Firm Sizes: Facts, Formulae, Fables and Fantasies

Robert Axtell†*
The Brookings Institution
Washington, DC 20036 USA

CSED Working Paper No. 44
February 2006

THE BROOKINGS INSTITUTION
THE JOHNS HOPKINS UNIVERSITY
CENTER ON SOCIAL AND ECONOMIC DYNAMICS

FIRM SIZES: FACTS, FORMULAE, FABLES AND FANTASIES

Robert Axtell†*

The Brookings Institution

Washington, DC 20036 USA

CSED Working Paper No. 44

February 2006

Version 4: September 2005

Abstract

Recently discovered *facts* concerning the size distribution of U.S. firms are recapitulated—in short, these sizes are closely approximated by the Zipf distribution, a Pareto (power law) distribution with exponent of unity. Interesting consequences of this result are then developed, having primarily to do with *formulae* for the distribution's moments, and difficulties of reasonably characterizing a 'typical' firm. Then, a leading candidate explanation for these data—the Kesten random growth process—is assessed in terms of its realism vis-à-vis actual firm growth. Insofar as it has fluctuations that are quite different in character from actual firm size variability, the Kesten and related stochastic growth processes qualify more as *fables* of firm growth than as credible explanations. Finally, new explanations of the facts are proposed by considering firms to be partitions of the set of all workers. Assuming all partitions to be equally likely, the observed distribution of firm sizes is hypothesized to be the distribution of block sizes in the most likely partitions. An alternative derivation of this distribution as a constrained optimization problem is also described. Given that these calculations involve unimaginably vast magnitudes, it seems just short of *fantastic* to consider them relevant empirically.

---

† Preliminary versions of this paper have been presented at Carnegie Mellon University, The Brookings Institution and George Mason University. A similar edition of this working paper is set to appear in *Power Laws in the Social Sciences*, Claudio Cioffi-Revilla, ed., Cambridge University Press, forthcoming in 2006. Comments from seminar participants have proven helpful and are gratefully acknowledged. I also thank John Nolan and Herbert Wilf for mathematical help on stable distributions and combinatorial partitions, respectively.

* Please address paper inquiries to Robert Axtell (raxtell@brookings.edu) 202-797-6020 or via post at 1775 Massachusetts Ave., N.W. Washington, DC 20036.

I A Brief History of the Firm Size Distribution

Gibrat (1931) inaugurated the systematic study of this subject, finding that the lognormal distribution well-described French industrial firms. Subsequently, similar analyses were carried out for other countries (e.g., Florence (1953) in the UK). Economic theorists joined the discussion with Simon's attempts to explain the skew and leptokurtic size distributions obtaining among publicly-traded firms, by recourse to stochastic growth models (Simon 1955; Simon and Bonini 1958; see also Steindl 1965). Such models contain essentially no elements of conventional economics (e.g., prices, profits), yet so well rationalize these extremely regular data that Simon was led to inveigh caustically against the neoclassical U-shaped cost curve explanation of firm behavior and 'optimal size' (Ijiri and Simon 1977: 7-11). Given the undeniable regularity of these data, discussion of skew firm size distributions entered into the developing field of industrial organization (IO), although accompanied by little of Simon's critique (e.g., Scherer 1970; 1980).

Over about a generation, from the early 1970s until now, IO progressively assimilated a game theoretic orientation, viewing the firm as a unitary and rational actor engaged in playing market games non-cooperatively. This led to the monotonic increase in strategic considerations in IO texts (e.g., Tirole 1988), crowding out many classical concerns, including empirical results like the firm size distribution. Indeed, the situation has evolved to such an extent that IO students today may never encounter any discussion of firm size distributions in a modern textbook (e.g., Shy 1995)!

Sutton (1998) is one of the few IO researchers who did not lose sight of skew firm size data as game theory conquered IO. His work uses firm size distribution data—across markets and industries—as important constraints on the kinds of game theoretic models that are worth writing down.

A related stream of economic research in which matters of firm sizes loom large is that of small business economics. Nearly ubiquitous in this area is some measure of what constitutes a 'small' firm or establishment, along with some other measure of the frequency of such entities. But this literature tends to be heavily descriptive and non-quantitative, and when it finally turns econometric, the kinds of specifications employed are seemingly much more informed by neoclassical assumptions than the skew character of the empirical data.

Outside the mainstream of economics, but a research orientation that explicitly recognizes the importance of firm sizes, is work on organizational ecology (e.g., Carroll and Hannan 2000). These efforts are further concerned with the closely related firm-specific variables of age and growth, as well as industry-specific life cycles. Unfortunately, explanations of firm size in this literature are often more qualitative than quantitative, and only weakly-related to data.

Today, facilitated by advances in information technology, ever more current and accurate data on firms are progressively becoming available. This represents an important scientific opportunity, but is simultaneously an ironic development given the state of the IO textbooks. This paper explores what these new data are telling us about the complete universe of U.S. firms—their sizes, dynamics, employment characteristics, revenue streams. We are particularly interested in two aspects of these data: (1) developing some measure of the 'typical' U.S. business firm; and (2) exploring alternative explanations for why the data are what they are. We shall find that (1) is surprisingly difficult, while (2) is almost too easy, at least in the sense that many candidate explanations present themselves. Taken together we are left with the distinct impression not that the firm size distribution is a statistical curiosity that is well understood, but rather that it is a deep and enduring mystery that gives up its secrets only grudgingly and, even then, only in code.

