Research in Economics 70 (2016) 536–557

Contents lists available at ScienceDirect

Research in Economics

journal homepage: www.elsevier.com/locate/rie

A statistical physics implementation of Coase's theory of the firm ☆

Sergey V. Buldyrev [a], Michael A. Salinger [b,*], H. Eugene Stanley [c]

[a] Department of Physics, Yeshiva University, New York, United States
[b] Department of Markets, Public Policy and Law, Questrom School of Business, Boston University, United States
[c] Department of Physics and Center for Polymer Studies, Boston University, United States

ARTICLE INFO

Article history:
Received 15 June 2016
Accepted 5 August 2016
Available online 30 August 2016

Keywords:
Theory of the firm
Firm growth
Statistical mechanics

ABSTRACT

We present a stochastic, dynamic model of firm growth that captures the essential features of Coase's theory of the firm and reproduces important statistical regularities in firm size and growth. For the model to generate these statistical regularities, the parameters must be tuned so that firms involved in "unrelated" activities evolve. Thus, at the same time that the model predicts the statistical properties of firm growth, it suggests that attempts to validate Coase's theory at the level of the individual firm might be futile. The model draws on models of critical phenomena from statistical physics, the motivation being that the observed statistical properties of firm growth are similar to the statistical properties of physical systems near their critical point.

1. Introduction

In this article, we present a statistical physics[1] model of firm size and growth that is motivated by Coase (1937) theory of the firm and that reproduces observed features of the statistical properties of firm size and growth. The essence of the statistical physics approach is to relate the macroscopic statistical properties of a system to a simple stochastic process followed by the large number of individual system components. In the tradition of Herbert Simon,[2] the assumptions about the behavior of individual firms reflect purposeful economic behavior, but we do not assume that firms optimize nor do we assume that the system is ever in equilibrium.

We propose two reasons – one positive and one negative – for revisiting the methodological debates raised by Simon both generally and specifically with respect to understanding the boundaries of the firm. The positive reason is that our work is motivated by a set of empirical results not available at the time of the earlier debates.[3] As we will describe in Sections 2 and 3, the distribution of firm growth rates have scaling properties that are similar to those of physical systems near their critical points. Once thought to be rare phenomena only observed under laboratory conditions, it is now accepted that critical phenomena naturally occur in a wide variety of contexts. An important tenet of modern physics is that a common model can be useful for studying diverse physical phenomena that fall into a common universality class.

☆ We have benefited from comments by L.A.N. Amaral, Ulrich Doraszelski, Shlomo Havlin, and Oliver Williamson. We gratefully acknowledge research support from NSF Grants SES-0113103, CMMI 11-25290, PHY 15-05000, and CHE-1213217. grant SES-0113103.
* Corresponding author.
[1] For an introduction, see Stauffer and Stanley (1996).
[2] For models of bounded rationality, see Simon (1982). For work specifically related to firm size and growth, see Simon and Bonini (1958) and Ijiri and Simon (1977).
[3] See Cyert and March (1963) for a defense of models that do not entail optimization. See Machlup (1967) for a defense of optimizing models.

http://dx.doi.org/10.1016/j.rie.2016.08.005
1090-9443/

Remarkably, the lattice models that have been used to understand a wide range of physical phenomena provide a natural framework to formalize Coase's theory of the firm. Much of Simon's work concerns explaining the shape of the distributions of firm size. The distribution of firm size has been observed to have an upper tail fatter than that of a Gaussian. Simon explored growth processes that generated the observed types of distributions. In contrast, we are starting with and seeking to explain a richer and more surprising set of empirical results. The statistical results are examples of "power laws," which Gabaix (2016) has recently argued are prevalent and stable in economic data and demanding of theoretical explanations.

The negative reason for considering statistical physics models is that progress in explaining the allocation of activity among firms using Coase's framework has been much slower than one might have expected.[4] One challenge has been to identify exactly what the theory predicts. Without being able to measure contractual and organization costs, the hypothesis that firms choose the most efficient contractual alternative has no empirical content. A further complication is that many situations that make contracting difficult might also create complications for internal organization. Since the theory states that the extent of integration depends on the *difference* between the two, any valid test of the theory might require measurements much more precise than we can hope to achieve. So far, the closest the literature has come to a generally accepted result is that vertical integration is likely when firms must invest in durable, transaction-specific assets. Under such circumstances, writing arms-length contracts preventing opportunistic behavior is difficult. While some support for this hypothesis exists, it is weaker than is generally acknowledged. At most, the evidence identifies asset specificity as one factor in the vertical integration decision in fairly narrowly defined cases, and vertical integration is just one aspect of the scope of firms.

The model we present below both predicts the statistical properties of firm size and growth and suggests that there are likely to be limits to our ability to understand with any precision which activities co-exist within the same firm. Ideally, a model of firm size and growth would explain which activities coexist within the same firm. For example, it would explain why until 2011 General Electric (GE) operated both the broadcast network NBC and sold jet engines. Even if we could precisely measure organizational and contracting costs, it seems unlikely that any model would yield such a prediction. While our model would not predict any precise combination, it would predict that pairs of largely unrelated activities coexist within the same firm with a much higher probability than one might initially expect. Even if NBC and GE's jet engines are not good fits, operating NBC might fit with manufacturing television sets. Manufacturing television sets might fit with manufacturing other electrical appliances, which in turn might fit with manufacturing electrical generation equipment. Finally, producing electrical generation equipment might fit with producing jet engines. That is, if one can construct a map of activities in which one can measure how related two activities are, then it is possible for there to be a path that connects two largely unrelated activities.

Using models originally designed to explain physical phenomena to understand economic or other social systems raises a fundamental issue. There are physical systems, like magnets, in which order (or "coordination") occurs in a large system with only local interactions. One atom can only directly interact with the atoms immediately surrounding it. With rational beings (who can anticipate the future) and low-cost communication and transportation, it is not clear that economic interactions must only be local in any sense (time, geography, or "product space"). Yet, within this model, the statistical properties of firm growth are similar to the statistical properties of physical systems with only local interactions. As a purely logical matter, this result does not prove that firms are boundedly rational. However, until someone develops a model of rational behavior that predicts observed behavior as well, the result provides support for the proposition that models of bounded rationality are useful.

The remainder of this paper is organized as follows. Section 2 discusses precedents for our work in economics. Section 3 concerns statistical physics: it presents some key concepts from statistical physics, explains why these models are appropriate for modeling firm size and growth, and discusses some of the issues involved in the precise choice of model. Section 4 describes the assumptions of the model, Section 5 present the results, and Section 6 contains some concluding comments.

2. Prior literature

2.1. The distribution of firm size

In *Inégalités Economiques*, Gibrat (1931) presented striking evidence that the distribution of firm size at different times and within different "populations" was approximately lognormal; and showed the lognormality could be generated by a process in which the distribution of growth rates is independent of initial size.[5] This work was an early example of the use of statistical physics in economics. The model was one of a stochastic process, and the objective of the model was to explain the shape of a distribution that emerged from it.

This result is much more remarkable than economists have generally acknowledged. Perhaps the reason it has not generated much recent interest is that systematic differences in firm sizes across markets are related to industry

[4] See Section 2C for a further discussion and for references.

[5] Our characterization of Gibrat's book is based on Sutton (1997).

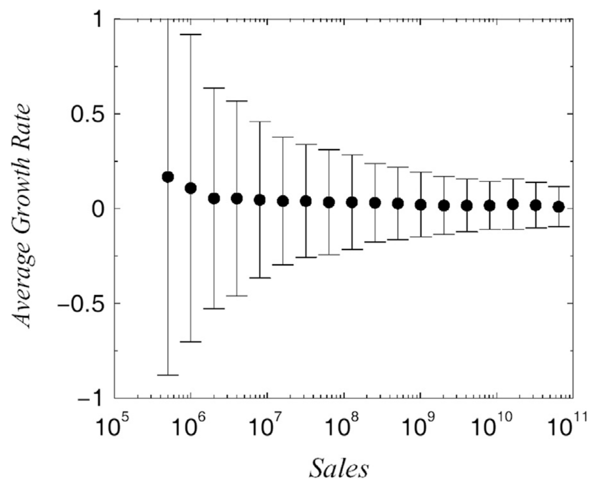


Fig. 1. Initial size and mean growth rates.

characteristics like production scale economies and the opportunities to advertise. If the distribution of firm sizes within an industry are driven by these factors, then the overall distribution of firm size would seem to be determined by the composition of output. As a consequence, there is no reason to believe that this distribution would remain stable over time or that it would necessarily be subject to modeling.

The attention that should appropriately be paid to the over-all distribution of firm size raises the same fundamental issue raised by Coase: what is the nature of the firm? Prior to Coase, analysis of firm size was based on the standard microeconomic model of the firm and therefore viewed firm size as being driven by demand, cost, and technology factors specific to the market in which the firm operated. Coase's point was that the analysis missed the essential feature of the boundaries of firms. The same issues arise in assessing different approaches to firm dynamics. Consider, in particular, whether the work of Pakes and his co-authors is the proper approach to understanding firm dynamics.[6] Like the work on firm size and growth that preceded the transactions cost economics (TCE) approach, their work is rooted in the microeconomic theory of the firm. Its contribution is to base the analysis on a model of oligopoly and to take explicit account of innovative activity on firm dynamics. As important as this work is, it is a model of business unit dynamics, not firm dynamics. To continue with the example of General Electric, that framework might be useful for understanding General Electric's jet engine business and the evolution of jet engine prices and technology over time. How General Electric's jet engine business fares certainly has some effect on the size of General Electric as a corporation. However, one simply cannot understand the size of General Electric without trying to understand the set of businesses it chooses to be in. General Electric is arguably an outlier by being in a diverse set of businesses, but virtually all large corporations consist of multiple business units that could conceivably be stand-alone firms.

One might argue that the relative neglect of the distribution of firm size is because there are no regularities to explain. Gibrat found that the log normal distribution fit the data well for different time periods and different regions at both the sector and individual industry level. Yet, subsequent work has found that no single distribution provides a suitable fit to all samples.[7] Whether or not the distributions within industries are lognormal has, however, little bearing on whether we should be interested in the approximate log normality of firm size in a multi-industry sector.[8]

The more fundamental question is whether there is compelling evidence that the overall distribution of firm size is log normal (or consistently has some other distribution). While the authors emphasized the deviations from log normality, the results in Stanley et al. (1995) can be interpreted as support for the approximate log normality of publicly-traded manufacturing firms. Using data on both publicly-traded and privately-held firms, Axtell (2001) found that the Pareto distribution fit the data better. In particular, within the entire population of firms, the modal firm has one employee. This feature of the distribution is simply not consistent with any distributional form with an interior mode.

Axtell's and Stanley et al.'s findings are not necessarily inconsistent because one concerns publicly-traded firms and one does not. To capture both distributions in a model, one would need to incorporate into it the decision to go public. This is beyond the scope of this paper.

[6] See, eg., Pakes and McGuire (1994), Ericson and Pakes (1995) and Berry et al. (1995).

[7] See Quandt (1966).

[8] Lucas (1978) presents a model of the over-all distribution of firms size that abstracts away from the industry-specific cost and demand factors that dominate most of the economics literature on firm size.

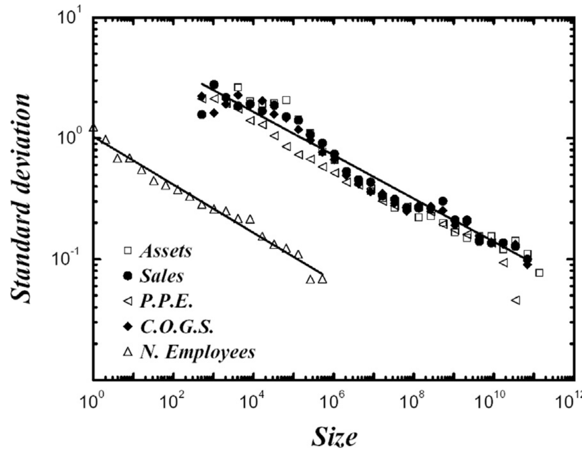


Fig. 2. Initial size and standard deviation of growth rate.

