YORK UNIVERSITY

Economics from the Top Down
Does Hierarchy Unify Economic Theory?

Blair Fix

A Dissertation
submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements
for the Degree of
Doctor of Philosophy

Graduate Program in Environmental Studies
York University
Toronto, Ontario

August 2018

Abstract

What is the unit of analysis in economics? The prevailing orthodoxy in mainstream economic theory is that the individual is the 'ultimate' unit of analysis. The implicit goal of mainstream economics is to root macro-level social structure in the micro-level actions of individuals. But there is a simple problem with this approach: our knowledge of human behavior is hopelessly inadequate for the task at hand. Faced with real-world complexities, economists are forced to make bold (and seldom tested) assumptions about human behavior in order to make models tractable. The result is theory that has little to do with the real world.

This dissertation investigates an alternative approach to economics that I call 'economics from the top down'. This approach begins with the following question: *what happens when we take the analytical focus off of individuals and put it into social hierarchy*? The effect of this analytical shift is that we are forced to deal with the realities of concentrated power. The focus on hierarchy leads to some surprising discoveries. First, I find evidence that hierarchical organization has a biophysical basis. I show that institution size (firms and governments) is strongly correlated with rates of energy consumption, and that the growth of institutions can be interpreted as the growth of social hierarchy. Second, I find that hierarchy plays an important role in shaping income and income distribution. I find that income scales strongly with *hierarchical power* (defined as the number of subordinates under one's control), and that hierarchical power affects income more strongly than any other factor measured. Lastly, using an empirically informed model of the hierarchical structure of US firms, I find that hierarchy plays a dominant role in shaping the income distribution tail.

These results hint that hierarchy can be used to unify the study of economic growth (understood in biophysical terms) and income distribution. I conclude by making the first prediction of how the concentration of hierarchical power should relate to the growth of energy consumption. This prediction sheds new light on the origin of inequality. While this 'top down' approach to economics is in its infancy, the results are encouraging. Focusing on hierarchy gives fresh insight into many of the important questions facing society — insight that cannot be obtained by focusing on individuals.

Acknowledgments

It is not easy to forge your own scientific path, let alone to declare that much of what has been written in your field needs to be rethought. Contrarian thinking often leads to isolation. Thankfully, I have not been isolated during my time at York, and much of this has to do with the work of Jonathan Nitzan. Together with Shimshon Bichler, Jonathan has created a path-breaking approach to political economy that has strongly shaped my thinking. But more than this, Jonathan has provided many opportunities for me to share my research, and has offered extremely useful feedback. For this I am grateful. I would also like to thank the 'capital as power' community. The many discussions on the web forum have been intellectually invigorating. Parts of this dissertation have benefited from discussions with Shai Gorsky and James McMahon.

I also owe a great debt to Charlie Hall, who took the risk of letting me (an unknown PhD student) write a book in his Energy Briefs series. The experience taught me a great deal about how to think like a scientist, and how to communicate effectively. I would also like to thank Ellie Perkins for advising me during my seven years at York, and for giving me the freedom to chart my own path. Thank-you also to Mark Thomas and Justin Podur for the many helpful comments. I am also grateful for financial support from the Social Sciences and Humanities Resource Council, the Ontario Graduate Scholarship Program, and the Lewis and Bennett Graduate Scholarship.

Lastly, thank-you to Grace and Garry Fix for the many hours of proof reading, and to Emily and Petra for the support along this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction: Economics from the Top Down

[I]n modern economics ... the ultimate unit of analysis is always the individual; more aggregative analysis must be regarded as only provisionally legitimate.

— Brennan and Tullock [1]

This dissertation offers a new approach to economic theory that I call 'economics from the top down'. To avoid confusion, this has nothing to do with 'trickle-down' economics, or with prescriptive economics of any kind. Instead, 'economics from the top down' is an investigative approach that is motivated by the following question: *what happens when we take the analytical focus off of individuals and put it into social hierarchy*?

By the standards of modern neoclassical orthodoxy, this is a heretical question. As Brennan and Tullock articulate, individuals are *the* basic unit of analysis in economics — full stop. But while it is an admirable goal to try to root complex social structure in the behavior of individuals, we are hopelessly far from being able to achieve this correctly. The problem is that we do not have an accurate model of human behavior (one that can make precise, falsifiable predictions). Without such a model, seeking to explain social structure in terms of the behavior of individuals is fraught with difficulty. To mitigate our ignorance about human behavior, bold assumptions must be made. But even if a model gives realistic results, its validity remains entirely hinged to the foundational assumptions, which often go untested [2].

E.O. Wilson, discussing a similar problem in biology, deserves to be quoted at length. To make this passage relevant to economics, simply replace the word 'biologists' with 'economists' and substitute any biological phenomena with social phenomena:

> Biologists, it has been said, suffer from physics envy. They build physics-like models that lead from the microscopic to the macroscopic, but find it diffi-

cult to match them with the messy systems they experience in the real world. Theoretical biologists are nevertheless easily seduced. ... Armed with sophisticated mathematical concepts and high-speed computers, they can generate unlimited numbers of predictions about proteins, rain forests, and other complex systems. With the passage to each higher level of organization, they need to contrive new algorithms, which are sets of exactly defined mathematical operations pointed to the solution of given problems. And so with artfully chosen procedures they can create virtual worlds that evolve into more highly organized systems. Wandering through the Cretan labyrinth of cyberspace they inevitably encounter emergence, the appearance of complex phenomena not predictable from the basic elements and processes alone, and not initially conceivable from the algorithms. And behold! Some of the productions actually look like emergent phenomena found in the real world.

Their hopes soar. They report the results at conferences of like-minded theoreticians. After a bit of questioning and probing, heads nod in approval: "Yes, original, exciting, and important — if true." If true ... if true. *Folie de grandeur* is their foible, the big picture their illusion. They are on the edge of a breakthrough! But how do they know that nature's algorithms are the same as their own, or even close? Many procedures may be false and yet produce an approximately correct answer. The biologists are at special risk of committing the fallacy of affirming the consequent: It is wrong to assume that because a correct result was obtained by means of theory, the steps used to obtain it are necessarily the same as those that exist in the real world. [3]

Wilson makes the problem extremely clear. But in the face of a paucity of knowledge, how do we proceed? How do we build social science theories that connect the macro level to the micro level? The method that I adopt in this dissertation is to partially chip away at the problem. I admit that we know very little about human behavior, so I do not attempt a 'bottom-up' approach [4]. That is, I do not attempt to explain macro-level social structure in terms of the behavior of individuals. Instead, I adopt a 'top-down' approach. I attempt to explain macro-level social structure in terms of *another* structure — social hierarchy. The name 'top down' serves two purposes. Firstly, it differentiates my method from the 'bottom-up' approach — often called methodological individualism [5, 6]. Secondly, the name 'top-down' nicely captures the focus on hierarchy. In a hierarchy, power flows from the *top down*. By focusing on the top-down structure of hierarchy, we implicitly put concentrated power at center stage. This is very different from the neoclassical approach, in which concentrated power is ignored, or even assumed not to exist [7–9].

But why focus specifically on *hierarchy*? Why not some other social structure? I choose hierarchy as my unit of analysis for a number of reasons. First,

hierarchy is *ubiquitous*. Hierarchy, I believe, is the basic building block for most (if not all) modern institutions, and it has deep roots that likely extend into prehistory [10, 11]. And in evolutionary terms, humans are but one of a vast number of social animals, virtually all of which use hierarchy as a method of social organization [12–17]. A second reason to focus on hierarchy is that it offers a simple way of studying the class structure of society. While many social scientists have stressed a focus on class structure [18–28], there is no consensus on how classes should be defined and studied. Hierarchy is useful because it provides a mathematically-generalizable form for defining and studying social class. Lastly, I am interested in hierarchy because it is conspicuously *absent* from mainstream economic theory, and thus its role in shaping social structure is poorly understood.

1.1   Summary of Findings

On the face of it, this dissertation is a sprawling journey through a wide variety of seemingly unrelated social phenomena. At various points, I investigate energy consumption, institution size, technological change, intra-firm income distribution, the different factors that affect income, personal income distribution, functional income distribution, and changes in income inequality over time. How are these things possibly related? The surprising finding in this dissertation is that all of these phenomena can be linked to social hierarchy. Let me explain how.

In Chapter 2, I explore the relation between energy consumption and institution size. I find that as energy consumption increases (both across space and across time), there is a systematic increase in institution size. Specifically, as energy consumption increases, self-employment declines, employment in large firms increases, average firm size increases, and government employment increases. I find evidence that these trends are indicative of a general increase in social hierarchy with energy consumption. Why is hierarchy related to energy consumption? I hypothesize that increasing energy consumption requires increasing the scale and complexity of technology, which in turn, requires greater social coordination. But according to the work of anthropologist Robin Dunbar, brain size places a key limit on primate group size. [29–31]. Dunbar's primate evidence predicts an average human group size of about 150 (Dunbar's number). Building on the work of Turchin and Gavrilets [32], I propose that hierarchy allows humans to sidestep this group-size limitation. A hierarchy's nested chain of command allows group size to grow without any corresponding increase in the

number of required social relations. This suggests that increasing hierarchical organization plays a central role in increasing energy consumption.

In Chapter 3, I turn the focus to personal income. The dominant paradigm in personal income distribution theory is that income stems from productivity. But this approach has a severe (but little discussed) problem: when differences in individual productivity are measured objectively (and not circularly), they are far too small to account for observed differentials in income. But if not productivity, then what explains differences in personal income? I propose that personal income is most strongly determined by *hierarchical power*. What is hierarchical power? I define it as the ability to influence subordinates within a hierarchical chain of command. I measure hierarchical power in terms of the number of subordinates under an individual's control. Using this metric, I find that relative income within firms scales strongly with hierarchical power. Using data for intra-firm promotions/demotions, I also find that *changes* in relative income within firms scale strongly with changes in hierarchical power. Lastly, I find that grouping individuals by hierarchical level (across firms) affects income more strongly than any other factor measured. This evidence suggests that hierarchy plays a key role in shaping personal income.

In Chapter 4, I keep the focus on income, but expand the scope of analysis. I conduct a general inquiry into how hierarchy affects income distribution. I build a hierarchical model that extrapolates the available firm-level data to create a large-scale simulation of the hierarchical structure of the United States economy over the last two decades. After showing that this model does a reasonably good job of reproducing the features of US income distribution, I use the model for a wide variety of analysis. This leads to three major findings.

First, I find that hierarchy plays a dominant role in shaping the tail of US income distribution. This is important, because the power-law tail of income distribution is a celebrated empirical regularity that is usually explained in individualistic terms [33–49]. In contrast, I find that the power law scaling of top incomes is likely caused by hierarchical organization. The second major finding is that hierarchy can be used to relate personal and functional income distribution. Drawing on Nitzan and Bichler's 'capital as power' hypothesis [8], I propose that earning capitalist income is a function of hierarchical power. I find that CEO pay evidence is consistent with this hypothesis. Moreover, a model that generalizes CEO pay trends accurately reproduces the distribution of capitalist income in the United States. Lastly, I investigate if the recent explosion in US top income shares can be understood in terms of a hierarchical redistribution of income. I find that a model implementing this hypothesis accurately reproduces several

Economic Growth ⟵······· Synthesis? ·······⟶ Income Distribution

Evidence for a Power Theory of Personal Income Distribution

Energy and Institution Size

A Hierarchy Model of Income Distribution
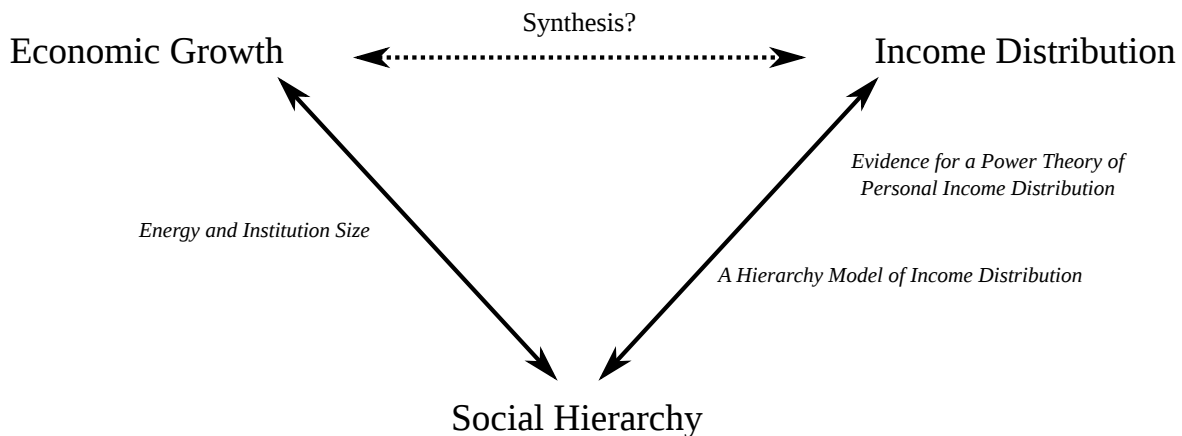
Social Hierarchy

Figure 1.1: A Glimpse of a Synthesis?

