

Complexity and Empirical Economics

Steven N. Durlauf¹

January 27, 2003

Abstract

This paper explores the state of interplay between recent efforts to introduce complex systems methods into economics and the understanding of empirical phenomena. The empirical side of economic complexity may be divided into three general branches: historical studies, the identification of power and scaling laws, and analyses of social interactions. I argue that, while providing useful “stylized facts,” none of these empirical approaches has produced compelling evidence that economic contexts exhibit the substantive microstructure or properties of complex systems. This failure reflects inadequate attention to identification problems. Identification analysis should therefore be at the center of future work on the empirics of complexity.

¹ Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI, 53706-1393. I thank the National Science Foundation, John D. and Catherine T. MacArthur Foundation, Institute for Research on Poverty and University of Wisconsin Graduate School for financial support. This paper reflects many conversations with Buz Brock related to the issues discussed here. Robert Axtell and Blake LeBaron have made useful comments on an earlier draft.

1. Introduction

This paper provides an overview of the empirical side of the growing literature on economic complexity. While still controversial, there is increasing interest and application in economics of ideas from complex systems theory. My goal is both to survey the main areas in which complexity and empirical economics have been connected as well as to evaluate the extent to which empirical economics has provided support for the assumptions that underlie and the implications that have been drawn from complex system approaches.

Most of the existing research on economic complexity has been theoretical, as is apparent from collections of papers such as Anderson, Arrow, and Pines (1988), Arthur, Durlauf, and Lane (1997) and Blume and Durlauf (2003). This is unsurprising, since a first goal of research on economic complexity has been the determination of the ways in which complex systems represent an extension as opposed to an alternative to standard (by which I mean neoclassical) economic theory. At the same time, economic complexity has from its inception been strongly motivated by a desire to explain substantive empirical phenomena. For example, Arthur's work on the dynamics of increasing returns was motivated by a desire to explain how one of a set of technologies (eg. a specific computer operating system or an arrangement of letters on a keyboard) came to dominate a market. Similarly, Krugman's (1996) work on complexity and economic geography was motivated by historical patterns of regional specialization.

While far more disjoint than the theoretical literature, empirical research has been part of the complexity research program almost from its inception. Currently, there are three main areas of work on the complexity/empirics interface. The first consists of historical studies. The study of economic complexity was in fact, originally championed to a large extent by economic historians in the context of empirical studies of path dependence in economic activity.² The second consists of the identification of data patterns that are consistent with some of the features of complex environments. A major

² Path dependence has generally been used to refer to environments in which a shock or a set of shocks has permanent effects on a system. There is no reason to regard path dependence as different from nonergodicity, a concept I describe below.

feature of this work has been the effort to identify where power laws, which represent a particular class of probability distributions, and scaling laws, which describe relationships between variables that appear to be independent of the scale of measurement, occur in various economic data series. This search has to a substantial extent been led by physicists as there exists a range of complex physical systems in which such laws are commonplace. A third area of work has focused on the study of social interactions. To a large extent, this work has eschewed an explicit connection to complexity; nevertheless a number of social interactions models, e.g. Brock and Durlauf (2001a,b,2003) and Glaeser, Sacerdote, and Scheinkman (1996), possess structures mathematically equivalent to certain complex systems. More important for the purposes of this paper, empirical work on social interactions has focused on the analysis of precisely the type of interdependences between individual actors that lie at the heart of the microstructure of complexity-based models.

My overall assessment of the empirical complexity literature is critical. The literature has succeeded in describing interesting historical episodes and performing original statistical calculations that are consistent with complex systems models as well as presenting a body of regression evidence that suggests the presence of the sorts of interdependences across individuals that are a hallmark of complexity. However, this evidence is far from decisive and is amenable to alternative interpretations. It is therefore unclear whether this work has provided evidence in support of economic complexity *per se*.

Section 2 of this paper reviews some of the properties of complex systems that one would expect to observe in a complex economic environment. Section 3 discusses historical studies of path dependence. Section 4 describes some of the recent work on power and scaling laws. Section 5 surveys some empirical studies of social interactions. Section 6 provides a summary and conclusions.

2. Empirics: General Considerations

In this section, I discuss some general properties of complex systems. My objective is not to define complexity per se (something the literature has grappled with for years and which is ultimately not important to the discussion), but to describe some salient features of such models. For my purposes, complex systems are those comprised of a set of heterogeneous agents whose behaviors are interdependent and may be described as a stochastic process. What distinguishes complexity from any environment with such interdependences (such environments of course include the Arrow-Debreu model and any number of evolutionary game theory environments) is the presence of the aggregate properties I describe. Following Durlauf (2001), four properties seem particularly relevant to social science contexts

- i. Nonergodicity. A system is nonergodic if the conditional probability statements that describe the system do not uniquely characterize the average or long run behavior of the system. A standard example of a nonergodic system is one where a shock at one point in time affects the long run state of the system.
- ii. Phase transition. A system exhibits a phase transition if it can undergo a qualitative change in its aggregate properties for a small change in its parameters. Phase transitions are commonplace in physical contexts. Water experiences a phase transition when its temperature moves below 0 degrees centigrade. Similarly, if one heats a magnetized piece of iron, there is a temperature above which magnetization disappears.
- iii. Emergent properties. Following ideas well described in Anderson (1972) and Crutchfield (1994), emergent properties are properties of a system that exist at a higher level of aggregation than the original description of a system. By this definition, ice is an emergent property of water as it is a property of how water molecules are collectively aligned, not of one molecule in isolation, whereas

the properties by which one molecule aligns with its neighbors is described at the level of the molecule. Similarly, magnetization is an emergent property as it derives from the alignment of spins of individual atoms in a common piece of iron.

- iv. Universality. A property is universal if its presence is robust to alternative specifications of the microstructure of the system. In physics, magnetization is universal in the sense that its presence in iron occurs for a range of different specifications of the interdependence of spins between individual atoms.