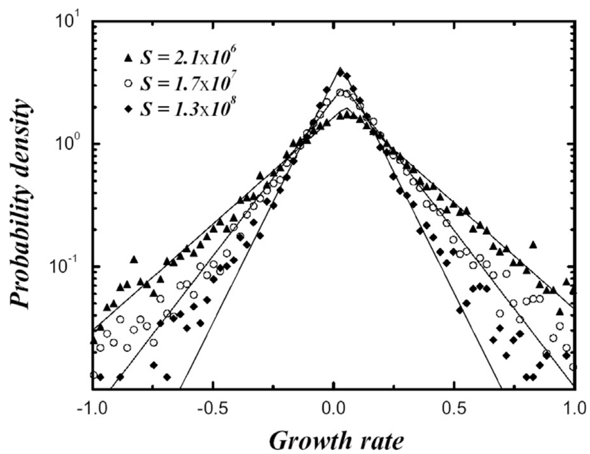


Fig. 3. Distributions of growth rates conditional on initial size.

2.2. The distribution of firm growth rates

Gibrat's empirical findings were that the distribution of firm size was lognormal. He then showed that such a distribution would result if the distribution of growth rates was identically distributed across firms. He did not study empirically whether the distribution of growth rates depends on size (and other factors), but others have since then.

The most studied relationship concerns mean growth rates and size. No clear consensus has emerged. Several studies[9] report that smaller firms grow, on average, faster than larger firms. Hall (1987) and Evans (1987) have addressed whether the negative relationship between firm size and mean growth rates could be attributed to attrition bias, since small firms that shrink are more likely to disappear than large ones. Both concluded that small firms grow faster on average than large firms even after accounting for attrition bias. In contrast, however, Samuels (1965), Prais (1976), and Dunne and Hughes (1994) found that larger firms, on average, grow faster.

Fig. 1 is similar to the plot of mean growth rates of Compustat manufacturing firms from Amaral et al. (1997).[10] It shows mean logarithmic growth rates of real sales for different size classes.[11] In the graph, the average growth rates are larger for the smallest size classes. A regression of growth on size would likely generate a negative and statistically significant slope coefficient. If one eliminates the smallest three size classes, however, the mean growth rate does not appear to be a

[9] See Hymer and Pashigan (1962) and Mansfield (1962).

[10] Fig. 4(b) in Amaral et al. (1997) shows mean growth rates for several different size measures. The approximate independence of mean growth rates and firm size holds for all the measures considered there.

[11] The data are Compustat firms with SIC codes 2000–3999 over the period 1974–1993. The data for all 19 years of growth rates are pooled after adjustment to 1987 constant dollars based on the GDP price deflator and then placed into bins according to initial size. The width of each bin on a logarithmic scale is 2 (meaning that the upper bound of each bin is twice the lower bound). The error bars represent one standard deviation.

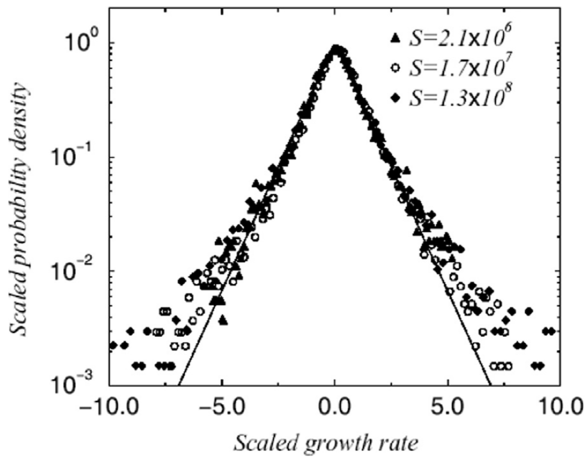


Fig. 4. Data collapse.

decreasing function of size. The higher average growth rates for the smaller firms could be attributable to attrition bias.[12] Even if it is not, however, the absence of any apparent relationship above the smallest size classes indicates that there is no strong relationship between mean growth rates and size. Scherer (1980) reached a similar conclusion as apparently did Sutton (1997).

As several studies have shown [Hart (1962), Hymer and Pashigan (1962), Singh and Whittington (1968, 1975), and Hall (1987)], the standard deviation of growth rates does decrease with size. The strength and nature of the relationship has not been appreciated until recently. Fig. 2 is a log–log plot of the standard deviation of growth rates against size (using several different measures of size) from Amaral et al. (1997).[13] It is nearly[14] linear over many orders of magnitude with a slope of approximately $-1/6$. Gibrat's law would imply a slope of 0. If firms are collections of business units with independent growth rates and if firm size was determined by the number of such units within the firm, then the slope would be $-1/2$.[15]

Work by economists on the distribution of firm growth has centered entirely on correlates with the first two moments of the distribution. In the analysis of complex systems in physics, however, it is common to study the entire distribution. Fig. 3 shows a plot of the probability density function of firm growth rates for three different size classes of firms first reported in Amaral et al. (1997).[16] As one would expect from the relationship between firm size and the standard deviation of growth rates, the widest distribution is for the smallest size class. What is striking about these distributions is that they have the same non-Gaussian shape. Fig. 4 contains what is called a “data collapse.” It shows the distributions for three different size classes, but each is scaled by the standard deviation.[17] The different distributions essentially lie on top of one another.

This result – i.e., the data collapse – suggests that the firm growth process may be an example of a dynamic system near its critical point. We will discuss critical phenomena in more detail in Section 3. For now, two points will suffice. First, the model we develop here is a statistical mechanical model of a system near its critical point. The result of the data collapse is the empirical justification for using such a model. Second, whatever the merits of the reasons for the economics profession largely dismissing the Gibrat/Simon tradition, these results provide a justification for reviving it. In broad terms, there were two arguments for a statistical mechanical approach to economics. First, there were interesting empirical regularities to explain. Second, similar kinds of regularities showed up in sufficiently diverse settings that the explanation could not plausibly reside in the details of the phenomenon. As Simon once argued, “no one supposes that there is any connexion [sic] between horse-kicks suffered by soldiers in the German army and blood cells on a microscopic slide other than that the same urn scheme provides a satisfactory abstract model of both phenomena.”[18] How persuasive this argument is turns on the extent to which the similarity of the empirical findings is truly surprising. Many phenomena have single-peaked distributions. Given such a distribution, there are a relatively small number of candidates for how to characterize them. If one examines many distributions on diverse phenomena and characterizes each according to whether they are normal,

[12] Hall (1987) and Evans (1987) concluded otherwise, but their conclusions rest on specific assumptions about the censoring process.

[13] This figure is Fig. 4(c) in Amaral et al. (1997). In the legend, “P.P.E.” stands for property, plant, and equipment, “C.O.G.S.” stands for cost of goods sold and “N. Employees” stands for number of employees. The results in Amaral et al. (1997) are slight refinements on those originally reported in Stanley et al. (1996).

[14] The plots in Fig. 2 do entail some deviations from linearity. However, the deviations would require at least a third degree polynomial. Moreover, the nature of the deviations is different for the different measures of firm size.

[15] See Schmalensee (1989), p. 995.

[16] In Fig. 3, the bins span a factor of 8, with the value in the legend being the lower bound. That is, the initial size for the data denoted by a triangle are for $8^6 < S_0 < 8^7$, where S_0 is real sales during the first year of the period.

[17] Specifically, the horizontal axis is $r_{scal} = \sqrt{2}[r_1 - \overline{r}_1(s_0)]/\sigma_1(s_0)$, where s_0 is the logarithm of initial size $\overline{r}_1(s_0)$ and $\sigma_1(s_0)$ are the mean and standard deviation of growth rates for firms in the size class in which s_0 falls. The vertical axis is $p_{scal} = \sqrt{2}\sigma_1(s_0)p(r_1|s_0)$.

[18] Simon (1982), p. 425.

exponential, or power law, of course many of them will fall into the same category even if they have nothing to do with each other.[19] The scaling relationships revealed in Figs. 2–4 are a more surprising set of results than having a similarly shaped distribution and are therefore more demanding of an explanation.

These empirical results not only provide a justification for developing a model to explain the statistical properties of firm growth, but also have implications for what statistical properties are most important to explain. The Jovanovic (1982) model of the nature of the firm predicts negative relationships between firm size and age on one hand and mean growth rates on the other. Yet, as Fig. 1 illustrates, the relationship between initial size and mean growth is weak; and firm age is an inherently ambiguous concept.[20] The model only predicts these results qualitatively. Moreover, the Jovanovic model is quite complicated. We question whether such weak empirical findings are worthy of such a sophisticated explanation. The data collapse and the strength of the size-standard deviation relationship are much more striking statistical properties of firm growth than the size-mean growth rate relationship and therefore more demanding of a theoretical explanation.

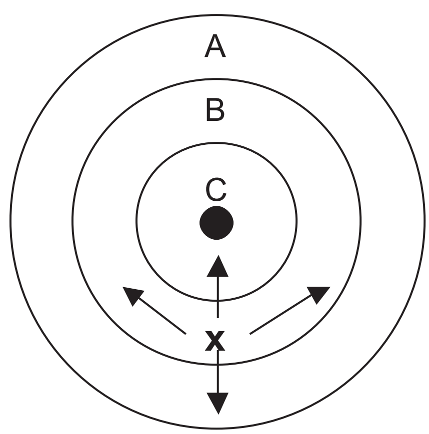
2.3. Transaction cost economics

2.3.1. The nature of the firm

Although the article was largely neglected for many years, Coase (1937) ultimately caused a revolution in the way economists approach questions of the boundaries of firms. Prior to the revival of interest in Coase's article in the mid-1970s, the theory of the firm centered around the role of scale economies in the production of a single good. Coase's insight was that firms supplant arms-length contracts as a mechanism for coordinating economic activity and that activities would be organized within a single firm when the costs of that form of organization are less than contracting costs. The ultimate effect of Coase's article was to divert the focus of analysis from production technology to transactions costs.

While Coase's article describes a new conceptual framework for understanding the boundaries of firms,[21] it does not lay out a complete formal model. The conclusion to the article does, however, provide some thoughts about the nature of such a model. The conclusion contains the following chart which represents a map of economic activity. The concentric circles each represent industries. Coase argues as follows[22]:

> Imagine an entrepreneur who starts controlling exchange transactions from *x*. Now as he extends his activities in the same product (B), the cost of organizing increases until at some point it becomes equal to that of a dissimilar product (A and C).



.

Within this conceptualization, a small area in the diagram represents a transaction.[23] The size of firm is determined by the number of transactions organized within it. A firm can grow either by increasing sales of its existing product, in which case it would occupy a greater fraction of circle B, or by expanding into a new product (C or A). An assumption implicit in this conceptualization is that for a firm currently selling product B, products A and C are the natural activities to expand into.

While Coase described a static equilibrium within this type of model, he anticipated that a more realistic version would be dynamic. As he argued:

---

[19] The difference among these is that, if x is a random variable, the tails are normal if they fall off at the rate of e^{-x^2}, exponential if they fall off at the rate of e^{-x}, and power law if they fall off at the rate of $x^{-\alpha}$.

[20] Lucent Technologies and Time Warner Cable are just two companies for which the age is not clear.

[21] Much of the modern literature on Coase's theory of the firm concerns vertical integration, so our formalization of Coase's observations about the general boundaries of firms might not conform to what many people consider to be "Coase's theory of the firm." (We are grateful to Oliver Williamson for this point.) Of course, the primary motives for our model are to explain a set of statistical regularities and to advance our understanding of the allocation of activity among firms. We find the similarities between this model and the conclusion of Coase's article to be striking, but they do not make the model inherently more compelling.

[22] Coase (1937) at p. 402.

[23] One might conceive of a transaction as being a point, but we describe it as a "small area" so that the number of transactions is finite.

> At the margin, the costs of organizing within the firm will be equal either to the costs of organizing in another firm or to the costs involved in leaving the transaction to be 'organized' by the price mechanism. Business men will be constantly experimenting, controlling more or less, and in this way, equilibrium will be maintained. This gives the position of equilibrium for static analysis. But it is clear that the dynamic factors are also of considerable importance, and an investigation of the effect changes have on the cost of organizing within the firm and on marketing costs generally will enable one to explain why firms get larger and smaller. We thus have a theory of moving equilibrium.[24]

What is striking about this section of Coase's article is that, putting aside the inessential detail of characterizing the space as a set of concentric circles, it describes a modeling approach that is similar to models currently used in statistical physics. In Section 4, we will develop a model that is drawn from statistical physics, captures the approach described in the conclusion to Coase's article, and predicts the empirical findings described in Section 2B.