This figure shows how I conceive the big-picture structure of this dissertation. Each of the three papers (Ch. 2-4) connects either biophysical economic growth or income distribution to social hierarchy. But this connection begs a question: are growth and income distribution also related? I explore this possibility in Chapter 5.

―――――――――――――――――――――――――

key trends in US income distribution. To summarize, hierarchy seems to play a central role in shaping the size, composition, and dynamics of top incomes.

1.2   A Glimpse of a Synthesis?

I have given this dissertation the inquisitive (and not declarative) subtitle "Does Hierarchy Unify Economic Theory?". As I see it, each of the three papers in this dissertation connects either biophysical economic growth or income distribution to social hierarchy. This hints at a connection between growth and income distribution themselves (see Fig. 1.1). It is the possibility of unifying these two phenomena that informs the dissertation subtitle.

The reader may be asking — what is biophysical economic growth? In short, it is the growth of the economy measured in biophysical rather than monetary terms. As discussed in the 'Methods' section below, I treat energy consumption as biophysical indicator of economic scale. Why? Energy is the life-blood of all non-equilibrium systems. The rate of energy flow limits the types of structure that a given system can achieve. As such, when I connect the growth of energy consumption to social hierarchy (Ch. 2), I view this as an implicit connection between biophysical economic scale and social hierarchy.

In Chapter 5, I use the cumulative results in the dissertation to offer the glimpse of a synthesis between biophysical economic growth and income distribution. The basic thinking is as follows. If social hierarchy increases with energy consumption, and hierarchy is a mechanism for concentrating power, it follows that power should become more concentrated as energy consumption increases. Furthermore, if concentrations of hierarchical power lead to concentrations of income (as found in Ch. 3 and 4), the growth of energy consumption should be associated with an increase in income inequality.

To make this prediction concrete, I use the results in Chapters 2-4 to build a model of how hierarchical power concentration might increase with energy consumption. If this model is correct (and there are many caveats), it indicates something surprising. It suggests that a society's first order of magnitude increase in energy consumption — from subsistent metabolic levels to agrarian levels — should correspond with a massive increase in the concentration of hierarchical power. After this initial transition, the model suggests that further increases in energy consumption (to industrial levels) should have little effect on power concentrations. Given the connection between power inequalities and income inequalities, this suggests that the transition from hunter-gatherer societies to agrarian societies should be associated with a substantial increase in inequality. And counter-intuitively (to me at least), all subsequent changes to biophysical economic scale should have little effect on inequality. Interestingly, recent archaeological evidence suggests that this is what actually occurred [50]. Hunter-gatherer societies had very little inequality, but the transition to agriculture brought levels of inequality that were comparable to modern, industrial societies. My analysis suggests that this non-linear trend owes to the non-linear scaling behavior of hierarchy itself.

To summarize, using hierarchy as the unit of analysis seems to be a fruitful way to do economic research. Hierarchy, it would seem, lies at the very heart of human social organization, and is related to many of the outstanding questions in economics (and social science in general).

1.3   Methods

The methods used in this dissertation bear little resemblance to what most people would recognize as 'economics'. Because my methods are so different, I want to make their intellectual origins explicitly clear. My approach has four main components, outlined below.

A Biophysical Approach to Economics

Put succinctly, a biophysical approach to economics means taking the laws of thermodynamics seriously. These laws outline the basic rules of energy transformation: (1) energy can neither be created nor destroyed; and (2) all energy transformation processes *must* incur losses. It is hard to overstate the scientific importance of these laws. Indeed, the physicist Arthur Eddington once remarked "if your theory is found to be against the [laws] of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation" [51].

The laws of thermodynamics imply that, without flows of energy, all roads lead to equilibrium. And thermodynamic equilibrium is a *boring* state. Most of the interesting things that scientists study are out of equilibrium, and are sustained by a constant flow of energy [52, 53]. Life is perhaps the most compelling example. All life on earth is united by a common struggle — a "struggle for free energy available for work" [54]. The ability to harness energy places key constraints on the structure of life, from the level of the cell [55], to the organism [56, 57], to the ecosystem [58].

Probably the first economist to take the laws of thermodynamics seriously was Nicholas Georgescu-Roegen [59]. Since Georgescu-Roegen's work in the 1970s, there has been growing interest in reformulating economic theory to have a biophysical basis [60–64]. By far the most popular approach is to reform neoclassical growth theory by adding energy as a third factor of production, beside labor and capital. A non-exhaustive list of scholars who have pursued this approach would include [63, 65–74].

I take the biophysical approach seriously. But unlike many other economists, my goal is *not* to use energy consumption to explain the growth of real GDP. In fact, I am not interested in economic output at all. Basic measurement issues (outlined below) conspire to make the objective measure of economic output impossible. Instead, I am interested in energy consumption in its own right. Because of its importance for sustaining non-equilibrium structure, I use energy consumption as a *biophysical indicator of economic scale*. (For more details about this approach, see [75]).

Addressing The Measurement Problem

Most of mainstream economic theory is prefaced on the idea that economic output is objectively measurable. This is true of neoclassical marginal productivity theory, which explains income distribution in terms of the output of labor and capital [76–81]. It is also true of neoclassical economic growth theory,

which assumes that the economy has a measurable, aggregate output [82, 83]. The curious thing, however, is that these theories are all derived using the assumption of a one-commodity economy. For instance, Giorgio Colacchio observes that "the only case consistent with the marginal productivity theory is that of a 'one-commodity' economy" [84]. Similarly, in formulating his canonical growth model, Robert Solow assumes: "There is only one commodity, output as a whole" [82]. Why do these theories begin with such a bizarre assumption? It is because this is the only condition under-which the comparison and aggregation of different outputs is possible.

The central problem is this: if we want to add or compare two or more things that are qualitatively different, we need a common unit of measurement. However, for each different choice of unit, our comparison (or aggregation) will yield different results. Giampietro et al. call this the "epistemological predicament associated with purposive quantitative analysis ... the observer always affects what is observed when defining the descriptive domain" [85].

Economists make matters worse by choosing price as a unit of comparison. This does two things. Firstly, it makes marginal productivity theory circular. Why? Output is supposed to explain income, but by using prices to aggregate/compare output, we are actually measuring output *in terms of income*. Secondly, the fact that prices change over time causes a host of measurement problems. Francis Edgeworth observes:

> If one great group of commodities varies pretty uniformly in one direction, and another in a different direction (or even in the same direction but in a markedly different degree), then the task of restoring the level of prices can no longer be regarded as a *purely objective* ... problem. (cited in [86], emphasis added)

Over the years, many authors have commented on one or more aspects of this measurement problem (a non-exhaustive list would include [87–93]). But while many critical economists are aware of the problem, few are willing to take the logical course of action. If heterogeneous output cannot be objectively compared or aggregated, then there is no sense in trying to measure it. As a result, we need to build theory that does not rely on the concept of economic output.

As far as I know, Jonathan Nitzan and Shimshom Bichler [8] were the first to arrive at this conclusion. They propose an approach to political economy that focuses entirely on differential (price-ratio) quantities rather than on 'real' output. Inspired by Nitzan and Bichler, I have made the decision to abandon the measurement of economic output. Instead, I do one of two things. When I

A Random network   **B** Scale-free network   **C** Hierarchical network
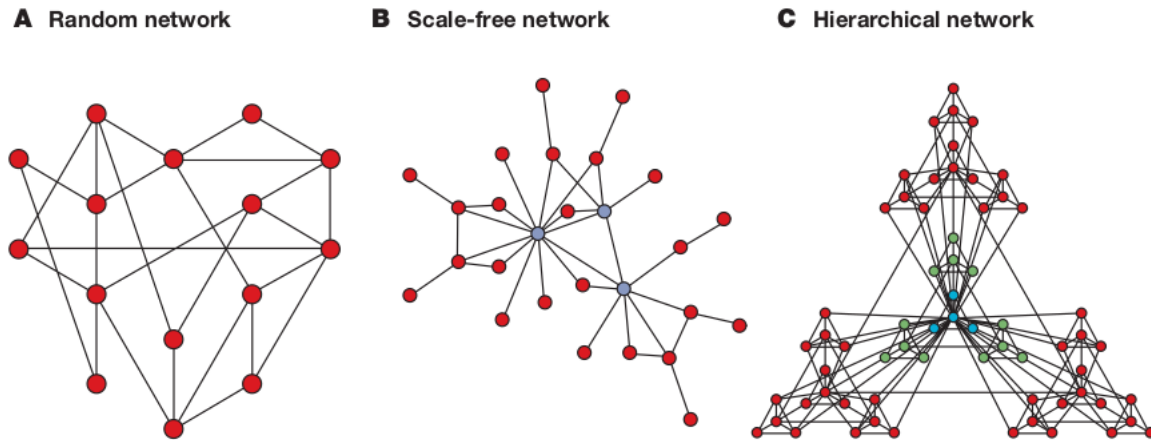
Figure 1.2: Different Forms of Networks

This figure (taken from Barabasi and Otvai [94]) shows three different types of networks. On the left is a random network, generated by adding edges between nodes at random. In the middle is a scale-free network. This name owes to the fact that there is no typical scale for the number of connections between nodes. Some nodes have many connections, some have very few. Lastly, the right panel shows a hierarchical network, which is characterized by a nesting structure.

---

want a measure of (biophysical) economic scale that is independent of monetary value, I use energy consumption per capita. Alternately, when I am interested in prices, I use differential ratios to allow comparisons.

Recognizing Ultra-sociality

There is a curious disconnect between how economists model humans, and how the more historical (and biological) oriented social sciences view our species. In economics, humans are treated as essentially asocial "globules of desire" [95]. Individuals exist purely to maximize their own utility. This asocial model is at odds with the rest of our scientific knowledge. Modern science recognizes that humans are but one form of primate, and all primates are social animals. Moreover, there is growing agreement that human sociality far surpasses our primate cousins. Rather than merely being social, humans are *ultra-social* [96–102]. This means that we form very large groups and are capable of cooperating with non-kin in ways that other primates cannot.

Taking ultra-sociality seriously means focusing on social connections between individuals. Network science offers a powerful way to do this [103]. We imagine individuals as 'nodes' in the network, and social relations as the 'edges'. My
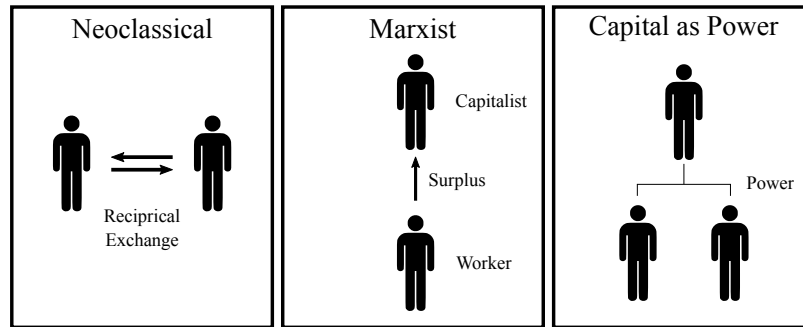
Figure 1.3: Idealized Units of Social Interaction

This figure shows my understanding of the basic units of social interaction adopted by Neoclassical and Marxist theory. Neoclassical theory is predicated on reciprocal exchange between utility maximizing parties. Marxist theory is predicated on the production of surplus by workers and its appropriation by capitalists. I propose that social (branching) hierarchy should be used as the unit of interaction for a capital as power approach to political economy. The premise is that a superior wields power over one or more subordinates.

---

focus on hierarchy is inspired by network science. As shown in Figure 1.2, a hierarchy is really just a particular type of network — one with a nested structure. A pure hierarchical network has a very important property. No matter where we begin, if we trace connections (going in only one direction) we will always end up in the same place [104]. To see how this works, think about a hierarchical chain of command. No matter which subordinate we begin on, if we move up the chain of command, we will always end at the same individual — the 'ruler'. A hierarchy is special type of network that concentrates power in the hands of the few. This property, I believe, is extremely important for understanding human social structure.

