1-\beta)}(e^*)^{(\beta+\tau)}$, where the skill threshold e^* is the solution to the following equation (which has no closed-form solution, except for some specific distributions of skills, like the Pareto):

$$\beta\left(e^*\right)^{(\beta+\tau)/\beta}\int_{e_m}^{e^*} dF(e) = (1-\beta)\int_{e^*}^{e_M} e^{(\beta+\tau)/\beta} dF(e)$$

We consider six alternative probability density functions of entrepreneurial skills in the population, $f(x)$, and, for each of them, we obtain the distribution of firm size, $f_s(x)$. In all cases (except for the uniform distribution) the support of the distribution of firm size is $[S_{\min}, +\infty)$, with $S_{\min} = h(e^*; w^*) = a(e^*)^b$, $a = \theta^{1/\beta}((1-\beta)w^*)^{-1/\beta}$ and $b = (\beta + \tau)/\beta$.

The probability density functions of the six distributions of skill, the corresponding distributions of firm

size, and the Zipf plot of the size distribution (a graph of the logarithmic transformation of the survival function as a function of the logarithmic of size), often used to represent the DFS, are represented graphically in Table 1. The continuation of Table 1 also shows the respective functional form for each of the six distributions of skill and of firm size.

Although the distributions of skills considered have *very* different shapes, the probability density function of the firm size distribution is strictly decreasing and convex in all cases (even in the case of the Uniform distribution of skills). Different distributions of skill turn into quite similar, at least from the visual observation, distributions of firm size. In fact, since in all cases the distribution is a decreasing and convex function of size, all distributions "resemble" a power law, and would be considered as such under a "loose" definition of power law distributions. However, technically, only in the case where the distribution of skills is a Pareto distribution, the distribution of firm size is indeed a Pareto, with lower bound $a\,(e^*)^b$ and power parameter $\frac{\alpha}{b}$. The Pareto distribution of skill is also the only one for which the Zipf plot of the distribution of firm size is a linear function. In the other size distributions, the logarithm of the survival function is a decreasing and concave (non-linear) function of the logarithm of size.[7]

In conclusion, since the size distribution of firms is a power transformation of the truncated upper tail of the distribution of skills in the population, and the upper tail of many distributions will be decreasing and convex with values of the random variable, the size distribution of firms will also be decreasing and convex with size. This would lead to the belief that the distribution follows a power law, although this will only be strictly true if the distribution of skills is also a power law. The importance of the distinction between "resemblance" and "true" power law distribution may depend on the research context. In the context of this paper, with an explicit model of the formation of firms and the match between entrepreneur-managers and firms in the economy, whether the distribution of the variables of interest are true power laws or not would be particularly relevant in testing predictions of, say, how the distribution of entrepreneurial skills in the population determines the distribution of size and performance variables.

---

[7] As complement, Appendix 2 shows the closed solution to the market equilibrium from occupational choices when the distribution of skill is Pareto.

4 Generalization

The expressions in [2] are valid for any distribution of skills and for any labor demand function. In order to obtain some general hints about the properties of the DFS, we must impose some reasonable conditions on the functions of labor demand $L(e)$ and distribution of skills, $F(e)$. From the way the skills of the entrepreneur enter into the production function, i.e., as part of the total factor productivity term, the labor demand will be increasing and convex in the entrepreneur's skill, $L = h(e; w), \frac{\partial h}{\partial e} > 0, \frac{\partial^2 h}{\partial e^2} > 0$. Moreover, for commonly used production functions (Cobb Douglas and CES production functions), the demand for labor is an increasing and convex power function of entrepreneurial skills $h(e; w) = a\, e^b$ (with $a > 0$ and $b > 1$), Medrano-Adan et al., 2015, 2019).

Since there are individuals with *extremely* high levels of skill (the superstars), we can reasonably assume, for modeling purposes, that the maximum skill in the population is unbounded $e_M \to +\infty$, and that the support of the distribution is $[e_m, +\infty)$. But then, its probability density function must be strictly decreasing and convex in the right tail (for sufficiently high values of e, see Lemma 1 in the Appendix), since probability values must be non-negative and the limit of the cumulative distribution must be equal to one.

To sum up, for commonly used production functions, the labor demand function will be increasing and convex in the level of skills; and, for reasonable distributions of skills, the probability density function is decreasing and convex. The following claim mathematically proves that, given these conditions, the probability density function of the DFS will be strictly decreasing and convex for all size values.

Claim 2. If $\frac{\partial h}{\partial e} > 0, \frac{\partial^2 h}{\partial e^2} > 0$, and $f'(x) < 0$, $f''(x) > 0$, for $x \geq e_1$, then $f'_S(s) < 0$ and (in general) $f''_S(s) > 0$ $(\forall s > S_{\min})$.

Proof of Claim 2: See Appendix 1.

The distribution of firm size inherits the properties of the right tail of the skill distribution: it will be (bounded from below and) strictly decreasing and convex, i.e., similar in graphical appearance to a power law (a power function with negative power parameter). On the other hand, even if the distribution of skills is bell-shaped (normal, lognormal, etc) the probability density of the DFS will be strictly

Table 1. Graphical representations of the distributions of skills and firm size, for six alternative distribution of skills.

Distribution of skills	Probability density function of skills, in the population (blue) and among entrepreneurs (orange), vs skills.	Probability density function of the distribution of firm size, as a function of size s (horizontal axis).	Zipf-plot of the firm-size distribution: $\ln[1 - F_s(s)]$ as a function of size s, in log scale
Pareto distribution			
Lognormal distribution			
Normal distribution			

Distribution of skills	Probability density function of skills, in the population (blue) and among entrepreneurs (orange), vs skills.	Probability density function of the distribution of firm size, as a function of size s (horizontal axis).	Zipf-plot of the firm-size distribution: $\ln[1 - F_s(s)]$ as a function of size s, in log scale
Exponential distribution			
Extreme value distribution			
Uniform distribution			

Distribution of skills	Probability density function of skills	Probability density function of the distribution of firm size, as a function of size s
Pareto Distribution $P(e_m, \alpha)$:	$f(x) = \alpha e_m{}^\alpha x^{-1-\alpha}$	$f_S(x) = \left(a e^{*b}\right)^{\frac{\alpha}{b}} \frac{\alpha}{b} x^{-1-\frac{\alpha}{b}}$
Lognormal Distribution $LN(\mu, \sigma)$:	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}\,x} \mathrm{Exp}\left[-\dfrac{(\log[x]-\mu)^2}{2\sigma^2}\right]$	$f_S(x) = \dfrac{1}{1-\frac{1}{2}Erfc\left[\frac{\mu-\log[e^*]}{\sigma\sqrt{2}}\right]} \dfrac{1}{bx}\dfrac{1}{\sqrt{2\pi}\sigma} \mathrm{Exp}\left[\dfrac{-1}{2\sigma^2}\left(\dfrac{1}{b}\log\left[\dfrac{x}{a}\right]-\mu\right)^2\right]$
Normal distribution $N(\mu, \sigma)$:	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \mathrm{Exp}\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$	$f_S(x) = \dfrac{1}{1-\frac{1}{2}Erfc\left[\frac{\mu-e^*}{\sigma\sqrt{2}}\right]} \dfrac{1}{ab\sigma\sqrt{2\pi}}\left(\dfrac{x}{a}\right)^{-1+\frac{1}{b}} \mathrm{Exp}\left[\dfrac{-1}{2\sigma^2}\left(\left(\dfrac{x}{a}\right)^{\frac{1}{b}}-\mu\right)^2\right]$
Exponential Distribution $Exp(\lambda)$:	$f(x) = \lambda\,\mathrm{Exp}[-\lambda x]$	$f_S(x) = \dfrac{\lambda e^{\lambda e^*}}{ab}\left(\dfrac{x}{a}\right)^{-1+\frac{1}{b}} \mathrm{Exp}\left[-\lambda\left(\dfrac{x}{a}\right)^{\frac{1}{b}}\right]$
Extreme value distribution $EV(\alpha, \theta)$:	$f(x) = \dfrac{1}{\alpha}\mathrm{Exp}\left[\dfrac{\alpha-x}{\theta}-\mathrm{Exp}\left(\dfrac{\alpha-x}{\theta}\right)\right]$	$f_S(x) = \dfrac{1}{1-\mathrm{Exp}\left[-\mathrm{Exp}\left(\frac{\alpha-e^*}{\theta}\right)\right]} \dfrac{1}{\theta ab}\left(\dfrac{x}{a}\right)^{-1+\frac{1}{b}} \mathrm{Exp}\left[\dfrac{\alpha}{\theta}-\dfrac{1}{\theta}\left(\dfrac{x}{a}\right)^{\frac{1}{b}}-\mathrm{Exp}\left(\dfrac{\alpha}{\theta}-\dfrac{1}{\theta}\left(\dfrac{x}{a}\right)^{\frac{1}{b}}\right)\right]$
Uniform distribution $U(x_m, x_M)$:	$f(x) = \dfrac{1}{(x_M-x_m)}$	$f_S(x) = \dfrac{1}{ab(x_M-e^*)}\left(\dfrac{x}{a}\right)^{-1+\frac{1}{b}} \quad \left(a\,e^{*b} \leq x \leq a\,x_M^b\right)$

Note: $Erfc[z] = 1 - \dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^z \mathrm{Exp}\left[-t^2\right]dt$, and Exp[] is the exponential function.

decreasing and convex (losing its bell shape because only those in the right tail of the distribution of skills become entrepreneurs) and will resemble a power law.

In fact, we can prove that if the labor demand function is a power function of skills, then the distribution of skills being a Pareto distribution is a necessary and sufficient condition for the DFS to be a power law.

*Claim 3. If the labor demand is a power function of skills, then the distribution of firm size is a power law **if and only if** the distribution of skills is Pareto.*

Proof of Claim 3: See Appendix 1.

Consequently, in the framework of occupational choice models, with individuals who differ in their levels of skill, the distribution of firm size will follow a power law **only if** both the labor demand function and the distribution of skills are power functions. For other labor demand functions and/or other distributions of skills, the distribution of firm size will **not strictly** follow a power law, although its graphical appearance may resemble a power law (probability density function will be decreasing and convex).

What are the a priori assumptions about the distribution of entrepreneurial skill in the population? One

reasonable proxy of the variable skill is the human capital of the working population. The OECD-PIAAC project measures the cognitive skills of the working population across OECD countries in a standardized way. The published reports on the results of the project show that in all countries the distribution of cognitive skills in the population is "bell-shaped" but non-symmetric (Broecke et al. 2017). Since the entrepreneurial input in the production function is in related with quality of the decisions and monitoring time, it is reasonable to expect that the variable entrepreneurial skill will be monotonically increasing with the level of cognitive skill, and then, its distribution will also be bell-shaped but non-symmetric. Specifically, in the empirical analysis (section 4), we will assume that the distribution of entrepreneurial skill in the population is lognormal and estimate the corresponding distribution of firm size that results from the market equilibrium.