2.3.2. Building on Coase's work

Google Scholar lists over 33,000 citations to Coase's article. A thorough review of the transactions cost economics literature would require a book. Much of this research starts from the premise that Coase's insight is fundamentally correct and that the key task for subsequent research is to identify and measure the respective costs of internal organization and arms-length contracts. Given the extent of the attention devoted to this issue over the last three decades, we submit that what is most striking is not what we have learned but, rather, how little has been established. This claim is not based just on our own reading of the literature. Summarizing the literature for *Handbook of Industrial Economics*, Williamson (1989) wrote, "recent and continuing headway notwithstanding, transaction cost economics maintains that our understanding of the economic institutions of capitalism … is very primitive."[25] More than 50 years after his article was published, Coase (1988) himself expressed disappointment at the rate of progress.

The 2001 *American Economic Review Papers and Proceedings* contained a set of papers about new empirical directions in the scope of the firm. In his paper in that session, Whinston (2001) wrote that "the strong association that … has [been] found between specificity and integration has made the TCE one of the great success stories in industrial organization over the last 25 years."[26] First of all, one might question how thoroughly established the point is. For many years, "Exhibit 1" supporting the importance of dedicated assets in integration decisions was General Motors' acquisition of Fisher Body. However, in 2000, the *Journal of Law and Economics* published an issue with four papers that challenged the conventional representation of that acquisition [Coase (2000), Freeland (2000), Casadesus-Masanell and Spulber (2000), and Klein (2000)]. If we are not sure of that case, one wonders how confident we can be about the other "documented" instances in which asset specificity led to integration. Even if we take the conventional reading of the literature at face value, all it establishes is that asset specificity is one factor in the vertical integration decision. We do not know that it is the most important factor. Finally, even a complete understanding of the integration decision would not constitute a complete theory of the firm.

One of the difficulties in finding evidence to support Coase's theory is that the predictions of the theory are not clear. For example, Whinston (2001) showed that the "property rights theory," which is one of the approaches to lending content to Coase's basic insight, yields a complicated set of predictions that would require very precise measurements to test. The issues that arose in the general controversy concerning the Fisher Body case also indicate that the predictions even about vertical integration are not so clear. Although his conclusions may be controversial, Freeland argued that the hold-up problems vertical integration is supposed to eliminate were actually more severe within the merged entity.

Even if one believes that our reading of the literature is overly critical of the progress made, it is clear that we are far from realizing what Coase originally stated as his hope for the framework he proposed. In his 1937 article, he wrote:

> Inventions which tend to bring factors of production nearer together, by lessening spatial distribution, tend to increase the size of the firm. Changes like the telephone and telegraph which tend to reduce the cost of organizing spatially will tend to increase the size of the firm. All changes which improve managerial technique will tend to increase the size of the firm.[27]

The vast majority of the work in TCE has focused on the most microscopic level of understanding individual organizations or even individual contracts. Since the publication of Coase's article, we have lived through tremendous changes in the technology of transactions that, based on the theory, we should be able to link to the distribution of firm size. The claims for success for the TCE are at the microanalytic level, not the macro-organizational level we ultimately hope for.

2.4. Other examples of statistical physics approaches in economics

Although by no means the dominant tradition in the field, there is a long history of statistical physics approaches to economics. In light of the increased interest in this approach in recent years, a thorough review is beyond the scope of this

[24] Coase (1937) at pp. 404–5.
[25] Williamson (1989), p. 136.
[26] Whinston (2002), p. 184.
[27] Coase (1937), p. 397.

paper. Instead, we begin by discussing statistical mechanical techniques within the first area of economics in which these techniques were applied: finance. We then briefly mention a variety of areas of economics in which scholars have recently used statistical mechanical techniques before concluding with a longer discussion of statistical mechanical approaches to the scaling relationships for firm growth described in Section 2.

2.4.1. *Financial prices*

Bachelier (1900) thesis was the seminal work in statistical physics. It laid the foundation for much subsequent work in continuous time finance, including option pricing. His work was based on the assumption that security returns follow a random walk.

Mandelbrot (1963) presented evidence that the distribution of many securities returns series were fatter-tailed than the Gaussian and argued for the wider use of the Pareto distribution in finance.

Mantegna and Stanley (1995, 2000) analyze stock returns on a wide variety of time scales (2 min up to 25 days) and develop a scaling relationship of stock returns at different time scales. This technique is particularly powerful for understanding the probability of very rare events because there are many more observations on 2-minute returns than on daily returns. Their analysis suggests that events like the dramatic market collapses such as those in October, 1987, and the market collapse of 2008–9 occur with greater frequency than is suggested by other models.

Other important recent applications of statistical physics to finance are LeBaron (1992) and Brock et al. (1992).

2.4.2. *Other applications*

Bak et al. (1993) and Scheinkman and Woodford (1994) build a model of inventory behavior at successive levels of production. Their model is an example of self-organized criticality and predicts large fluctuations in aggregate output. Krugman (1996) further builds on this analysis.

Zipf's law for city size states that a log–log plot of city size for rank yields a straight line with a slope of -1. See Zipf (1949), Mandelbrot (1965), and, for a more recent analysis, Gabaix (1999).

Lee et al. (1998) show that the statistical properties of growth rates for the GDP of countries have scaling properties that are very similar to the one described above for company growth. That is, on a log–log scale, the distribution of growth rates for countries with similar-sized GDP's is tent shaped, and a log–log plot of the standard deviation of the growth rate against country size yields a nearly linear relationship with a slope of approximately $-1/6$.[28]

In a sweeping set of papers, some solely authored and some co-authored, Steve Durlauf has explored statistical physics approaches to social interactions. See Blume and Durlauf (2001), Brock and Steven (2001), and Durlauf (1996a, 1996b, 1999, 2001).

2.4.3. *Scaling relationships in firm size and growth*

The work presented below is related to Buldyrev et al. (1997), Amaral et al. (1998), Axtell (1999), and Sutton (2002) in that it is designed to explain the same basic set of empirical findings.

Buldyrev et al. (1997) present a model of organizational hierarchy in which decisions get passed down through successive layers. At each decision point in the hierarchy, a manager can either accept the decision of his immediate superior or reject it and make his own independent decision.

Amaral et al. (1998) describe a model of Gibrat-type growth processes at the business unit level and firms consisting of multiple units with uncorrelated growth processes. This model predicts the key empirical findings about firm growth for a wide variety of parameters.

Axtell (1999) presents a model in which firms are teams of individuals who select effort levels and share the output. Holding effort levels constant, adding individuals (or collections of individuals) to a firm increases output. As firms become larger, however, each individual has more of an incentive to shirk because of the sharing rule. The growth dynamics come from having individuals randomly joining and leaving firms. This model implies a Pareto distribution of firm size and the scaling relationships in growth dynamics described above.

Sutton (2002) presents a model in which firms consist of independent, equally-sized units of varying sizes. He assumes that all partitions of the firm are equally likely. For example, a firm of size 4 could consist of four one-unit divisions, two one-unit divisions and one two-unit division, two two-unit divisions, one three-unit and one one-unit division, or one four-unit division. This model is similar to the Amaral et al. model in that it views firms as consisting of units whose growth rates are independent of each other. An important difference between this paper and the others (as well as ours) is that Sutton only analyzes the size-standard deviation relationship rather than the entire shape of the distribution. In addition, he argues that the size-standard deviation relationship varies over time; and he interprets our characterization of the relationship as a bound. In this regard, his work is related to his other work on bounds relationships in industrial economics (Sutton, 1997).

Riccaboni et al. (2008) present a model that is related to Amaral et al. (1998). In their model, firms consist of multiple business units with the number of units varying by firm. The distributions of growth rates of business units are independent and identically distributed; but, since larger firms have more distinct business units than do smaller firms, the standard deviation of the growth rates is smaller for larger firms. If all business units were the same size, in which case the variation

[28] Remarkably, Plerou et al. (1999) find somewhat similar results for university research funding, but with a scaling exponent of $-1/4$ rather than $-1/6$.

in firm size would be due entirely to variation in the number of independent, identically-sized business units within the firm, then the relationship between the logarithm of the standard deviation of firm growth rates with a slope of -0.5. However, the condition that would give rise to a slope of -0.5 would not be stable dynamically. Even if one started with firms consisting of equal-sized business units, the random growth process applied to individual business units would over time cause variation in the size of business units and, as a result, larger firms would tend to have both more and larger business units. Riccaboni et al. show that for a range of reasonable parameters, the slope of the log-log relationship between firm size and the standard deviation of growth rates would generally fall within the range of -0.14 to -0.20.

Given that there are already five models that predicting the same empirical findings, one might question the need for a sixth. However, there is no reason to believe that these findings sufficiently identify the "true" model of the firm. We can reject models that are not consistent with empirical findings, but we cannot accept a model just because it is consistent with them. The primary feature that distinguishes the model we present below is that it is explicitly designed to address the question of which activities are organized within a given firm.

3. Critical phenomena

3.1. Two physical metaphors

There are two examples of critical phenomena in the physical world that provide important insights into the model we are going to develop and the rationale for it. The first is magnets, which serve as a metaphor for the emergence of order from random processes in the absence of any macroscopic coordination. Magnetization occurs when a sufficiently large fraction of the atoms in a material acquire a common orientation. Since atoms orient randomly, it was at one time considered a mystery as to how magnetization could occur. The commonly accepted resolution of the paradox is the Ising model in which the orientation of one atom affects the orientation of the atoms immediately surrounding it.

In a one-dimensional system, local interactions would not be sufficient to generate common orientations of atoms far apart from each other. Let p be the probability that two adjacent items have the same orientation. Now, consider two atoms N atoms away from each other. The probability that the two atoms and all those in between them have the same orientation is p^N, which becomes exponentially small as N becomes large. In a higher dimension system, the probability of an unbroken path of atoms with a common orientation between two distant atoms is not necessarily exponentially small. Even though the probability of an unbroken chain along a particular transmission path shrinks exponentially, the number of possible transmission paths grows by the product of an exponential and a power law. The exponential decay in the probability of an unbroken chain along a particular transmission path and the exponential component of the growth of the number of paths cancel each other out, leaving the power law component of the growth in the number of transmission paths as the equation that governs the relationship between the distance between two atoms and the probability of an unbroken path of atoms with the same orientation connecting them.

There is a particular interest in magnets around their critical point. Heating a magnet causes it to lose its magnetic strength. The critical temperature is the temperature above which a magnet loses all of its magnetization. Moreover, if one heats a magnet to a temperature below its critical temperature and then cools the magnet off to its original temperature, the original strength of the magnetization is restored. If one heats a magnet above its critical point, however, then the magnetization is lost even if the temperature is subsequently reduced. As will be discussed below, the behavior of a magnet around its critical temperature has unique statistical properties.

The second physical metaphor is earthquakes. What one would like from a theory of earthquakes is a prediction of when and where earthquakes will occur and their magnitude when they do. A more modest objective is to know the risk of earthquakes of a given size in a given area. Is an earthquake that would cause Los Angeles to drop into the ocean a once in 10^3 year risk, a once in a 10^6 year risk, or a once in a 10^9 year risk? It is inherent in rare phenomena that direct observation of their frequency takes a very long time. Because earthquakes have the statistical scaling properties associated with critical phenomena, one uses observations on the (higher) frequency of smaller earthquakes to project the risk of larger ones. Earthquakes stand as a metaphor for what a theory of the firm can and cannot do. Just as it is not possible to predict exactly when and where the next big earthquake will hit, it may not be possible to predict with precision what activities will fall together within a given firm. Still, as with earthquakes, it might be possible to accomplish the more modest but nonetheless substantial task of understanding the distribution of size and growth rates.

An important difference between magnets on the one hand and earthquakes on the other is that the former reach a critical point only under controlled conditions whereas earthquakes are an example of naturally occurring critical phenomenon or "self-organized criticality." Other examples of self-organized criticality are sand piles and invasion of liquid into porous disordered media. The hypothesis that the allocation of activity among firms is a critical phenomenon requires that the economy self-tune to such a point.

3.2. *Statistical properties of critical phenomena – scaling*

As described above, the critical point of a magnet is easily observed. A magnet placed over a set of thumbtacks picks up some number of them. Heat the magnet and, initially, some of the tacks fall off. Heat it enough and they all fall off. The temperature where they all fall off is the critical temperature.

There are many other examples of physical systems that are now considered to be examples of critical phenomena even though they do not have the striking physical manifestation of a magnet at its critical point. Why? How does something get characterized as a critical phenomenon without the striking physical manifestation of a magnet or water? The answer to this question is key to our claim that we should think of the allocation of activity within firms as an example of a critical phenomenon, which is in turn key to our choice of how to model firm size and growth.