Each of these properties, as it manifests itself in a complex system, is of potentially great importance in understanding socioeconomic phenomena. For example, the presence of multiple pure Nash equilibria in models with social interactions (Brock and Durlauf (2001)) or other complementarities (Cooper (1999)) is often interpretable, in stochastic contexts, as a form of nonergodicity. This may be seen when one interprets the behavior of each agent as a stochastic process that represents his optimal decision given the decisions of others; as such the individual behavior descriptions constitute conditional probabilities and the equilibrium of the system is a joint probability structure compatible with these conditional measures.³ Similarly, Schelling's (1971) demonstration that complete segregation can be the steady state configuration of a dynamic sequence of location decisions even when all agents prefer some degree of integration is an example of how emergence may be found in a social context.

Of course, the presence of these properties in an economic system does not imply that it is complex. As argued in Blume and Durlauf (2001), the first welfare theorem of economics is an example of an emergent property as well. Similarly, the independence of the first welfare theorem from the particulars of preferences and endowments makes it a universal property in this sense. Hence one can see that empirical evaluations of complexity may be subject to identification problems at even this very general level.

³ See Brock and Durlauf (2001b) Section 2 for a more formal discussion of this relationship as well as for an extended development of such a model.

These complex systems properties provide a useful benchmark in evaluating empirical work on complexity and will be used to evaluate the different empirical aspects of economic complexity.

3. Economic History and Technological Evolution

Economic historians have been at the forefront of efforts to identify an important empirical role for complexity. In particular, economic history has provided a range of examples that have been purported to demonstrate path dependence. The most famous empirical example of path dependence is the QWERTY keyboard configuration studied by David (1985). David argues that the adoption of the QWERTY keyboard is a consequence of a set of decentralized, uncoordinated decisions by agents in an environment in which there are strong network externalities. This led to the keyboard being locked-in as a technological standard, despite the fact that a more efficient alternative was available, the Dvorak keyboard. As such, the market dominance of the QWERTY keyboard is one of several potential long run standards that could have emerged as the standard; under a different sequence of realizations of shocks early in the process, the Dvorak keyboard would have emerged as the keyboard standard.⁴ As such, the process describing the evolution of the keyboard standard is nonergodic. As articulated by David, the QWERTY story has much in common with the model of technology adoption developed in Arthur (1989).

The QWERTY example has been subjected to very strong attacks by Liebowitz and Margolis (1990,1995). These authors argue that there is no reason to interpret the evolution of keyboards as an example of inefficient and path dependent technological lock-in. One thrust of this work has been to challenge the claim that the Dvorak keyboard is superior to QWERTY, either in terms of speed or in terms of ergonomics. They further argue that the QWERTY standard emerged out of a process that was far more competition-driven than David has suggested. David's rebuttals to these attacks

⁴ To be more precise, with positive probability a sequence of shocks could have occurred which resulted in a Dvorak keyboard standard.

(1997,2000) have not dealt with the specific evidence presented by Liebowitz and Margolis that the QWERTY example is inaccurate. Rather, his response has largely amounted to 1) arguing that Liebowitz and Margolis employ a faulty notion of path dependence, 2) citing other examples of technological lock-in such as pesticides (Cowan and Gunby (1995)), nuclear reactors (Cowan (1990)), or railway track gauges (Puffert (2002)), or 3) arguing that given the presence of network externalities, and other “frictions” (relative to the Arrow-Debreu world with the attendant first welfare theorem), the burden of proof should be on Liebowitz and Margolis to demonstrate that the QWERTY adoption is efficient and not path dependent, not for path dependence advocates to show that it was not.

The QWERTY controversy and related studies of path dependence unfortunately illustrate the limits of historical studies as much as they provide insight into economic complexity. To some extent, this is a function of the absence of a “new round” of empirical work or even debate that addresses the factual questions raised by Liebowitz and Margolis. Further, much of the often acrimonious discussion of QWERTY is irrelevant. Criticisms of the definition of path dependence as appear in the economic history literature suffer from the Socratic error of arguing that one cannot determine whether something is an instance of a class without a complete definition of the class; in this case it is certainly possible to determine whether the evolution of the typewriter keyboard standard is an instance of path dependence without addressing all aspects of the definition of path dependence.

Beyond the absence of progress in the analysis of QWERTY, there are also difficulties with drawing strong findings from the evidence that has been presented. Specifically, there has been inadequate attention to what is meant by counterfactuals in evaluating technology adoption.⁵ This is clearest when one assesses the impact of arguments on whether the Dvorak keyboard is or is not more efficient than QWERTY. Much of the controversy is based on examining efficiency studies that were done at different points in time in order to come up with an overall assessment of the relative merits of the competitors. However, the “true” relative efficiency of the two keyboards is

⁵ This is an issue raised in David (2000), but as will be apparent, I believe it applies to all participants in the QWERTY controversy.

only indirectly relevant in the sense that what presumably matters in arguing for nonergodicity in technology adoption is the joint evolution of available information and the emergence of a particular keyboard standard from the decisions of various economic actors.

For example, suppose that a new, unambiguously better keyboard were discovered today. My reading of their work suggests that Liebowitz and Margolis would not argue that this superior alternative would inevitably be adopted; adoption would depend on the costs of learning the new keyboard, fixed costs to producing new keyboards, etc. as compared to the benefits (and attendant implications for profit opportunities) that the new keyboard would provide. On the other hand, nonadoption would hardly be a compelling example that “history matters.” Nonergodicity of the type argued for in the technology literature is deeper than the claim that if network externalities and other costs to adoption are strong enough, then an inferior technology will survive the development of a superior competitor. David, (1986, pg. 43) for example, refers to the importance of “...‘historical accidents,’ which is to say, by the particular sequencing of choices made close to the beginnings of the process. It is there that essentially random, transient factors are most likely to exert great leverage...” The nonadoption of a new keyboard as I have described in my counterfactual hardly fits with this conception.