Occupational choice models explain the heterogeneity among firms via the heterogeneity in the indivisible skill of their respective entrepreneur-managers. Other studies highlight the heterogeneity observed in the total factor productivity, TFP, of firms and production plants (Moral Benito 2018; Decker et al. 2018), and still others attribute the heterogeneity in TFP, at least in part, to the high dispersion observed in the quality of management within and across countries around the world (Bloom and Van

Reenen, 2007). Although no direct connection has yet been established, the observed heterogeneity in TFP and in management skills among firms within a country and across countries could just be the reflection of differences in the construct variable entrepreneurial skill among those compared. In fact, the skill input in Eq. [3] above is part of the TFP term of the production function, together with the general productivity parameter θ. Ortín-Ángel and Vendrell-Herrero (2010), Vendrell-Herrero et al. (2014) explicitly use the TFP of firms as a measure of entrepreneurial talent. Medrano-Adán et al. (2019) provide evidence that the observed distribution of TFP among Spanish firms in Moral Benito (2018) is very much in line with the distribution predicted from an occupational choice model with values of the parameters calibrated for the Spanish economy.

Crawford et al. (2015) explain the power law distribution of entrepreneurs' outcomes, such as size of firms, as the result of power law distributed inputs grouped into "endowment" (i.e., human capital of entrepreneurs), "engagement" (i.e., entrepreneur's working time), "expectations" (i.e., future growth projections of the venture) and "environments" (i.e., sales in different industries) variables. Occupational choice models can explain the link between input and outcome variables in the market equilibrium for input variables that condition the occupational choice of working as entrepreneur or as employee, and that enter as inputs of the entrepreneur in the production function. The endowment, engagement, expectations, and environment variables could all eventually be part of the vector of variables z mentioned above that could condition the occupational choice. Endowment variables such as human capital of the entrepreneur, together with engagement variables such as working time of the entrepreneur, would qualify as entrepreneurial inputs of the production function and be part of the "skill" variable. However, expectations and environment variables as inputs of the entrepreneur in the production function would be more difficult to justify.

In any case, the occupational choice theory supports the view that there will be a link between the distribution of entrepreneurial inputs and the distribution of entrepreneurial outputs and formally derives what such link will be. In the context of the model the necessary condition for the outcome being distributed as a power law is that the input be distributed as a power law too. There is supporting evidence of the skewedness of the distribution of entrepreneurial input and output variables but as Joo et al. (2017) show not all right-skewed distributions that

"resemble" power laws are in fact "true" power law distributions. Stochastic additive and multiplicative shocks (Andriani and McKelvey, 2009) explanations of the distribution of entrepreneurial and organizational variables, and theories that explain the emergence of outliers in such distributions (Crawford, 2018) will predict links between entrepreneurial input and output variables that at this point occupational choice models could not explain.

5 Empirical tests of the distribution of firm size

In this section, we use the Axtell (2001) data about the size distribution of firms, from the US census, to test some predictions from the theory and to compare the results of the empirical analysis with those obtained directly when assuming that the distribution of firm size follows a power law. The exposition will be divided in two parts. First, we follow the conventional approach of fitting a power law distribution to the data and applying some conventional test of model specification to reject, or not, the null hypothesis that the firm size distribution variable in the USA follows a power law distribution. Second, we estimate the distribution of firm sizes predicted from our occupational choice model, assuming that the distribution of entrepreneurial skills in the population is lognormal, and compare the goodness of fit with that obtained when assuming that the variable is distributed according to a power law.[8] The use of Axtell's firm size data from the USA is because it is easily available and because Axtell's paper is probably the most-cited paper supporting the view that the firm size variable is distributed according to a power law. (In Appendix 3, we provide complementary evidence with Spanish data.)

5.1 Test of the null hypothesis: The distribution of firm size is a power law

We first replicate the conventional empirical analysis of the distribution of firm size, ignoring the predictions from the theory above. Under the null hypothesis that

---

[8] There is extensive research comparing the goodness of fit of power law and of other distribution functions, for example lognormal and exponential ones, to actual data on the distribution of economic and non-economic variables (Clauset et al., 2009; Joo et al., 2017). The difference with the research reported here is that we compare the goodness of fit of a distribution predicted by a theoretical model with the goodness of fit of a power law.

firm size, measured by number of employees L, follows a Pareto distribution, with power parameter $\alpha > 0$ and minimum size $L_{min} > 0$, the probability density function, the cumulative distribution function, and the survival function of the size variable are given by:

$$f(x) = \alpha\, L_{min}^{\alpha}x^{-1-\alpha}, \quad F(x) = 1 - L_{min}^{\alpha}x^{-\alpha},$$
$$SF(x) = 1 - F(x) = L_{min}^{\alpha}x^{-\alpha} \quad (\alpha > 0, x \geq L_{min})$$

Taking logarithms of $f(L)$ and $SF(L)$ we have the log-linear functions,

$$\ln[f(x)] = \ln\left[\alpha L_{min}^{\alpha}\right] - (1+\alpha)\ln[x] \tag{4.a}$$

$$\ln[SF(x)] = \alpha\ln[L_{min}] - \alpha\ln[x] \tag{4.b}$$

Axtell (2001) fitted Eq. [4.b] to US Census binned data on firm size, concluding that the distribution of firm sizes in the USA follows a power law, with a power parameter close to one (Zipf distribution). The 1997 US Census data on the distribution of firm size that Axtell used in testing the power law is reported in the first three columns of Table 2. With binned data, Eq. [4.b] is written as,

$$\ln\left[1 - \frac{N_i}{N}\right] = c_0 - \alpha\ln[L_i] \tag{5}$$

where N is the (total) number of firms and N_i denotes the cumulative number of firms with L_i employees or fewer.

The Newey-West Least Squares, LS, estimation of [5] with the data from Table 2 gives the following results: estimated slope $\hat{\alpha} = 1.0598$ (SE = 0.0555, p value = 0.0000); estimated constant $\hat{c}_0 = 0.7125$ (SE = 0.3745, p value = 0.0862); and R-squared of 0.987. Axtell (2001) reported practically the same estimated value for the slope parameter $\alpha = 1.059$ (the estimated constant is not reported), and the same high R-squared, and from these results concluded: "the power law distribution well describes the data" (page 1819).

Is this sufficient to conclude that the distribution of firm size in the USA is a power law? The answer could be yes if we only want to test that the distribution "resembles" a power law, but not to test that the distribution is a "true" power law. First, consider the constant c_0 with informative value since it provides an estimate of the lower bound of the distribution or minimum firm size. From the regression results, $\hat{L}_{min} = \text{Exp}[\hat{c}_0/\hat{\alpha}] = 1.96$. With this minimum, the probability of observing firms with 1 employee would be close to zero. However, from Table 2, more than 20% of firms in the USA have one

employee. The introduction of the restriction that the minimum firm size is one in the estimation, implies setting the constant c_0 in model [5] equal to zero:

$$\ln\left[1 - \frac{N_i}{N}\right] = -\alpha\ln[L_i] \tag{6}$$

The estimated slope from [6] is $\hat{\alpha} = 0.9782$ (SE = 0.0342 and $R^2 = 0.979$), close to 1. With the restriction of a minimum size equal to one, the distribution of the firm size continues to "resemble" a power law.

From an econometric point of view, a very high R-squared is not sufficient for a correct model specification. The residuals show an inverted-U pattern for both models, [5] and [6], which suggests a better fit for a nonlinear model, as can be seen in Fig. 2, which also plots the residuals from the log-quadratic model [7] and from the DFS predicted from the occupational choice model, Eq. [10] below. In fact, the Ramsey RESET specification test[9] (Ramsey, 1969) rejects the null hypothesis of correct specification for models [5] and [6] with p value of 0.0005.

Furthermore, the columns in the right-hand side of Table 2 show the number of firms predicted in each bin of the size variable from a power law distribution with parameter values equal to those estimated from eqs. [5] and [6], and those predicted from a Zipf distribution, with power parameter values equal to 1 and the estimated by maximum likelihood,[10] $\hat{\alpha}_{ML} = 0.501547$. The prediction errors are substantial in all cases, although they tend to concentrate differently, depending on the estimation method (in the upper tail in the ML estimation, and in the lower tail in the other estimations). It is worth noticing that the residuals are very low (consistent with R-squared higher than 0.979) in

---

[9] The Ramsey Regression Equation Specification Error Test (RESET), Ramsey (1969), is a general specification test, which tests whether nonlinear combinations of the fitted values help explain the dependent variable. If the null hypothesis that all coefficients of non-linear combinations of the fitted values are zero is rejected, then the model suffers from misspecification (Wooldridge, 2019; Green, 2012).

[10] With binned data, the likelihood function can be approximated by $\overline{L}() \prod\limits_{i=1}^{13} [F(L_i) - F(L_{i-1})]^{n_i}$. If the firm size distribution follows a Zipf distribution, then $F(L_i) = \frac{1}{\zeta(1+\alpha)} \sum\limits_{k=1}^{L_i} k^{-1-\alpha}$. The log-likelihood function to maximize is $Log[L] = N\ln\left(\frac{1}{\zeta(1+\rho)}\right) + \sum\limits_{i=1}^{13} n_i\ln\left[\sum\limits_{k=1}^{L_i} k^{-1-\rho} - \sum\limits_{k=1}^{L_{i-1}} k^{-1-\rho}\right]$ and the optimal solution is $\hat{\rho}_{ML} = 0.501547$.

Table 2 Axtell's (2001) US Census 1997 data on firm size distribution, and predicted number of firms in size classes, from different distributions and estimated parameters

Observed[*]			Predicted values from …			
Size class		N firms	Zipf distribution		Estimated Pareto distribution	
Lower limit	Upper limit L_i	Census data N_i	$\alpha = 1$ [1]	ML estimate[2] $\widehat{\alpha}_{ML} = 0.5015$	OLS[3] $\widehat{L}_{min} = 1.96$	OLS[4] $\widehat{L}_{min} = 1$
0	1	1,026,469	2,931,388	1,850,102	0**	1578811**
2	4	1,611,601	1,241,769	1,239,623	2,559,392	2,000,707
5	13	1,342,582	431,743	732,460	1,613,776	850,202
14	40	575,228	144,664	423,438	451,632	261,591
41	121	190,236	48,250	244,140	136,146	86,393
122	364	53,513	16,084	140,720	42,012	29,175
365	1093	14,903	5361	81,107	13,063	9924
1094	3280	4909	1787	46,748	4072	3384
3281	9841	1657	596	26,944	1271	1155
9842	29,524	610	199	15,530	397	394
29,525	88,573	178	66	8951	124	135
88,574	265,720	48	22	5159	39	46
265,721	797,161	6	7	2973	12	16
797,162	1E+20	0	4	4046	5	8
	Total, N	4,821,940	4,821,940	4,821,940	4,821,940	4,821,940

[*]Axtell's (2001) US Census 1997 observed data on firm sizes. Source: http://www2.econ.iastate.edu/tesfatsi/USFirmSizesAreZipfDistributed.RAxtell2001.pdf

**Since the Pareto distribution is continuous, we consider that the number of firms with 1 employee corresponds to values of the Pareto distribution in the interval (0.5, 1.5]

[1] Predicted values for a population size $N = 4,821,940$ from a Zipf Distribution with $\alpha = 1$ (Zipf law)

[2] Predicted values from a Zipf distribution with maximum likelihood estimated parameter $\widehat{\alpha} = 0.501547$

[3] Predicted values for the same population size and Pareto distribution with estimated parameters (ordinary least squares) $\widehat{\alpha} = 1.0598$, $\widehat{L}_{min} = 1.96$

[4] Predicted values from a Pareto Distribution for the estimated parameters $\widehat{\alpha} = 0.978$, $\widehat{L}_{min} = 1$. Slope parameter estimated with robust least squared estimation method and restricted $L_{min} = 1$

the log-log models [5]–[6], where the residuals are defined as $\widehat{u}_i = \ln\left[1 - \frac{N_i}{N}\right] - (\widehat{c}_0 - \widehat{\alpha}\ln[L_i])$, while they are "large" when calculated with the original, un-transformed, values of the size variable, $\widehat{u}'_i = N_i - \widehat{N}_i$. This evidence casts doubt on the rightness of using measures of goodness of fit, i.e., R^2 values, of the log-log model specification (survival function) to test whether the distribution of firm size is a power law, or not.