Capital as Power

It is hard to overstate the importance of Nitzan and Bichler's [8] 'capital as power' framework to my approach. To begin with, Nitzan and Bichler offer a compelling critique of both the neoclassical and Marxist approaches to political economy. The problem, they argue, is that the prerequisite units simply do not exist. Neoclassical theory is based on the concept of reciprocal exchange in which individuals maximize utility. But utility is unobservable, even in principle. Marxist theory, on the other hand, is based on the concept of surplus value. Workers create value, which is then appropriated by capitalists. But like utility, Nitzan and

Bichler convincingly argue that surplus value cannot (even in principle) be measured. Why? It is based on the non-existent unit of 'socially-necessary abstract labor time'.

Nitzan and Bichler argue that political economy needs a fresh start — a "ctrl-alt-del" [105]. I find this boldness liberating — it unburdens us of centuries of dead-end theoretical baggage. So what is the way forward? Nitzan and Bichler argue that it involves focusing on the relation between power and monetary value. I agree. I take this focus on power (and value) and merge it with a focus on hierarchy. My contribution to capital as power is to add an idealized unit of social interaction — the power-relation between a superior and subordinates within a hierarchy (see Fig. 1.3). Of course, this is not the only type of social relation that humans engage in; rather, it is one that has received too little attention from political economists.

1.4   Layout

This dissertation consists of the three self-contained papers:

1. Energy and Institution Size
2. Evidence for a Power Theory of Personal Income Distribution
3. A Hierarchy Model of Income Distribution

'Energy and Institution Size' has been published in PLOS ONE [106], and 'Evidence for a Power Theory of Personal Income Distribution' is currently under review at the Journal of Economic Issues.

A note to the reader. The writing of these three papers spans a significant period of time, while the overarching theme that unifies them has only recently become clear to me. As such, each paper makes little reference to others. I leave the discussion of connections for the conclusion in Chapter 5.

References

1. Brennan G, Tullock G. An economic theory of military tactics: Methodological individualism at war. Journal of Economic Behavior & Organization. 1982;3(2-3):225–242.

2. Leontief W. Theoretical assumptions and nonobserved facts. American Economic Review. 1971;61(1):1–7.

3. Wilson EO. Consilience: The unity of knowledge. New York: Random House; 1999.

4. Epstein JM, Axtell R. Growing artificial societies: social science from the bottom up. Washington, D.C.: Brookings Institution Press; 1996.

5. Hodgson G. Behind methodological individualism. Cambridge Journal of Economics. 1986;10(3):211–224.

6. Hodgson GM. Meanings of methodological individualism. Journal of Economic Methodology. 2007;14(2):211–226.

7. Brown C. Is there an institutional theory of distribution? Journal of Economic Issues. 2005;39(4):915–931.

8. Nitzan J, Bichler S. Capital as Power: A Study of Order and Creorder. New York: Routledge; 2009.

9. Peach JT. Distribution and economic progress. Journal of Economic Issues. 1987;21(4):1495–1529.

10. Price TD, Feinman GM. Foundations of social inequality. vol. 1. New York: Springer Science & Business Media; 1995.

11. Price TD, Feinman GM, editors. Pathways to power: New Perspectives on the Emergence of Social Inequality. New York: Springer; 2010.

12. Barroso FG, Alados CL, Boza J. Social hierarchy in the domestic goat: effect on food habits and production. Applied Animal Behaviour Science. 2000;69(1):35–53.

13. Guhl AM, Collias NE, Allee WC. Mating behavior and the social hierarchy in small flocks of white leghorns. Physiological Zoology. 1945;18(4):365–390.

14. Kondo S, Hurnik JF. Stabilization of social hierarchy in dairy cows. Applied Animal Behaviour Science. 1990;27(4):287–297.

15. Meese GB, Ewbank R. The establishment and nature of the dominance hierarchy in the domesticated pig. Animal Behaviour. 1973;21(2):326–334.

16. Sapolsky RM. The influence of social hierarchy on primate health. Science. 2005;308(5722):648–652.

17. Uhrich J. The social hierarchy in albino mice. Journal of Comparative Psychology. 1938;25(2):373.

18. Dahrendorf R. Class and class conflict in industrial society. Stanford: Stanford University Press; 1959.

19. Giddens A. The class structure of the advanced societies. Hutchinson London; 1973.

20. Marx K. Capital, Volume I. Harmondsworth: Penguin/New Left Review; 1867.

21. Reardon SF, Bischoff K. Income Inequality and Income Segregation. American Journal of Sociology. 2011;116(4):1092–1153. doi:10.1086/657114.

22. Robinson RV, Kelley J. Class as conceived by Marx and Dahrendorf: Effects on income inequality and politics in the United States and Great Britain. American Sociological Review. 1979; p. 38–58.

23. Weber M. Economy and society: An outline of interpretive sociology. Univ of California Press; 1978.

24. Wodtke GT. Social Class and Income Inequality in the United States: Ownership, Authority, and Personal Income Distribution from 1980 to 2010. American journal of sociology. 2016;121(5):1375–1415.

25. Wright EO. Approaches to class analysis. Cambridge University Press; 2005.

26. Wright EO. Class structure and income determination. vol. 2. New York: Academic Press; 1979.

27. Wright EO, Perrone L. Marxist class categories and income inequality. American Sociological Review. 1977;42(1):32–55.

28. Wright EO. Race, class, and income inequality. American Journal of Sociology. 1978;83(6):1368–1397.

29. Dunbar RI. Neocortex size as a constraint on group size in primates. Journal of Human Evolution. 1992;22(6):469–493.

30. Dunbar RI. Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences. 1993;16(04):681–694.

31. Dunbar RI. The social brain hypothesis and its implications for social evolution. Annals of human biology. 2009;36(5):562–572.

32. Turchin P, Gavrilets S. Evolution of complex hierarchical societies. Social Evolution and History. 2009;8(2):167–198.

33. Angle J. The surplus theory of social stratification and the size distribution of personal wealth. Social Forces. 1986;65(2):293–326.

34. Champernowne DG. A model of income distribution. The Economic Journal. 1953;63(250):318–351.

35. Chatterjee A, Chakrabarti BK, Chakraborti A. Econophysics and sociophysics: trends and perspectives. Milan: John Wiley & Sons; 2007.

36. Chatterjee A, Chakrabarti BK. Kinetic exchange models for income and wealth distributions. The European Physical Journal B. 2007;60(2):135–149.

37. Gabaix X, Lasry JM, Lions PL, Moll B. The dynamics of inequality. Econometrica. 2016;84(6):2071–2111.

38. Hegyi G, Neda Z, Santos MA. Wealth distribution and Pareto's law in the Hungarian medieval society. Physica A: Statistical Mechanics and its Applications. 2007;380(1):271–277.

39. Kitov IO. Mechanical model of personal income distribution. arXiv preprint arXiv:09030203. 2009;.

40. Nirei M, Souma W. A two factor model of income distribution dynamics. Review of Income and Wealth. 2007;53(3):440–459.

41. Nirei M. Pareto distributions in economic growth models. IIR Working Paper WP#09-05. 2009;.

42. Mandelbrot B. The Pareto-Levy law and the distribution of income. International Economic Review. 1960;1(2):79–106.

43. Pianegonda S, Iglesias JR, Abramson G, Vega JL. Wealth redistribution with conservative exchanges. Physica A: Statistical Mechanics and its Applications. 2003;322:667–675.

44. Silva AC, Yakovenko VM. Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983–2001. EPL (Europhysics Letters). 2004;69(2):304.

45. Simon HA. On a class of skew distribution functions. Biometrika. 1955;42(3/4):425–440.

46. Scheffer M, Bavel Bv, Leemput IAvd, Nes EHv. Inequality in nature and society. Proceedings of the National Academy of Sciences. 2017; p. 201706412. doi:10.1073/pnas.1706412114.

47. Rutherford RSG. Income distributions: a new model. Econometrica: Journal of the Econometric Society. 1955;23(3):277–294.

48. Yakovenko VM, Rosser Jr JB. Colloquium: Statistical mechanics of money, wealth, and income. Reviews of Modern Physics. 2009;81(4):1703.

49. Wold HO, Whittle P. A model explaining the Pareto distribution of wealth. Econometrica, Journal of the Econometric Society. 1957;25(4):591–595.

50. Kohler TA, Smith ME, Bogaard A, Feinman GM, Peterson CE, Betzenhauser A, et al. Greater post-Neolithic wealth disparities in Eurasia than in North America and Mesoamerica. Nature. 2017;551(7682):619–622. doi:10.1038/nature24646.

51. Eddington AS. The Nature of the Physical World. New York: The MacMillan Company; 1928.

52. Chaisson E. Cosmic Evolution: The Rise of Complexity in Nature. Cambridge, Mass.: Harvard University Press; 2002.

53. Kondepudi DK, Prigogine I. Modern Thermodynamics: from Heat Engines to Dissipative Structures. Chichester: John Wiley & Sons; 1998.

54. Boltzmann L. Populare Schriften. SEVERUS Verlag; 2011.

55. Lane N, Martin W. The energetics of genome complexity. Nature. 2010;467(7318):929–934.

56. Ahlborn BK. Thermodynamic Limits of Body Dimension of Warm Blooded Animals. Journal of Non-Equilibrium Thermodynamics. 2005;25(1):87–102. doi:10.1515/JNETDY.2000.006.

57. McNab BK. Resources and energetics determined dinosaur maximal size. Proceedings of the National Academy of Sciences. 2009;106(29):12184–12188. doi:10.1073/pnas.0904000106.

58. Odum HT. Systems Ecology; an introduction. New York: Wiley; 1983.

59. Georgescu-Roegen N. The entropy law and the economic process. Cambridge, MA: Harvard University Press; 1971.

60. Giampietro M, Mayumi K, Sorman A. The Metabolic Pattern of Societies: Where Economists Fall Short. New York: Routledge; 2012.

61. Victor P. Managing without growth : slower by design, not disaster. Northampton, Mass.: Edward Elgar; 2008.

62. Giampietro M, Mayumi K, Sorman A. Energy Analysis for a Sustainable Future: Multi-Scale Integrated Analysis of Societal and Ecosystem Metabolism. New York: Routledge; 2013.

63. Cleveland C, Costanza R, Hall C, Kaufmann R. Energy and the US economy: a biophysical perspective. Science. 1984;225(4665):890–897.

64. Cleveland C, Kaufmann RK, Stern DI. Aggregation and the role of energy in the economy. Ecological Economics. 2000;32(2):301–318.

65. Ayres R, Warr B. Accounting for growth: the role of physical work. Structural Change and Economic Dynamics. 2005;16(2):181–209.

66. Ayres RU, Warr B. The economic growth engine: how energy and work drive material prosperity. Cheltenham, UK: Edward Elgar Publishing; 2010.

67. Beaudreau BC. Energy and Organization: Growth and Distribution Re-examined. Westwood, CT: Greenwood Publishing Group; 1998.

68. Hall C, Lindenberger D, Kummel R, Kroeger T, Eichhorn W. The Need to Reintegrate the Natural Sciences with Economics. BioScience. 2001;51(8):663–673.

69. Hannon B, Joyce J. Energy and technical progress. Energy. 1981;6(2):187–195.

70. Kummel R. The impact of energy on industrial growth. Energy. 1982;7(2):189–203.

71. Kummel R. Energy as a factor of production and entropy as a pollution indicator in macroeconomic modelling. Ecological Economics. 1989;1(2):161–180.

72. Kummel R, Strassl W, Gossner A, Eichhorn W. Technical progress and energy dependent production functions. Journal of Economics. 1985;45(3):285–311.

73. Kummel R, Lindenberger D, Eichhorn W. The productive power of energy and economic evolution. Indian Journal of Applied Economics. 2000;8(2):1–26.

74. Kaufmann RK. A biophysical analysis of the energy/real GDP ratio: implications for substitution and technical change. Ecological Economics. 1992;6(1):35–56.

75. Fix B. Rethinking Economic Growth Theory from a Biophysical Perspective. Hall C, editor. New York: Springer; 2015.

76. Clark JB. The Distribution of Wealth. New York: Macmillan; 1899.

77. Jevons WS. The theory of political economy. London: Macmillan and co.; 1879.

78. Marshall A. Principles of Economics. London: Macmillan; 1890.

79. Menger C. Principles of Economics. New York: University Press; 1871.

80. Walras L. Elements d'economie politique pure, ou, Theorie de la richesse sociale. F. Rouge; 1896.

81. Wicksteed PH. An Essay on the Co-ordination of the Laws of Distribution (1932 Edition). London: London School of Economics; 1894.