II The Facts

In previous work (Axtell 2001) I have reported that the distribution of U.S. firm sizes closely follows the Pareto distribution with exponent near unity, i.e., the Zipf distribution.[1] Noting the size of a firm by S, with s referring to some specific size, the tail CDF of this distribution is

$$\Pr\left[S \geq s\right] = \left(\frac{s_0}{s}\right)^{\alpha}, \tag{1}$$

where s_0 is the minimum size and α is a parameter. This distribution is a classical example of a power or scaling law. The empirical data for 1997 are shown binned as a PDF in figure 1 below, with size measured by employees, along with a line having α = 1.059.[2]

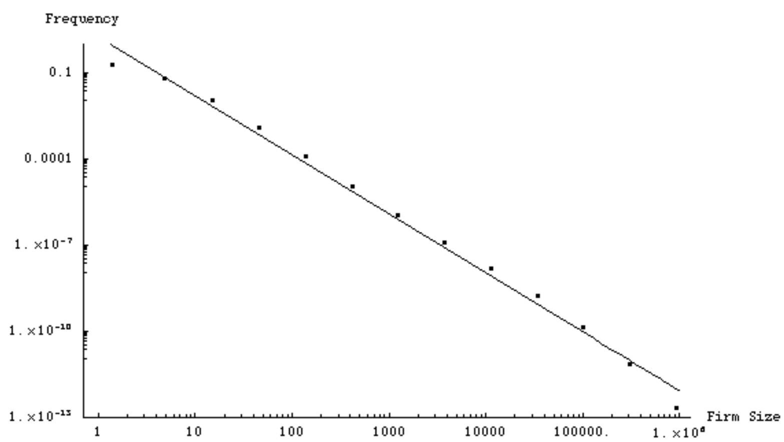

Figure 1: PDF of U.S. firm sizes, 1997 Economic Census data

Thee Zipfian character of this distribution is robust to alternative definitions of firm size (e.g., firm receipts).

Data of such vast regularity are highly unusual in the social sciences. Only at the extremes of the support do the data depart in any systematic way from the distribution. Indeed, there are relatively too few very small and very large firms in the data. Such deviations are often

[1] The Zipf distribution is usually considered a discrete distribution; more on this below.
[2] The origin of these data are tax filings and, for reasons of confidentiality, only binned data are available. The kinds of statistical procedures used on these data are therefore not generally commensurate with other papers in this volume that analyze raw data.

termed 'finite size effects' in the literature on scaling laws, meaning simply that the size of real firms is bounded both from above and below, and so it is precisely at these limits of support where deviations occur.

That power laws describe the upper tail of the firm size distribution has been known for some time (e.g., Simon and Bonini 1958; Ijiri and Simon 1977), but it was always thought that there existed some minimum size below which the number of firms decreased (e.g., Mandelbrot 1997: 204). Clearly this is not the case as the mode of this distribution is a single employee. That one distribution describes essentially the entire firm size distribution suggests that a uniform process of firm growth and decline is operational over all firm sizes. In §IV below we will describe such processes. But first we turn our attention to some of the peculiar properties of this extremely skew distribution.

III The Formulae

Zipf distributed random variables (RVs) are quite different from the kinds of RVs on which students of probability cut their teeth. Zipf RVs do not follow the central limit theorem (CLT), but rather are, asymptotically, members of the class of so-called *stable* distributions. The origin of this violation of the CLT is the non-existence of the second and higher moments of this distribution. Furthermore, for firms the first moment is barely meaningful. These facts are recounted in this section along with various other formulae relevant to Zipf and near-Zipf distributions.

1 Moments: Non-Existence of Second, Limited Utility of First

The Pareto distribution is sufficiently skew that all moments beyond the α^{th} do not exist. To see this, we start with the Pareto PDF

$$f(s) \equiv \frac{\alpha s_0^\alpha}{s^{\alpha+1}} \tag{2}$$

and compute its m^{th} moment

$$\int_0^\infty s^m f(s) ds = \alpha s_0^\alpha \int_0^\infty s^{m-1-\alpha} ds = \frac{\alpha s_0^{m-\alpha}}{\alpha - m} \tag{3}$$

Clearly, this expression is only meaningful for $\alpha > m$.

From the 1997 firm size data, the average firm size is 19.00 (105,299,123 employees/5,541,918 firms). Using (3) together with the empirical estimate of α = 1.059 gives a mean firm size of approximately 17.95, in reasonable agreement with the empirical value.

The moments of any finite data will always exist. Since we are dealing with α just larger than 1, non-convergence of the second moment simply means that as additional data are added to partial computations of the variance there is divergence from previous estimates. An example of this is shown in figure 2 below for two distinct partial sum paths through 10^4 Zipf distributed random variates (i.e., α = 1).