The last specification test is equivalent to one of possible omitted explanatory variables. The log-linear transformation of the survival function, Eq. [5], could be generally written as a Taylor's second-order approximation of an unknown functional form, as follows.

$$\ln\left[1 - \frac{N_i}{N}\right] = c_0 - \alpha\ln[L_i] + \delta(\ln[L_i])^2 \quad (7)$$

The power law distribution is a special case of [7] with $\delta = 0$. If the estimated value of δ is different from zero, the condition from the power law that the logarithm of the survival function is a linear function of the size variable in logarithms, would be violated. The fit of [7] to the US Census data gives estimated values of $\widehat{a} = 0.680$ (SE = 0.0715) and $\widehat{\delta} = -0.030$ (SE = 0.00584 and p value = 0.0006). Therefore, the hypothesis of $\delta = 0$ is

Fig. 2 Residual Plot. Residuals $\widehat{u}_i = \ln[1-N_i/N] - \widehat{\ln}\left[1-\widehat{F}(L_i)\right]$ from the estimation of the (transformed) logarithms of survival Functions, Eqs. [5], [6], [7], and [10]. Horizontal axis: firm size measured by the logarithm of the number of employees. Vertical axis: Values of the residuals $\widehat{u}_i = y_i - \widehat{y}_i$, where $y_i = \ln[1 - N_i/N]$ and the fitted values $\widehat{y}_i$ from the estimations of Eqs. [5], [6], [7], and [10] are shown in the Notes of Fig. 3
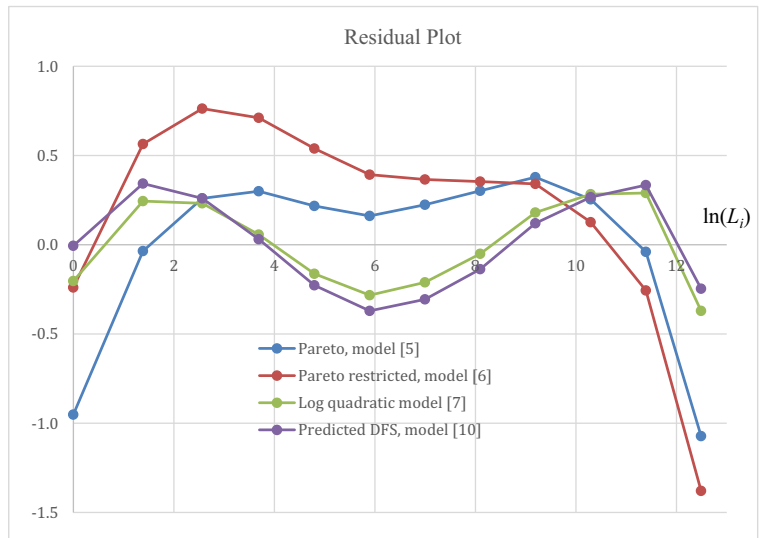
rejected at a high level of significance (higher than 99.9%). The hypothesis of log-linearity between the value of the survival function and the value of the size variable is rejected. Consequently, the specification tests do not support the conclusion that the data on the distribution of firm size in the USA follows a power law.

5.2 Estimation of firm size distribution predicted by the model

We now estimate the distribution of firm size predicted by the occupational choice model and lognormal distribution of entrepreneurial skills in the population. From Section 2, the probability density function of the distribution of firm size is given by:

$$f_S(x) = \frac{1}{1 - \frac{1}{2} Erfc\left[\frac{\mu - \log[e^*]}{\sigma\sqrt{2}}\right]} \frac{1}{bx} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{Exp}\left[\frac{-1}{2\sigma^2}\left(\frac{1}{b}\log\left[\frac{x}{a}\right] - \mu\right)^2\right]$$

With probability cumulative function,

$$F_S(x) = \frac{Erfc\left[\frac{b\mu + \ln(a) - \ln(x)}{b\sigma\sqrt{2}}\right] - Erfc\left[\frac{\mu - \ln(e^*)}{\sigma\sqrt{2}}\right]}{2 - Erfc\left[\frac{\mu - \ln(e^*)}{\sigma\sqrt{2}}\right]}$$

where $Erfc[z] = 1 - \frac{2}{\sqrt{\pi}}\int_0^z \mathrm{Exp}[-t^2]\,dt$. Or, equivalently, as

$$F_S(x) = \frac{1}{2 - 2c_3}\left(Erfc\left[\frac{c_1 - \ln(x)}{c_2}\right] - 2c_3\right) \tag{8}$$

where $c_1 = b\mu + \log[a]$, $\qquad c_2 = b\sigma\sqrt{2}$, $\qquad$ and $c_3 = \frac{1}{2} Erfc\left[\frac{\mu - \log[e^*]}{\sigma\sqrt{2}}\right]$.

The parameters (c_1, c_2, c_3) of the cumulative probability distribution F_s depend on the values of the primitive parameters of the model (a, b, μ, σ, e^*). For example, knowing the parameters of the distribution of skills (μ, σ), the other parameters of the model could be calculated as follows: $e^* = Exp\left(\mu - \sqrt{2}\sigma Erfc^{-1}[2c_3]\right)$, $b = \frac{c_2}{\sqrt{2}\sigma}$, and $a = Exp(c_1 - b\mu)$.

For the estimation of the parameters of the cumulative probability distribution from the binned data on firm sizes in the USA (third column of Table 2), Eq. [8] is written as,

$$\frac{N_i}{N} = \frac{1}{2 - 2c_3}\left(Erfc\left[\frac{c_1 - \ln(L_i)}{c_2}\right] - 2c_3\right) \tag{9}$$

Or in logarithms of the survival function, $1 - F_s$,

$$\ln\left[1 - \frac{N_i}{N}\right] = \ln\left[\frac{1}{2 - 2c_3}\left(2 - Erfc\left[\frac{c_1 - \ln(L_i)}{c_2}\right]\right)\right] \tag{10}$$

where, as before, N is the (total) number of firms and N_i denotes the accumulated number of firms with L_i employees, or fewer.

We estimate the parameters (c_1, c_2, c_3) of the predicted-by-the-model distribution from the

Table 3 Estimations of the distributions of firm sizes (from Axtell's US Census 1997 data): predicted by the model (under lognormal skills) versus Pareto DFS

Estimation of the (untransformed) cumulative probability function. Non-linear least squares.					
	SR	R^2	c_1	c_2	c_3
Predicted DFS [9] restricted[(2)] (5% managers)	0,007495	0,9891	−7529	6099	0,950
	SR	R^2	L_{min}	α	
Pareto DFS[(3)]	0,012034	0,9825	0,6896	0,5637	
Estimation of the (transformed) logarithm of the survival function. Non-linear least squares.					
	SR	R^2	c_1	c_2	c_3
Predicted DFS [10] restricted[(4)] (5% managers)	0,7451	0,9963	−6290	5067	0,950
	SR	R^2	L_{min}	α	
Pareto DFS, model [5]	2,639,034	0,9871	19,587	10,598	

[(1)] Predicted DFS refers to the distribution of firm sizes obtained from our theoretical model, Claim 1, under the assumptions that the distribution of skills in the population is Lognormal(μ, σ) and that the labor demand function is a power function of skills, $L(e; w) = a\,e^b$

[(2)] The corresponding endogenous parameters are $a = 2.32E{-}13$, $b = 10.782$, $\mu = 2\,\sigma$, $= 0.4$, $e^* = 14.267$

[(3)] We directly fit the cumulative distribution function of a Pareto: $N_i/N = 1 - (L_{min}/L_i)^\alpha$

[(4)] The corresponding endogenous parameters are $a = 3.1E{-}11$, $b = 8.95687$, $\mu = 2\,\sigma$, $= 0.4$, $e^* = 14.267$

untransformed model, Eq. [9], and from the log-log model, Eq. [10], for the purpose of illustrating the relevance of the logarithmic transformation for the distribution of the values of the residuals.[11] Since both models, [9] and [10] are non-linear, we estimate them by nonlinear least squares (NLS). Since the proportion of firms with employees in the sample data is 5%, we introduce this restriction, c_3 = 1 − 0.05 = 0.95 in the estimation of the other two parameters. The restriction has two purposes. One, to take advantage of information available from sources external to the model, and the other to estimate the model with the same degrees of freedom as when estimating the parameters, two, of the power law distribution, thus making the results more comparable.

To compare with the results from model [9], it is interesting to fit the data to the (untransformed) distribution function of the Pareto distribution

$$\frac{N_i}{N} = 1 - \left(\frac{L_{min}}{L_i}\right)^{-\alpha} + u_i \tag{11}$$

The NLS estimates of α and $L_{\min}$ are $\widehat{\alpha} = 0.5637$ and $\widehat{L}_{\min} = 0.6896$.

The results of these estimations using non-linear least square estimation methods are presented in Table 3, together with the estimated values of the parameters of the power law, again with both untransformed (model [11]) and transformed (model [5]) values of the variables.

Figure 3 shows the plot of observed values of the survival function,[12] $\ln[1 - N_i/N]$, and fitted values of the four logarithmic models analyzed: the log-linear models from the Pareto distribution, [5] and [6], the log (nonlinear) quadratic model, [7], and the logarithm of the survival function of the DFS predicted from the occupational choice model, Eq. [10]. The superior fit of the logarithmic nonlinear functional form, [7] and [10], is evident.

For the distribution of firm sizes predicted by the occupational choice model and a lognormal distribution of skill, the estimated values of the parameters in the estimation with untransformed variables and in the estimation with log-transformed values of the variables (survival function) are very similar. On the contrary, when we fit the data to a Pareto distribution of firm sizes, as if the power law were the true distribution of the size variable, the estimated values of the parameter vary between the two estimations: estimated parameter α

---

[11] We minimize the sum of squares of residuals in the "original" variable, number of firms, $\widehat{u}_i = (N_i/N) - \widehat{F}_s(L_i) = \frac{N_i}{N} - \frac{1}{2-2\widehat{c}_3}\left(Erfc\left[\frac{\widehat{c}-\ln(L_i)}{\widehat{c}_2}\right] - 2\widehat{c}_3\right)$. This contrasts with the approach commonly followed, where the residuals are defined in the log-log model of the survival function, $\widehat{u}_i = \ln[1 - N_i/N] - \widehat{\ln}\left[1 - \widehat{F}_S(L_i)\right]$.