The sort of nonergodicity that arises in complex systems and that can be attributed to historical accidents in the sense of David is driven by the interplay of decisions and information along a dynamic path. Without a specific description of these dynamics, any historical analysis which focuses on limiting behavior will necessarily be unpersuasive. For this reason, suggesting that there should be a presumption in favor of findings of inefficient technological lock-in are unpersuasive, since along such paths, one needs to carefully account for selection pressures, just as in any evolutionary context, in order to meaningfully discuss the likelihood of efficiency or inefficiency, the possibility for multiple steady states or multiple meta-stable states⁶. The evolutionary game theory

⁶Meta-stable states are states of a system that, although not traps in that the system eventually leaves them, if entered then with high probability are not left for extremely long periods of time. Some of the historical studies of path dependence would seem to be

literature has made clear that presumptions one way or the other about long run convergence to efficient outcomes depend on a range of details of the environment under study;⁷

Beyond the specific argumentation of the particular historical inquiries, historical studies of path dependence may be faulted in terms of the links that are drawn between the empirical finding of path dependence and models of economic complexity. The logic of path dependence, in particular in the context of technological standards, derives entirely from the presence of network externalities, as well illustrated in work such as Farrell and Saloner (1985). The various studies of path dependent technologies do not illustrate any principle deeper than this. Now, while path dependence is equivalent to nonergodicity, at least as the term is employed in the literature and so may be consistent with a complex systems interpretation, it does not appear to be a particularly strong test. After all, coordination failure models evince a host of equilibria which are Pareto rankable and locally stable, but one would not want to equate such models with complexity per se. While the dynamic stories associated with technology standards are narrated in a way to suggest the presence of the sort of micro-structure associated with complex systems, there is in fact little direct or indirect evidence that this type of micro-structure is actually present. Put differently, it is unclear whether technology lock-in says anything more than that there may be large costs to changing an established technology. Notice that there is also an identification problem with respect to the source of such costs. The important generative mechanism in complex systems is the feedbacks between the decisions of individual actors. Yet this particular mechanism has not been empirically identified in the QWERTY case, and my own reading of other historical studies is that they do not do so either.

None of this means either that the historical evidence of nonergodicity or path dependence is incorrect or that the body of historical evidence is inconsistent with a complex systems perspective on the economy. Rather, my general conclusion is that

better conceptualized as examples of meta-stability; does one really want to argue that current nuclear reactor technology choice is totally irreversible in the sense required by nonergodicity?

⁷ See Samuelson (1997) for a valuable overview.

such studies are a useful source of ideas. But these studies do not currently represent persuasive evidence in support of the complex systems perspective.

4. Power Laws and Scaling Laws

A second area of empirical work on economic complexity has attempted to identify the presence in economic data of certain statistical properties that are associated with complex systems. In particular, this work has attempted to identify power and scaling laws. A random variable x_i is said to obey a power law if it has an associated distribution function $F_x(\gamma)$ such that

$$1 - F_x(\gamma) \sim \gamma^{-\alpha}$$

It is easy to see that the associated probability density function for x may be expressed as $f_x(\gamma) \propto \gamma^{-(1+\alpha)}$, so processes that obey power laws are Pareto distributed.

The most famous example of a power law is Zipf's law relating the frequency of objects and their sizes. Zipf's law is usually thought of in terms of a rank size rule. Letting $x_{(r)}$ denote the r 'th order statistic of a series x (so that $x_{(1)}$ is the largest value in the series, $x_{(2)}$ the second largest value, etc.) the series obeys the rank size rule if $rx_{(r)} = K$ for some constant K . The rank size rule can be shown to imply that $1 - F_x(\gamma) \sim \gamma^{-1}$, see Adamic (2003) for a simple derivation, and so has the canonical form of a power law. Zipf (1949) argued that many phenomena, ranging from the distributions of words in texts to the distribution of the number of species in genera to city sizes exhibit such properties. Recent research has focused on the identification of Zipf-type properties in a range of socioeconomic data. Important examples include Axtell (2001) on firm sizes and Gabaix (1999) on city sizes.

The study of power laws is usually conducted within the context of scaling laws. Scaling laws refer to relationships between statistics of a system that are qualitatively

scale-independent, in the sense that, modulo changes in some parameter values, the same relationship is found at different scales. Typically, these laws are expressed in linear form, although this may require transforming variables in ways that are unintuitive, at least to economists. The linear form of a power law is derived by taking logarithms,

$$\log f_x(\gamma) = c - (1 + \alpha) \log \gamma ;$$

indeed this expression is canonical in the scaling literature. Much of the empirical work on power and scaling laws explicitly searches for linear or piecewise linear relationships. Further, assessments of the presence of such laws are done visually, in essence by computing a graph of one variable against another and inspecting. Power and scaling laws appear to be ubiquitous in complex physical systems (at least with respect to their mathematical idealizations).

Within the physics community, there has emerged a subfield known as “econophysics” in which a major research activity is represented by efforts to find power and scaling laws in different socioeconomic data sets. While much of this work focuses specifically on power laws, it has also considered other probability distributions. This literature is well surveyed in the recent book by Mantegna and Stanley (2000). The primary focus of this research has been financial time series, apparently because of the large quantities of data available at high frequencies which permits the evaluation of cross-section distributions over different time horizons. However, a range of other data sets have also been explored. H. Eugene Stanley is arguably the leading figure in this empirical research program; work by him and coauthors includes findings such as

- i. *Per capita output.* Canning et al (1998) find that there is an approximately linear relationship between the log of the variance of residuals in per capital real output and the log of the level of output, i.e. if per capita output growth for country i in year t is decomposed as $g_{i,t} = \delta_i + \psi_t + r_{i,t}$ where δ_i is a country-specific fixed effect and ψ_t is a time fixed effect, then, denoting per capita output of i as y_i , then approximately,

$$\log \sigma_{r|y_i} \sim \kappa - .15 \log y_i$$

- ii. *Stock price fluctuations.* Gabaix et al (2002) have argued in favor of a cubic power law for stock price returns. Using individual stock price data, they compute normalized stock returns $r'_{i,t} = \frac{r_{i,t} - r_i}{\sigma_{r,i}}$ where r_i is the mean return on the stock and $\sigma_{r,i}$ its standard deviation. They find that for return horizons ranging from 15 minutes to 1 day, returns approximately (at least for large γ) obey a cubic power law, i.e.