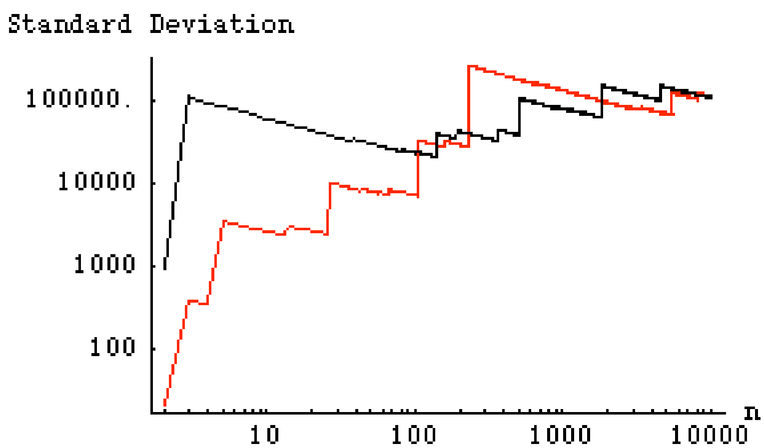

Figure 2: Non-convergence of the second moment of firm sizes

Note in this figure that each sample path ends up at the same point, but each has a unique trajectory, and is clearly diverging, not converging.

For the value of α we have for firms, the first moment *is* well defined, but just barely. For if α were unity then it too would be undefined. For any finite sample and α close to 1, it is an open question as to how reliable the mean value will be, i.e., whether there are enough data from which to build up a reliable estimate of it. Indeed, the average size itself is scaling (Feller 1971), implying that the distribution of mean values also follows a power law. Essentially, the mean value can range very widely over the support. This peculiar property in the context of a different data set[3] has been termed the 'nobody knows anything' principle, since it implies that it is very difficult to make accurate predictions about the 'typical' value of such processes (De Vany and Walls 2002; De Vany and Walls 2004). This should make us skeptical of any econometric work based on the time series of average firm size (e.g., Simon 1957; Lucas 1978; Kumar, Rajan et al. 1999).

Because of the nonexistence of the second moment of the Pareto and Zipf RVs we are working with here, the conventional CLT does not hold, for a requirement of the CLT is a finite second moment. Rather, it is the so-called *stable* RVs that are more closely related Pareto ones, at least asymptotically (Samorodnitsky and Taqqu 1994). Stable RVs have their own CLT: suitably normalized sums of iid stable RVs are stably-distributed. Such RVs assault our Gaussian-trained intuition.

2 *Alternative Measures of Location and Dispersion*

Given the limited utility of the mean in the case of firm sizes, and the irrelevance of the variance, here we investigate other measures of location (i.e., central tendency) and dispersion (i.e., inequality).

The modal value of a Pareto distribution is s_0 which we fix at 1, in accord with firm employee data. The median, M, is

$$M = s_0 2^{\frac{1}{\alpha}}. \tag{4}$$

Using the value of α above, $M \approx 1.89$, and since we are measuring size in terms of employees let us call this 2, which is also what it would be in the idealized case of the pure Zipf distribution, i.e., $\alpha = 1$ with $s_0 = 1$. This is reasonably different from the median estimated from the actual data, which is between 3 and 4.

The geometric mean, G of the Pareto distribution always exists and can be written in terms of the parameters as

[3] For Hollywood movies De Vany (2004) has found stable distributions of profits with $\alpha \sim 1.26$, which is somewhat less skew than the Zipf distribution.

$$G = s_0 \exp\left(\frac{1}{\alpha}\right). \tag{5}$$

For the empirical α, $G \approx 2.57$. It would be slightly higher—$G = e \approx 2.72$—in the case of a true Zipf distribution. The actual geometric mean is unknown since the raw firm size data are unavailable.

The harmonic mean, $H = (E[S^{-1}])^{-1}$ exists for the Pareto distribution and can be represented symbolically as

$$H = s_0\left(1 + \frac{1}{\alpha}\right). \tag{6}$$

For U.S. firm sizes $H \approx 1.94$, while it is 2 for the Zipf distribution.

Moving on to measures of dispersion, the Gini index, g, is a standard measure of inequality, more commonly encountered in the context of income and wealth distributions. $g \in [0, 1]$ with low values representing little dispersion and high values the reverse. In the case of the Pareto having $\alpha > 1$, it amounts to

$$g = \frac{1}{2\alpha - 1}.$$

For our firm size data, $g \approx 0.894$, which is very high indeed—much larger than Gini indices typical of income data, and larger even than estimates for U.S. wealth—further evidence of the extreme skewness of firm sizes.

3 Representation: Finite vs. Infinite Support

In practice, the support of empirical data is always finite. This leads to consideration of the so-called 'truncated Pareto distribution' (Johnson, Kotz et al. 1994: 608).[4] For support on $[s_0, s_\infty]$, it has CDF

$$\Pr\left[S \le s_i\right] = \frac{1 - \left(\frac{s_0}{s_i}\right)^\alpha}{1 - \left(\frac{s_0}{s_\infty}\right)^\alpha} = \frac{s_\infty^\alpha}{s_i^\alpha}\,\frac{s_i^\alpha - s_0^\alpha}{s_\infty^\alpha - s_0^\alpha}. \tag{7}$$

When $s_\infty \gg s_0$ this expression becomes

[4] Truncated zeta distributions have also been considered (e.g., Mandelbrot 1997: 202).