[12] The plot of Fig. 3 is referred to as the Zipf plot. See Stanley et al. (1995) for an early application to the distribution of firm size.
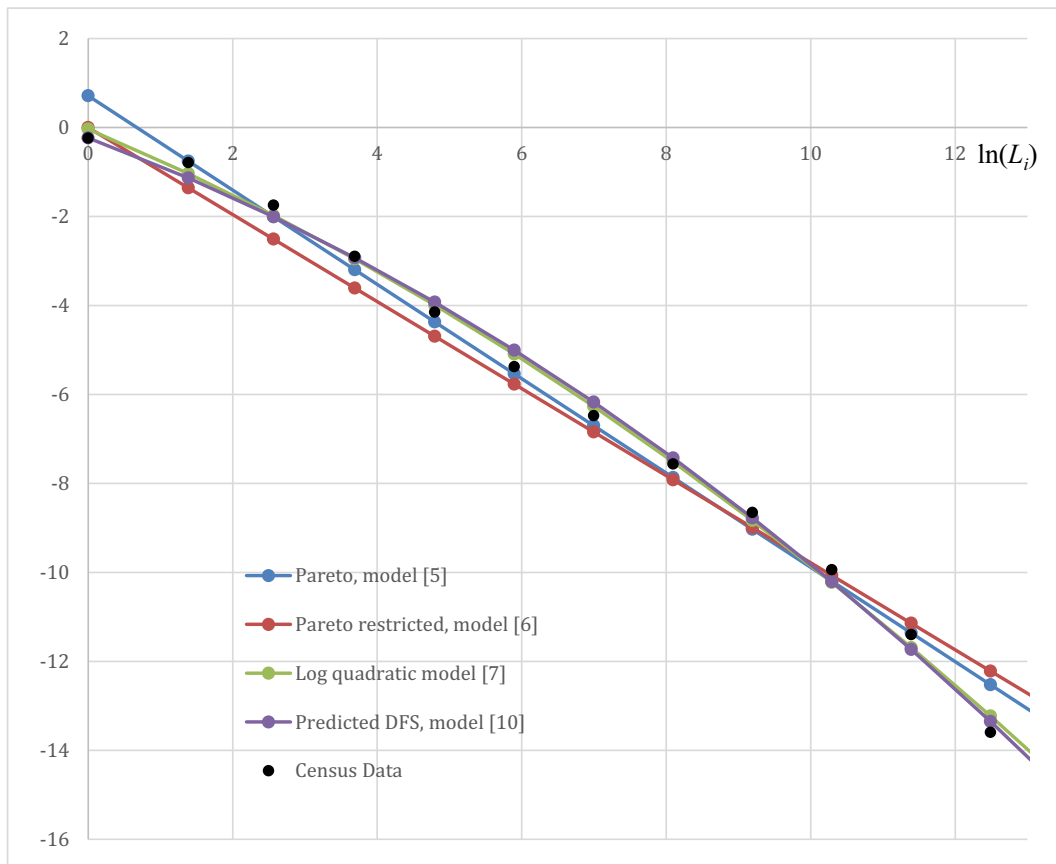
Fig. 3 Observed data on firm size distribution (Axtell's 2001 US Census 1997) and fitted values from the Logarithms of the Survival Functions, Eqs. [5], [6], [7], and [10]. Vertical axis: Observed CENSUS data $y_i = \ln[1 - N_i/N]$ and fitted values of the Logarithm of the Survival Functions, $\widehat{y}_i = \widehat{\ln}[1 - N_i/N] = \widehat{\ln}\left[\widehat{SF}(L_i)\right]$, from the estimations of Eqs. [5], [6], [7], and [10]. Fitted values from model [5]: $\widehat{y}_i = \widehat{c}_0 - \widehat{\alpha}\ln[L_i]$, with $\widehat{\alpha} =$ 1.0598, $\widehat{c}_0 =$ 0.7125. Fitted values from model [6]: $\widehat{y}_i = -\widehat{\alpha}\ln[L_i]$, with $\widehat{\alpha} =$ 0.9782. Fitted values from model [7]: $\widehat{y}_i = \widehat{c}_0 - \widehat{\alpha}\ln[L_i] + \widehat{\delta}(\ln[L_i])^2$ with $\widehat{c}_0 = -0.0359$, $\widehat{\alpha} = 0.6801$, and $\widehat{\delta} = -0.0301$. Fitted values from model [10]: $\widehat{y}_i = \ln\left[1 - \frac{1}{2(1-0.95)}\left(2 + Erfc\left[\frac{\widehat{c}_1 - \ln(L_i)}{\widehat{c}_2}\right]\right)\right]$, with $\widehat{c}_1 = -6.290$, $\widehat{c}_2 = 5.067$

equal 0.5637, with untransformed values of the variables, and equal to 1.0598 with the transformed ones.

Table 4 shows the fitted number of firms for each size class, obtained from the estimations reported in Table 3, which correspond to the DFS predicted by the occupational choice model, Eqs. [9] and [10], and to the DFS when assuming it follows a power law, Eqs. [5] and [11]. Table 4 also shows the 1997 US Census data on firm sizes used by Axtell (2001). The fitted values obtained from the DFS predicted by the occupational choice model are closer to the observed data than those obtained from the estimated power laws, in both cases: when directly estimating the cumulative distribution function, and when estimating the logarithm of the survival function.

Figure 4 displays the observed values of the cumulative number of firms and values of the fitted (untransformed) CDFs corresponding to the Pareto distribution, Eqs. [11], and the DFS predicted by the occupational choice model, Eq. [9]. While Fig. 3 shows the fitted values for logarithms of the survival function, Fig. 4 displays the fitted values from directly fitting the data to the cumulative distribution functions.

Finally, Fig. 5 depicts the residual plot of models [9] and [11]. Both, Table 4 and Fig. 5, show that the prediction errors from the estimated theoretical model are smaller in each and all size classes than the errors from the estimations of a power law.

Although these results do not prove that the "true" distribution of skills in the population is lognormal, they

Table 4 Axtell's (2001) US Census 1997 data on firm size distribution and predicted number of firms in size classes, from different distributions and estimated parameters

Size class	1997 census data*	Predicted values from …			
		Estimated (untransformed) cumulative distribution function		Estimated (transformed) logarithm of the survival function	
Upper limit L_i	Numbers of firms, n_i	Predicted DFS[1], Eq. [9]	Pareto DFS[2] Eq. [11]	Predicted DFS[3], Eq. [10]	Pareto DFS[4], model [5]
1	1,026,469	922,343	911,428	1,005,439	0*
4	1,611,601	2,032,276	2,120,530	2,266,412	2,559,415
13	1,342,582	938,438	868,903	901,424	1,613,742
40	575,228	480,667	432,266	390,772	451,635
121	190,236	242,425	226,901	162,743	136,150
364	53,513	116,501	121,137	62,850	42,014
1093	14,903	52,779	65,031	22,247	13,065
3280	4909	22,459	34,976	7189	4073
9841	1657	8965	18,823	2118	1271
29,524	610	3356	10,132	568	397
88,573	178	1178	5454	139	124
265,720	48	388	2936	31	39
797,161	6	120	1580	6	12
1,00E+11	0	47	1843	1	5
Total, N	4,821,940	4,821,940	4,821,940	4,821,940	4,821,940

*Source: http://www2.econ.iastate.edu/tesfatsi/USFirmSizesAreZipfDistributed.RAxtell2001.pdf

[1] Predicted values from the estimated cumulative distribution function predicted by the occupational choice model, assuming a lognormal distribution of skill (Eq. [9]), and with the restriction that the number of entrepreneurs-firms with employees is 5% of the working population ($c_3 = 0.95$). The estimated coefficients by Non-Linear Least Squares are $\hat{c}_1 = -7.529$, $\hat{c}_2 = 6.099$

[2] Predicted values from the estimated cumulative distribution function assuming a Pareto distribution of firm sizes, Eq. [11]. The estimated coefficients, by non-linear least squares are $\hat{\alpha} = 0.5637$, $\hat{L}_{min} = 0.690$

[3] Predicted values from the estimated logarithm of the Survival Function predicted by the occupational choice model, assuming a lognormal distribution of skill (Eq. [10]), and with the restriction that the number of entrepreneurs-firms with employees is 5% of the working population ($c_3 = 0.95$). The estimated coefficients by Non-Linear Least Squares are $\hat{c}_1 = -6.290$, $\hat{c}_2 = 5.067$

[4] Predicted values from the estimated logarithm of the survival function assuming a Pareto distribution of firm sizes (Eq. [5]), with estimated parameters (ordinary least squares) $\hat{\alpha} = 1.0598$, $\hat{L}_{min} = 1.96$

**Zero predicted firms with size lower than 2

do show that (i) the distribution of firm size predicted by the occupational choice model, assuming a lognormal distribution of skill, fits the data better than a power law; and (ii) the fact that a visual observation of the empirical distribution of values of the outcome variable, in this case the size of the firm, suggests that the distribution "resembles" a power law, is not sufficient for the actual distribution being a power law. Having a theoretical model of what determines the distribution of the variable

firm size helps to discern when the apparent and the true distributions will coincide.

To sum up, the fact that the log-quadratic model [7] and DFS predicted by the occupational choice model, Eqs. [9] and [10], fit the data significantly better than the Pareto distribution, together with the results of the Ramsey RESET tests, the observation of the (non-random) residuals' plots, and the statistical significance of the quadratic term in model [7], are all evidence that
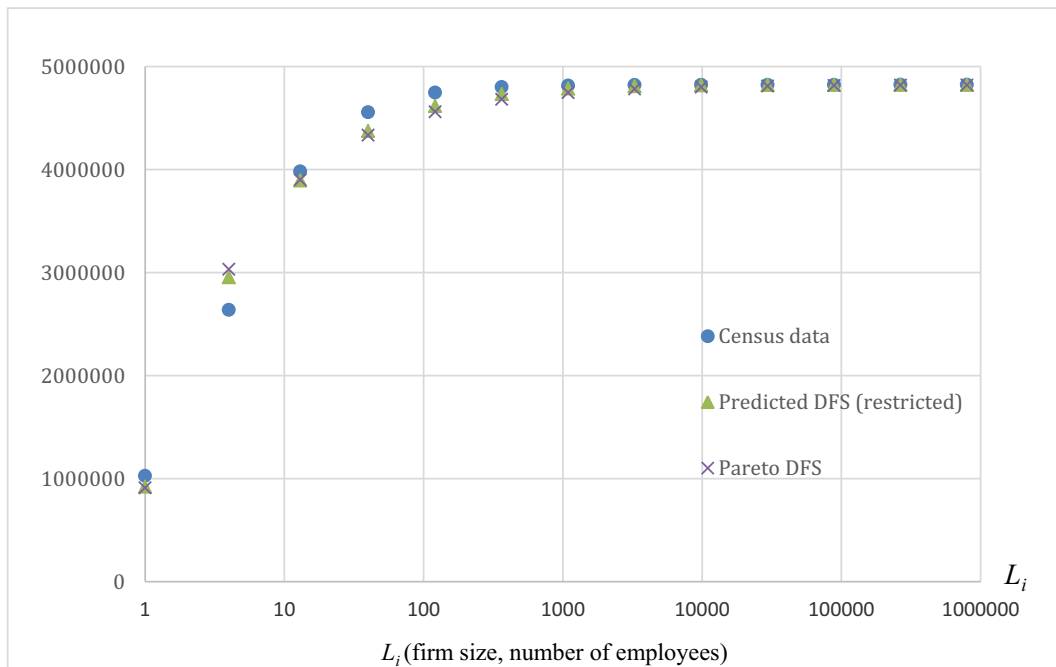
Fig. 4 Observed data, N_i, on firm size distribution (Axtell's 2001 US Census 1997) and fitted values, $\widehat{N}_i$, from the *untransformed* Cumulative Distribution Functions, Eqs. [9] and [11]. Vertical axis: observed, N_i, and fitted values, $\widehat{N}_i = N\widehat{y}_i = N\,\widehat{F}(L_i)$, of the cumulative number of firms with L_i employees or less. $N =$ 4,821,940. Fitted values from estimation of Eq. [9]: $\widehat{N}_i = N\widehat{y}_i$ with $\widehat{y}_i = \frac{1}{2(1-0.95)}\left(2 + Erfc\left[\frac{-7.5285-\ln(L_i)}{6.099}\right]\right)$. Fitted values from estimation of Eq. [11]: $\widehat{N}_i = N\widehat{y}_i$ with $\widehat{y}_i = 1-\left(\frac{\widehat{L_{min}}}{L_i}\right)^{-\widehat{\alpha}}$, $\widehat{\alpha} =$ 0.564 and $\widehat{L}_{min} = 0.690$. Horizontal axis: number of employees, L_i

question Axtell's conclusion that the distribution of firm size in the USA is a power law.