82. Solow RM. A contribution to the theory of economic growth. The Quarterly Journal of Economics. 1956;70(1):65–94.

83. Swan TW. Economic growth and capital accumulation. Economic Record. 1956;32(2):334–361.

84. Colacchio G. Marginal Product of Labor; 2018. Available from: https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/labor-marginal-product.

85. Giampietro M, Allen TFH, Mayumi K. The epistemological predicament associated with purposive quantitative analysis. Ecological Complexity. 2006;3(4):307–327.

86. Vining DR, Elwertowski TC. The relationship between relative prices and the general price level. The American Economic Review. 1976;66(4):699–708.

87. Felipe J, Fisher FM. Aggregation in production functions: what applied economists should know. Metroeconomica. 2003;54(2–3):208–262.

88. Felipe J, Holz CA. Why do Aggregate Production Functions Work? Fisher's simulations, Shaikh's identity and some new results. International Review of Applied Economics. 2001;15(3):261–285.

89. Harcourt GC. On the Cambridge, England, Critique of the Marginal Productivity Theory of Distribution. Review of Radical Political Economics. 2015;47(2):243–255. doi:10.1177/0486613414557915.

90. Landefeld JS, Parker RP. BEA's chain indexes, time series, and measures of long-term economic growth. Survey of Current Business. 1997;77(5):58–68.

91. Nitzan J. Inflation as restructuring. A theoretical and empirical account of the US experience. McGill University; 1992.

92. Pullen J. The marginal productivity theory of distribution: a critical history. London: Routledge; 2009.

93. Robinson J. The production function and the theory of capital. The Review of Economic Studies. 1953;21(2):81–106.

94. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature reviews genetics. 2004;5(2):101.

95. Veblen T. Why is economics not an evolutionary science? The quarterly journal of economics. 1898;12(4):373–397.

96. Farley JC. The continuing evolution of ultrasocial economic organization. Behavioral and Brain Sciences. 2016;39.

97. Gowdy J, Krall L. The ultrasocial origin of the Anthropocene. Ecological Economics. 2013;95:137–147.

98. Pagel M. Evolution: adapted to culture. Nature. 2012;482(7385):297.

99. Richerson PJ, Boyd R. The evolution of human ultra-sociality. In: Indoctrinability, ideology, and warfare: Evolutionary perspectives. New York: Berghahn Books; 1998. p. 71–95.

100. Tomasello M. The ultra-social animal. European journal of social psychology. 2014;44(3):187–194.

101. Turchin P. The Puzzle of Human Ultrasociality: How Did Large-Scale Complex Societies Evolve? In: Cultural Evolution: Society, Technology, Language, and Religion. Cambridge, MA: MIT Press; 2013. p. 61–74.

102. Williams LA, Bliss-Moreau E. Humans are ultrasocial and emotional. Behavioral and Brain Sciences. 2016;39.

103. Barabási AL. Network science. Cambridge: Cambridge university press; 2016.

104. Mones E, Vicsek L, Vicsek T. Hierarchy measure for complex networks. PloS one. 2012;7(3):e33799.

105. Bichler S, Nitzan J. Capital as power: Toward a new cosmology of capitalism. Real-World Economics Review. 2012;(61):65–84.

106. Fix B. Energy and Institution Size. PLOS ONE. 2017;12(2):e0171823. doi:doi:10.1371/journal. pone.0171823.

Chapter 2

Energy and Institution Size

Abstract

Why do institutions grow? Despite nearly a century of scientific effort, there remains little consensus on this topic. This paper offers a new approach that focuses on energy consumption. A systematic relation exists between institution size and energy consumption per capita: as energy consumption increases, institutions become larger. I hypothesize that this relation results from the interplay between technological complexity and human biological limitations. I also show how a simple stochastic model can be used to link energy consumption with firm dynamics.

2.1 Introduction

Throughout the last century, there has been a recurrent desire to connect human social evolution to changes in energy consumption [1–4]. The motivation is simple: the laws of thermodynamics dictate that any system that exists far from equilibrium must be supported by a flow of energy [5]. Since human societies are non-equilibrium systems, it follows that energy flows ought play an important part in social evolution. However, it has proved difficult to move from grand pronouncements based on the laws of thermodynamics to a *quantitative* understanding of the relation between energy use and social evolution [6]. This paper offers a contribution to such a quantitative understanding.

This paper is concerned with one particular aspect of social change: the growth in size of the *institutions that control human labor*. While such institutions have taken many forms throughout history, in the modern era, the control of human labor is dominated by two institutions: the *business firm* and *government*. In this paper, institution *size* refers to the amount of human labor (i.e employment) controlled by an organization. Under this definition/metric of in-

stitution size, I demonstrate that a pervasive, positive correlation exists between institution size and energy use per capita.

I pursue two avenues for understanding the relation between energy and institution size. The first approach draws on the rich history of stochastic modelling within firm size theory. Stochastic (random) models have been successfully used to link firm *dynamics* to the overall firm size distribution. Yet there is little understanding of what drives variations in firm dynamics. Using data on firm age and firm size to constrain a stochastic model, I demonstrate that firm dynamics are likely related to rates of energy consumption, and I offer a prediction of what this relation should look like.

The second approach is more speculative, and aims to offer a general explanation of why rates of energy consumption are related to institution size. I propose two factors that mediate this relation: *technological scale* and *social hierarchy*. I hypothesize that increases in energy consumption involve a trend towards the use of technologies that are larger and more complex. These increasingly large technologies require the coordination of greater numbers of people. Given the limitations of the human brain [7], I argue that large-scale social coordination is most easily achieved through social *hierarchy* [8] and that firms and government are specific manifestations of this hierarchy.

This paper is organized as follows. After a brief review of the strengths and weaknesses of various theories of institutional size (Sec. 2.1.1), Section 2.2 discusses the empirical evidence connecting energy consumption with institution size. Section 2.3 then uses a stochastic model to further illuminate the relation between energy use and firm dynamics. Finally, Section 2.4 presents and tests a series of hypotheses linking institution size to technological scale and social hierarchy.

2.1.1   Theories of Institutional Size

Theories of institution size can be divided into two classes: those that concern themselves with the *causes* of institutional growth ('why' theories) and those that do not ('how' theories). 'How' theories have met with great empirical success, while 'why' theories have struggled to offer explanations that are testable.

All 'how' theories of institutional size can be traced back to the work of the French economist Robert Gibrat, who discovered that the rate of growth of business firms seemed to be *independent* of their size [9]. While later investigation found this 'law of proportional effect' to be only approximately true — growth rate variance tends to decline with size [10–12] — it has led to a rich history of

stochastic firm growth models [13, 14]. The basic principle is that firm growth is treated probabilistically. Each firm is submitted to a series of random shocks that make it grow (or shrink) over time. When applied to large numbers of firms, the result is a firm size *distribution*. The surprising finding is that these purely random models can very accurately predict the functional form of real-world firm size distributions (see Appendix A.6).

Despite their success, 'how' theories are not particularly satisfying because they do not explain *why* institutions grow. Unfortunately, theories that *do* attempt to explain the cause of institution growth often rely on unmeasurable variables, and as a result, are untestable.

The theory of the firm has been dominated by Ronald Coase's *transaction cost* approach. According to Coase, "... a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organizing in another firm" [15]. Unfortunately, transaction costs have been notoriously difficult to define (let alone measure), rendering Coasian theory untestable [16, 17]

Other theories propose that management talent is the driver of firm growth. For instance, Robert Lucas assumes that the firm size distribution results from "allocat(ing) productive factors over managers of different ability so as to maximize output" [18]. Yet Lucas concedes that the causal factor in this model — the talent of managers — is "probably unobservable". Despite this problem, Lucas's theory remains popular [19, 20].

Still other theories propose that firm growth is the result of a resource-driven competitive advantage [21, 22]. Unfortunately, this approach has struggled to stipulate exactly how a particular resource is transformed into a value-creating competitive advantage. Priem and Butler argue that the 'resource-based view' advances a theory of value that is tautological — resources create *value* because they are (among other things) *valuable* [23].

In terms of measurability, theories of government size have faired no better than theories of firm size. One approach is to apply the rational-choice model to the behavior of voters. Government size is treated as a reflection of the preferences of utility maximizing voters [24, 25]. However, without an objective measure of individuals' internal preferences, this theory is untestable.

Another approach is to assume that government bureaucracies (or government as a whole) are self-serving entities that attempt to maximize their budgets, but are restrained by voters and/or an institutional framework such as the constitution [26, 27]. While maximizing behavior is one of the fundamental postu-

lates of neoclassical economics, the hypothesis that humans maximize external pay-offs has been falsified [28].

The lack of measurable variables has consistently plagued 'why' theories of institution size. If a new theory is to be successful, it must demonstrate a connection between institution size and some universally measurable quantity. Energy consumption is just such a quantity.

2.2   Energy and Institution Size: Empirical Evidence

To study the relation between energy and institution size, I compare variations in energy use per capita to variations in the size of firms and government over both space and time. For firms, I investigate how changes in the base, tail and mean of the firm size distribution are related to changes in energy use per capita. I use self-employment data to investigate the base of the firm size distribution (relying on the assumption that self-employer firms are very small). To investigate the tail of the firm size distribution, I look at the employment share of the largest firms. To quantify the relative size of government, I measure the government share of total employment.

Comparison of these institution size metrics with energy use per capita are shown in Figures 2.1-2.3. Figure 2.1 shows international trends (each colored line represents the path through time of a specific country), while Figure 2.2 shows time-series data for United States. Figure 2.3 (which focuses only on firms) merges data from Figures 2.1-2.2 and adds US *sectoral* and *subsectoral* level data. Although this synthesis merges data that are not identically defined (see Fig. 2.3 caption), the result is clear: the inclusion of sectoral data serves to extend (by two orders of magnitude) the trends found at the national level. In the case of small firms and mean firm size, the inclusion of sectoral data also increases the regression strength.

To summarize our findings, the evidence in Figures 2.1-2.3 suggests the following 'stylized' facts. As energy use per capita increases:

1. The small firm employment share *declines*;
2. The large firm employment share *increases*;
3. The mean firm size *increases*;
4. The government employment share *increases*.

Findings 1-3 suggest that increases in energy consumption are associated with a shift in employment from small to large firms. This indicates that the firm
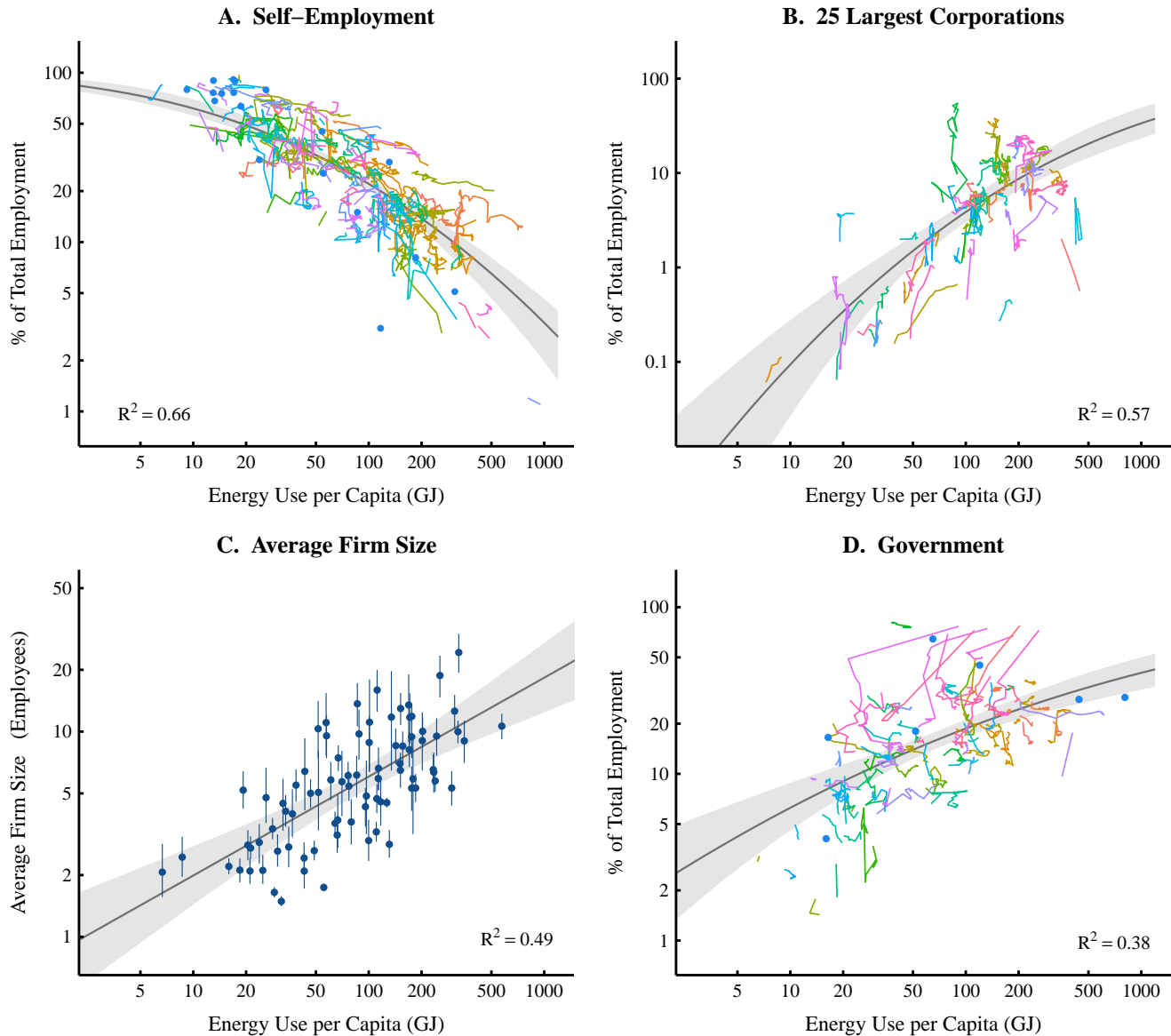
Figure 2.1: Institution Size vs. Energy Use per Capita at the International Level