$$\Pr\left[S \le s_i\right] = 1 - \left(\frac{s_0}{s_i}\right)^{\alpha} \tag{8}$$

which is equivalent to (1). This last condition certainly holds for the firm size data—$s_\infty \sim 10^6 \gg s_0 \sim 1$—so we can use (1) with little error. For firm size measured by receipts, $s_\infty \sim O(\$10^{11}) \gg s_0 \sim O(\$10^5)$ and this condition continues to hold. For the near-Zipfian case of $\alpha \sim 1$, (7) can be written

$$\Pr\left[S \le s_i\right] = \frac{s_\infty}{s_i}\frac{s_i - s_0}{s_\infty - s_0}$$

The moments of the truncated Pareto distribution always exist and the first two can be shown to be

$$\mu = \frac{\alpha}{\alpha - 1}\frac{s_0 s_\infty^\alpha - s_0^\alpha s_\infty}{s_\infty^\alpha - s_0^\alpha}$$

$$\sigma^2 = \frac{\alpha}{2-\alpha}\frac{s_0^\alpha s_\infty^2 - s_0^2 s_\infty^\alpha}{s_\infty^\alpha - s_0^\alpha} - \left(\frac{\alpha}{\alpha - 1}\frac{s_0 s_\infty^\alpha - s_0^\alpha s_\infty}{s_\infty^\alpha - s_0^\alpha}\right)^2.$$

For $s_\infty \gg s_0$ these become

$$\mu \approx \frac{\alpha}{\alpha-1}\left(s_0 - s_0^\alpha s_\infty^{1-\alpha}\right) \tag{9}$$

$$\sigma^2 \approx \frac{\alpha}{2-\alpha} s_0^\alpha s_\infty^{2-\alpha} - \left(\frac{\alpha}{\alpha-1}\left(s_0 - s_0^\alpha s_\infty^{1-\alpha}\right)\right)^2. \tag{10}$$

Note that in this limiting case, i.e., of $s_\infty \gg s_0$, the PDF (8) does not depend on the upper support limit, s_∞. However, this parameter is present in the expressions for the moments, (9) and (10), where its effect is to create a deviation from the comparable expressions for the non-truncated Pareto distribution. In particular, if we estimate the mean firm size from the data using $s_\infty = 10^6$, we obtain 9.95, substantially below the actual value of 19.0. However, as one increases s_∞ progressively the estimate of the average increases monotonically until, in the limit of $s_\infty \uparrow \infty$ one obtains the value associated with the non-truncated distribution, i.e., 17.95.

The median of the truncated Pareto distribution can be derived as

$$M = 2^{\frac{1}{\alpha}}\frac{s_0 s_\infty}{\left(s_\infty^\alpha + s_0^\alpha\right)^{1/\alpha}}. \tag{11}$$

Clearly this specializes to (4) in the limit of s_∞ » s_0. Using the values of parameters for the 1997 data gives M ~ 1.92.

A notion related to finite support is that of having large expanses of the support that are unpopulated. This is not the case with the Pareto distribution, but is true of a class of distributions due to Haight (Johnson, Kotz et al. 1993: 469-471). The application of such distributions to firm size data may be a fertile area for future research.

4 Representation: Discrete vs. Continuous Support

For firm size measured in terms of employees, it is natural to conceive of firm size, S, as a discrete RV, and the firm size distribution as a discrete distribution. In such circumstances a Zipf-like probability density function is the so-called zeta distribution, having PMF

$$\Pr\left[S = s_i\right] = c s_i^{-(\alpha+1)}$$

for $s_i \geq s_0$, the minimum size.[5] The constant, c, is related to the Riemann zeta function, defined as

$$\zeta(k) \equiv \sum_{i=1}^{\infty} i^{-k},$$

such that

$$c = \left[\sum_{i=1}^{\infty} s_i^{-(\alpha+1)}\right]^{-1} = \frac{1}{\zeta(\alpha+1)}.$$

The m^{th} moment of this distribution is (Johnson, Kotz et al. 1993: 466)

$$\mu_m = \frac{\zeta(\alpha - m + 1)}{\zeta(\alpha + 1)},$$

for $m < \alpha$; for $m \geq \alpha$; the moment is infinite. For the estimated α, the mean is 11.0, too small by a considerable margin. Almost certainly this is due to having too few of the smallest firms in the actual data, at least with respect to the number required by the zeta distribution.

[5] This is also known as the discrete Pareto distribution (Johnson, Kotz et al. 1993: 466).

5 Alternative Notions of the 'Typical Firm': Florence Median

Another property of the Zipf distribution we wish to explore has to do with the discrepancy between the typical firm with respect to all firms and the typical firm with respect to all workers. Remember that the average firm size in 1997 was 19.0, computed on the basis of comprehensive data on firms, but it could also have been obtained from a random sample of firms.[6] But what if we poll not the firms but workers in the firms, and ask them the size of their firm? If we average over all employees are we going to get the same result? That is, if we select workers with a uniform probability and ask them what size their firm is, do we get 19.0 again? The answer is 'no' and dramatically so, for the typical employee works in a firm much larger than 19. One way to see this is to compute the median of the distribution of *employee weighted* firm sizes, i.e., of the PDF *sf(s)*, suitably normalized. This turns out to be

$$\frac{(\alpha - 1)s_0^{\alpha-1}}{s^{\alpha}}$$

for $\alpha > 1$. Compared with (2), note that this expression is just a Pareto distribution with an exponent larger than that of the underlying firm size distribution by unity. Having a smaller exponent means that this distribution is even *more* skew than the firm size one!