$$1 - F_r(\gamma) \sim \gamma^{-3}$$

- iii. *Firm growth.* Amaral et al (1997) examine annual growth rates for US manufacturing companies. They find that the conditional density of firm growth rates r given the log of its size at the beginning of the year, s , is well approximated by an exponential distribution,

$$f(r|s) \sim \frac{1}{\sqrt{2}\sigma_{r|s}} \exp\left(-\frac{\sqrt{2}|r - E(r|s)|}{\sigma_{r|s}}\right)$$

The power law and scaling literature has identified a number of interesting statistical properties of different economic data series. As such, it has made a valuable contribution in identifying a range of “facts” that should help constrain theoretical modeling. Scaling laws are, in the context of complex systems, emergent properties, and so their presence would appear to speak to the empirical relevance of complexity. To the extent that the findings of scaling laws in very different data sets are believable, this can and has (e.g. Stanley et al (2001)) been interpreted as evidence of universal properties in economic data. However, the implications of this new literature for economic complexity

are still very unclear. The reason for this is that literature on power and scaling laws has yet to move beyond the development of statistical measures to the analyses of model comparison and evaluation⁸. In other words, many of the empirical claims in this literature concerning the presence of a particular law in some data set fail to adequately address the standard statistical issues of identification and statistical power. Hence, it is difficult to conclude that the findings in this literature can allow one to infer that some economic environment is complex.

Interpretative difficulties with respect to findings of scaling laws exist on several levels. First, it is clear that a number of the findings asserted in the literature are based on weak evidence. This can be seen quite clearly in the context of claims by various analyses to have found Zipf's law. For example, as in Gabaix (1999), evidence for Zipf's law can amount to conducting a regression of city size on city rank and arguing that a regression coefficient of -1 (since low rank order is associated with large population) and high goodness-of-fit verify that Zipf's law is present. However, the finding that regressions of rank size on the log of population provides a high R^2 is by itself obviously uninformative since the rank order is by construction inversely associated with the city sizes to begin with. As for the finding of a slope near -1 the fact that data are constructed to produce a negative coefficient makes the finding far less informative than it might appear; again, one does not know what to expect for data generated under plausible alternatives to Zipf's law.

Similar issues arise in Axtell's (2001) analysis, which is based on a regression of the log of firm size frequencies against firm size. So long as the density of firm size is monotonically decreasing over its range, this regression relates one monotonically decreasing series to another, which essentially guarantees a non-zero regression coefficient. Once again, it is unclear how non-Zipf data will perform under the procedure. Now, each author could legitimately argue that the goodness of fit of their regressions is sufficiently high that it cannot be explained by the fact that the two series under study are monotonic. But in order to make this response persuasive, it is necessary to have a benchmark as to what level of goodness of fit one will expect under alternative

⁸ See the insightful paper by Brock (1999) for an extended evaluation of the scaling literature, one that has strongly influenced my subsequent discussion.

distributions. Further, it is interesting to note that there appears (when one inspects the graphs of the series in these papers) some deterioration of the goodness of fit near the extreme points of the range of sizes. Axtell suggests that for his context this reflects truncation effects, but it might also be evidence of misspecification.

Other issues may be raised as well. For example, some of the data transformations that appear in this literature are questionable. In studies of city sizes, small cities are usually dropped as they do not fit the law; but it is unclear this is a legitimate practice. In regressions of log probability against log size, one can imagine (as for Axtell (2001)) that the relationship between the two is sensitive to the way observations are aggregated to create observations in the regression; yet such issues have not been studied. Combining specific issues like these with the more general problems of interpreting the exercises, I believe that most of the claims in support of Zipf's law are overstated. The empirical Zipf's law literature needs to formally address the power of different procedures to discriminate between evidence of Zipf's law and various alternatives and to deal with issues of data mining with the same care as other empirical literatures.

Identification problems again appear in the context of financial market studies. An important critique of work on scaling is due to LeBaron (2001). He shows how a number of simple stochastic volatility models can produce similar scaling and power law phenomena to those that have been reported in the econophysics literature, at least in the sense that the data plots of the distribution of returns (for example the logs of the distribution function of returns against the log of returns) from a stochastic volatility model appear similar to those found in the Dow. LeBaron's paper makes clear that use of visual inspections of log/log probability plots to uncover power laws can lead to very misleading inference.

Additional insight into the limitations of the scaling law literature may be obtained when one examines the commentary generated by LeBaron's paper. In particular, the commentary by Stanley and Plerou (2001) is very informative. Stanley and Plerou reject LeBaron's claims essentially by arguing that their graph of stock return data revealed a power law that was mathematically impossible under LeBaron's specification. But this criticism seems to miss the point that for the data set LeBaron

analyzed, the model he analyzed produces very similar figures to those found empirically. The fact that LeBaron's model performs less well (at least visually), when one moves from approximately 28,000 to 1,000,000 observations may reflect the need for a more complex stochastic volatility model than the 3-factor model he in fact employs. The main point to emphasize is that the power law/scaling literature has yet to develop appropriate methodologies for model comparison exercises.

There are good reasons to suppose that the issue of power matters in practice. As discussed in a recent survey by Mitzenmacher (2002), it is quite difficult to distinguish whether a process obeys a power law as opposed to a log normal distribution, which would mean the log probability density of the process is

$$\log f_x(\gamma) = -\log x - \log \sqrt{2\pi}\sigma_x - \frac{(\log x - E(x))^2}{2\sigma_x^2}$$

As Mitzenmacher (2002) observes, one can rewrite this expression as

$$\log f_x(\gamma) = -\frac{(\log x)^2}{2\sigma_x^2} + \left(\frac{E(x)}{\sigma_x^2} - 1 \right) \log x - \sqrt{2\pi}\sigma_x - \frac{E(x)^2}{2\sigma_x^2}$$

Comparing this expression with $\log f_x(\gamma)$ associated with a power law, it is apparent that

it is quite difficult to distinguish a power law from a log normal unless $\frac{(\log x)^2}{2\sigma_x^2}$ is

sufficiently large. Hence, log normal processes can exhibit approximate power laws for part of the range of x ; discriminating between power laws and the log normal will depend on the extreme tail of the empirical density of x . Yet my reading of the literature is that one often sees a deterioration of the goodness of fit in the extreme tail of empirical densities. Further, estimates of the tail are likely to be relatively inaccurate due to a paucity of observations. These considerations suggest that much more work needs to be done to develop persuasive evidence that power laws are as common in economic data as has been claimed.