Systems near a critical point exhibit statistical behavior that is strikingly different from systems that are not near a critical point. Imagine taking two magnets of the same size but made of different materials, for example, iron and nickel. One can apply the same magnetic field to the two magnets, and examine the strength of the magnetization as a function of the temperature. One can think of the strength of the magnetization as the number of thumbtacks the magnet would pick up. If one were to plot the number of thumbtacks against the temperature for the two materials, the number of thumbtacks at a given temperature, the over-all shape of the relationships, and the critical temperature would be different.

There would, however, be one feature of the graphs that would be the same. Let T be the temperature and T_c be the critical temperature. Near the critical point, the curve would be asymptotically vertical. In addition, if one were to plot the log of the number of thumbtacks against the log of $(T-T_c)/T_c$, both graphs would be linear over approximately three orders of magnitude and *they would have the same slope*. That is, even though nickel and iron are much different, both the linearity on a log–log scale and the slope of the relationship are common or universal. Moreover, this exponent would apply not just to nickel and iron but to a wide variety of magnetic materials. It is widely accepted in physics that these universal features of systems are mathematical properties of systems of strongly interacting sub-units and do not depend on many of the details of the systems that one might expect to matter. It is this principle that makes us believe that there might be universal features of the statistical properties of the growth of quite diverse firms.

There is another feature of systems near the critical point that is of interest. Again, consider the nickel and iron magnets. Suppose one were to partition the magnets first into successive powers of 2 (starting with 2^{10} so that one could determine a distribution). Since the distribution of tacks on the magnet would not be perfectly uniform, there would be some distribution of tacks on the different parts of the magnets. Away from the critical point, the distributions would not necessarily have the same shape. The distributions on the smallest scale are not normal. The larger scale objects are the sums of the smaller scale objects. If the realizations for different objects were independent of each other, the distribution of the larger scale objects would be more nearly normal because of the central limit theorem. Near the critical point, however, the distribution of the tacks on different scales would have a common, non-normal shape. Moreover, this shape would be common across nickel and iron.

The statistical technique used to detect this property is the data collapse. To do so, one measures the distribution of some quantity at different scales, then normalizes the distribution by subtracting the mean and dividing by the standard deviation, and then plots the distributions on a common curve. If they collapse onto a single curve, then the distribution is said to be "scale-free".

These two features of critical phenomena - the linearity on a log-log scale of a key parameter over many orders of magnitude and the data collapse - are the statistical features used to identify critical phenomena and they are precisely what we have found with the economic data on the growth of firms.

3.3. *Modeling critical phenomena*

There are a wide variety of critical phenomena for which some form of lattice model is used. The earliest example is the Ising model of a magnet. That model consists of a two-dimensional lattice in which atoms on the lattice have one of two orientations (labeled as positive or negative). The orientation of each atom is chosen randomly. Interactions are introduced by having the probability that a particular atom switch orientations depend on whether the neighboring atoms have the same orientation.[29] Within such a model, a critical point is when the parameter reflecting the interaction among the orientations is tuned just to the point where there are clusters of common orientation spanning the entire lattice.[30] At this point, the "spin–spin" correlation function, which relates the probability that two sites have the same orientation to their distance apart on the lattice, follows a power law function of the distance.[31]

[29] More precisely, let $S(x,y)$ be the "spin" of the node (x,y) where the spin takes on the value either $+1$ or -1, k be the Boltzmann constant and T be temperature. Then the energy at that node is defined as $E(x,y)=-J\,S(x,y)\,[S(x+1,y)+S(x-1,y)+S(x,y+1)+S(x,y-1)]$ and J is the interaction parameter. In a particular iteration of the model, the probability that node (x,y) flips its orientation is 1 if the $E(x,y)$ is positive and $e^{E(x,y)/kT}$ if $E(x,y)$ is negative. See Baxter (1982).

[30] Let $K=J/kT$, where k is the Boltzmann constant, T is temperature, and J is the interaction parameter defined above. The model is tuned to its critical point when T is set so that $\sinh(2K)=1$. See Baxter (1982).

[31] Let $g(r)$ be the correlation of the spins of two points that are a distance r apart. In general, $g(r)e^{-r/\xi}r^{-\eta}$ where $\xi=(T-T_c)^{-\nu}$ is the "characteristic scale," and ν and η are critical exponents. When T is far away from T_c, the exponential term dominates and the correlation function is exponential (and

While the Ising model is suitable for understanding magnets, physicists have modified lattice models to reproduce other examples of critical phenomena. In magnetization, all (or nearly all) atoms end up with the same orientation. The analogy in the economic problem we are modeling would be for all activity to be organized within the same firm. That is not what we observe. The interactions between adjacent sites in the model are necessary to get all the sites to have the same orientation, but they are not necessary to get critical behavior.

Arguably the simplest model of critical phenomena is a percolation model, which is similar to the Ising model but without correlations in the orientation of adjacent nodes. In percolation models, sites have one of two orientations and a cluster is a connected set of sites with the same orientation. Percolation occurs when a single cluster spans the entire lattice, and the critical point represents the boundary between where percolation does and does not occur.

The model we use is a generalization of percolation known as a Potts model. Percolation models allow sites to take on one of two possible states. Potts models allow for any number of possible states. For the problem we are addressing, we believe that a minimum of three states is necessary.

It is important to clearly understand the relationship between the economic problem we are modeling and the models that we are adapting from physics. We are not simply taking a model used in natural science and calling it an economic model. Rather, we are using a lattice model because we believe that it is appropriate for exploring Coase's theory of the firm. We are not choosing a Potts model because we think that the allocation of activity in a firm is like the physical problems addressed by those models. Rather, we have made the assumptions that strike us as being natural for understanding firms, and these assumptions turn out to bear a resemblance to Potts models. As a result, we are able to use existing results about Potts models. In particular, we are able to draw on previous work in selecting the parameters in a Potts model that give critical behavior.

The simplest Potts models do not self-tune to a critical point. A slightly more complicated class of these models, known as "invasion percolation" models, do. In order to justify modeling the behavior of firms as critical phenomena, however, we need to show that the behavior of the model at the critical point reproduces observed empirical phenomena more closely than does the model away from its critical point. To do so, we need a model that does not self-tune. Extending our work to a self-tuning model that resembles invasion percolation models remains a subject for future research.

4. The model

The model we present below is designed to capture the spirit of the dynamic extension of Coase's theory described both in the conclusion to his article and in Section 2-A-1 above. Such a model must have certain essential features. First, the model must assume the existence of a set of transactions (or potential transactions) that are candidates to be organized within firms.[32] The size of a firm is then the quantity of transactions that it organizes. Second, a firm can organize more transactions (i.e., it can grow) either by selling more of the product it currently sells or by expanding into new products. Third, there has to be some sense in which some activities are more natural fits for each other than others. Fourth, the model must have some on-going stochastic process. Otherwise, the model would ultimately reach a steady state in which no further evolution would occur. Fifth, it must embody both an advantage to organizing activities that fit well together in a single firm and a disadvantage of having activities that do not fit well together within a single firm. Without the latter feature, the model would predict that all activity would coexist within a single firm.

4.1. General setup

As we described in Section 3-B-1, Coase represented the space of economic activity as a set of concentric circles, with each circle representing an industry. The assumption about a product space captures the idea that some activities are more natural fits for each other than others. While sufficient for the broad conceptualization that he was putting forward in 1937, concentric circles are not a natural way to represent the space for computer simulation. A far simpler approach is to represent the space of possible transactions as a lattice, which accomplishes the same objective as the circles. Points near each other on the lattice are better fits for each other than points far away from each other.

With Coase's conceptualization, expanded output of an existing product is represented by taking up a bigger portion of a circle in which the firm already operates, and producing a new product is represented by moving into a different circle. Even with a two-dimensional lattice, one can give the model a similar interpretation. Each row can represent a product. Thus, if a firm starts by operating at a single node, it is selling one unit of a particular product. If it expands horizontally, it sells more units of the product. If it expands into a different row, it sells a different product. Within this conceptualization, each column represents a single customer. A vertical movement then means selling additional items to existing customers. Note that

(*footnote continued*)
therefore linear on a semi-log scale). With T near T_c, the second term dominates and we get $g(r) \sim r^{-\eta}$ (and therefore linear on a log–log scale.) Again, see Baxter (1982).

[32] We use the term "organized" in deference to the Coasian tradition, in which it is assumed that including different activities within the same firm results in some coordination among them. We do not explicitly model the nature of the coordination, however.

within this conceptualization, the first two requirements listed above are satisfied: each node represents a possible transaction, and firm growth occurs when more nodes are included in the same firm.

As we will describe in more detail below, each firm starts out as a single node. Firms then engage in a search process in which they try to organize additional activities. When they do so, they examine points or collections of points (i.e., firms) on their borders.

We assume that each node is either active or inactive. Because each node represents the sale of a particular product to a particular individual and because not all customers buy all products, the possibility of inactive sites is a realistic feature of the model. It also provides an opportunity to place a stochastic element in the model. Specifically, we allow for a probability that an inactive site becomes active and a probability that an active site becomes inactive. (We label these probabilities p_b and p_d, respectively, with the subscripts "b" standing for "birth" and "d" standing for "death.") These probabilities play an important role in the model. The growth of firms can be thought of as coming from two distinct sources. One is the sort of growth that would occur for a business unit within the neo-classical model of the firm. This growth arises from changes in demand, factor costs, and technology. The other is from mergers and divestitures. The probabilities with which inactive sites become active and vice versa are the source of the first type of growth in the model. Another important feature of the model is that each active site has an orientation, which we label "positive" and "negative." As with the distinction between active and inactive sites, this feature both adds an element of realism and an opportunity to introduce a stochastic element to the model. This feature captures the notion that while some companies are in closely related activities, they are nonetheless poor strategic fits. For example, AOL merged with Time-Warner in 2000. AOL was an Internet service provider and Web portal that needed content for its site, and Time-Warner had content that it knew it wanted to be able to distribute over the Internet. While the fit between the two companies may have seemed natural at the time, the merger has gone down in business history as a failure. Even companies in the same business can be poor strategic fits. A prominent example might be Daimler and Chrysler.[33] These examples are not isolated events.[34] Our model allows for a probability in each round that the orientation of a transaction changes. (We denote this probability as p_f, where the subscript "f" denotes "flip.") Again, we believe that this is a realistic feature. An example is Lucent Technologies, which was the manufacturing division that AT&T spun off in 1996. When AT&T was split up in 1984 as a settlement of the United States government's antitrust suit against it, the long distance division and the equipment manufacturing division (Western Electric) were kept together. AT&T continued to supply the local telephone companies that had been part of the old AT&T. Immediately after divestiture, the telecommunications arm of AT&T did not compete directly with these companies. Two major developments changed that. The first was the development of cellular telephone markets. AT&T entered those markets with its purchase of McCaw Communications. At that time, there were two cellular providers in each market with one of them being the incumbent local telephone company. The other was a set of deregulatory trends that culminated in the Telecommunications Act of 1996. Other companies, including long distance companies, were allowed to compete in local markets, and the Act provided for a process under which local telephone companies would ultimately be allowed to offer long distance service. Thus, AT&T found increasingly that its equipment customers were reluctant to purchase from it because they were its service competitors.

We assume that the orientation of transactions affects what firms form (and break up) through the profit function:

$$\pi_i = |P_i - N_i|^{\alpha} - k(P_i + N_i)^2, \tag{1}$$

where π_i is the profits of firm i, P_i is the number of positively oriented sites in firm i, N_i is the number of negatively oriented sites, $1 < \alpha \leq 2$, and $k \ll 1$. Firms explore transactions and collections of transactions (firms) on their border and decide to include them based on whether the combined profits go up.[35] For example, suppose $\alpha = 1.5$ and $k = 0.1$. Consider the merger of two firms (call them 1 and 2), each of which has one node with a positive orientation. Let π_i be the stand-alone profits of firm i and π_c be the combined profits if they merge. We then have $\pi_1 = \pi_2 = 1^{1.5} - 0.1 \cdot 1 = 0.9$ and $\pi_1 + \pi_2 = 1.8$. If they merge, $\pi_c = 2^{1.5} - 0.1 \cdot 2^2 \approx 2.43$. By way of contrast, suppose that firms 1 and 2 had opposite orientations. Then, each would still have pre-merger profits of 0.9. If they were to merge, their profits would then become $\pi_c = 0^{1.5} - 0.1 \cdot 2^2 = -0.4$. Because they have different orientations, a merger would lower their profits.