Näively, one might think that the mean of this employee-weighted size distribution could be readily computed. Unfortunately, the mean of this distribution is in essence the variance of the original firm size distribution, which does not exist. So characterization of a 'typical firm' weighted by employee size will be even more difficult in this case.

Since the median always exists, we can compute it for this employee-weighted firm size distribution. This was apparently first done empirically by Florence (1953), and is referred to by some writers as the *Florence median* (Pryor 2001). Given our lack of access to raw data we cannot determine this quantity exactly, but from Small Business Administration firm size tabulations it appears that it is between 500 and 1000 employees. For the Pareto distribution with unbounded support we can use (4) to compute the median, with α reduced by unity to account for the employee weighting. Doing so yields a Florence median for firms in 1997 as M = 126,500, which is too large by a considerable margin. Alternatively, if we use the truncated Pareto distribution then the Florence median, computed by (11) with $s_\infty = 10^6$, is M = 254, which is much closer to the empirical value.

For a Zipf distribution of firm sizes, the employee-weighted firm size distribution has some peculiar properties. For example, the number of employees in each size 'bin' is equal when the bins are logarithmically sized. To see this we use logs of base b and compute

$$\int_{b^x}^{b^{x+1}} sf(s)ds = \int_{b^x}^{b^{x+1}} s\frac{s_0}{s^2}ds = s_0\left[\ln\left(b^{x+1}\right) - \ln\left(b^x\right)\right] = s_0 \ln\left(b\right)$$

which is independent of x and thus constant. In the firm size data this condition is not strictly met. Specifically, for an exponent of 1.059 instead of unity, a discrepancy is introduced such that

$$\int_{b^x}^{b^{x+1}} s\frac{\alpha s_0^\alpha}{s^{2.059}}ds \propto b^{-0.059x}$$

This expression is slowly decreasing in x, meaning there are somewhat more employees in the small size bins than in the large ones. This pattern is approximately confirmed in the data.

These results suggest that in formulating tax and other business policies, democratic governments should give at least as much consideration to firms having between 1 and 10 employees as to those having 10^5 - 10^6 workers.

6 *Disaggregation into Industries*

The distributions of firm sizes within industries, according to

[6] Assuming, that is, that the data were sampled properly from a skew distribution!

various classifications, are well-known to be skew (Schmalansee 1989), but the Pareto distribution does not seem to well-describe all or even most industries (Quandt 1966). Consider the following alternative. Assume there is some idiosyncratic distribution of firm sizes; then, if the distribution of *industry sizes* is Zipfian, under a wide variety of conditions, the overall size distribution will be as well. Unfortunately, this hypothesis is rejected by the data. Asplund (1998) advanced the similar notion that it is *market size* that is Zipf distributed, but this appears to also be false empirically.

Sutton (1997; 1998) has argued that within markets and submarkets there are exponential bounds to firm size, with parametric heterogeneity across markets and industries. He has pioneered the 'bounds' approach to firm size, which results in the hypothesis that exponential distributions bound intra-market and potentially intra-industry skewness, depending on whether markets are independent.

We can use Sutton's hypothesis to motivate the existence of the Zipf distribution at the aggregate level. Let Ψ be a RV representing the exponential distribution of firm sizes in a particular industry, having PDF

$$\psi\left(s;\bar{s}\right) = \frac{1}{\bar{s}}\exp\left(-\frac{s}{\bar{s}}\right),$$

were s stands for size and $\bar{s}$ is the average size in the industry, a parameter. Now, if the distribution of average sizes across industries, A, is a random variable distributed according to a Pearson type V distribution, with parameters λ and β, i.e.,

$$a\left(\bar{s};\lambda,\beta\right) = \frac{\lambda^{\beta}}{\Gamma\left(\beta\right)\bar{s}^{1+\beta}}\exp\left(-\frac{\lambda}{\bar{s}}\right),$$

then the overall size distribution will follow a modified Pareto law, since

$$f\left(s;\lambda,\beta\right) = \int_0^{\infty} a\left(\bar{s};\lambda,\beta\right)\psi\left(s;\bar{s}\right)d\bar{s} = \beta p^{\beta}\left(\frac{1}{\lambda+s}\right)^{1+\beta}.$$

Preliminary analyses of U.S. data reveals that the distribution of average industry sizes is very well-fit by the Pearson V, as shown in figure 3 below for SIC 4 digit industries in 1997. The largest average firm size exceeds 11,000 employees (department stores) in these data.