Stanley and Plerou (2001) raise another issue that is important in assessing empirical work on scaling laws. They note that models that produce scaling laws typically only do so as an asymptotic limit, i.e. the scaling law becomes a good approximation for the probability distribution $F_x(\cdot)$ only when the argument of the function is large. This raises unresolved questions of how to empirically evaluate evidence of scaling laws when the laws themselves are only an approximation. This suggests an interesting parallel between the scaling law literature and the macroeconomic literature on model calibration, where the formal testing of models has been replaced with qualitative evaluations of whether a model is a good approximation. Many researchers (including myself) regard model calibration as very problematic as it does not have clear criteria by which a model can be concluded to be falsified. This same problem of ill-defined evaluative criteria arises in the scaling literature.

Beyond these specific problems, the empirical literature may be faulted for failing to adequately address issues of heterogeneity. As pointed out in Brock (1999) (who makes a number of important arguments about heterogeneity), scaling laws are generally analyses of unconditional objects such as the distribution of a series or the distribution of a series given one variable (size). From the perspective of much of empirical economics, such unconditional objects are not natural objects of inquiry since differences in individual characteristics are usually very pronounced. Two obvious examples of this are the cross-country growth behavior, where a primary focus of empirical study has been the identification of the full range of sources of heterogeneity between countries (see Durlauf and Quah (1999) for a survey) and the study of treatment effects on individual outcomes, where heterogeneity in who receives treatments and in the effects of treatments are of first order importance in evaluating a given policy intervention (cf. Heckman (2000,2001) and Manski (1995)).⁹

⁹ Some of the econophysics models that generate power laws are in fact driven by heterogeneity in the sense that an important feature of them is the presence of a random parameter that varies across agents; Gabaix (1999) uses one model of this type to generate Zipf's law behavior; a similar model is described in Sornette (1998). However, this is different than accounting for heterogeneity that reflects economically meaningful, and possibly measurable, differences.

Why would heterogeneity across the objects under study matter in interpreting scaling findings? Again, the issue comes down to identification and power. One of the consistent findings in scaling studies is the presence of thick tailed densities; in fact power laws are a leading example of such densities. Heterogeneity can produce such thickness. One way to think about heterogeneity is from the perspective of mixture distributions; failing to account for heterogeneity is tantamount to failing to account for the mixing distribution that characterizes a latent parameter that differs across agents. As is well known from the literature on mixture distributions, cf. Lindsay (1995, chapter 2), mixtures generally possess thicker tails than the underlying component densities from which they are generated. So, especially for small data sets (Axtell (2001), for example uses fewer than 15 observations in his Zipf's law analysis), one can well imagine that findings of thick tailed distributions could be an artifact of the heterogeneity of the data and that statistical claims that the data support a particular density are likely to suffer from serious problems of statistical power.

Second, the empirical literature on scaling laws is difficult to interpret because of the absence of a compelling set of theoretical models to explain how the laws might come about. This is very much the case if one examines the efforts by physicists to explain findings of scaling laws in socioeconomic contexts. Within physics, it is an accepted practice to identify patterns in data and then look for stochastic processes which can explain them. As such, one very typically finds that the economic model proposed to explain some empirical findings by econophysicists amounts to taking a stochastic process known to generate the phenomenon and labeling elements of it as individual actors.¹⁰ In economics, in contrast, the search for models to explain an empirical finding is much more restricted in that such searches are limited to models in which the behavior of individual agents may be understood as deriving from the interplay of preferences,

¹⁰ Bouchaud (2001) is a very readable overview of this type of theorizing.

constraints and beliefs.¹¹ For this reason, efforts by physicists to explain scaling laws have generally been regarded as failures from the perspective of social scientists.¹²

Is this negative assessment fair? I believe that it is. The econophysics approach to economic theory has generally failed to produce models that are economically insightful. If one wants to know how a policy change will affect the volatility of some aggregate, it is necessary to specify how the policy affects the behavior of the agents whose actions produce the aggregate. This specification requires that individual behavior be understood through the interplay of preferences, constraints and beliefs that is very typically missing from econophysics models. Notice that my criticism is not that econophysicists fail to employ models based on particular rationality assumptions or models that assume frictionless markets (both common criticisms of economics in that community). My criticism is that they do not use models that adequately respect the purposefulness of individual behavior.

I am unaware of any theoretical argument that would imply that the presence of universal properties to systems of interacting agents can allow one to avoid developing a detailed (and plausible) specification of individual behavior in order to analyze counterfactuals such as the effects of different policies. I also believe that the absence of a body of economically sensible models that generate scaling laws has implications for the assessment of empirical findings of scaling laws. With respect to the broad (albeit less formal) question of how to interpret evidence, it is very reasonable to downplay the finding of scaling laws when there does not exist a body of economic theory that suggests such a finding is reasonable.

The weaknesses of the behavioral assumptions that underlie complex systems models of economic phenomena are particularly apparent in the context of financial markets models. Many such models, a good example of which is Lux and Marchesi (1999), have been constructed to produce fat tailed distributions in returns and long

¹¹ There is also a strong tendency in the econophysics community to denigrate the body of existing economic theory, leading both to a misunderstanding of that theory as well as a failure to integrate complex systems perspectives into the theory. Instead, one sees theoretical models proposed that all too often make little sense to a social scientist.