This figure shows how different metrics of institution size vary with energy consumption per capita. Panels A-C analyze variations in firm size by looking at the base, tail, and estimated mean of the firm size distribution. Panel D analyzes variations in government size. In order to show as much evidence as possible, panels A, B and D are a mix of time series and scatter plot. Lines represent the path through time of individual countries while points represent a country with a single observation. Error bars in panel C represent the 95% confidence interval of mean firm size estimates. Variations in self-employment, large-firm, and government employment share vs. energy are modelled with log-normal cumulative distribution functions. Mean firm size vs. energy is modelled with a power law. Grey regions indicate the 99% confidence region of each model. For sources and methodology, see Appendix A.1.
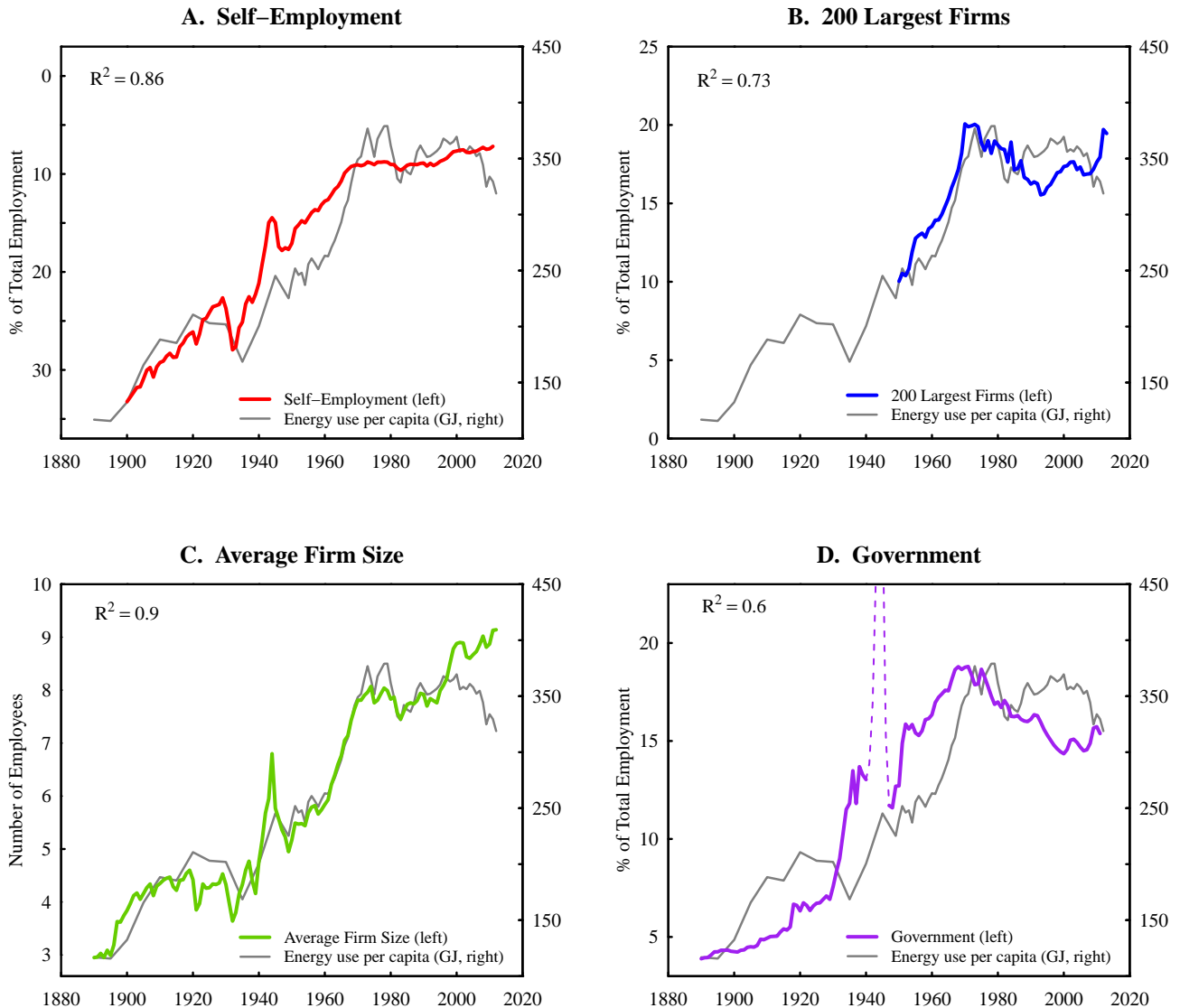
Figure 2.2: Institution Size vs. Energy Use per Capita in the United States

This figure shows the trends for various measures of institution size in the United States over the last century. Trends mirror those found at the global level. As energy consumption per capita increases, self-employment rates decline (panel A, note reverse scale), the large firm employment share increases (panel B), mean firm size increases (panel C), and the government employment share increases (panel D). Note that government regressions exclude World War II (dotted line). For sources and methodology, see Appendix A.1.

A. Small Firms and Self–Employment

B. Large Firms

C. Average Firm Size
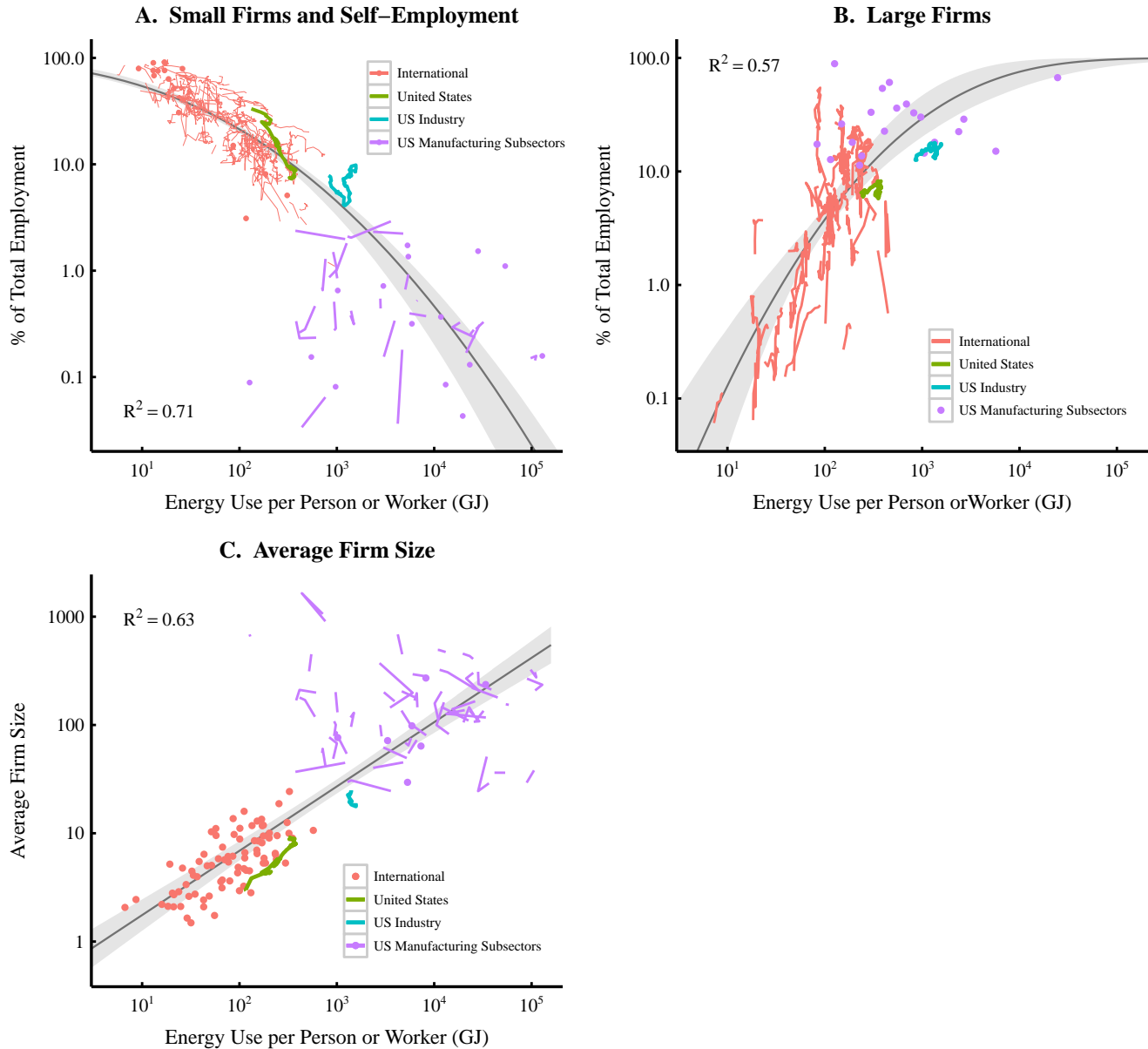
Figure 2.3: Synthesizing Evidence — Firm Size vs. Energy Use per Person or Worker

This figure combines data from 3 different units of analysis (nations, sectors, and subsectors) to offer a comprehensive picture of the relation between firm size and energy use per capita (or per worker). 'US Industry' consists of construction and manufacturing sectors, while 'US Manufacturing Subsectors' are the smallest subdivisions of the manufacturing sector. At the national level, energy use is measured per *person*, while at the sectoral level, it is measured per *worker*. In panel A, self-employment data (for nations and US Industry) is merged with the data for the employment share of firms with 0-4 employees in US manufacturing subsectors. In panel B, data for the employment share of the largest 25 firms (for nations and US Industry) is merged with data for the employment share of firms with more than 5000 employees in US manufacturing subsectors. Panel C shows mean firm size data at the national and sectoral level. Grey regions indicate the 99% confidence region of each model. For sources and methodology, see Appendix A.1.

size distribution becomes more *skewed* as energy consumption increases. In Appendix A.3, I demonstrate that this shift (at the national level) can be accurately modelled in terms of the changing exponent of a power law distribution.

Assuming a correlation between energy use and GDP, then the evidence presented here is consistent with previous research that has focused on the relation between firm size and GDP per capita [18, 20, 29–31]. However, my focus here on energy use (rather than GDP) is intentional: it is part of a larger effort to ground economic theory in the laws of thermodynamics [32], and to root empirical analysis in biophysical (rather than monetary) phenomena [33–36].

Following the long-standing division in institution size theory between 'how' and 'why' theories, I adopt two separate approaches for understanding the relation between institution size and energy consumption. The first approach deals with the 'how' question: *how* exactly do changes in firm size occur? To answer this question, I use a stochastic model to illuminate the relation between energy use and firm dynamics. The second approach deals with the more difficult 'why' question: *why* is institution size related to energy consumption. To answer this question, I investigate the relation between energy, technological change, and social coordination.