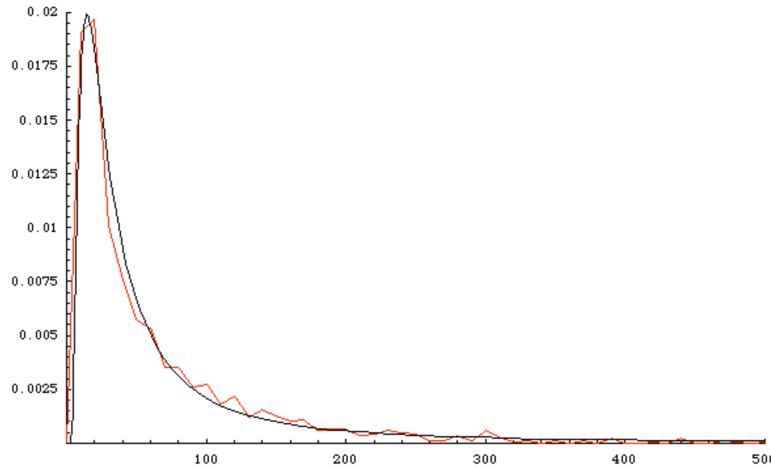

Figure 3: Distribution of U.S. average industry size for SIC 4 digit data in 1997

Overall, this result—that firm sizes are approximately exponentially distributed (Sutton's hypothesis) and that average firm size across industries is Pearson type V distributed, gives us a better understanding of how the Zipf distribution arises. However, in essence this simply pushes the question of the origin of skew sizes back one level, to the origin of industry sizes.

IV The Fables

Beginning with Gibrat (1931), and continuing with Simon (1955), Steindl (1965) and others, explanations of firm sizes have been proferred based on purely random processes. These have come to be known as stochastic growth models. Each of these generates skew distributions, the lognormal in the case of the Gibrat process, the Yule distribution in the case of the Simon model, and a Pareto in the work of Steindl. In order to get the size distribution to agree with empirical data, these models must be 'tuned' parametrically. This is not problematical conceptually, although it is viewed as a drawback by some (e.g., Krugman 1996).

Alternatively, a model that does not require parameter adjustment to produce the Zipf distribution is the so-called Kesten process (1973). It has been proposed as an explanation of city sizes (Gabaix 1999) and firm sizes (Axtell 2001). Call $\gamma(t)$ a random growth rate at time *t*. The Kesten process is essentially the Gibrat random growth process with a lower bound on size. Call $S_i(t)$ the size of the i^{th} entity at time *t*. Then its size at time *t+1* is

$$S_i(t+1) = \max\left[s_0, \gamma(t)S_i(t)\right]$$

Analysis reveals that this process yields a Pareto distribution of sizes, with exponent near unity.

However, some aspects of this process seem to be quite unrealistic. Specifically, since the number of firms is held constant in this process, it follows that all the variation in average size is due to fluctuations in the total size of the system. Now, this may be a reasonable assumption for city formation models, where the number of cities is more or less fixed, but for firm formation it is completely inappropriate.

There are many variations on this basic firm growth process that are more plausible, involving firm entry and exit, serially correlated growth rates, growth rate variance that changes with firm size, and so on. Indeed, there is a large 'equivalence class' of models that yield skew distributions of firm size, but none of these models contain any significant amount of economic reasoning, suggesting that random growth processes may not be the whole story.

What is needed instead is a model in which entry and exit are endogenous. Such a model, written at the agent level, exists and closely reproduces the empirical data (Axtell 1999). What seems like a better assumption in this case is to use a constant (or slowly growing) population and permit the number of firms to ebb and flow. This moves the origin of fluctuations in the average firm size to variations in the denominator, i.e., the number of firms, since the population is constant.

V The Fantasies

If it is really possible to explain the firm size distribution without recourse to economics, as the fable of random growth suggests, it might be possible to come up with other non-economic explanations that are perhaps less fabulous. One such explanation is the subject of the present section. While this idea is quite simple to state in words, it turns out to involve incredible calculations, computations so fantastic that this explanation will, at first blush, seem little more than fantasy.

1 Set Partitions: Firms Composed of Unique Workers

Consider the set of all people hired as employees in the U.S. in a particular year, and call this set E, with $|E| = n$. A partition, P, of E is a set of non-empty, pairwise disjoint subsets of E having the property that their union is E; $|P| = k$. Such subsets are commonly called blocks, classes or parts. Symbolically, the blocks B_i, with $i \in \{1, 2,\ldots, k\}$, form a valid partition P of E when

$$B_i \cap B_j = \emptyset \forall i \neq j \text{ with } \bigcup_{i=1}^{k} B_i = E$$

We will consider each block to be a firm—employees are only permitted to have one job—and a partition to be the set of firms in an economy, with the union of all firms being the set of all employees.

For a given set, E, the number of partitions into k blocks is given by the Stirling numbers of the second kind, $S(n,k)$. Expressinos for $S(n, k)$ are known in both recursive and closed form (Bogart 2000: 100, 127):

$$S(n,k) = \sum_{i=0}^{n-1} \binom{n-1}{i} S(i,k-1) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n$$

The asymptotic behavior of $S(n, k)$ has been known since at least the time of Laplace (Odlyzko and Richmond 1985). For each n the Stirling numbers are unimodal in k, as illustrated in figure 4 for $n = 100$.