¹² Brock (1993) is a deep analysis of how rigorous physics models may be adapted to socioeconomic contexts without sacrificing basic economic logic, suggesting that these efforts need not inevitably fail.

memory in volatility, two features whose empirical salience LeBaron (2001) has questioned. Regardless of whether these properties are or are not present in financial data, a common problem with these models is that their logic does not embody what I believe are fundamental features of financial markets. In particular, these models often fail to embody behavioral assumptions that ensure the nonexistence of arbitrage opportunities. In my view, no arbitrage constitutes the most fundamental empirical feature of asset markets and so is properly a building block of the modern theory of finance. The standard theory of finance has little to say about fat tails and long memory in volatility; the theory is not inconsistent with these features but rather does not provide any insight as to why they might occur. A successful advance of the current theory of finance should therefore show how to augment a model that reflects no arbitrage in ways that explains these additional features. Instead, behaviors that will eliminate arbitrage opportunities are often not built into these models and presence or absence of arbitrage opportunities is given inadequate attention in the analysis of equilibrium asset returns produced by the model. It is not an advance, in my view, to specify an environment in which arbitrage opportunities go unexploited; this is simply bad microeconomics¹³

The economics literature has also failed to provide a compelling body of theoretical models to explain scaling laws.¹⁴ A partial exception to this claim is Gabaix (1999) where Zipf-type phenomena are interpreted using a stylized but by no means uninterpretable behavioral framework. One potential approach is to employ versions of models such as Brock and Durlauf (2001a,b, 2003) which have well-specified microeconomic structures yet produce probabilistic descriptions of aggregate behavior that are mathematically isomorphic to complex systems (in this case, statistical mechanics) models and investigate whether they can produce scaling-type phenomena.

¹³ For example, Lux and Marchesi (1999) argue that excess holding returns in their model are consistent with white noise, but the statistical calculation they use for this claim was developed to identify the presence of long memory and has unknown statistical (in particular) power properties. There are many available tools for conducting a far stronger analysis.

¹⁴ There are some classic older papers especially Champernowne (1948) that provide models of thick tailed income distributions, but even this paper generally subsumes the economic reasoning to the judicious choice of assumptions to desired shape for the tail of the income distribution. However, the Champernowne's model is much richer than those found in the econophysics literature.

Without advances of this type, the various empirical findings on power laws and scaling are likely to continue to be regarded as curiosities.

5. Social Interactions

The final literature that is important in empirically assessing economic complexity concerns social interactions. Studies of social interactions have generally not directly addressed the complex system properties described in Section 2. On the other hand, it is a literature that has directly addressed the assumptions that underlie the complexity literature. Specifically, this literature has attempted to measure interdependences in behavior across individuals. As such, it contrasts with the scaling law literature, which focuses exclusively on outcomes and the historical literature, whose discussion of interdependences is generally somewhat sketchy. With respect to the question of nonergodicity, this literature is potentially very important since it is the strength of these interdependences that determines whether nonergodicity holds. Similarly, phase transitions are driven by the presence of nonlinearities in interdependences which this work can in principle identify. For some cases, such as discrete choice or duration models, nonlinearities are usually intrinsic as the behaviors of others affects the probability of an action by one person, so the empirical issue is the form these nonlinearities take.

One part of the social interactions literature has explicitly attempted to model the influence of groups on individual behavior. In such work, a researcher typically constructs a probability model to predict individual behavior given a set of individual level and group level control variables. Individual level variables include factors such as family and parental characteristics. Group level variables include factors such as the percentage of others in an individual's reference group who choose a certain behavior and the group level average of some characteristic such as educational attainment.¹⁵

¹⁵ Other work is more historical in nature as it identifies contexts where group influences seem to determine aggregate behaviors. A nice example is Young and Burke (2001) who

This literature has explored a wide range of contexts. A leading area of work concerns residential neighborhood effects; this literature has focused on how peers as defined by geographic proximity, sometimes differentiated by ethnic group as well, influence a range of social and economic factors. Evidence of neighborhood based social interactions has been found in the context of teenage fertility (Crane (1991) and South and Baumer (2000)), use of government-based programs to help the poor (Aizer and Currie (2002), Bertrand, Luttmer, Mullainathan (2000)) and crime (Sirakaya (2002)). While much of this work has focused on individual level data, there has also been a literature that has attempted to measure the presence of social interactions using aggregate data: examples include Glaeser, Sacerdote, and Scheinkman (1996) and Topa (2001). This literature has identified interdependence effects in a wide range of contexts; some of this evidence is surveyed in Brock and Durlauf (2001) and Durlauf (2003).

As in the previous cases, there are reasons why the empirical evidence that has been generated on social interactions is not decisive. One problem with much of the empirical social interactions literature is a lack of attention to distinguishing between different sources of social effects. As argued in a seminal contribution, Manski (1993), social effects may either be contextual, i.e. predetermined characteristics of a group may affect individual choice, or endogenous, i.e. represent direct feedbacks between agents. Manski (1993) provides a class of linear models in which it is impossible to identify contextual versus endogenous effects; this follows from what Manski has named the “reflection problem” whereby endogenous effects and contextual effects are perfectly correlated. Brock and Durlauf (2001b) and Moffitt (2001) provide additional analysis of this issue. In particular, as developed in Brock and Durlauf (2001b), identification of the parameters associated with different types of effects requires prior information that there exist individual determinants of behavior whose group level averages are not contextual effects and/or nonlinearities in the behavioral process which are strong enough to allow one to rule out endogenous and contextual effects from being collinear. While the various econometric studies have received frequent citation in the empirical literature, their main ideas have not been incorporated into empirical practice. As a result, different studies

show how the terms of tenant farmer contracts in Illinois exhibit bunching that is strongly suggestive of social interaction effects.

employ different measures of group effects with little attention to the question of distinguishing different effects from one another.

Further, much of the empirical social interactions literature has failed to deal with the problem of endogeneity in group memberships. In the case of residential neighborhoods, one would expect that individuals residing in the same neighborhood will have similarities with respect to unobservable characteristics. If correlation in unobservables is not accounted for, then estimates of neighborhood effects will be biased as neighborhood-level controls can proxy for these unobservables. Exceptions to this generalization include Aaronson (1998) and Evans, Oates, and Schwab (1992). Interestingly, these studies come to very different conclusions, with Aaronson concluding neighborhood effects are present whereas Evans, Oates and Schwab finds that evidence of such effects vanishes when one uses instrumental variables to account for endogeneity. The disparity in findings is not surprising given the different data sets and statistical methodologies employed. Neither paper formally models neighborhood membership and subsequent behavior as a joint process, so each may be neglecting important information with respect to identifying neighborhood effects; see Brock and Durlauf (2001b, 2003) for further discussion. In an important recent paper, Ioannides and Zabel (2002) attempt to use the information associated with self-selection into neighborhoods to identify social interaction effects and find that this information is quite useful.