Because $\alpha > 1$, the first term in the profit function captures synergies among transactions with the same orientation. The second term captures diseconomies of large scale organization. Because k is small, the first term dominates provided that a firm is sufficiently small and has a sufficiently large proportion of transactions with the same orientation. However, the larger firms get, the more the second term comes into play.

With this profit function, two firms with different net orientations never merge. Whether or not two firms of the same net orientation merge depends both on firm size and on how pure the orientations are. If α is strictly less than 2, then larger potential merger partners must have "purer"[36] orientations than smaller ones for mergers to be profitable. Note that with

[33] Within our model, it is transactions that fit or do not fit together, not companies per se. Companies are collections of transactions and, as such, are either net positive or net negative. (Of course, they can be completely balanced, but such companies end up breaking up in our model.)

[34] See Ravenscraft and Scherer (1987).

[35] In the Axtell (1999) model of firm growth, firms consist of workers who all get equal shares of the profits. Individuals decide which firm to join based on their own income. A firm is not allowed to exclude a worker who wishes to join even if the effect is to lower the profits of the workers currently in the firm.

[36] That is, the proportion of nodes with the same orientation must be higher.

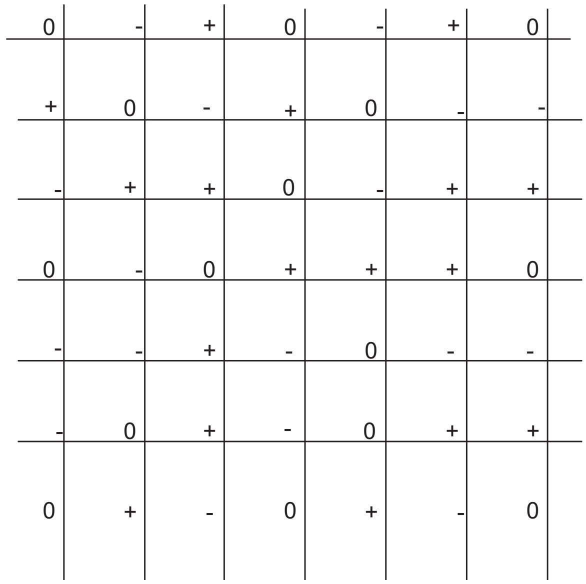


Fig. 5. Model initialization.

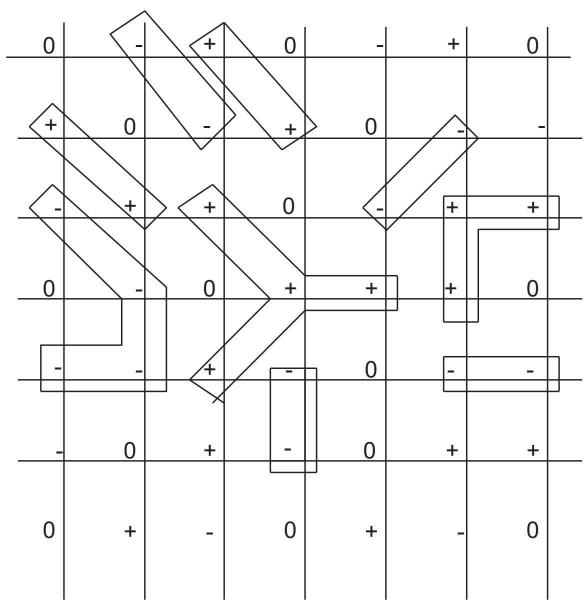


Fig. 6. Model after one merger round.

$\alpha < 2$, the profit function reaches a maximum when the firm is completely specialized (all P's or all N's) and the size of the firm is given by

$$S_i = \left(\frac{\alpha}{2k}\right)^{\frac{1}{2-\alpha}}, \tag{2}$$

where $S_i = P_i + N_i$.[37]

A few observations about the assumption of a two-dimensional lattice are in order. First, the primary reason for using two dimensions is expositional. The computer program that implements the simulations allows for more dimensions. Second, however, with one dimension representing consumers and the other representing a space of final products, the

[37] If $\alpha = 2$, then two profitable firms with the same orientation always merge. Provided that the parameters are set so that firms are only profitable if they have very high-proportions of transactions with the same orientation, then the model still places a limit on firm size. With the probability of flipping orientations, very large firms with sufficiently pure orientations would be rare.

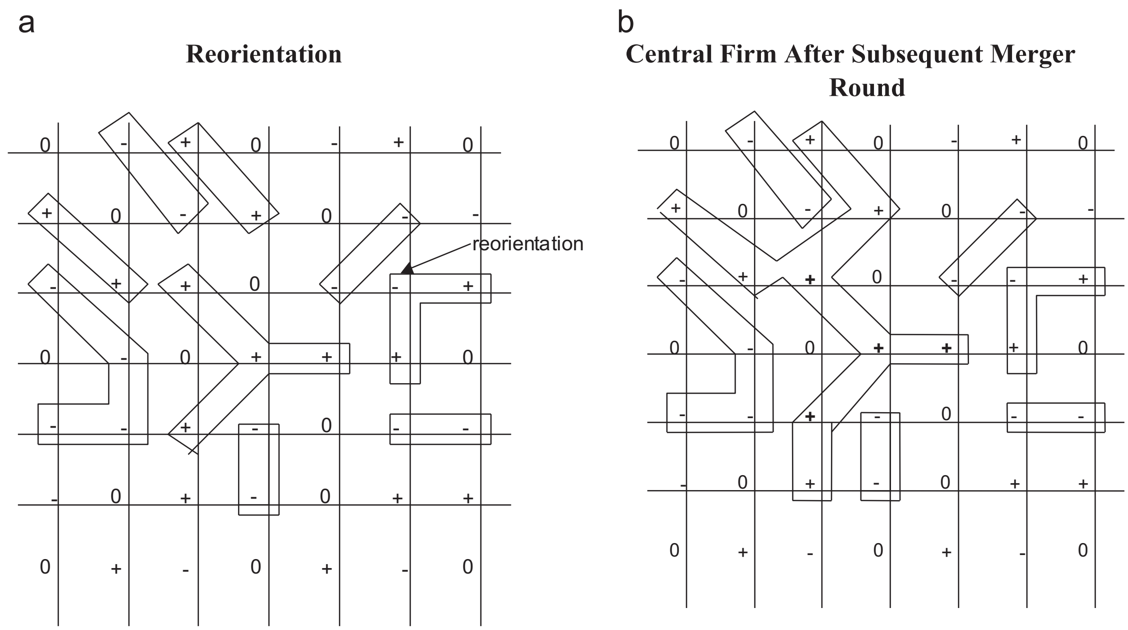


Fig. 7. Effect of reorientation on merger choice. (a) Reorientation. (b) Central firm after subsequent merger round.

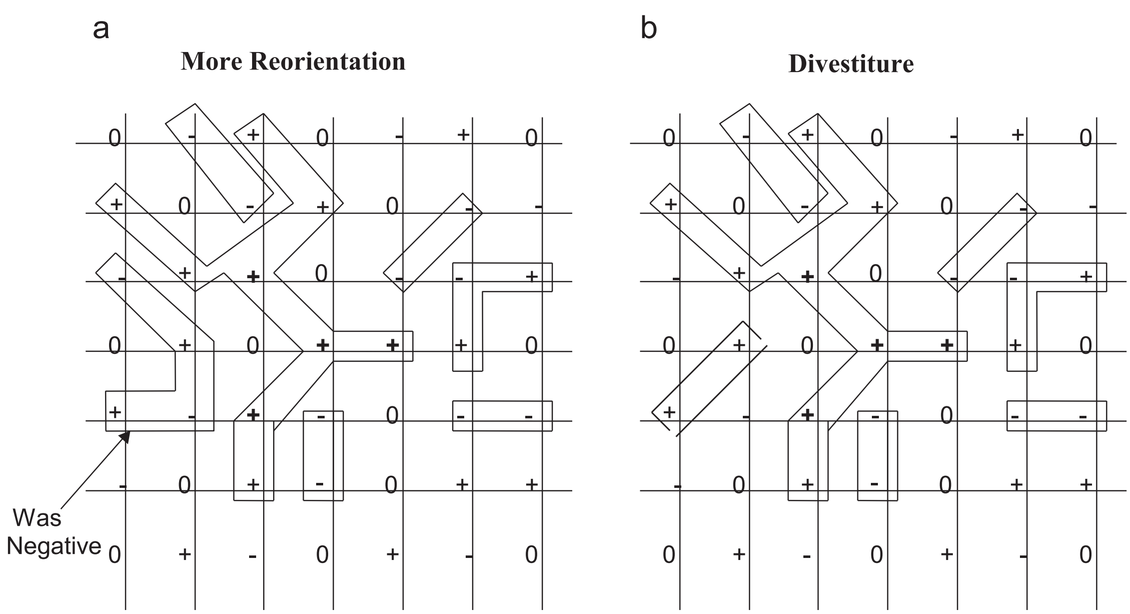


Fig. 8. Reorientation and divestiture. (a) More reorientation. (b) Divestiture.

two-dimensional model does not capture the vertical integration issues that have played so prominently in the literature on the scope of firms. One could incorporate vertical integration into the model by adding a third dimension that represents a set of inputs. To give a three dimensional model such an interpretation, it would not be appropriate to assume that nodes are all independently active or inactive. Rather, the nodes representing the inputs into a particular final sale would either all have to be active or all inactive depending on whether the node representing the final sale is active.[38] (The linked nodes would not, however, have to have the same orientation.) Third, while a lattice with more dimensions might seem more realistic, experience with models of this sort for three-dimensional physical systems suggests that two dimensions can yield quite good results. One dimensional systems do not, however.[39]

[38] Note also that the lattice structure would embody the assumption of a natural sequence for vertical integration to occur.

[39] Both the general issues of interpretation and the specific issues of the dimensionality of the lattice in this model are analogous to the issues that arise in models of product differentiation. One of the two primary approaches to modeling product differentiation is the address model originally due to

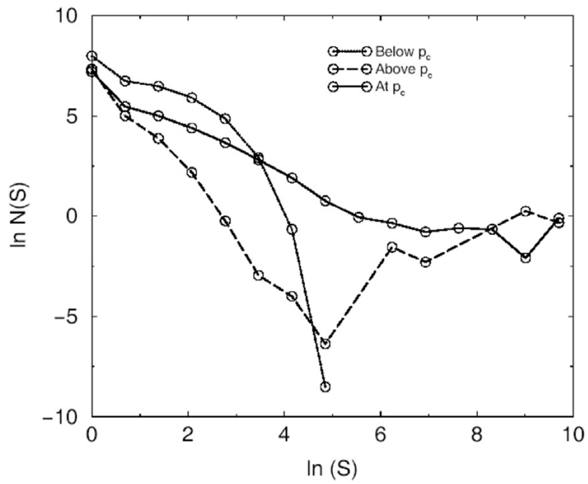


Fig. 9. Firm size distribution.

4.2. Simulation details

We begin the model by randomly assigning each node of a 200×200 lattice[40] to be inactive, active with a positive orientation, or active with a negative orientation. Fig. 5 illustrates a possible initialization of the model for a small segment of the lattice. This step happens only once.

The remaining steps are repeated. Of these, Step 1 is a "merger"[41] step. In random order, each firm exams each of its neighbors for possible inclusion in the firm.[42] The decision to include the candidate depends on whether the value of Eq. (1) with inclusion exceeds the sum of the values of the equation for the separate entities if inclusion does not occur. For the initialization shown in Fig. 5, Fig. 6 shows a possible state of the model after one merger round.

When one firm examines multiple neighbors, the merger decisions can depend on the order in which they are considered. We assume that the firm considers them in random order.

The behavior we assume is purposeful; mergers occur only if they increase joint profits. But it is arguably myopic. First of all, it does not allow for one firm to buy another and then divest the units that have opposite orientations. Second, at each step, a firm considers only its immediate neighbors without regard to the possibility that an acquisition would open up subsequent acquisitions.