2.3   The 'How' Question: Energy and Firm Dynamics

Beginning with the work of Gibrat [9] and later Simon and Bonini [37], stochastic models have been successfully used to explain the functional form of the firm size distribution in terms of firm *dynamics*. The implication of these models is that changes in average firm size occur through changes in firm dynamics. Given the connection between energy consumption and firm size, it follows that firm dynamics ought to vary with changes in energy consumption.

Ideally, we would look at this relation directly by investigating international variations in the firm growth rate distribution and comparing them to variations in energy consumption. Unfortunately, data constraints make such a comparison difficult. Calculating international firm growth rate distributions would require longitudinal data for a large, representative sample of firms in many countries. I am not aware of the existence of any such data at the present time. However, we can use what little data *is* available to make inferences about the relation between energy and firm dynamics.

Firm age data provides an indirect window into firm dynamics. If we assume that new firms start at a small size, then we can infer the historic rate of growth of any firm, given its current age and size (i.e. a new, large firm likely grew

rapidly, while an old, small firm likely grew slowly). Figure 2.4A shows how firm age is related to rates of energy consumption per capita. The dataset used here (the GEM database) does not report firm age directly. Instead, it reports whether or not a firm is under 42 months of age. I use this data in Figure 2.4A to calculate the fraction of firms that are under 42 months of age. This fraction tends to *decline* as energy use per capita increases.

This data clearly hints that a systemic relation exists between energy consumption and firm dynamics. In the following section, I use a stochastic model to make specific predictions about the form of this relation.

2.3.1   A Stochastic Model

The essence of all stochastic firm models is that growth is treated probabilistically. Each firm begins with some arbitrary initial size L_0. After every discrete time interval, the firm is subjected to a series of random 'shocks' (x_i) that perturb it from its initial size. In our model, these shocks are drawn randomly from a Laplace distribution. At any point in time, each firm's size $L(t)$ is equal to the initial size times the product of all shocks (Eq. 2.1). If the time interval is years, then each shock can be interpreted as the annual growth rate (in fractional form).

$$L(t) = L_0 \cdot x_1 \cdot x_2 \cdot ... \cdot x_t \tag{2.1}$$

This basic Gibrat model is unstable unless additional stipulations are added (see Appendix A.5). I add a reflective lower bound that disallows firms from shrinking below the size $L = 1$ (this is sometimes called the Keston process [38–40]). As long as firm growth rates have a *downward* drift, the model will produce a stable firm size distribution. Using this model requires the following assumptions:

1. The firm size distribution is a power law.
2. Firm growth rates are independent of size.
3. New firms are all born at size $L = 1$.
4. The firm birth rate is equal to the firm death rate.
5. Firm growth rates come from a *Laplace* distribution.
6. The firm size distribution exists in an equilibrium.

Assumption 1 is necessary because the model produces a power law distribution (see Appendix A.6). Recent studies have found that firm size distributions
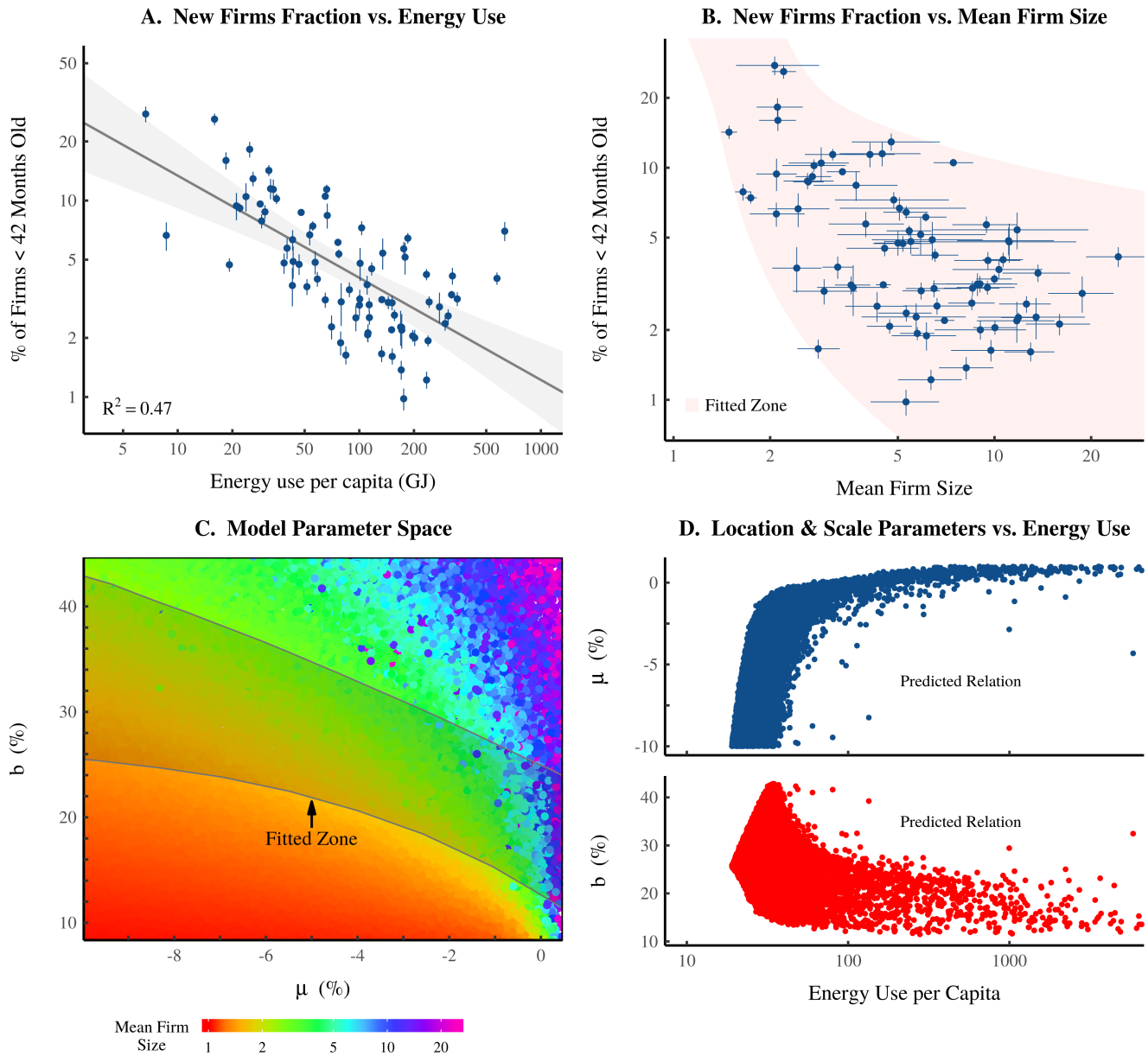
Figure 2.4: Using Firm Age Data to Estimate International Firm Dynamics

This figure demonstrates how firm age and mean size data can be used to restrict the parameter space of a stochastic model. This allows predictions to be made about the relation between energy use and firm dynamics. Panel A shows the country-level relation between the fraction of firms under 42 months old vs. energy use per capita (the grey region indicates the 99% confidence region of the regression). Panel B shows the country-level relation between the fraction of firms under 42 months old and mean firm size (error bars indicate 95% confidence intervals). The 'Fitted Zone' in Panel B shows the age-size relation produced by a stochastic model with a parameter range specifically chosen to capture the empirical data. Panel C shows the model's parameter space with the resulting mean firm size indicated by color. Using the regressed relation between mean firm size and energy use per capita (Fig. 2.1C), modelled mean firm size is then transformed into an estimate for energy use per capita. The resulting relation between μ and b vs. energy use per capita (for data in the fitted zone only) is plotted in panel D.

in the United States [41] and other G7 countries [42] are approximately power laws. Less is known about developing countries. In Appendix A.3, I demonstrate that the international data shown in Figure 2.1 is largely consistent with variations in a power law distribution, as are variations in the US firm size distribution over the last century.

Assumption 2 is a property of most stochastic firm growth models, and dates back to the work of Gibrat [9], who first found evidence that firm growth rates were independent of size. Since then, some studies have found that growth rate volatility tends to decline as firm size increases [10–12]). For the purposes of this model, I neglect this real-world complexity for the following reasons. First, firm growth rate studies use datasets (like Compustat) that are extremely biased towards large firms. Very little is known about the growth rates of small firms. In Appendix A.4, I use the Compustat database (which is very biased towards large firms) to estimate how growth rates might vary with size in a non-biased sample. I find that declines in growth rate volatility are likely important for only a small minority of the largest firms. Furthermore, it is quite possible that the rate at which volatility declines with firm size varies by country and/or through time. However, good data (on which to base a model) is unavailable. Faced with this lack of knowledge, I choose to make the simplifying assumption that firm growth rates do not vary with size.

Assumptions 3 and 4 give meaning to the reflective lower bound. We can interpret this boundary as a firm birth/death zone. Any firm that passes below $L = 1$ is assumed to have 'died'. The reflection then represents the 'birth' of a new firm of size $L = 1$. Since all firms that 'die' are immediately 'reborn', this mechanism assumes that the firm birth rate equals the firm death rate. This interpretation of the model allows firm age to be defined as the period since the last reflection. In the real world, new firms are obviously not all born at size one; however, evidence suggests that they are much smaller than established firms [43, 44].

Regarding assumption 5, it is well established that the firm growth distribution has a tent-shape that can be modelled with the Laplace distribution [45, 46]. A Laplace (or double exponential distribution) has a sharper peak and fatter tails than a normal distribution. Various theories have been proposed to explain this phenomenon [47, 48]; however the causes of this growth rate distribution are exogenous to the current model.

Assumption 6 justifies testing the model against empirical data. Given some arbitrary initial conditions, the model will always approach a stable firm size distribution that is a function of only the growth rate distribution (provided

that the stability conditions are met). Prior to arriving at equilibrium, there is *no relation* between the growth rate distribution and the firm size distribution (since any initial condition is possible). The equilibrium assumption justifies the link between growth rates and the firm size distribution.

2.3.2   Estimating Variations in Firm Dynamics

The goal of this analysis is to estimate how firm dynamics (i.e. growth rate distributions) change with levels of energy consumption per capita. This estimation involves three steps. First, we must use appropriate empirical data to restrict the parameter space of the model. Second, we analyze how this parameter space relates to mean firm size. Finally, we extrapolate, from mean firm size, the relation between model parameters and energy use per capita.

Modelled growth rates are determined by the Laplace probability density function below, where μ and b are the location and scale parameters, respectively.

$$p(x) = \frac{1}{2b}e^{-|(x-\mu)/b|} \tag{2.2}$$

The parameter μ indicates the most probable growth rate, while b corresponds to growth rate volatility (larger b indicates greater volatility). Because μ and b are *free* parameters, we must use appropriate empirical data to restrict their range.

To do this, I use the empirical relation between the proportion of firms under 42 months of age and mean firm size (Fig. 2.4B). A range of model parameters is chosen so that the resulting stochastic model produces the 'fitted zone' in Figure 2.4B. The corresponding parameter space of the model is shown in Figure 2.4C, with fitted zone parameters indicated by the shaded region. Equilibrium mean firm size for each μ and b coordinate is indicated by color.

The final step in the analysis is to use the regressed relation between mean firm size and energy use per capita (Fig. 2.1C) to estimate energy consumption levels from modelled mean firm sizes (for data within the fitted zone only). We can then plot the resulting predicted relation between model parameters and energy use per capita (Fig. 2.4D).

Our restricted stochastic model predicts the following: (1) μ should *increase* non-linearly with energy consumption; and (2) b should *decrease* non-linearly with energy consumption. In general terms, the model predicts that average firm growth rates should increase with energy consumption, while volatility should

decline. This result represents a definitive prediction about how firm dynamics should vary with rates of energy consumption. Future empirical work can determine if this prediction is correct.

2.4   The 'Why' Question: Energy, Technology and Hierarchy

Any attempt to explain why institutions grow must first settle on the appropriate scale: do we attempt to explain why *individual* institutions grow, or do we concern ourselves only with changes in *average* size? The former is almost certainly a futile task, much like offering a general theory to explain why individual species go extinct. The answer is almost certainly, "It is complex". Species go extinct because of the complicated relation between their physiological characteristics and their environment. Likewise, individual institutions grow/shrink because of the complex relation between their characteristics and their environment (both biophysical and social).