6 Conclusion and implications

The occupational choice theory has been, surprisingly, missing from explanations of the so-called "pervasiveness" of power law distributions among organizational and entrepreneurial variables, including the distribution of firm size, even though the pioneering paper of Lucas (1978) explicitly models the distribution of firm size. Occupational choice models go beyond the purely stochastic or chance explanation of the distribution of the organizational variables and restore the value added of management theory to explain differences in size and performance across firms. The paper establishes a correspondence between the distribution of skill in the subset of entrepreneurs and the distribution of firm size. The distribution of firm size turns out to

be a truncated non-linear transformation of the distribution of skills (convex power transformation in our case). In fact, in the context of the occupational choice model, the paper provides conditions for the distribution of skills and for the production technology, under which the distribution of firm size in the market equilibrium will be a "true" power law (a result that could be extended to heterogeneity in input variables other than skill).

6.1 Implications for theory

The explanation of the distribution of firm size and other organizational/entrepreneurial variables as a market equilibrium from competing profit-maximizing entrepreneurs, rather than more or less complex processes of stochastic growth, will shift the researcher's attention towards characteristics of these teams (their production technology, internal organization, relation to external markets, particularly financial markets), and to modeling the input of the entrepreneur-manager in the output
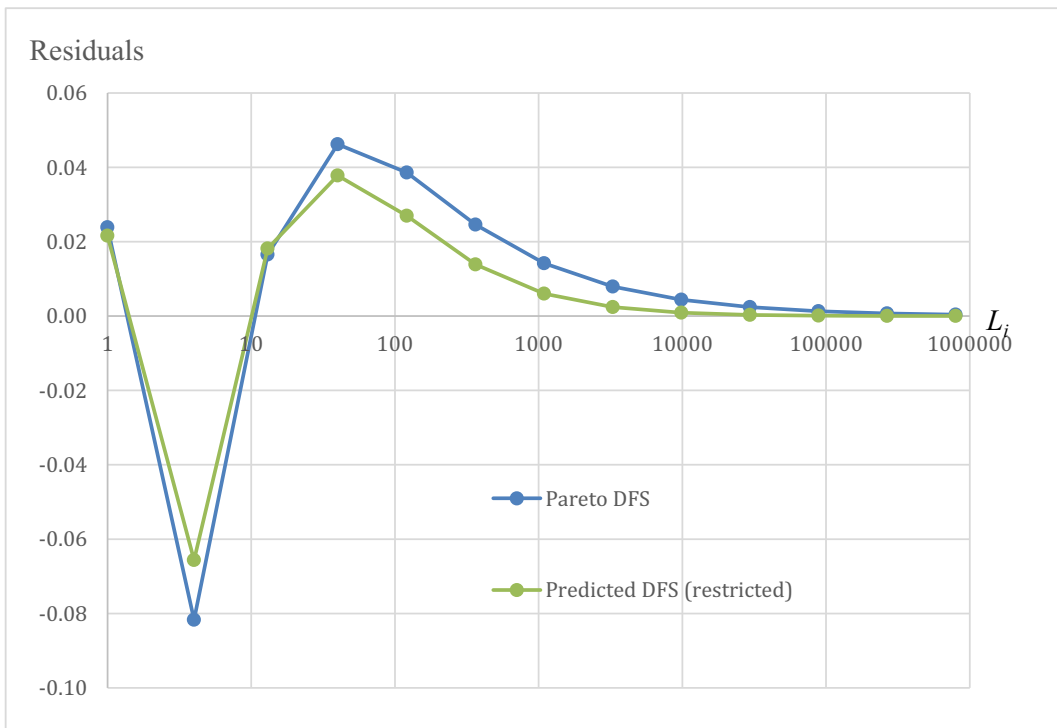
Fig. 5 Residual Plot. Residuals $\widehat{u}_i = (N_i/N) - \widehat{F}(L_i)$ from the estimation (by non-linear least squares) of the untransformed cumulative distribution functions, Eqs. [9] and [11]. Horizontal axis: firm size measured by the number of employees. Vertical axis: Values of the residuals $\widehat{u}_i = y_i - \widehat{y}_i$, where $y_i = N_i/N$ and the fitted values $\widehat{y}_i$ from the estimations of Eqs. [9], and [11] are shown in the Notes of Fig. 4

of the team, when explaining observed regularities in the distribution of such variables. So far, such regularities have led to a recognition from the outset that the "social world seems to be organized according to power law distributions" (Crawford et al., 2015: 705 and references therein) and, from this result, the research focus has narrowed to the investigation of causal processes from complexity science that can generate such power law distributions (Andriani and McKelvey, 2009). Occupational choice models then provide microeconomic team production explanations of power law-like distribution of organization variables such as firm size, as requested by Axtell (2001, 2006), that complement other approaches, where the micro-organization of production is treated as a "black box."

In the occupational choice models of Lucas (1978) and Rosen (1982), from which we draw in this paper, the power law-like distribution observed in entrepreneurial output variables such as size and profit of firms, and therefore the difference between low performing and extremely high performing firms, are very much related to the way the entrepreneurial input enters into

the production function, from the job to the firm levels. In the market equilibrium, there is a lower bound in skill that determines the size and profit of the smaller firm with employees and a continuum of firm sizes and profits as a function of the skills of the respective entrepreneur-managers. Therefore, occupational choice models can explain the whole range of size and performance of firms, a feature of the theories of entrepreneurship that are particularly valued (Crawford et al., 2015). Two forces are in place: one from scale economies of skills that push towards concentration of more resources under the direction of fewer higher-skilled entrepreneurs; the other, from organizational size diseconomies (a la Coase and Penrose) that limit the growth of the firm and push towards dispersion of production in more numerous and smaller firms. Entrepreneurship theories should then further investigate these two forces and the other factors that intervene in the final outcome mediating effects (shape of the distribution of skills, parameters of the production technology, cost of capital, neutral technological progress).

6.2 Implications for empirical research

The theoretical findings have implications for empirical analysis that examine the pervasiveness of the power law distribution across economic variables, and in particular for research on the distribution of firm size. For example, Clauset et al. (2009) empirically find that the hypothesis of a power law distribution cannot be rejected for values of the respective variable above an empirically determined minimum value. A purely empirical analysis cannot determine the relevance of this result because there is no theory about what the minimum value of the variable *should* be. In the occupational choice model, the market equilibrium determines the size of the smallest firm and this result should be accounted for when testing for power law distributions. The same argument applies when testing whether the size distribution predicted from the model is rejected, or not, by the empirical data.

Theoretical and empirical advances to distinguish between distributions that "resemble" power law and "true" power law are relevant for conducting empirical tests of proposed theories, particularly if the results can have policy implications. For example, a power law distribution of firm size implies that the firm size elasticity of the survival function is constant, while the model presented in the paper predicts that the absolute value of the elasticity will be monotonically increasing with size. Size-conditioned public policies for firms can be different in one case and in the other.

The paper highlights the difference between comparing the goodness of fit to a dataset of a normal distribution and of a power law distribution (as in Clauset et al., 2009 and Joo et al., 2017), and testing the hypothesis that the distribution is a true power law. In the latter case, the necessary and sufficient condition for the distribution being a power law is that the log of the survival function is a linear function of the logarithm of the value of the variable. Therefore, the strong empirical test of a power law should be formulated and implemented as a model specification test, in line with the test of Eq. [7] above. The text also points out the limitations of using the R^2, goodness of fit, as a criterion to test for a power law, as well as the precautions that should be taken when working with binned data. The supplementary on-line document discusses in more detail some methodological issues that arise in the empirical estimation of the distribution of firm size that could be extended to other economic variables.

Having a theoretical model of team production and management from which to make predictions about the endogenous distribution of relevant organizational variables, for example firm size, allows for enriching interactions between the theory and the evidence in empirical research. For example, the parameter of the explanatory variable of the log-linear transformation of the survival function can be directly related to values of the parameters of the model, particularly parameters of the production technology and the internal organization of firms. To make economic sense, the values of these parameters must be within certain bounds, something that can be tested as a way of strengthening the conclusions from the empirical research. Comparative static analysis around the market equilibrium of changes in the distribution of size resulting from changes in the parameters of the model can help to explain differences in the distribution of size of firm across countries and/or changes in the distribution in one country, over time. At a higher level, the model may predict a distribution of the endogenous values of the relevant organization variable that can be directly tested in the empirical analysis, as we do in this paper. Once again, the theory will impose restrictions on the empirical distribution, for example on the choice of the lower bound of the values of the random variable (minimum size of the firm predicted by the model and observed minimum size, in our case).

6.3 Implications for management and policy

The occupational choice models restore the value of the management input in determining the size and performance of firms. The distribution of entrepreneurial skill in the population will be reasonably given in the midterm. However, the model identifies another variable that managers control and that can alter the market equilibrium from occupational choices, namely the diseconomies of organizational size. In the model, these diseconomies have to do with the intensity of supervision of employees by the entrepreneur at the job level (Rosen, 1982). More generally, organizational size diseconomies are related to the limits to the growth-size of the firm that Penrose (1959) attributed to the (fixed) management input, to management costs when the entrepreneur takes the place of the price system in the direction of resources (Coase, 1937), and to costs from "loss of control" in hierarchical organizations. Managers must be aware that the internal organization of firms, for

example more delegation of decision power and less intensity of supervision in environments of higher goal congruence, will mitigate the organizational size diseconomies and will affect the equilibrium distribution of firm size, with more output concentrated in fewer large firms. The model is stylized on the details about the internal organization of firms (for example, the number of hierarchical levels and the role of intermediate managers) but it is sufficient to highlight the importance—for the performance of the firm—of organizational design variables, jointly with strategic ones.