Step 2 is a reorientation step. Each inactive node becomes active with probability p_b. If it does become active, it has an equal chance of having a positive or negative orientation. Each active node flips its orientation with probability p_f and dies with probability p_d. Fig. 7a shows a possible state of the lattice after a reorientation round. Fig. 7b shows the subsequent merger step for the central firm in the lattice (as of the end of the previous round). Note that this central firm does not acquire the firm in which the reorientation occurred, but it would have if the orientation had not occurred.

Once reorientation occurs, previously profitable firms can become unprofitable. Step 3 is a divestiture step. The first part of this step is to compute the profits of each firm using (1). We then break up firms with negative profits as follows. First, we randomly select a unit within the firm with an orientation opposite to the main orientation of the firm. We then find the cluster of similarly oriented units that are connected to a unit and then spin that cluster off into a new firm. We then check whether the original firm

(*footnote continued*)
Hotelling (1929). A difficulty with this type of model is that the space is often hard to measure and that, in both theoretical and empirical work, it is necessary to assume fewer dimensions than the "true" space actually has. Yet, these models prove useful because the practical alternatives, such as the variety models of Spence (1976) and Dixit and Stiglitz (1977), ignore product characteristics altogether and assume that the substitutability between products is only a function of their market shares. Indeed, statistical physics approaches to social and economic phenomena have analogs to these variety models. For example, models of Zipf's law for cities assumes that the attraction of new activities to cities is merely a function of city size, not of specific characteristics like geography or weather. Just as the address models make it possible to capture the assumption that the extent of substitutability among products depends on their characteristics, so the lattice model captures the obvious point that some activities fit more naturally together within firms than do others.

[40] One of the challenges of exploring scale-free processes on a lattice is that both the units (i.e., the nodes) of the lattice and the maximum size impose constraints on the scales that can be observed. In such models, scale-free features (like the linearity on a log–log scale of the size-standard deviation relationship) tend to break down at both extremes. We show below how to assess whether such deviations are merely artificial features of the model's scale.

[41] In some cases, "merger" can be interpreted as the outcome of competition. Under this interpretation, rather than having Firm A acquire Firm B, Firm A competes for B's customers and wins. For large-scale acquisitions, this interpretation is arguably stretched in that Firm A competes for all of B's customers (and for none of the customers of another firm with a different orientation).

[42] Each point on the lattice has eight neighboring points. A firm (A) is considered a neighbor of another firm (B) if any point in A is a neighbor of any point in B.

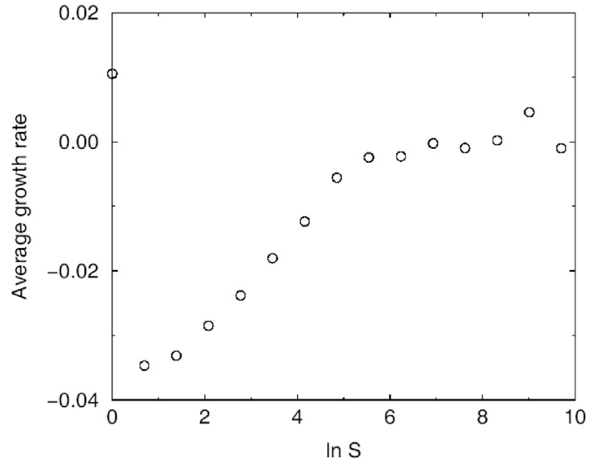


Fig. 10. Firm size and mean growth rates.

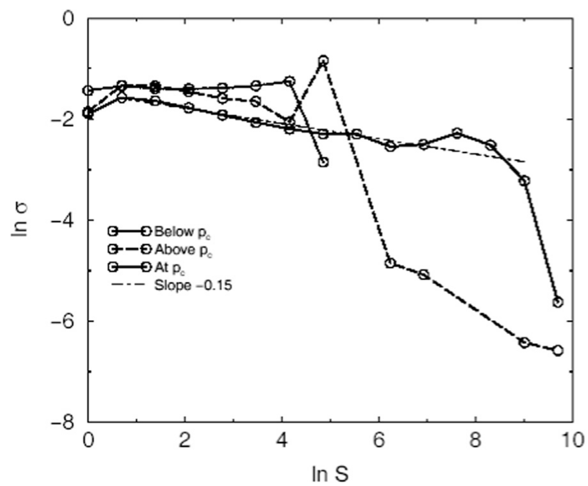


Fig. 11. Scaling property of the standard deviation.

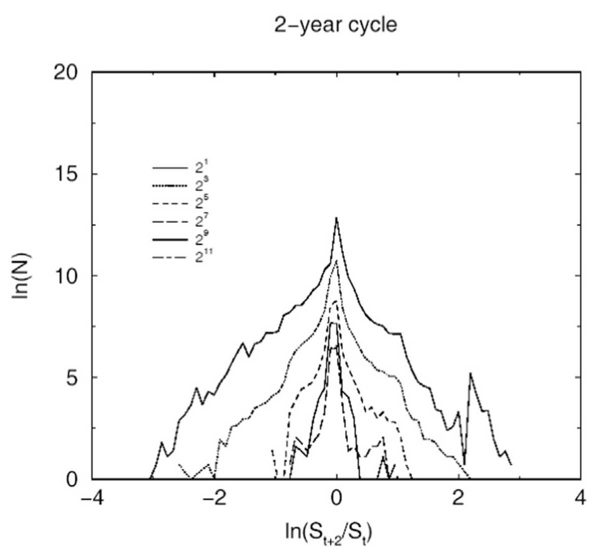


Fig. 12. Distribution of growth rates conditional on firm size.

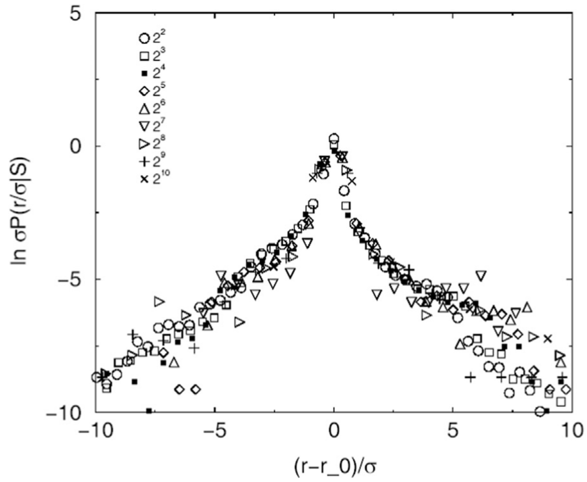


Fig. 13. Data collapse.

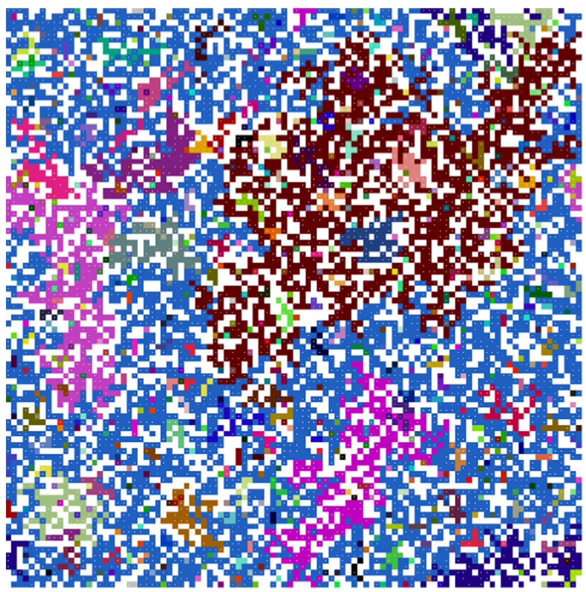


Fig. 14. Allocation of activity within firms.

remains connected. If not, we split it into several disconnected firms, using the majority rule to define their orientations. Fig. 8a shows a possible subsequent reorientation round, and Fig. 8b shows the break-up that would ensue.

As with the merger stage, behavior at the divestiture stage is purposeful, but it is not optimizing behavior. If the firm were to behave optimally, it would always be looking to shed activities with the wrong orientation. Instead, the behavior can be characterized as "satisficing." As long as the company is doing sufficiently well, it does not explore divestiture. It is only when results become unsatisfactory that it reoptimizes its configuration.

Step 4 is to calculate growth rates and record the distribution of size and growth rates for different size classes. Steps 1–4 are then repeated until the distribution of firm sizes is stable. We then iterate the model 20,000 times and record the results to construct the distributions.

In simulation models of this sort, it is standard to report a standard deviation associated with each estimated result. Without proving analytically that the model converges in some formal statistical sense, we cannot know whether it does. If such an analytical result were obtainable, however, we would not need to rely on simulations. As a practical matter, each run of the model reaches a point in which the change in results from additional blocks of iterations is very small compared to the differences across runs. Thus, even if the model does converge statistically, the number of iterations would have to increase by orders of magnitude to get substantially more precision.

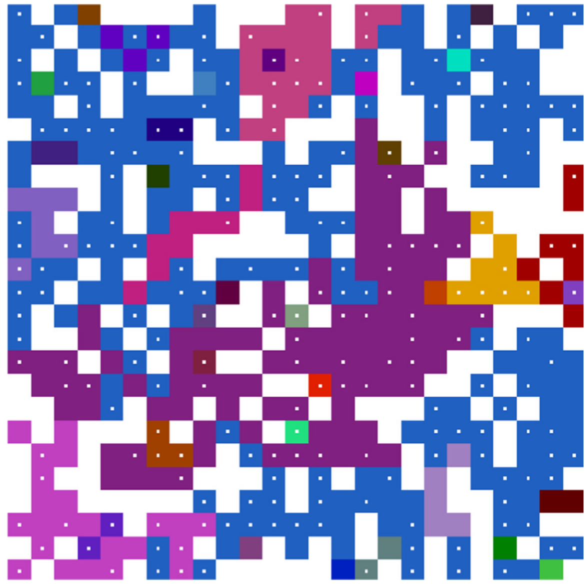


Fig. 15. Central region of lattice.

Some explanation for how we determine that the model has converged is in order.[43] We run the model 5 separate times, recording various features of the model at each iteration. We break each run up into 400 successive sets of 50 iterations. Let $X(i,j)$ be the average number of firms across 50 iterations of the number of firms of size 9–16 nodes in the jth block of run i (I $\in$ {1,2,3,4,5}). The model is assumed to converge when the increments $|X(i,j - X(i,j\text{-}1)|$ are much smaller than:

$$\sigma_j = \frac{\sqrt{\sum_{i=1}^{5} \frac{\left(X(i,j) - \overline{X}(.,j)\right)^2}{4}}}{\sqrt{5}} \tag{3}$$

A key issue of interest is whether the firms that emerge are natural firms in that they consist of closely related activities. A measure from statistical physics that captures this idea is the fractal dimension. To determine the fractal dimension, first calculate the average squared radius for each firm

$$R_i^2 = \frac{\sum_{i=1}^{S_i} (x_i - \overline{x})^2 + \left(y_i - \overline{y}\right)^2}{S_i}, \tag{4}$$

where x_i and y_i are the coordinates of activity i and $\overline{x}$ and $\overline{y}$ are the average of the coordinates of the active points in the firm. Then, group the firms into size classes, calculate the average value of R_i^2, and calculate $\frac{\Delta \ln(\overline{R}_j^2)}{\Delta \ln(\overline{S}_j)}$, where $\overline{R}_j^2$ and $\overline{S}_j$ are the mean values in size class j. If m is the slope, the fractal dimension is 2 /m.[44]

5. Results

The empirical findings that motivate this model are typical of a system near its critical point. To illustrate the importance of criticality, we simulate the model both with parameters tuned to the critical point and with parameters not tuned to the critical point. As we will see, the model more nearly reproduces the empirical findings for the tuned parameters.

A critical point is one in which firms exist at all size scales up to firms that span the entire lattice. The four parameters in the model are k, p_d, p_b, and p_f. We can select any values for the first three and then choose p_f to be near the critical point. To understand the relationship between p_f and criticality, suppose that p_f were 0, which would mean that a site would never change its orientation (unless it died and was reborn). If so, the economy tends to evolve into two firms, each having nearly all the sites of the same orientation in the economy.[45] In contrast, if p_f is particularly high, then firms cannot grow to span the entire space. The reorientation process ultimately forces firms to break up before they get too large.

[43] For a further discussion of the computation of standard errors and of convergence in models of this sort, see Binder (1986), pp. 1-45.