The very success of stochastic firm growth models — in which *randomness* is the explanatory mechanism — suggests that the individual institution is not the appropriate domain for a 'why' explanation. Rather, we should be concerned with *groups* of institutions. This decision effectively bars the traditional toolbox of economic theory, which is to construct models based on simple postulates about the behavior of individual entities (consumers, firms, governments, etc.). Instead, we must rely on qualitative reasoning, tested against quantitative empirical evidence.

My explanation of the energy versus institution size relation builds on the 'social brain' hypothesis proposed by Dunbar [49]. According to this hypothesis, the size of the human brain inherently limits our ability to maintain social relations. As Tuchin and Gavrilets note, social hierarchy offers a way around this limit [8]. Within a hierarchy, an individual must maintain relations with only his direct superior and direct subordinates. This means that a hierarchically organized group can grow in size *without* a corresponding increase in the number of required social relations. I argue that firms and governments are simply the modern embodiment of social hierarchy, and are used as tools of social coordination.

To connect social coordination to energy consumption, I explore the connection between energy use and technological scale. I argue that increases in energy consumption are associated with the use of increasingly large technologies. The construction, operation, and maintenance of these larger technologies, in turn, requires greater social coordination.

I formalize this reasoning in the joint hypotheses below. The order of these hypotheses is meant to show a line of reasoning, not necessarily a direction of causality.

Hypotheses

    A. Increases in per capita energy consumption are accomplished (in part) through increases in technological *scale*.

    B. Increases in technological scale require increases in *social coordination*.

    C. Humans have a *limited* capacity to maintain social relations. Hence, egalitarian social coordination has strict limits.

    D. *Social hierarchies* allow the scale of social coordination to grow without a corresponding increase in the number social relations.

    E. Institutions (firms and governments) are dedicated social hierarchies.

In the following sections, I review the empirical evidence in support of each of these hypotheses.

2.4.1 Energy, Technological Scale and Social Coordination

My focus on technology (hypothesis A) is motivated both by theoretical arguments and by the empirical results in Fig. 2.3.

From a theoretical (thermodynamic) perspective, energy 'consumption' is best thought of as a *conversion* process. For most organisms, this energy conversion process occurs *within* the body via cellular metabolism. Humans are unique among all other organisms in that we have developed many inorganic ways of harnessing energy *outside* our bodies. This inorganic energy consumption necessarily involves the use of man-made energy converters that transform primary energy into forms useful to humans. We call these man-made energy converters 'technology'. Since energy use is fundamentally related to technology, it makes sense to explore the ways in which technology relates to institution size.

On the empirical side, the fact that firm size scales with energy consumption both at the national and *sectoral* level (Fig. 2.3) hints that technology mediates this relation. Unlike nation-states, which are defined by geographic boundaries, economic *sectors* are defined by a particular type of *activity*. Similar activities tend to use similar technologies. This is especially true as we move to the

smallest manufacturing subsectors. With names like *Sawmills* (NAIC 321113), *Petroleum Refineries* (NAIC 32411), and *Iron Foundries* (NAIC 331511), these subsectors are practically *defined* by the technologies they use. This suggests that differences in energy use between such subsectors are related to differences in the technologies employed.

To illuminate the relation between energy and technology, consider the definitional statement that energy per capita (E_{pc}) is equal to total energy consumption (E) divided by population (P):

$$E_{pc} = \frac{E}{P} \tag{2.3}$$

Let us now define N as the total number of energy converters in society. By multiplying by N/N, we can rearrange equation 2.3 to give:

$$E_{pc} = \frac{E}{N} \cdot \frac{N}{P} \tag{2.4}$$

Equation 2.4 indicates that energy use per capita is a function both of technological *scale* (E/N, average capacity per energy converter) and technological *density* (N/P, the number energy converters per capita).

In terms of social coordination, there is a fundamental difference between increasing energy consumption through technological density versus technological scale: the former is a *decentralized* process, while the latter requires *centralization*. Increasing energy use per capita through technological density involves independent changes in the behaviour of individuals, meaning it is an atomistic process. However, increasing energy consumption through technological scale requires the centralization of resources and human labor. Thus, it requires increases in social *coordination*.

As an example of a technological *density* process, consider the spread of household appliances (which are a type of end-use energy converter). The invention and widespread adoption of technologies such as the refrigerator, washer, dryer, microwave oven, and dishwasher vastly increased the number of energy converters per capita. At least on the consumer end (not the production end) this process was highly decentralized — individuals *independently* added more electronic devices to their lives.

As an example of a technological *scale* process, consider the changing scale of the industrial technologies shown in Table 2.1. Relative to their early prototypes, these technologies have undergone increases in scale by factors of one hundred (tanker ships) to factors of over a *million* (electric power plants). These

Table 2.1: Scale Increase of Various Industrial Technologies

Type	Early Prototype	Largest Today	Unit	Scaling Factor
Electric Power Plant	0.0125	2 2500	megawatts	1.80×10^6
Oil Refinery	5.5	1 240 000	barrels per day	2.24×10^5
Aluminium Smelter	5.7	1 060 000	tonnes per year	1.86×10^5
Internal Combustion Engine	0.75	107 390	horsepower	1.43×10^5
Mining Excavator	380	2 324 0000	cubic meters per day	6.12×10^4
Blast Furnace	0.3	5 500	cubic meters	1.83×10^4
Tanker Ship	1809	260 859	gross tonnage	1.44×10^2

This table shows the size of 7 selected industrial technologies at their earliest stage of development ('Early Prototype') and at the largest scale existing today. Column 5 shows the scaling factor between the largest and early technologies (largest/early). Technologies are ranked in descending order of scaling factor. For data sources, see Appendix A.1.

---

changes in technological scale necessarily involve the increasing coordination of human labor. For instance, the largest oil refinery in the world, located in Jamnagar, India, employs 2500 people on site [50]. Rather than acting autonomously (like the users of consumer electronics), these individuals must coordinate their actions over a wide range of different tasks. This suggests that increases in technological scale require an increase in social coordination.

But to what degree are increases in energy use per capita actually achieved through increases in technological scale? Given the complexity of technological change, this question is difficult to answer at a general level (for all technologies). Instead of a general test of hypothesis A, I present here a case study of electricity production and consumption in the United States (Fig. 2.5A-B). The results of this case study indicate that increases in technological scale have played an important role in meeting increases in per capita electricity use over the last century.

Figure 2.5A shows how the indexed change in US electricity use per capita relates to the indexed change in mean power plant size (as measured by nameplate capacity). Over the last 100 years, the two series tracked together quite closely, with both electricity use and power plant size increasing rapidly between 1920 and 1980 and plateauing thereafter. How important was this change in technological scale for meeting per capita demand? To answer this question, Figure 2.5B plots the indexed ratio of mean power plant size to electricity use per capita. This ratio indicates the fraction of electricity use per capita growth
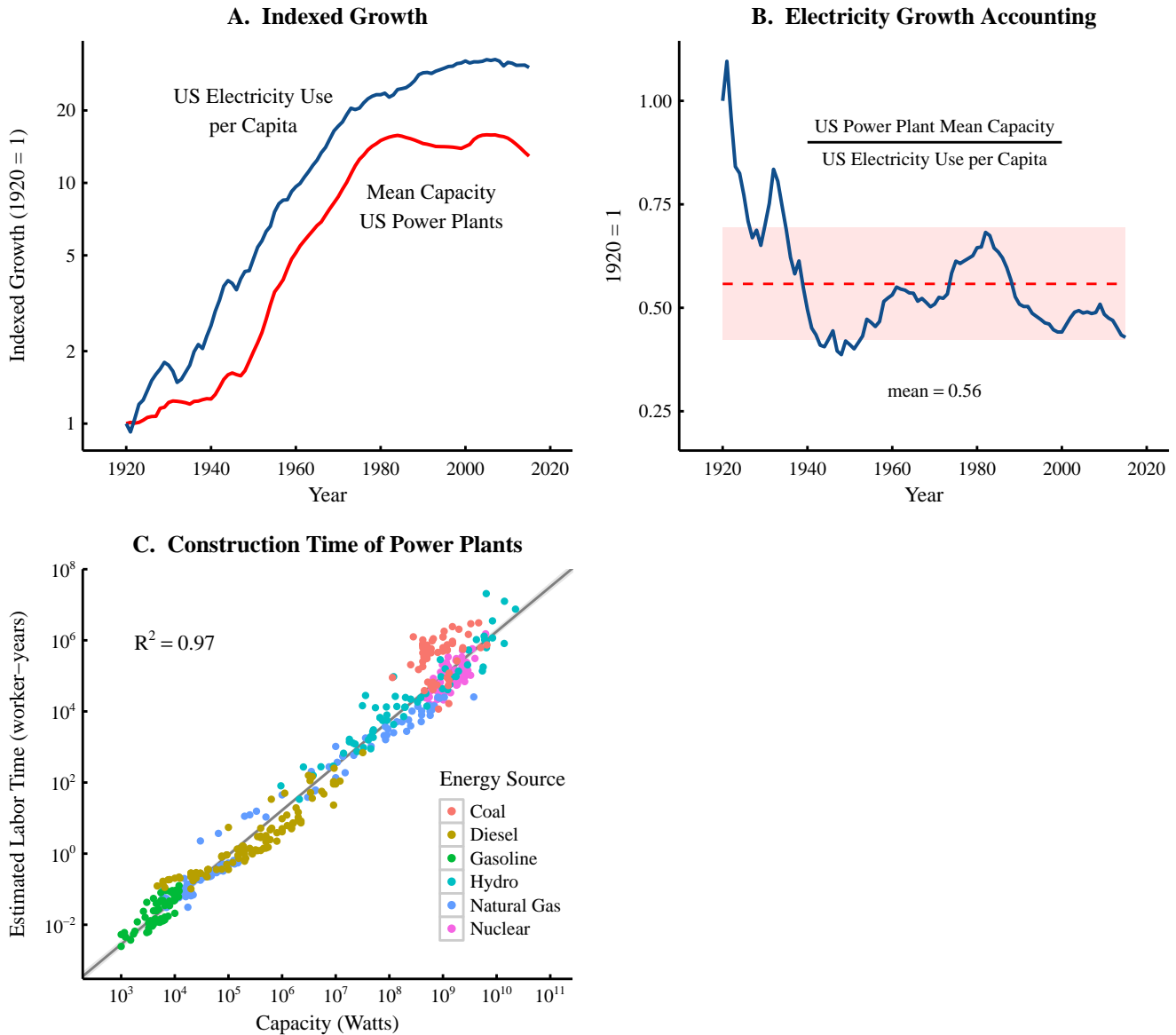
Figure 2.5: Technological Scale and Social Coordination in Electricity Generation

Panel A shows the time-series relation between the mean capacity of US power plants and US electricity use per capita. Both series are indexed to 1 in the year 1920 in order to show relative growth. Power plants tend to get larger as electricity use per capita increases increases. Panel B shows the fraction of US per capita electricity use growth (since 1920) that was met by increases in mean plant size. The dashed line indicates the mean over the period 1920-2015, while the shaded region shows the standard deviation. Panel C shows the relation between power plant capacity and the estimated construction labor time. The entire range of electricity generation technology is included in this plot — from the smallest gasoline generators to the largest hydroelectric power plants. Different primary energy sources are indicated by color. Data is modelled with a power law. Grey regions indicate the 99% confidence region of the regression. For sources and methodology, see Appendix A.1.

that was met by increases in power plant capacity. Between 1920 and 2015, increases in power plant capacity accounted for roughly *half* of the total increase in electricity use per capita.

In the US electricity generation sector, increases in technological scale obviously played a major role in meeting increases in per capita electricity consumption. Was this increase in scale accompanied by a corresponding increases in the scale of social coordination (hypothesis B)? Answering this questions requires that we first define what we mean by the 'scale' of social coordination, and specify how this relates to a given technology.

I define the 'scale' of social coordination as the number of people required to construct, maintain, and operate a specific technology. For measurement purposes, however, I limit my analysis only to *construction* labor time. This decision is driven primarily by data availability (and lack thereof). For the most part, published power plant data focuses almost exclusively on *costs*, and primarily on the cost of *construction*. Fortunately, with a few simplifying assumptions, construction cost data can be used to estimate construction labor time. I use this latter metric to quantify the scale of social coordination associated with a given power plant.