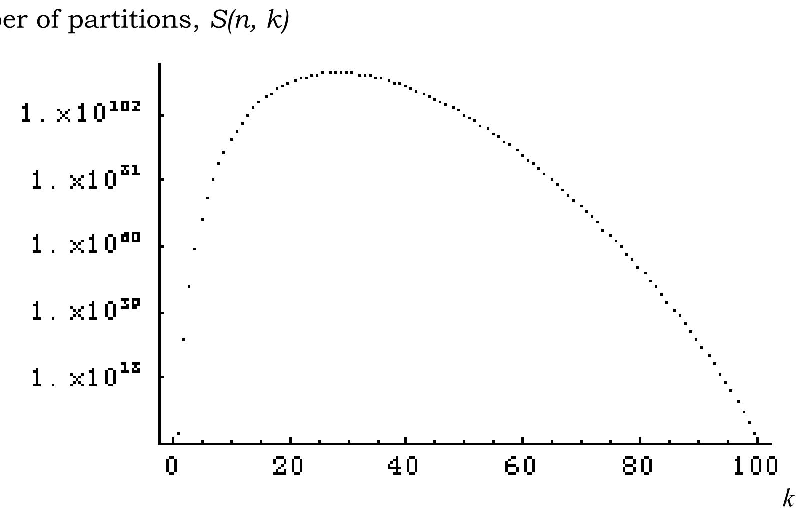

Figure 4: Stirling number of the second kind, $S(n, k)$, for $n = 100$ and $k \in \{1, 100\}$; note the unimodal character of this function

Note that the ordinate in this figure is in logarithmic coordinates.

We can represent $k*$, the number of partitions (firms) that extremizes $S(n, k)$ for fixed n (number of employees) as follows.

$$k^* = \arg\max_k \left[S(n,k)\right]$$

Unfortunately, we cannot explicitly compute it for n corresponding to the number of workers in the U.S., for the quantities involved are simply too vast. However it is known that k^* is non-decreasing in n. It turns out that there is an explicit formula for a tight bound on k^* in the combinatorics literature (Canfield and Pomerance 2002):

$$k^* \in \left\{\left\lfloor e^r - 1\right\rfloor, \left\lceil e^r - 1\right\rceil\right\}$$

where r solves

$$re^r = n$$

Evaluating this expression at $n = 105 \times 10^6$, the total number of employees in the U.S. in 1997, yields $k^* \sim 6{,}699{,}470$, so the average firm size, n/k^* is 15.7. This is not too far from the empirical value of 19.0 Furthermore, over the past decade the average firm size has expanded slightly as both the workforce and the number of firms have grown. Such a trend is evident in this formalism as well, for as n increases the value k^* increases, albeit more slowly, and so n/k^* increases as well. It would be interesting to know the distribution of block sizes at the optimal value of k^*, but apparently this is not well understood. For random integer partitions (as opposed to set partitions) this is known (Vershik and Yakubovich 2001) and is essentially exponential (Gibbsian).

2 Maximum Entropy Partitions

Aspects of this formalism are related to the following purely statistical considerations. Define the fraction of all firms having size *i* as p_i. Now, extremize the so-called non-extensive entropy of the probability distribution, i.e.,

$$S_q \propto \frac{1 - \sum_{i=1}^{k} p_i^q}{1 - q}$$

where *c* is a constant, such that

$$\sum_{i=1}^{k} p_i = 1 \text{ and } \sum_{i=1}^{k} i p_i = n$$

The first constraint merely preserves probability and the second constraint keeps the number of workers constant. It turns out that the equilibrium firm size distribution is then

$$p_i \propto \left[1 - (1-q)i\right]^{-1/(1-q)}.$$

For $q = 2/3$, $1/(1\text{-}q) = 2$, and so the Zipf law is recovered. More work needs to go into understanding the meaning of *q* and the economic content, if any, of this formalism.

VI Summary and Conclusions

It has been argued that the *facts* on firm sizes are partially understood through *formulas*, partially through *fables* of stochastic growth, and possibly via the *fantasy* of randomly partitioning workers into firms. Today there is a wealth of new firm data coming on-line, data that should help substantially to restrict the class of industrial organization theories to those that are empirically relevant.

While much progress has been achieved in better understanding the firm size distribution and its implications for models, we have just begun to scratch the surface of its ultimate policy relevance. High on the policy agenda is to assault common conceptions of the 'typical firm' in order to help decision-makers craft policy that helps the entire economy and not just some particular slice of the overall firm size distribution.

Curiously, intra-country distributions of city sizes in modern industrial nations appear to nearly coincide with the distribution of firm sizes—see Gabaix (1999) for a recent analysis. This leads naturally to questions as to whether the existence of one of these distributions necessarily determines the other, or if they are jointly determined by some third process. In recent work with Richard Florida we demonstrate that skew firm sizes can generate skew city sizes.

While dual focus on small businesses and large industries may be a good strategy, given the current state of understanding of business dynamics, it is unlikely to be the optimal strategy in the long run. For the growth of business firms seems to be largely and interestingly independent of size, meaning that the intrinsic dynamics of firms are really the result of accreting employees one at a time. When this process is fully understood it will be meaningless to have either a Small Business Administration or a Large Business Adminstration, just as it would make little sense to have two Departments of Health and Human Services, one for people who are shorter than average and one for people who are taller than average. Rather, a deep understanding of business firms will surely have as its cornerstone the notion that there are universal aspects to firm growth, just as there are universal features in tree growth, across species, regions, habitats and climates. No two trees are identical, as are no two firms, but all trees use photosynthesis to turn sunlight into biochemicals, just as all firms use technology to turn investment into profits. Once the large-scale features of firm growth are more fully understood we can begin unpacking the complex micro-machinery of organizational form and function that makes firms alive.