My own view is that the social interactions literature nevertheless contains the strongest overall evidence that the forces that produce complexity are in fact present. One reason for this is that there is an immense amount of historical, ethnographic, and social psychological evidence that supports the belief that social interaction effects are important. Another reason is that despite the problems that exist with the various empirical studies, this literature contains much tighter connections between theory and econometrics and much more attention to issues of statistical power and specification than the other empirical literatures that bear on complexity. However, this evidence is still sufficiently weak that the social interactions literature cannot be said to have demonstrated the presence of the sorts of microfoundations that underlie complex economic systems.

6. Summary and Conclusions

As my brief discussion has indicated, each of the main parts of the empirical literature on economic complexity suffers from serious weaknesses. Historical studies have provided good reasons to believe that technological lock-in can occur, but at least in the case of the QWERTY keyboard, have not shown that lock-in has anything to do with complexity per se. The study of power and scaling laws has opened new areas of empirical inquiry, but are difficult to interpret when potential heterogeneities and forms of temporal dependence are ignored. The empirical literature on social interactions makes the important contribution of focusing on the interdependences that are the essential feature of complex environments, but has generally failed to adequately distinguish between feedback and other social effects as well as from a failure to adequately address issues of endogeneity of the groups through which interactions occur. None of these criticisms shows that economic complexity is a mirage, but implies that evidence that has been adduced of the value of the complexity perspective is far weaker than its advocates have claimed.

From the perspective of econometrics, as my discussion of the specific research literatures has shown, identification problems are endemic to the current empirical literature on complexity. The disparate empirical strategies that have been employed to provide evidence on economic complexity have yet to integrate theoretical models of complexity with data analysis in such a way as to show how a given aggregate property is associated with the interactions between agents in a way that allows for a plausible finding that a given environment is in fact complex.

How might the empirical literature on economic complexity be rendered more persuasive? One possibility is for empirical work to follow the path of structural estimation. Models such as those developed in Brock and Durlauf (2001a,b,2003) are based on the sort of mathematical structures that one finds in many complex systems, specifically those of statistical mechanics. At the same time, the models are constructed in such a way that the behavioral descriptions are also interpretable as likelihood functions. Structural econometrics of this type has proven invaluable in the evolution of

a wide range of theories of microeconomics and macroeconomics. Most important, structural approaches will allow for direct feedbacks between empirics and theory as empirical findings suggest ways to modify theories and theoretical models guide empirical exercises.

In addition, it is clear that there is a need for the development of methods to allow for more fruitful evaluation of the basic statistical facts that are the hallmark of complex systems. An important example of how this might be accomplished is due to Ioannides and Overman (2002). Ioannides and Overman show how tests for Gibrat's law, i.e. the independence of the rate of growth from initial size, may be used as a test of Zipf's law for city sizes. Their analysis both provides much more powerful evidence of Zipf's law than has typically appeared and is also able to identify limits to the accuracy of the Zipf's law across the city size distribution. This is an extremely promising approach and indicates how some of the empirical claims made concerning economic complexity can be strengthened.

The limitations of the current empirical literature are not surprising given the immaturity of economic complexity as a paradigm. The current generation of empirical work has been produced by advocates of complexity and has been motivated by a belief that the approach matters. I suspect that economic complexity, like many other ideas, will experience a dialectical empirical development, as an initial wave of supportive studies is criticized, leading to the development of more persuasive evidence. Such a process is critical to the long term viability of economic complexity as it will not become a major component of economic reasoning until a tight connection between theoretical work and empirics is developed. Unless such a connection is achieved, even an openminded scholar will be justified in taking the Scottish legal option of concluding that the importance of complexity in understanding socioeconomic phenomena is "not proven."

References

Aaronson, D., (1998), "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," *Journal of Human Resources*, XXXII, 915-946.

Adamic, L., (2003), "Zipf, Power-laws and Pareto: A Ranking Tutorial," mimeo available at <http://ginger.hpl.ho.com/shk/papers/ranking/ranking/html>.

Aizer, A. and J. Currie, (2002), "Networks or Neighborhood? Correlations in the Use of Publicly-Funded Maternity Care in California," *NBER Working Paper no. 9209*.

Amaral, L., S. Buldyrev, S. Havlin, P. Maass, M. Salinger, H. E. Stanley, and M. Stanley, (1997), "Scaling Behavior in Economics: The Problem of Quantifying Company Growth," *Physica A*, 1-24.

Anderson, P., K. Arrow and D. Pines, eds., (1988), *The Economy as an Evolving Complex System*, Redwood City: Addison Wesley.

Arthur, W. B., (1989), "Increasing Returns, Competing Technologies and Lock-In by Historical Small Events: The Dynamics of Allocation Under Increasing Returns to Scale," *Economic Journal*, 99, 116-131.

Arthur, W. B., S. Durlauf, and D. Lane, eds., (1997), *The Economy as an Evolving Complex System II*, Redwood City: Addison-Wesley.

Axtell, R., (2001), "Zipf Distribution of US Firm Sizes," *Science*, 293, 1818-1820.

Blume, L. and S. Durlauf, (2003), eds., *The Economy as an Evolving Complex System III*, in process.

Bouchaud, J.P., (2001), "Power-Laws in Economy and Finance: Some Ideas from Physics," *Quantitative Finance*, 1, 105-112.

Brock, W. (1993), "Pathways to Randomness in the Economy: Emergent Nonlinearity and Chaos in Economics and Finance," *Estudios Economicos*, 8, 3-55.

Brock, W. (1999), "Scaling in Economics: A Reader's Guide," *Industrial and Corporate Change*, 8, 3, 409-446.