In the market equilibrium from occupational choices, where entrepreneurs compete for control of production resources, the total output produced is maximized, within the total resource constraints. Policy makers should then be concerned about the actual working of the market for entrepreneur-managers, including the market for corporate control, and the financial markets that provide funds to talented but non-wealthy would-be entrepreneurs. Concentration forces in markets, driven by efficiency factors, i.e., more efficient firms, and firms managed by more skilled entrepreneurs concentrating a high volume of resources, with the corresponding concentration of power and profits, may raise policy concerns in terms of limits to competition and/or increasing income inequality. In other circumstances, the policy concerns may focus on the other side of the income distribution and public authorities may make policy decisions regulating labor markets, for example with the introduction of minimum wages. Caution should be taken, however, so that regulations and government policies do not create opportunities for rent seeking that attract talent that, in the absence of these opportunities, would be allocated to productive entrepreneurial tasks (Murphy et al. 1991). Then, a better understanding of the efficiency-driven determinants of the distribution of firm size and the possible distributional consequences of the free-market equilibrium results, as well as the potential opportunity costs of diverting talented individuals from productive entrepreneurial tasks, will be relevant for better-informed public policies that must consider the tradeoff between economic efficiency and other social goals.

6.4 Future research

The occupational choice model as formulated in this paper is highly stylized, static, and based on strong assumptions, mainly because we want to produce some

robust predictions on the results. Moreover, the occupational choice model is not the only way to open the "black box" of microeconomics that can explain the regularities observed in the distribution of important organization variables, such as the size and performance of firms. Therefore, from the perspective of this paper, future research should be directed mainly towards generalizing the assumptions of the model and testing for the robustness of old and new predictions. In this respect, the proposed occupational choice model should not be viewed as an alternative to complexity science in the explanation of regularities observed in important organization variables, but as a complementary one. Incorporating complex stochastic dynamics from out-of-equilibrium situations in extensions of the static occupational model, including the possibility of changes in the distribution of skill over time from rational individual decisions on investment in human capital, could be a very fruitful area of research. For example, Knudsen et al. (2017) model the dynamics of convergence to the Cournot-Nash target in an oligopoly market with frictions; the same set-up could be replicated to model the convergence to the size of a firm in the occupational equilibrium and/or convergence to the size distribution equilibrium.

In addition to being static, and only looking at market equilibrium solutions, i.e., ignoring out-of-equilibrium results, another limitation of the paper is that all heterogeneity among individuals that determines the occupational choice is summarized in the single variable of entrepreneurial skill. We outline in the text how, conceptually, the paper could be generalized to accommodate situations with multiple sources of heterogeneity; more particularly, the analysis could be generalized to differentiate between the endowment of individuals of skill to perform operational tasks, and skill to perform entrepreneurial tasks, in line with Jovanovic (1994). Fuchs-Schündeln and Schündeln (2005) find supporting evidence that risk aversion is an important variable in explaining the occupational choice of working as employee or as self-employed, in line with the theoretical prediction of Kihlstrom and Laffont (1979). Thus, the model could be extended to include heterogeneity in skill and in risk aversion in the same theoretical model. In a similar vein, the analysis could be generalized to solve for the market equilibrium when the performance of entrepreneurs improves with more balanced, multiple skills, rather

than a single one (Lazear, 2004). Another realistic assumption in future generalizations of the results could be to allow for the possibility that individuals "learn" about their entrepreneurial skills from trial and error and from noisy signals of performance from firms that they establish, as in Jovanovic (1982), Hopenhayn (1992), and Melitz (2003), with the extension to open economies. The extended model would expand the outcome from entrepreneurial decisions to entry and exit of firms, and to entrepreneurs that fail or succeed.

Appendix 1. Proofs of the claims

Lemma 1. If x is a random variable with support $[x_m, +\infty)$, then there exists an $x^* > 0$ such that the probability density function (PDF) is strictly decreasing and convex for all $x > x^*$.

Proof. The demonstration is made by *reductio ad absurdum* (reduction to absurdity). Consider a random x variable with support $[x_m, +\infty)$ (x_m may be any real number or $-\infty$).

The probability density function satisfies that $\int_{-\infty}^{+\infty} f(x)dx = 1, f(x) \geq 0 \; \forall x$. If $f(x)$ is not decreasing in the right tail (and the support is $[x_m, +\infty)$), then there exist $\delta > 0$ and $x^* > 0$, such that $f(x) > \delta$ for all $x > x^*$, but then $\int_{x^*}^{+\infty} f(x)dx > \int_{x^*}^{+\infty} \delta dx = +\infty$, which is absurd. Consequently, $f(x)$ is decreasing in the right tail; more precisely, there exists x^* such that $f(x)$ is strictly decreasing for all $x > x^*$. Moreover, if $f(x)$ is concave for all $x > x^*$, then there is a value x^{**} such that $f(x^{**}) = 0$, which is absurd, since we are assuming that $f(x) > 0$ for all $x \in [x_m, +\infty)$. Therefore, $f(x)$ must be strictly decreasing and convex in the right tail. ∎

Proof of claim 1 In equilibrium, the individuals who become entrepreneurs are those with skill levels higher or equal to e^*, so that the distribution of skills in the group of entrepreneurs is just the left-truncated distribution of skills in the population. Therefore, its support is $[e^*, e_M]$ for e_M finite, and $[e^*, +\infty)$ otherwise; and its probability cumulative function and density probability function are characterized by.

$$F_{ent\,skills}(e) = \frac{F(e)-F(e^*)}{F(e_M)-F(e^*)}, \quad f_{ent\,skills}(e) = \frac{f(e)}{F(e_M)-F(e^*)} \quad (\forall e \geq e^*)$$
(*).

since, for any continuous random variable X with support $[x_m, x_M]$ and probability density function $g(x)$ and cumulative distribution function $G(x)$, the truncated distribution of X for the values in $[a, b]$ ($x_m \leq a < b \leq x_M$) is characterized by $g_{truncated}(x) = \frac{g(x)}{G(b)-G(a)}$ and $G_{truncated}(x) = \frac{G(x)-G(a)}{G(b)-G(a)}$ (for any x such that $a \leq x \leq b$).

On the other hand, Firm size S, measured by number of employees, is a function of the entrepreneur's skills e, $S = h(e; w^*)$, for $e \geq e^*$; so that the distribution of firm size is just the transformation (by "h") of the distribution of the entrepreneurs' skills. And, for any continuous random variable X with probability density function $f_X(x)$ and cumulative distribution function $F_X(x)$, and $Y = h(X)$, where h is invertible (and monotonous and differentiable), the probability cumulative function and the density probability function of Y are characterized by $F_Y(y) = F_X(h^{-1}(y))$ and $f_Y(y) = f_X(h^{-1}(y))\frac{dh^{-1}(y)}{dy}$. By applying these expressions to the distribution of entrepreneurs' skills, expression (*) above, and $S = h(e; w^*)$, we directly obtain that the probability cumulative function and PDF of S are characterized by

$$F_S(s) = \frac{F(h^{-1}(s))-F(e^*)}{F(e_M)-F(e^*)},$$

$$f_S(s) = \frac{f(h^{-1}(s))}{F(e_M)-F(e^*)}\frac{dh^{-1}(s)}{ds} \quad \forall s \geq S_{min}$$

Where $S_{min} = h(e^*; w^*)$. ∎

Proof of claim 2. On the one hand, for reasonable production functions, the (individual) labor demand will be increasing and convex in the entrepreneur's skill, $L = h(e; w), \frac{\partial h}{\partial e} > 0, \frac{\partial^2 h}{\partial e^2} > 0$, so that its inverse, $h^{-1}()$, will be strictly increasing and concave, $\frac{dh^{-1}(s)}{ds} > 0$ and $\frac{d^2h^{-1}(s)}{ds^2} < 0$. On the other hand, for reasonable

distributions of skills, with support $[e_m, +\infty)$, there exists e_1 such that $f'(x) < 0$ and $f''(x) > 0$, for $x \geq e_1 \geq e_m$. So that, by assumption,

$$\frac{dh^{-1}(s)}{ds} > 0, \frac{d^2h^{-1}(s)}{ds^2} < 0, f'(x) < 0, f''(x) > 0. \ (*)$$

From [2], the probability density function of the distribution of firm size is $f_S(s) = \frac{f(h^{-1}(s))}{F(e_M)-F(e^*)}\frac{dh^{-1}(s)}{ds}$ and its first derivative is

$$\frac{df_S(s)}{ds} = \frac{1}{F(e_M)-F(e^*)}\left[f'(L^{-1}(s))\frac{dL^{-1}(s)}{ds}\frac{dL^{-1}(s)}{ds} + f(L^{-1}(s))\frac{d^2L^{-1}(s)}{ds^2}\right]$$

It directly follows, from the assumptions $(*)$, that $f'(h^{-1}(s)) < 0$, $\frac{dh^{-1}(s)}{ds} > 0$, $f(h^{-1}(s)) > 0$ $\frac{d^2h^{-1}(s)}{ds^2} < 0$, and then both summands in the above expression of $\frac{df_S(s)}{ds}$ are negative; i.e. $f_S(s)$ is strictly decreasing.

On the other hand, the second derivative is given by

$$\beta\theta^{\frac{1}{\beta}}\left(\frac{1-\beta}{w^*}\right)^{\frac{(1-\beta)}{\beta}}(e^*)^{\frac{(\beta+\tau)}{\beta}} = w^*,$$

$$\int_{e_m}^{e^*} dF(e) = \left(\frac{\theta(1-\beta)}{w^*}\right)^{\frac{1}{\beta}}\int_{e^*}^{e_m} e^{(\beta+\tau)/\beta}dF(e)$$

$$\frac{d^2f_S(s)}{ds^2} = \frac{1}{F(e_M)-F(e^*)}\left[f\}(h^{-1}(s))\left(\frac{dh^{-1}(s)}{ds}\right)^3 + \right.$$
$$+ 2f'(h^{-1}(s))\left(\frac{dh^{-1}(s)}{ds}\right)\frac{d^2h^{-1}(s)}{ds^2} +$$
$$+ f'(h^{-1}(s))\frac{dh^{-1}(s)}{ds}\frac{d^2h^{-1}(s)}{ds^2} +$$
$$\left. + f(h^{-1}(s))\frac{d^3h^{-1}(s)}{ds^3}\right]$$

From the assumptions, expression $(*)$ above in this proof, it is obvious that the three first summands in the above expression are positive. The fourth one is also positive for labor demands of power function (or polynomial) type. Consequently, $f_S''(s) > 0$. ∎

Proof of claim 3. If the individual labor demand is a power function, $h(e; w) = a_1\ e^{b_1}$, then

$$h^{-1}(s) = (s/a_1)^{1/b_1} \text{ and } \frac{dh^{-1}(s)}{ds} = \frac{1}{b_1a_1}\left(\frac{s}{a_1}\right)^{(1-b_1)/b_1}.$$

Sufficient condition: (if $f(x)$ is a power law, then $f_S(x)$ is also a power law).