References

Asplund, M. (1998). On the Size Distributions of Firms and Markets. Department of Economics working paper. Stockholm, Sweden, Stockholm School of Economics.
Axtell, R. L. (1999). The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions. Working paper. Santa Fe, N.M., Santa Fe Institute.
Axtell, R. L. (2001). "Zipf Distribution of U.S. Firm Sizes." Science **293**(5536): 1818-20.
Bogart, K. P. (2000). Introductory Combinatorics. New York, N.Y., Harcourt Academic Press.
Canfield, E. R. and C. Pomerance (2002). "On the Problem of Uniqueness for the Maximum Stirling Number(s) of the Second Kind." INTEGERS: Electronic Journal of Combinatorial Number Theory **2**: #A01.
Carroll, G. R. and M. T. Hannan (2000). The Demography of Corporations and Industries. Princeton, N.J., Princeton University Press.
De Vany, A. (2004). Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry. New York, N.Y., Routledge.
De Vany, A. and D. Walls (2002). "Does Hollywood Make Too Many R-rated Movies? Risk, Stochastic Dominance, and the Illusion of Expectation." The Journal of Business **75**(3): 425-451.
De Vany, A. and D. Walls (2004). "Motion Picture Profit, the Stable Paretian Hypothesis, and the Curse of the Superstar." Journal of Economic Dynamics and Control **28**(6).
Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Volume II. New York, N.Y., John Wiley & Sons.
Florence, P. S. (1953). The Logic of British and American Industry: A Realistic Analysis of Economic Structure and Government. Chapel Hill, N.C., The University of North Carolina Press.
Gabaix, X. (1999). "Zipf's Law for Cities: An Explanation." Quarterly Journal of Economics **114**(3): 739-767.
Gibrat, R. (1931). Les Inegalieies Economiques; Applications: Aux Ineqalities des Richesses, a la Concentartion des Entreprises, Aux Populations des Villes, Aux Statistiques des Families, etc., d'une Loi Nouvelle, La Loi de l'Effet Proportionnel. Paris, Librarie du Recueil Sirey.
Ijiri, Y. and H. A. Simon (1977). Skew Distributions and the Sizes of Business Firms. New York, N.Y., North-Holland.
Johnson, N. L., S. Kotz and N. Balakrishnan (1994). Continuous Univariate Distributions. New York, N.Y., John Wiley & Sons.
Johnson, N. L., S. Kotz and A. W. Kemp (1993). Univariate Discrete Distributions. New York, N.Y., John Wiley & Sons.

Kesten, H. (1973). "Random Difference Equations and Renewal Theory for Products of Random Matrices." Acta Mathematica **CXXXI**: 207-248.
Krugman, P. (1996). The Self-Organizing Economy. New York, N.Y., Blackwell.
Kumar, K. B., R. G. Rajan and L. Zingales (1999). What Determines Firm Size? Working paper. Chicago, Ill., Graduate School of Business, University of Chicago.
Lucas, R. E., Jr. (1978). "On the Size Distribution of Business Firms." Bell Journal of Economics **9**: 508-523.
Mandelbrot, B. (1997). Fractals and Scaling in Finance: Discontinuity, Concentration and Risk. New York, N.Y., Springer-Verlag.
Odlyzko, A. M. and L. B. Richmond (1985). "On the Number of Distinct Block Sizes in Partitions of a Set." Journal of Combinatorial Theory A **38**: 170-181.
Pryor, F. L. (2001). "Will Most of Us be Working for Giant Enterprises by 2028." Journal of Economic Behavior and Organization **44**(4): 363-382.
Quandt, R. (1966). "On the Size Distribution of Firms." American Economic Review **56**: 416-432.
Samorodnitsky, G. and M. S. Taqqu (1994). Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Washington, D.C., Chapman & Hall.
Scherer, F. M. (1970). Industrial Market Structure and Economic Performance. Chicago, Rand McNally.
Scherer, F. M. (1980). Industrial Market Structure and Economic Performance. Chicago, Rand McNally.
Schmalansee, R. (1989). Inter-Industry Studies of Structure and Performance. Handbook of Industrial Organization, Volume II. R. Schmalansee and R. D. Willig. New York, N.Y., North-Holland. **2**.
Shy, O. (1995). Industrial Organization: Theory and Applications. Cambridge, Mass., MIT Press.
Simon, H. A. (1955). "On a Class of Skew Distribution Functions." Biometrika **42**: 425-440.
Simon, H. A. (1957). Models of Man: Social and Rational. New York, N.Y., John Wiley & Sons, Inc.
Simon, H. A. and C. Bonini (1958). "The Size Distribution of Business Firms." American Economic Review **48**(4): 607-617.
Steindl, J. (1965). Random Processes and the Growth of Firms. New York, N.Y., Hafner Publishing Company.
Sutton, J. (1997). "Gibrat's Legacy." Journal of Economic Literature **XXXV**(1): 40-59.
Sutton, J. (1998). Technology and Market Structure. Cambridge, Mass., MIT Press.
Tirole, J. (1988). The Theory of Industrial Organization. Cambridge, Mass., MIT Press.

Vershik, A. M. and Y. Yakubovich (2001). "The Limit Shape and Fluctuations of Random Partitions of Naturals with Fixed Number of Summands." Moscow Mathematical Journal **1**(3): 457-468.