Brock, W. and Durlauf, S., (2001a) "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235-260.

_____(2001b), "Interactions-Based Models," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics 5*, Amsterdam: North-Holland.

Brock, W. and S. Durlauf, (2003), "Multinomial Choice with Social Interactions," mimeo, University of Wisconsin.

Canning, D., L. Amaral, Y. Lee, M. Meyer, and H. E. Stanley, (1998), "Scaling the Volatility of GDP Growth Rates," *Economic Letters*, 60, 335-341.

Champernowne, D., (1948), "A Model of Income Distribution," *Economic Journal*, 63, 318-351.

Cowan, R., (1990), "Nuclear Power Reactors: A Study in Technological Lock-In," *Journal of Economic History*, 50, 541-567.

Cowan, R. and P. Gumby, (1996), "Sprayed to Death: Path Dependence, Lock-In, and Pest Control Strategies," *Economic Journal*, 106, 521-542.

Crane, J., (1991), "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology*, 96, 1226-1259.

David, P., (1985), "Clio and the Economics of QWERTY," *American Economic Review*, 75, 332-337.

David, P., (1986), "Understanding the Economics of QWERTY: The Necessity of History," in *Economic History and the Modern Economist*, W. Parker, ed., Oxford: Blackwell.

David, P., (1997), "Path Dependence and the Quest for Historical Economics: One More Chorus for the Ballad of QWERTY," *Oxford University Discussion Papers in Economic and Social History* no. 20.

David, P., (1999), "At Last, A Remedy for Chronic QWERTY-Skepticism!," *Stanford University Economics Department Working Paper* no. 99-025 .

David, P., (2000), "Path Dependence, Its Critics, and the Quest for 'Historical Economics'," *Stanford University Economics Department Working Paper* no. 00-020.

Durlauf, S., (2001), "A Framework for the Study of Individual Behavior and Social Interactions," *Sociological Methodology*, 31, 47-87, 2001.

Durlauf, S., (2003), "Groups, Social Influences and Inequality: A Membership Theory Perspective on Poverty Traps," mimeo, University of Wisconsin.

Evans, W., W. Oates, and R. Schwab, (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*, 100, 5, 966-991.

- Farrell, J. and G. Saloner, (1985) "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, 16, 70-83.
- Feldmann, A. and W. Whitt, (1996), "Fitting Mixtures of Exponentials to Long-Tail Distributions to Analyze Network Performance Models," mimeo, AT&T Laboratory.
- Gabaix, X., (1999), "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 739-767.
- Glaeser, E., B. Sacerdote, and J. Scheinkman, (1996), "Crime and Social Interactions," *Quarterly Journal of Economics*, CXI, 507-548.
- Farrell, J. and G. Saloner, (1985) "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, 16, 70-83.
- Ioannides, Y, (2001), "Topologies of Social Interactions," mimeo, Tufts University.
- Heckman, J. (2000), "Causal Parameters and Policy Analysis: A 20th Century Perspective," *Quarterly Journal of Economics*, 115, 1, 457-497.
- Heckman, J., (2001), "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 673-748.
- Ioannides, Y. and H. Overman, (2002), "Zipf's Law for Cities: An Empirical Examination," *Regional Science and Urban Economics*, forthcoming.
- Katz, M. and C. Shapiro, (1986), "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, 822-841.
- Krugman, P., (1996), *The Self-Organizing Economy*, Oxford: Basil Blackwell.
- LeBaron, B., (2001), "Stochastic Volatility as a Simple Generator of Power Laws and Long Memory," *Quantitative Finance*, 1, 621-631.
- Liebowitz, S. and S. Margolis, (1990), "The Fable of the Keys," *Journal of Law and Economics*, 22, 1-26.
- Liebowitz, S. and S. Margolis, (1995), "Path Dependence, Lock-In, and History," *Journal of Law, Economics, and Organization*, 11, 205-226.
- Lindsay, B., (1995), *Mixture Models: Theory, Geometry, and Applications*, Hayward: Institute for Mathematical Statistics.
- Lux, T. and M. Marchesi, (1999), "Scaling and Criticality in a Stochastic Multi-Agent Model," *Nature*, 397, 498-500.

Manski, C., (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-542.

Manski, C., (1995), *Identification Problems in the Social Sciences*, Cambridge: Harvard University Press.

Mantegna, R. and H. E. Stanley, (2000), *Introduction to Econophysics*, New York: Cambridge University Press.

Mitzembacher, M, (2002), "A Brief History of Generative Models for Power Law and Log Normal Distributions," mimeo, Harvard University.

Puffert, D., (2002), "Path Dependence in Spatial Networks: The Standardization of the Railway Track Gauge," *Journal of Economic History*. 39, 282-314.

Samuelson, L, (1997), *Evolutionary Games and Equilibrium Selection*, Cambridge: MIT Press.

Schelling, T., (1971), "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1, 143-186.

Sirakaya, S. (2002), "Recidivism and Social Interactions," mimeo, University of Wisconsin.

Sornette, D., (1998), "Linear Stochastic Dynamics with Nonlinear Fractal Properties," *Physica A*, 250, 295-314.

South, S. and E. Baumer, (2000), "Deciphering Community and Race Effects on Adolescent Premarital Childbearing," *Social Forces*, 78, 4, 1379-1408.

Stanley, H. E., L. Amaral, P. Gopikrishnan, and V. Plerou, (2000), "Scale Invariance and Universality of Economic Fluctuations," *Physica A*, 283, 31-41.

Stanley, H. E. and V. Plerou, (2001), "Scaling and Universality in Economics: Empirical Results and Theoretical Interpretation," *Quantitative Finance*, 1, 563-567.

Topa, G., (2001), "Social Interactions, Local Spillovers, and Unemployment," *Review of Economic Studies*, 68, 2, 261-295.

Young, H. P. and M. Burke, (2001), "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture," *American Economic Review*, 91, 559-573.

Zipf, G., (1949), *Human Behavior and the Principle of Least Effort*, Cambridge: Addison-Wesley.