On the one hand, if the distribution of skills follows a power law, then its probability density function may be

written as $f(x) = Cx^{-1-\alpha}$ (where $e_M = +\infty$, $C = \alpha e_m^{\alpha}$, $F(e_M) - F(e^*) = e_m^{\alpha}(e^*)^{-\alpha}$ and $F(e^*) = -e_m^{\alpha}(e^*)^{-\alpha}$), and from [3], the probability density function of the distribution of size is given by

$$f_S(s) = \frac{f(h^{-1}(s))}{F(e_M)-F(e^*)}\frac{dh^{-1}(s)}{ds}$$

$$= \frac{C\left((s/a_1)^{1/b_1}\right)^{-1-\alpha}}{[F(e_M)-F(e^*)]b_1a_1}\left(\frac{s}{a_1}\right)^{(1-b_1)/b_1}$$

$$= C'\left(\frac{s}{a_1}\right)^{-1-(\alpha/b_1)}$$

which is a power law as well, but with constant parameter $C' = \frac{C}{[F(e_M)-F(e^*)]b_1a_1}$ and power parameter (α/b_1).

Necessary condition: (if $f_S(x)$ is a power law, then $f(x)$ must be a power law as well).

First, remember that the DFS follows a power law if and only if the elasticity of its probability density function is constant (and let us denote this elasticity κ_S):

$$\eta_S = \frac{s}{f_S(s)}\frac{df_S(s)}{ds} = \kappa_S$$

Let us assume that this elasticity is constant and equal to κ_S. From [3], $f_S(s) = \frac{f(h^{-1}(s))}{F(e_M)-F(e^*)}\frac{dh^{-1}(s)}{ds}$, so that the above elasticity is given by

$$\eta_S = \frac{s}{f_S(s)}\frac{df_S(s)}{ds}$$

$$= \frac{s\frac{1}{F(e_M)-F(e^*)}\left[f'(h^{-1}(s))\frac{dh^{-1}(s)}{ds}\frac{dh^{-1}(s)}{ds} + f(h^{-1}(s))\frac{d^2h^{-1}(s)}{ds^2}\right]}{\frac{f(h^{-1}(s))}{F(e_M)-F(e^*)}\frac{dh^{-1}(s)}{ds}}$$

$$= \eta_S = \frac{sf'(h^{-1}(s))}{f(h^{-1}(s))}\frac{dh^{-1}(s)}{ds} + \frac{s\frac{d^2h^{-1}(s)}{ds^2}}{\frac{dh^{-1}(s)}{ds}}$$

Remember that, if labor demand is a power function, $h(e; w) = a_1\ e^{b_1}$, then $h^{-1}(s) = (s/a_1)^{1/b_1}$, $h^{-1}(s) = \left(\frac{s}{a_1}\right)^{1/b_1}$, $\frac{dh^{-1}(s)}{ds} = \frac{1}{b_1a_1^{1/b_1}}s^{(1/b_1)} - 1$ and $\frac{d^2h^{-1}(s)}{ds^2} = \frac{(1-b_1)/b_1}{b_1a_1^{1/b_1}}s^{(1/b_1)-2}$. Note also that $h^{-1}(s) = b_1s\frac{dh^{-1}(s)}{ds}$ or equivalently, $\frac{dh^{-1}(s)}{ds} = \frac{1}{b_1a_1^{1/b_1}}s^{(1/b_1)-1} = \frac{1}{b_1s}\left(\frac{s}{a_1}\right)^{(1/b_1)} = \frac{1}{b_1s}h^{-1}(s)$.

By substituting $\frac{dh^{-1}(s)}{ds} = \frac{1}{b_1a_1^{1/b_1}}s^{(1/b_1)-1}$, $\frac{d^2h^{-1}(s)}{ds^2} = \frac{(1-b_1)/b_1}{b_1a_1^{1/b_1}}s^{(1/b_1)-2}$, and $h^{-1}(s) = b_1s\frac{dh^{-1}(s)}{ds}$, into the last expression of the last elasticity of η_S we get

$$\eta_S = \frac{s f'\left(h^{-1}(s)\right)}{f\left(h^{-1}(s)\right)} \frac{h^{-1}(s)}{b_1 s} + \frac{s \frac{(1-b_1)/b_1}{b_1 a_1^{1/b_1}} s^{(1/b_1)-2}}{\frac{1}{b_1 a_1^{1/b_1}} s^{(1/b_1)-1}}$$

$$= \frac{1}{b_1} \frac{h^{-1}(s)}{f\left(h^{-1}(s)\right)} f'\left(h^{-1}(s)\right) + \frac{1-b_1}{b_1}$$

Let us now make the change of variable $x = h^{-1}(s)$. Then,

$$\eta_S = \frac{1}{b_1} \frac{x}{f(x)} f'(x) + \frac{1-b_1}{b_1}$$

The DFS follows a power law if this elasticity is constant,

$$\eta_S = \frac{1}{b_1} \frac{x}{f(x)} f'(x) + \frac{1-b_1}{b_1} = \kappa_S \Leftrightarrow$$
$$\frac{x}{f(x)} f'(x) = b_1(\kappa_S + 1) - 1$$

This is a simple differential equation, and the "unique" solution is $f(x) = C_1 x^{b_1(\kappa_S+1)-1}$, which is the density probability function of a Pareto distribution (for an appropriate value of the constant C_1).

Alternatively, we can prove that $f(x)$ must be a power law as follows. If labor demand is a power function and the distribution of size follows a power law, $f_S(s) = C_1 s^{-1-k}$, then, from [3]

$$f_S(s) = C_1 s^{-1-k}, \quad \text{and} f_S(s)$$

$$= \frac{f\left(h^{-1}(s)\right)}{F(e_M) - F(e^*)} \frac{dh^{-1}(s)}{ds} \Rightarrow C_1 \ s^{-1-k}$$

$$= \frac{1}{[F(e_M) - F(e^*)] b_1 a_1} \left(\frac{s}{a_1}\right)^{(1-b_1)/b_1} f\left((s/a_1)^{1/b_1}\right)$$

After rearranging some terms, we get

$$f\left((s/a_1)^{1/b_1}\right) = C_1 [F(e_M) - F(e^*)] b_1 a_1^{-k} \left((s/a_1)^{1/b_1}\right)^{(-1-kb_1)}$$

So that the probability density function of the distribution of skills may be written as (after changing variables, $x = (s/a_1)^{1/b_1}$)

$$f(x) = C x^{(-1-kb_1)}$$

which is a power law as well (but with power parameter $k b_1$ and constant/scale parameter $C = C_1[F(e_M) - F(e^*)]b_1 a_1^{1-k}$. ∎

Appendix 2. A closed equilibrium solution from occupational choices and power law distribution of entrepreneurial skills

The Pareto distribution of skills has the advantage that it allows for closed solutions to the occupational choice equilibrium, which facilitates the comparative static analysis of changes in the DFS to changes in the parameters of the model. In this appendix, we present the market equilibrium solution from occupational choices for the simple production function $Q = \theta \ e^{(1+\beta)} L^{1-\beta}$ and a distribution of entrepreneurial skill in the working population that follows a Pareto distribution (with minimum value e_m and power parameter α). It can be shown that equilibrium exists and is unique for values of the parameters that satisfy the condition[13] $(\alpha - 1)\beta > 1$. The level of skill for which the individual is indifferent between working as employee or working as an entrepreneur and the equilibrium salary are given by (see Medrano-Adán et al. (2019) for the details):

$$e^* = e_m \left(\frac{\alpha-\beta-1}{(\alpha-1)\beta-1}\right)^{1/\alpha} \quad w^* = \theta(1-\beta)^{1-\beta} \beta^\beta e_m^{\beta+1} \left(\frac{\alpha-\beta-1}{(\alpha-1)\beta-1}\right)^{(\beta+1)/\alpha}$$

The distribution of firm size is (also) a Pareto distribution with power parameter $\frac{\alpha\beta}{(\beta+1)}$ and minimum size value $S_{\min} = h(e^*; w^*) = a(e^*)^b = \frac{(1-\beta)}{\beta}$. The probability density function and cumulative distribution function are given by (for $x \geq S_{\min} = (1-\beta)/\beta$)

$$f_S(x) = \frac{\alpha\beta}{(\beta+1)} S_{\min}^{\frac{\alpha\beta}{(\beta+1)}} x^{-1-\frac{\alpha\beta}{(\beta+1)}}, \qquad F_S(x) = 1 - \left(\frac{S_{\min}}{x}\right)^{\frac{\alpha\beta}{(\beta+1)}}$$

The density and distribution functions only depend on the power parameter of the distribution of skills α, and on the parameter of the organization technology, β. We can easily prove the following analytical results:

1. The average firm size, given by $\overline{S} = \frac{\alpha(1-\beta)}{(\alpha-1)\beta-1}$, is strictly decreasing in α, the concentration of the distribution of skills, and in the organizational size diseconomies parameter, β.

2. The minimum firm size is strictly decreasing in β,

---

[13] This result follows from the property that if $g(x)$ and $h(x)$ are power functions, then $h(g(x))$ is also a power function. In particular, the relationship between skills of the entrepreneur and profit-maximizing number of employees, $L = h(e; w^*)$, will be a power function when the production technology is a linear homogeneous production function, for example a Cobb-Douglas or a CES.

3. The power parameter of the distribution function of firm size $\frac{\alpha\beta}{(\beta+1)}$ is strictly increasing in both α and β.

4. The quantiles of the firm size distribution $Quantile$ $(q) = \frac{(1-\beta)}{\beta}(1-q)^{-\frac{1+\beta}{\alpha\beta}}$ are strictly decreasing in both α and β).

The average size of firms in the occupational choice equilibrium increases with lower organizational size diseconomies, i.e., with lower intensity of time of the entrepreneur in the supervision of the employees under direction, lower β, and with higher dispersion of skills in the population, lower α. Lower organizational size diseconomies, lower β, also implies larger minimum firm size, i.e., lower number of micro firms in the market equilibrium. The proportion of employees in larger firms, for example in the top 1% quantile of the size distribution, decreases with α and with β.

The survival function of the DFS, the complementary of the cumulative distribution function, is a power function:

$$SF_S(x) = 1 - F_s(x) = \left(\frac{S_{\min}}{x}\right)^{\frac{\alpha\beta}{(\beta+\tau)}} \quad (\forall x \geq S_{\min} = (1-\beta)/\beta)$$

So that its plot in logarithmic scale (the so-called Zipf plot) is linear; there is a negative linear relationship between the log of the proportion of firms with L or more employees and the number of employees L.

$$\ln[1 - F_S(x)] = \frac{\alpha\beta}{(\beta+\tau)}[\ln(S_{\min}) - \ln(x)]$$

The DFS satisfies the property of power law distributions of constant elasticity, equal to $\frac{\alpha\beta}{(\beta+1)}$, between the proportion of firms above a value of the variable L and the value of L. In the Zipf distribution the power parameter would be equal to 1, $\frac{\alpha\beta}{(\beta+1)} = 1$, which implies that $(\alpha - 1)\beta = 1$. For values of the parameters satisfying this condition, the occupational choice equilibrium would not have a finite equilibrium (average size of firms tends to infinity, for example). In other words, for economically meaningful occupational choices equilibrium, the parameters of the model must satisfy $(\alpha - 1)\beta > 1$ and the power parameter must be greater than one. Axtell (2001) estimated a power parameter $\frac{\alpha\beta}{(\beta+1)} = 1.059$, greater than one, although, as we show in the main text, we reject the null hypothesis that the US distribution of firm size really follows a power law.