[44] The fractal dimension is constant in the limit of infinitely large objects exactly at criticality. In theory, m is the slope coefficient from the regression of $\ln(R_i^2)$ on $\ln(S_i)$. The fractal dimension for any finite object is affected by the finite size effects. Within the finite lattice, the relationship is not linear on a double logarithmic scale. As a result, we use the averages from successive bin sizes to calculate the slopes (and, in turn, the fractal dimension) conditional on size.

[45] This assumes that the lattice size is small enough that the total number of sites with a single orientation is smaller than the maximal firm size.

In the results reported below, we take $k=10^{-6}$, $p_b=0.04$ and $p_d=0.01$. For these values, the critical point is when $p_f \approx 0.02$. We report results for three values of p_f:0.001, 0.02, and 0.1.[46]

Fig. 9 shows the frequency distribution of firm sizes for the three parameter values. To get a sense of the scales in the graph, note that there are 40,000 nodes in a 200×200 lattice. Given $p_b=0.04$ and $p_d=0.01$, a total of 32,000 are active in the steady state – 16,000 with a positive orientation and 16,000 with a negative orientation. Thus, the maximum possible firm size in the lattice is approximately 16,000 or 9.7 on a natural log scale. With respect to the vertical scale, we compute the frequencies by pooling 20,000 iterations of the model. We then divide by the number of iterations so that the vertical scale is frequency per iteration. Since we use a log scale, a value of zero means an average of one firm per iteration in that size class.

For $p_f=0.001$, which is below percolation,[47] the vast majority of firms have a size less than 10 and virtually none have a size greater than $e^4 \approx 55$. The shape of the distribution on a log–log scale is a parabola, which means that the distribution is log normal.

For $p_f=0.1$, the size of virtually all firms is less than 10. In addition, however, there is on the order of 1 or 2 very large firms per iteration. With these parameters, there tends to be one firm that captures nearly all the sites of a given orientation. Activity does not necessarily concentrate into two firms, one with each orientation, because it is generally not possible to get connected clusters that span the entire space.

With $p_f=0.02$, which is near the percolation threshold, we observe firms at all scales, and the shape of the distribution is approximately linear on the log–log scale over the range $\ln(S)=1$ to $\ln(S)=8$.[48] This distribution is generally consistent with Axtell's finding of a Zipf firm size distribution for the economy as a whole.

Because of the uncertainty about whether the size distribution should be lognormal or a power law, we believe that the distribution around the percolation threshold or below it can be viewed as being consistent with the data. The size distribution above the percolation threshold is not.

Fig. 10 is a plot of the mean of the logarithmic growth rate versus the logarithm of initial size for $p_f=0.02$. For initial sizes of from 2 to 64, the mean growth rates are negative and increasing with size.[49] For initial sizes of 128 and above, the mean growth rates are very near 0 and approximately independent of size. The approximate independence of mean growth rates and size for the larger firms is in accord with empirical findings.

The upward slope for the smaller firms might appear to contradict empirical findings, but these mean growth rates are very small. The standard deviations of the growth rates for the smaller firms are approximately -0.2.[50] Consider, for example, a firm of size two. If it shrinks and survives, its size is cut in half. On a log scale, its growth rate is -0.69. If it grows, its size must increase to at least three, which is a logarithmic growth rate of 0.41, and it could conceivably grow much more. The mean growth rate for that size class of about -3.5% is small relative to this scale. Moreover, a combination of the granularity inherent in a lattice model and survivor effects could give rise to these small effects. Again, consider a firm with an initial size of two. If its orientation remains pure, it has a high probability of merging with a neighbor. If the neighbor evaluates mergers first, however, then it does not survive and its positive growth does not enter the computation of the mean growth rate. On the other hand, if one of its sites flips its orientation, the firm becomes unprofitable. It will not be acquired by or acquire another firm. It is likely to split up and, because it survives, its negative growth gets included in the mean growth rate.

Fig. 11 is a log–log plot of the standard deviation of the (logarithmic) growth rate against initial size. For $p_f=0.001$, the standard deviation of the growth rate is roughly independent of size (except for the largest size class for which a standard deviation could be computed). The result that the mean and standard deviations of the growth rates are independent of size are consistent with the approximate log normality of firm size for those parameter values. We have already ruled out $p_f=0.1$ based on the size distribution that it generates. Still, note that the standard deviation-size relationship that it generates cannot be easily characterized and is certainly not the approximately linear relationship observed in the data. With the parameter tuned to the critical value, however, the size standard deviation relationship is approximately linear over the range $e-e^9$ with a slope of -0.15.[51] Fig. 11 is a primary justification for our conclusion that the model near its critical

[46] We have, of course, explored the model with different parameters. To get the sorts of results we get here, it is important to keep the ratio of p_b–p_d such that a sufficient fraction of the sites are active. If too few sites are active, then percolation cannot occur because it is not possible to have a connected cluster. Provided that the ratio is large enough, then we can find a range of p_f that generates critical behavior.

[47] A counterintuitive feature of these results is that the lower p_f is below percolation while the higher p_f is above percolation. One might have expected that the high probability of changing orientations would be what prevents a cluster from evolving to span the entire space. Instead, the barrier to percolation seems to be that the death process causes firms to become disconnected before they become very large. The puzzle is why this same factor does not prevent the growth of large firms with higher flipping probabilities. The resolution may be that the firms that evolve with low-reorientation probabilities have more "thin" links that can be broken by the death of a single unit.

[48] The lattice places limits on the scale of firms that can be observed in the model, and the behavior of the model near both extremes will tend to deviate from the middle range.

[49] The mean growth rate of firms with an initial size of 1 is positive. This result is trivial since 1 is the minimum firm size and these growth rates are for surviving firms. A firm with initial size of 1 cannot both survive and have a negative growth rate.

[50] See Fig. 11.

[51] We estimate the slope with a regression but exclude the first and last points, which are distorted by finite size effects. A unit in the lattice cannot shrink without disappearing and, on the upper end, the dimensions of the lattice impose a constraint on how large firms can grow. Even without the last point, there is a "bump" near the upper end of the graph. This deviation from linearity is another finite size effect that is typical of lattice models of percolation. See Stauffer and Aharony (1994), chapter 2.7.

value is more consistent with our empirical findings than the model away from its critical value. Not only does it reproduce an approximately linear relationship on double logarithmic scale, but the scaling exponent is approximately the same as the empirical estimates.

Fig. 12 shows a plot of the distribution of growth rate conditional on firm size for $p_f = 0.02$. Particularly with the small size classes, because the amount of activity gained or lost must be an integer, the histograms get artificially "bumpy." Looking at two growth cycles instead of one partly alleviates this problem. Thus Fig. 12 shows results for a two-period cycle.[52] There are 10 different size classes ranging from an initial size of 4 to an initial size of 4096.[53] The more spread-out distributions are for the smallest size classes. To see whether the distributions have a common shape, we rescale each distribution by its standard deviation.

Fig. 13 shows the resulting data collapse. From approximately 5 standard deviations below the mean to 5 standard deviations above the mean, the curves are remarkably close to each other, which is in accord with the actual data.[54] The data do not appear to be exactly tent-shaped. There is more concentration in the middle of the distribution; however, both sides of the distribution do have an apparently long linear region. As we will discuss further in the conclusions, this model should be viewed as preliminary. Ultimately, we hope to construct a related model of that reproduces the tent shape. For now, though, we note that this model produces a common, fat-tailed shape for the distribution of growth rates for the different size classes.

Fig. 14 shows a "picture" of the entire lattice. In it, each color represents a firm. Fig. 15 shows a 25×25 portion of Fig. 14 on a larger scale. Sites with a positive orientation have a white dot in the middle whereas sites with a negative orientation do not.[55] The blue firm, which is the percolating cluster,[56] occupies 187 of the sites in Fig. 15. Of these, 113 have a positive orientation and 74 have a negative orientation. Within this square 182 have positive orientation.

There are two ways in which the composition of the firms shown in Figs. 14 and 15 are inefficient. The size of the blue firm in this region is about equal to the number of sites with positive orientation. Thus, focusing on this region and taking as given that it is efficient to have a firm of about the size of the blue firm, the positively oriented sites would be the transactions that it would be efficient to have within the firm. The firm that actually exists (i.e., the blue firm) has a much different composition. This "inefficiency" arises for historical reasons. Sites currently in the firm with negative orientations did fit at one point. Positively oriented sites might have gotten included in different firms because they previously had different orientations. This model does not explicitly include reorganization costs, but they are the implicit explanation for why continuous reorganization to create efficient firms does not occur.

The second way in which the firms are inefficient is that they are relatively dispersed. Rather than looking like coherent masses, they more closely resemble gerrymandered Congressional districts. There are firms with the same orientation as the blue firm that are nearer the center of both figures than are the points on the edge. One could construct a firm with the same size and net orientation as the blue firm and that spans a smaller space. This point is not unique to the percolating cluster. It applies to smaller firms as well. The total span of the organizations is greater than is necessary given the amount of activity coordinated within them.

The mathematical measure of whether firms are excessively dispersed is the fractal dimension. Based on the method of successive slopes, the fractal dimension ranges from 1.6 for firms of size 2 up to 1.9 for firms of size 128 and above.[57]

It is this finding that should make us skeptical that Coase's theory of the firm can yield reliable predictions about what activities will coexist within the same firm. With the tuned parameters, activities that are quite far apart from each other exist within the same firm while closer activities with the same orientation are part of other firms. Consider three activities, A, B, and C. Suppose A and B are far apart from each other but in the same firm. C is close to A and has the same orientation. Why is it that A and C do not end up in the same firm? The resolution of the paradox is that C is already in another firm. The entire firm might have an orientation different from C's or, alternatively, it might have a sufficiently mixed orientation so that a merger would lower joint profits. In a different simulation of the model, A and C might have ended up together. As the firms evolved, however, they ended up being separate.

6. Conclusions

In this paper, we have constructed a stochastic, dynamic model that captures the spirit of Coase's theory of the firm. The model is consistent with most of the observed statistical regularities about firm growth and it predicts relatively unrelated activities will coexist within a given firm. Of course, two related activities (i.e., sites near each other on the lattice) are more likely to exist within the same firm in this model than are two unrelated activities. As a result, the model suggests that it might be possible to ascertain statistically significant determinants of integration. However, the nature of the firms that

[52] With a two-year cycle, we estimate a scaling coefficient of -0.22 for the relationship between the logarithm of size and the logarithmic growth rate.

[53] The bins for initial size are in powers of 2. We do not show the distributions for initial sizes of 1 or 2. For an initial size of 1, the firm cannot simultaneously shrink and survive. A firm with an initial size of 2 could shrink and survive, but the entire left hand side of the distribution would be concentrated at a single point. The largest bin size is 2^{12}.

[54] The collapse of the actual data is actually over a smaller range: ± 2 standard deviations.

[55] These dots are present in Fig. 15 as well, but they are too small to see.

[56] A percolating cluster has a site on all four boundaries.

[57] One might argue that 1.9 is not much less than 2.0. This is to a large extent a consequence of using a two-dimensional model. It is known that in a two-dimensional percolation model, the fractal dimension of the percolating cluster is 1.9. It is 2.5 in a three-dimensional model and 4 in a six-dimensional model.

evolve certainly complicate the task of measuring the determinants of integration. Moreover, success in determining the factors that make it more likely that activities will coexist within the same firm cannot be expected to explain with much reliability the exact composition of the firms we observe.

The model only reproduces the empirical size-standard deviation relationship and a common, fat-tailed growth rate distribution across size classes when the model is tuned to the critical point. We could only demonstrate this point with a model that does not self-tune to the critical point. Ultimately, though, a model of self-organized criticality is needed to make it plausible that we should view the allocation of activity within firms as an example of self-organized criticality. Moreover, such a model should ideally reproduce the shape of the size distribution better than this model does. As noted above, we do not think it will be difficult to modify the model so that it self-organizes.

Whatever model of firm size and growth becomes accepted, it should be judged by its ability to explain statistical regularities of firm size and growth. It is important, though, to focus on the right empirical regularities. The relationship between firm size and the standard deviation of the growth rate is a stronger relationship than the relationship between size and the mean growth rate, and it more likely is the source of insights into the nature of the firm.

More than one model can predict the same empirical findings. Buldyrev et al. (1997), Amaral et al. (1998), Axtell (1999), Sutton (2002), Fu et al. (2005), Yamasaki et al. (2006), and Riccaboni et al. (2008), contain different models that all predict the same set of empirical facts. Ultimately, the choice among them should be based on their success in explaining further empirical findings. In the interim, though, the choice is likely to be based in part on the plausibility of the assumptions. The assumptions in the model in this paper are rooted in the Coasian approach to understanding the scope of firms and are therefore useful in helping sort out exactly what the testable implications of that approach are.