Appendix 3. Test of the power law with Spanish data on firm size

For robustness purpose we present empirical results additional to those in section 4 on the power law distribution of firm size, now with Spanish data. The Spanish official statistical office, INE, publishes annual binned data from 1999 till 2019 on the distribution of firm size (see https://www.ine.es/jaxiT3/Tabla.htm?t=299&L=0), with number of firms in each of the 11 size classes with limits {1, 2, 5, 9, 19, 49, 99, 199, 499, 999, 4999, more than 5000}. For each of the years we estimate Eqs. [5] and [7] in the main text and perform the Ramsey RESET and omitted variable specification tests.

For each of the 21 years considered (1999–2019), the estimated value of δ is in the interval $(-0.0133, -0.0328)$, and the p value of the test of the mull hypothesis that $\delta = 0$ is always lower than 0.00604 (with mean value equal to 0.00070). The Ramsey RESET test gives similar results, with p values lower than 0.0051 (with mean value equal to 0.0006). Table 5 shows a summary of the estimation results: the hypothesis of $\delta = 0$ is rejected at a high level of significance (higher than 99%) for all the years.

Consequently, the specification tests do not support the conclusion that the data on the distribution of firm size in Spain follows a power law, similar to the result in the main text for US firm size data.

Table 5 Estimations of the distributions of firm size with Spanish data: log-linear and log-quadratic model: selected years in the period 1999 to 2019

Year	Log-linear model [5]		Log-quadratic model [7]			
	$\widehat{\alpha}$	R^2	$\widehat{\alpha}$	$\widehat{\delta}$	p value (δ)	R^2
2000	1.173	0.9951	0.918	−0.0282	0.006054	0.9982
2005	1.156	0.9953	0.860	−0.0328	0.000012	0.9995
2010	1.128	0.9972	0.912	−0.0240	0.000041	0.9996
2015	1.088	0.9977	0.889	−0.0221	0.000000	0.9999

References

Aguinis, H., Gomez-Mejia, L., Martin, G., & Joo, H. (2018). CEO pay is indeed decoupled from CEO performance: charting a path for the future. *Management Research: Journal of the Iberoamerican Academy of Management, 16*(1), 117–136. https://doi.org/10.1108/MRJIAM-12-2017-0793.

Andriani, P., & McKelvey, B. (2009). From Gaussian to Paretian thinking: causes and implications of power laws in organizations. *Organization Science, 20*(6), 1053–1071. https://doi.org/10.1287/orsc.1090.0481.

Axtell, R. (2001). Zipf distributions of U.S. firm sizes. *Science, 293*(5536), 1818–1820. https://doi.org/10.1126/science.1062081.

Axtell, R. (2006). Firm sizes: Facts, formulae, fables and fantasies. Center on Social and Economic Dynamics Working Paper, No. 44. https://doi.org/10.2139/ssrn.1024813.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics, 122*(4), 1351–1408. https://doi.org/10.1162/qjec.2007.122.4.1351.

Broecke, S., Quintini, G., & Vandeweyer, M. (2017). Explaining international differences in wage inequality: Skills matter. *Economics of Education Review, 60*, 112–124. https://doi.org/10.1016/j.econedurev.2017.08.005.

Clauset, A., Rohilla Shalizi, C., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51*(4), 661–703. https://doi.org/10.1137/070710111.

Coase, R. H. (1937). The nature of the firm. *Economica, 4*, 386–405. https://doi.org/10.1111/j.1468-0335.1937.tb00002.x.

Crawford, G. C. (2018). Skewed Opportunities: How the distribution of entrepreneurial inputs and outcomes Reconceptualizes a research domain. Proceedings-Academy of Management. https://par.nsf.gov/biblio/10058144.

Crawford, G. C., Aguinis, H., Lictenstein, B., Davidsson, P., & McKelvey, B. (2015). Power law distributions in entrepreneurship: implications for theory and research. *Journal of Business Venturing, 30*, 696–713. https://doi.org/10.1016/j.jbusvent.2015.01.001.

Davis, S., & Henrekson, M. (1999). Explaining national differences in the size and industry distribution of employment. *Small Business Economics, 12*(1), 59–83. https://doi.org/10.1023/A:1008078130748.

Decker, R., Haltiwanger, J. R., Jarmin, R., & Miranda, J. (2018). Changing business dynamism and productivity: shocks vs. responsiveness. *NBER Working Paper*, 24236. https://doi.org/10.3386/w24236https://www.nber.org/papers/w24236.

Denrell, J., Fang, C., & Liu, C. (2014). Perspective - chance explanations in the management sciences. *Organization Science, 26*(3), 923–940. https://doi.org/10.1287/orsc.2014.0946.

Ferrante, F. (2005). Revealing entrepreneurial talent. *Small Business Economics, 25*(2), 159–174. https://doi.org/10.1007/s11187-003-6448-6.

Fuchs-Schündeln, N., & Schündeln, M. (2005). Precautionary savings and self-selection: evidence from the German reunification 'experiment'. *The Quarterly Journal of Economics, 120*(3), 1085–1120. https://doi.org/10.1093/qje/120.3.1085.

Gabaix, X. (2016). Power laws in economics: an introduction. *Journal of Economic Perspectives, 30*(1), 185–206. https://doi.org/10.1257/jep.30.1.185.

Gabaix, X., & Landier, A. (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics, 123*(1), 49–100. https://doi.org/10.1162/qjec.2008.123.1.49.

Geroski, P. A. (2000). The growth of firms in theory and practice. In N. Foss & V. Mahnke (Eds.), *Chapter 8 of Competence, Governance and Entrepreneurship* (pp. 168–186). Oxford, ISBN 10.0198297173/ISBN13:9780198297178: Oxford University Press.

Gibrat, R. (1931). *Les Inégalités économiques*. Paris: Recueil Sirey.

Green W. H. (2012). *Econometric Analysis* (7th Edition). London, England: Pearson Education Limited.

Henrekson, M., & Johansson, D. (1999). Institutional effects on the evolution of the size distribution of firms. *Small Business Economics, 12*(1), 11–23. https://doi.org/10.1023/A:1008002330051.

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica, 60*(5), 1127–1150. https://www.jstor.org/stable/2951541. https://doi.org/10.2307/2951541.

Ijiri, Y., & Simon, H. A. (1967). A model of business firm growth. *Econometrica, 35*(2), 348–355. https://doi.org/10.2307/1909116 https://www.jstor.org/stable/i332630.

Joo, H., Aguinis, H., & Bradley, K. J. (2017). Not all nonnormal distributions are created equal: Improved theoretical and measurement precision. *Journal of Applied Psychology, 102*(7), 1022–1053. https://doi.org/10.1037/apl0000214.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica, 50*(3), 649–670. https://doi.org/10.2307/1912606 https://www.jstor.org/stable/1912606.

Jovanovic, B. (1994). Firm formation with heterogeneous management and labor skills. *Small Business Economics, 6*(3), 185–191. https://doi.org/10.1007/BF01108287.

Kihlstrom, R., & Laffont, J. J. (1979). A general equilibrium theory of firm formation based on risk aversion. *Journal of Political Economy, 87*(4), 719-748. https://doi.org/10.1086/260790

Knudsen, T., Levinthal, D. A., & Winter, S. G. (2017). Systematic differences and random rates: reconciling Gibrat's law with firm differences. *Strategy Science, 2*(2), 111–120. https://doi.org/10.1287/stsc.2017.0031.

Lazear, E. P. (2004). Balanced skills and entrepreneurship. *The American Economic Review, 94*(2), 208–211. https://doi.org/10.1257/0002828041301425.

Lucas, R. (1978). On the size distribution of business firms. *The Bell Journal of Economics, 9*(2), 508–523 https://www.jstor.org/stable/3003596.

McKelvey, B. (2004). Toward a complexity science of entrepreneurship. *Journal of Business Venturing, 19*, 313–341. https://doi.org/10.1016/S0883-9026(03)00034-X.

Medrano-Adan, L., Salas-Fumás, V., & Sanchez-Asin, J. J. (2015). Heterogeneous entrepreneurs from occupational choices in economies with minimum wage. *Small Business Economics, 44*, 597–619. https://doi.org/10.1007/s11187-014-9610-4.

Medrano-Adán, L., Salas-Fumás, V., & Sanchez-Asin, J. J. (2019). Firm size and productivity from occupational choices. *Small Business Economics, 53*(1), 243–267. https://doi.org/10.1007/s11187-018-0048-y.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725. https://doi.org/10.1111/1468-0262.00467.

Moral Benito, E. (2018). Growing by learning: firm-level evidence on the size-productivity nexus. *SERIEs, 9*(1), 65–90. https://doi.org/10.1007/s13209-018-0176-2.

Murphy, K., Shleifer, A., & Vishny, R. (1991). The allocation of talent: implications for growth. *The Quarterly Journal of Economics, 106*(2), 503-530. https://doi.org/10.2307/2937945

Ortín-Ángel, P., & Vendrell-Herrero, F. (2010). University spin-offs vs. other NTBFs: productivity differences at the outset and evolution. In *Searle Center on Law, Regulation, and Economic Growth. Working Papers, 2010–027*. Chicago: Northwestern University.

Penrose, E. (1959). *The theory of the growth of the firm*. New York: John Wiley.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological), 31*(2), 350–371 http://www.jstor.org/stable/2984219.

Roper, S. (1998). Entrepreneurial characteristics, strategic choice and small business performance. *Small Business Economics, 11*, 11–24. https://doi.org/10.1023/A:1007955504485.

Rosen, S. (1981). The economics of superstars. *The American Economic Review, 71*(5), 845–858 https://www.jstor.org/stable/1803469.

Rosen, S. (1982). Authority, control, and the distribution of earnings. *The Bell Journal of Economics, 13*(2), 311–323. https://doi.org/10.2307/3003456.

Schumpeter, J. A. (1934). *The theory of economic development: an inquiry into profits, capital, credit, interest and the business cycle*. Cambridge, MA: Harvard University Press.

Shepherd, D. A., Wennberg, K., Suddaby, R., & Wiklund, J. (2019). What are we explaining? A review and agenda on initiating, engaging, performing, and contextualizing entrepreneurship. *Journal of Management, 45*(1), 159–196. https://doi.org/10.1177/0149206318799443.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika, 42*(3/4), 425–440. https://www.jstor.org/stable/2333389. https://doi.org/10.2307/2333389.

Stanley, M. H. R., Buldyrev, S. V., Havlin, S., Mategna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firms. *Economic Letters, 49*(4), 453–457. https://doi.org/10.1016/0165-1765(95)00696-D.

Storey, D. (1994). *Understanding the small business sector*. London: Routledge.

Vendrell-Herrero, F., González-Pernía, J. L., & Peña-Legazkue, I. (2014). Do incentives matter to promote high technology-driven entrepreneurial activity? *International Entrepreneurship and Management Journal, 10*(1), 43–66. https://doi.org/10.1007/s11365-011-0181-4.

Wooldridge, J. M. (2019). *Introductory econometrics: a modern approach* (7th Edition). Cincinati, O: South Western Cengage Learning.