References

Amaral, Luis A. Nunes, Buldyrev, Sergey V., Havlin, Shlomo, Leschorn, Heiko, Maas, Phillipp, Salinger, Michael A., Stanley, Eugene H., Stanley, Michael H.R., 1997. Scaling behavior in economics: I. Empirical results for company growth. J. Phys. I Fr. 7 (4), 621–633. April.

Amaral, Luis A. Nunes, Buldyrev, Sergey V., Havlin, Shlomo, Salinger, Michael A., Stanley, Eugene H., 1998. Power law scaling for a system of interacting units with complex internal structure. Phys. Rev. Lett. 80 (7), 1385–1388. February 16.

Axtell, Robert, 1999. The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions, mimeo.

Axtell, Robert L., 2001. Zipf distributions of US firm sizes. Science 293 (5536), 1818–1820. September 7.

Bachelier, Louis, 1900. Théorie de la spéculation. Ann. De. l'Éc. Norm. Superiure.

Bak, Per, Chen, Kan, Sheinkman, José, Woodford, Michael, 1993. Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics. Richerche Econ. 47 (1), 3–30. March.

Baxter, R.J., 1982. In: Exactly Solved Models in Statistical MechanicsAcademic Press, London.

Berry, Steven, Levinsohn, James, Pakes, Ariel, 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890. July.

Binder, K. (Ed.), 1986. Monte Carlo Methods in Statistical Physics, Springer-Verlag, Berlin.

Blume, Lawrence E., Durlauf, Steven N., 2001. The interactions-based approach to socioeconomic behavior. In: Yound, H. Peyton, Durlauf, Steven (Eds.), Social Dynamics, MIT Press, Cambridge, MA, pp. 15–44.

Brock, William, Steven, N. Durlauf, 2001. Discrete choice with social interactions. Rev. Econ. Stud. 68 (2), 235–260. April.

Brock, William, Lakonishok, Josef, LeBaron, Blake, 1992. Simple technical trading rules and the stochastic properties of stock returns. J. Financ. 47 (5), 1731–1764. Dec.

Buldyrev, Sergey V., Amaral, Luis A. Nunes, Havlin, Shlomo, Leschorn, Heiko, Maas, Phillipp, Salinger, Michael A., Stanley, Eugene H., Stanley, Michael H.R., 1997. Scaling behavior in economics: II. Modeling of company growth. J. de. Phys. I 7 (4), 635–650. April.

Casadesus-Masanell, Ramon, Spulber, Daniel, 2000. The fable of fisher body. J. Law Econ. 43 (1), 67–104. April.

Coase, Ronald H., 1937. The nature of the firm. Economica 4 (16), 386–405. Nov.

Coase, Ronald H., 1988. The nature of the firm: origin, meaning, influence. J. Law, Econ., Organ., Spring 4 (1), 3–47. Reprinted in Williamson, Oliver E. and Winter, Sidney G. (eds.). The Nature of the Firm: Origins, Evolution, and Development. New York: Oxford University Press, 1993, pp. 34–74.

Coase, Ronald H., 2000. The acquisition of fisher body by general motors. J. Law Econ. 43 (1), 15–32. April.

Cyert, Richard, March, James, 1963. A Behavioral Theory of the Firm, Englewood CliffsPrentice-Hall, NJ.

Dixit, Avinash K., Stiglitz, Joseph E., 1977. Monopolistic competition and optimum product diversity. Am. Econ. Rev. 67 (2), 297–308. March.

Dunne, Paul, Hughes, Alan, 1994. Age, size, growth, and survival: UK companies in the 1980's. J. Ind. Econ. 42 (2), 115–140. June.

Durlauf, Steven N., 1996a. A theory of persistent income inequality. J. Econ. Growth 1 (1), 75–93. March.

Durlauf, Steven N., 1996b. Neighborhood feedbacks, endogenous stratification, and income inequality. In: William, A., Barnett, Giancarlo, Gandolfo, Hillinger, Claude (Eds.), Dynamic Disequilibrium Modeling: Theory and applications: Proceedings of the Ninth International Symposium in Economic Theory and Econometrics, Cambridge University Press, Cambridge, UK, pp. 505–534.

Durlauf, Steven N., 1999. The memberships theory of inequality: ideas and implications. In: Brezis, Elise S., Temin, Peter (Eds.), Elites, Minorities and Economic Growth, Elsevier Science, Amsterdam, North-Holland, pp. 3–16.

Durlauf, Steven N., 2001. A Framework for the Study of Individual Behavior and Social Interactions, March (available at 〈〈http://www.ssc.wisc.edu/econ/archive/wp2016r.pdf〉〉).

Evans, David S., 1987. Tests of alternative theories of firm growth. J. Political Econ. 95 (4), 657–674. August.

Ericson, Richard, Pakes, Ariel, 1995. Review of Economic Studies 62 (1), 53–82.

Freeland, Robert F., 2000. Creating holdup through vertical integration: fisher body revisited. J. Law Econ. 43 (1), 33–66.

Fu, Dongfeng, Pammolli, F., b Riccaboni, M., Matia, K., Yamasaki, K., Stanley, H.E., 2005. The growth of business firms: theoretical framework and empirical evidence. Proc. Natl. Acad. Sci. 102 (52), 18801–18806.

Gabaix, Xavier, 1999. Zipf's law for cities: an explanation. Q. J. Econ. 114 (3), 739–768.

Gabaix, Xavier, 2016. Power laws in economics: an introduction. J. Econ. Perspect. 30 (1), 185–206.

Gibrat, Robert, 1931. Les inégalités économiques; applications: aux inégalités des richesses, à la concentration des enterprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel. Librairie du Recueil Sirey, Paris.

Hall, Bronwyn, 1987. the relationship between firm size and firm growth in the U.S. manufacturing sector. J. Ind. Econ. 35 (4), 583–606.

Hart, Peter, 1962. The size and growth of firms. Econ. N. S. 29 (113), 29–39.

Hotelling, Harold, 1929. Stability in competition. Econ. J. 39 (153), 41–57.

Hymer, Stephen, Pashigan, Peter, 1962. Firm size and rate of growth. J. Political Econ. 70 (6), 556–569.

Ijiri, Yuji, Simon, Herbert, 1977. Skew Distributions and Functions. Amsterdam, North Holland.
Jovanovic, Boyan, 1982. Selection and the Evolution of Industry. Econometrica: Econometrica 50 (3), 649–670.
Klein, Benjamin, 2000. Fisher-general motors and the nature of the firm. J. Law Econ. 43 (1), 105–142.
Krugman, Paul, 1996. In: The Self-Organizing EconomyBlackwell Publishers, Cambridge.
LeBaron, Blake, 1992. Some relations between volatility and serial correlations in stock returns. J. Bus. 65 (2), 199–219.
Lee, Y., Amaral, Luis, A.N., Canning, David, Meyer, Martin, Stanley, H. Eugene, 1998. Universal features in the growth dynamics of complex organizations. Phys. Rev. Lett. 80 (15), 3275–3278.
Lucas Jr., Robert E., 1978. On the size distribution of business firms. Bell J. Econ. 9 (2), 508–523.
Machlup, Fritz, 1967. Theories of the firm: marginalist, behavioral, and managerial. Am. Econ. Rev. 57 (1), 1–33.
Mandelbrot, Benoit, 1963. The variation of certain speculative prices. J. Bus. 36 (4), 394–419.
Mandelbrot, Benoit, 1965. A Class of Long-Tailed Probability Distributions and the Empirical Distribution of City Sizes.
Mansfield, Edwin, 1962. Entry, Gibrat's law, innovation, and the growth of firms. Am. Econ. Rev. 52 (5), 1023–1051.
Mantegna, Rosario, Stanley, H. Eugene, 1995. Scaling behaviour in the dynamics of an economic index. Nature 376 (6535), 46–49.
Mantegna, Rosario, Stanley, H. Eugene, 2000. In: An Introduction to Econophysics: Correlations and Complexity in FinanceCambridge University Press, Cambridge, UK.
Pakes, Ariel, McGuire, Paul, 1994. Computing markov-perfect nash equilibria: numerical implications of a dynamic differentiated products model. Rand J. Econ. 25 (4), 555–589.
Plerou, Vasiliki, Amaral, Luís A. Nunes, Gopkrishnan, Parameswaran, Meyer, Martin, Stanley, H. Eugene, 1999. Similarities between the growth dynamics of university research and of competitive economic activities. Nature 400 (6743), 433–436.
Prais, S.J., 1976. In: The evolution of giant firms in Britain. Cambridge University Press, Cambridge.
Quandt, Richard E., 1966. On the size distribution of firms. Am. Econ. Rev. 56 (3), 416–432.
Ravenscraft, David M., Scherer, F.M., 1987. In: Mergers, Sell-Offs, and Economic EfficiencyBrookings Institution, Washington.
Riccaboni, Massimo, Pammolli, F., Buldyrev, S.V., Ponta, L., Stanley, 2008. The size variance relationship of business firm growth rates. Proc. Natl. Acad. Sci. 105 (50), 19595–19600.
Samuels, J.M., 1965. Size and Growth of Firms. Rev. of Econ. Stud. 32 (2), 105–112.
Scheinkman, José, Woodford, Michael, 1994. Self-organized criticality and economic fluctuations. Am. Econ. Rev. 84 (2), 417–421.
Scherer, F.M., 1980. Industrial Market Structure and Economic Performance, 2nd edition. Houghton Mifflin, Boston.
Schmalensee, Richard, 1989. Inter-industry studies of structure and performance. Schmalensee, R., Willig, R.D. (Eds.), Handbook of Industrial Economics, 2. , North Holland, Amsterdam, pp. 951–1009.
Simon, Herbert A., 1982. In: Models of Bounded RationalityMIT Press, Cambridge.
Simon, Herbert, Bonini, C.P., 1958. The size distribution of business firms. Am. Econ. Rev. 46 (4), 607–677.
Singh, Ajit, Whittington, Geoffrey, 1968. In: Growth, Profitability, and ValuationCambridge University Press, London.
Singh, Ajit, Whittington, Geoffrey, 1975. The size and growth of firms. Rev. Econ. Stud. 42 (1), 15–26.
Spence, Michael, 1976. Product differentiation and welfare. Am. Econ. Rev. 66 (3), 407–414.
Stanley, Michael H.R., Buldyrev, Sergey, V., Havlin, Shlomo, Mantegna, Rosario N., Salinger, Michael A., Stanley, H. Eugene, 1995. Zipf plots and the size distribution of firms. Econ. Lett. 49 (3), 453–457.
Stanley, Michael H.R., Amaral, Luis A.N., Buldyrev, Sergey V., Havlin, Shlomo, Leschhorn, Heiko, Maass, Phillipp, Salinger, Michael A., Stanley, H. Eugene, 1996. Scaling Behaviour in the growth of companies. Nature, 379 (February 29), 804-806.
Stauffer, D., Aharony, A., 1994. In: Introduction to Percolation Theory2nd edition. Taylor and Francis, London.
Stauffer, Dietrich, Stanley, H. Eugene, 1996. In: From Newton to Mandelbrot: A Primer in Theoretical Physics with Fractals for the Personal Computer2nd Enlarged Edition Springer, Berlin.
Sutton, John, 1997. Gibrat's legacy. J. Econ. Lit. 35 (1), 40–59.
Sutton, John, 2002. The Variance of Firm Growth Rates: The 'Scaling' Puzzle.
Williamson, Oliver E., 1989. Transaction cost economics. Schmalensee, R.L., Willig, R.D. (Eds.), Handbook of Industrial Economics, 1. , North Holland, Amsterdam, pp. 135–182.
Whinston, Michael D., 2001. Assessing the property rights and transaction-cost theories of firm scope. Am. Econ. Rev. 91 (2), 184–188.
Yamasaki, Kazuko, Matia, K., Buldyrev, S.V., Fu, D., Pammolli, F., Riccaboni, M., Stanley, H.E., 2006. Preferential attachment and growth dynamics in complex systems. Phys. Rev. E 74 (035103R), 1–4.
Zipf, George, 1949. In: Human Behavior and the Principle of Least EffortAddison-Wesley, Cambridge, MA.