To estimate construction labor time from costs, I first note that by the rules of double-entry accounting, all costs eventually become someone's *income*. If we assume that all income accrues to labor (i.e. we neglect capitalist income) then we can divide the total cost of a project by an estimate of the average wage to obtain a rough estimate of the total labor time involved. I use GDP per capita as a measure of average income, giving equation 2.5 as my method for estimating labor time.

$$\text{Labor Time} \approx \frac{\text{Total Cost}}{\text{GDP per capita}} \tag{2.5}$$

Although this method contains some implicit bias/error, I show in Appendix A.7 that it is unlikely that this bias/error affects the integrity of the results (largely due to the vast size range of power plant studied here).

Figure 2.5C applies this method to estimate the construction labor time of approximately 500 different power plants and generators. The capacity of these plants/generators ranges over 7 orders of magnitude — from the smallest gas-powered generator (1000 watts) to the largest hydroelectric dams (the 22.5 gigawatt Three Gorges Dam). Different energy sources are indicated by color. The results show a strong scaling relation between plant capacity and construc-

tion labor time. This indicates that the scale of social coordination necessary to build a power plant is strongly related to the plant's energy conversion capacity.

To summarize, our case study of the electricity generation sector is consistent with both hypothesis A and B. We find that increases in power plant scale have played an important role in meeting increases in US per capita electricity consumption (hypothesis A). Furthermore, we find that power plant size is strongly related to construction labor time — our measure of the scale of social coordination (hypothesis B).

Admittedly, a case study of a single technology represents limited evidence. However, the vast scaling of the other technologies shown in Table 2.1 indicates that this line of reasoning has promise. To continue my arguments, I will assume that the findings of this case study can be generalized to many other technologies. The result (we assume) is a that increases in energy consumption require a generalized increase in the scale of human social coordination. The question, then, is how is this coordination accomplished?

2.4.2   Social Coordination and Human Biology

Social coordination can conceivably be achieved in many different ways (customs, markets, institutions, etc.). Thus, an increase in social coordination does not necessarily imply an increase in firm and government size. Why, then, have these institutions increased in size as energy consumption increases? Hypotheses C-E propose a chain of reasoning explaining why institutions are the most effective way of organizing large groups of people. The key to this reasoning is hypothesis C: humans have a *limited* ability to maintain social relations.

The evidence for this hypothesis comes primarily from the work of anthropologist Robin Dunbar, who has uncovered a startling relation between primate brain size and mean group size [7]: primate species with larger brains (as measured by the relative size of the neocortex) tend to live in larger groups. Dunbar has developed this finding into what he calls the *social brain hypothesis*: "primates evolved large brains to manage their unusually complex social system[s]" [49].

The implication of Dunbar's findings is that the size of the human brain places limitations on the number of social relations that an individual is able to maintain. Dunbar uses his primate data to predict a mean human group size of about 150. While this number should be considered exploratory, Dunbar notes that early egalitarian societies had group sizes around this order of magnitude [51].

A key feature of egalitarian organization is that any member of a group may

maintain relations with any other member of the group. Thus, the number of possible social relations increases linearly with group size. Given the hypothesized limitations in the human ability to maintain social relations, it follows that egalitarian social organization is not an effective method for coordinating large numbers of people.

One way of increasing group size beyond Dunbar's number is to organize groups in a way that *limits* human interaction. Turchin and Gavrilets note that this is a key feature of social hierarchies, which are characterized by a treelike chain of command [8]. Within a hierarchy an individual must maintain social relations only with his direct superior and direct inferiors. Thus, hierarchy allows group size to grow without any corresponding increase in the number of human relations (hypothesis D).

As evidence for this line of reasoning, Turchin and Gavrilets demonstrate that a strong correlation exists between the population of historical agrarian empires and the number of administrative (hierarchical) levels within their respective governments. Similarly, Hamilton et al. find a strong relation between population size and the number of hierarchical levels with various hunter-gatherer societies [52]. This evidence suggests that social hierarchy is a common tool used for increasing the scale of social coordination.

2.4.3   Hierarchy and Institution Size

Social hierarchies have taken many different forms at different points in human history. For instance, in many pre-state societies, social hierarchy took the form of the chiefdom. In middle-ages Europe, the feudal manor was the principle unit of hierarchy. In the modern era, I argue that business firms and governments are the principle unit of social hierarchy (hypothesis E). To test this hypothesis, I focus only on firms.

The implication of hypothesis E is that increasing firm size constitutes an *investment in social hierarchy*. If this reasoning is correct, then mean firm size should be an indicator of the relative 'top heaviness' of a society. Why? Hierarchies tend to become more top heavy as they become larger — the fraction of individuals in the upper echelons tends to grow as the size of the hierarchy increases. Thus, if firms are the modern embodiment of social hierarchy, then mean firm size should be related to the relative size of the upper social echelon.

Since the upper echelons of a hierarchy are almost exclusively involved in managing the activities of other people, it seems sensible to use the *management* profession as a metric for the size of this top cohort. Thus, if hypothesis E is
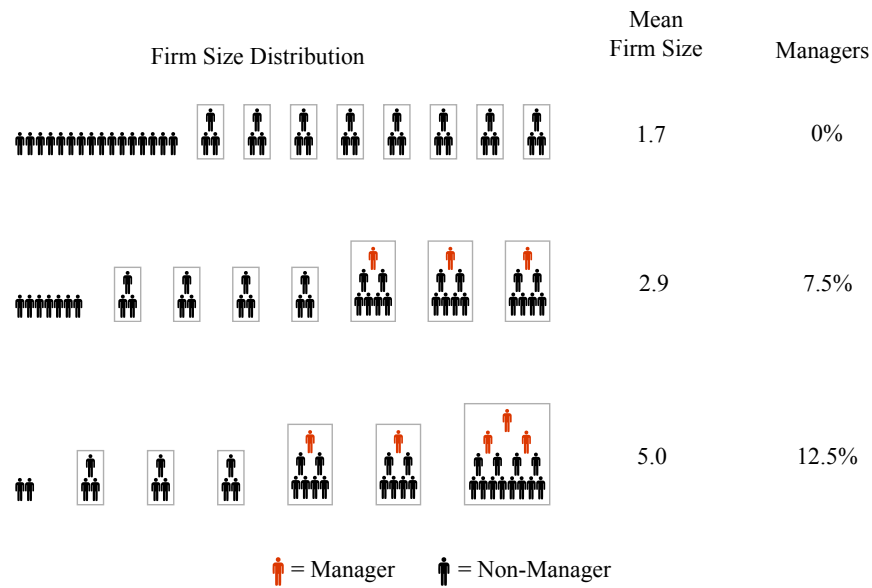
Figure 2.6: The Growth of Management as a Function of the Firm Size Distribution

This figure graphically demonstrates how the management fraction increases with firm size (assuming firms are 'ideal hierarchies'). Firms are indicated by boxes (with the exception of single-person firms) with a worker's hierarchical position shown vertically. The span of control — defined as the size ratio between adjacent hierarchical levels — is constant for all firms. In this picture, the span of control is 2. Managers (red) are assumed to be all individuals in and above the third hierarchical level. To maintain simplicity, this graphic does not use a power law firm size distribution.

---

correct, we expect that increases in mean firm size should be associated with an increase in the employment share of managers.

To refine this prediction, I develop a hierarchical firm model of society (Fig. 2.6) based on the following assumptions:

1. All firms are 'ideal' hierarchies with a single span of control.
2. All individuals in and above the third hierarchical level are considered 'managers'.
3. The firm size distribution is a power law.

Why assume that management begins at the third hierarchical level? Obviously, individuals within the lowest hierarchical level have no management responsibilities. Those in the second hierarchical level can be thought of as
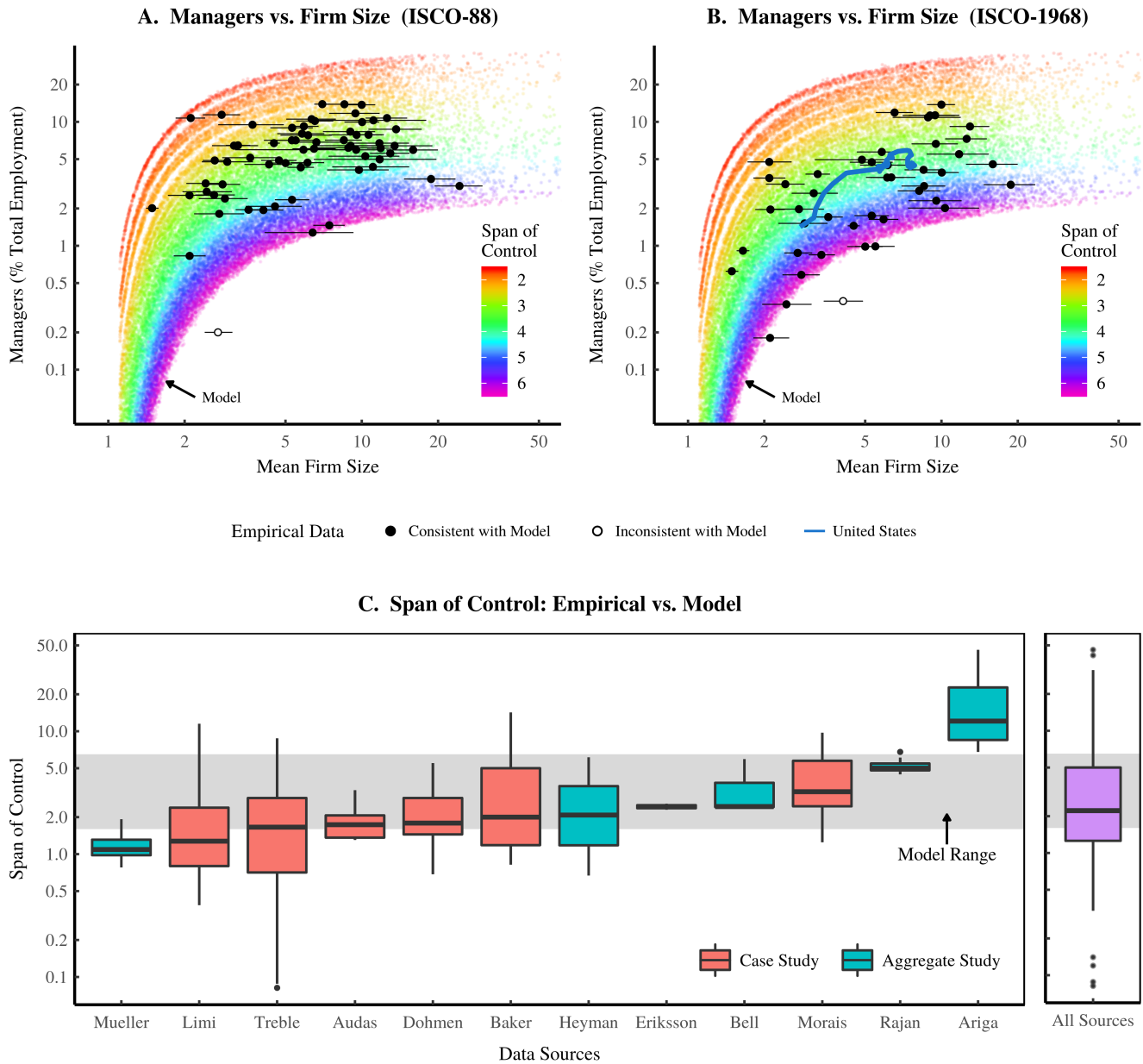
Figure 2.7: Testing the Hierarchical Model of the Firm Using Managment Share of Total Employment

Panels A and B plot the country-level relation between the management fraction and mean firm size. Modelled data is also shown in the background, with the span of control indicated by color. Panels A and B use different (incommensurable) classification methodologies for 'management'. Panel A uses ISCO-88 (which includes legislators, senior officials and managers) while panel B uses ISCO-1968 (which includes administrative and managerial workers). Error bars indicate the 95% confidence intervals for mean firm size. Panel C compares the span of control range from the model to the span distribution found by 12 different empirical studies. Red boxplots indicate case studies, and show the span of control distribution within a *single firm*. Blue boxplots indicate aggregate studies and show the span of control distribution across many different firms. The span of control distribution across all 12 studies is shown on the right. For sources and methodology, see Appendix A.1.

'working supervisors' — individuals who have some supervisory responsibilities but who spend a majority of their time engaged in 'production' [53]. I assume that individuals in and above the third hierarchical level are devoted mostly to managing the work of others.

This model predicts that the management fraction of employment should grow non-linearly with firm size, eventually approaching an asymptote defined only by the span of control. If the span of control is s, then the asymptote occurs at $1/s^2$ (see Appendix A.8 for the details of this calculation).

In Figure 2.7 I test this model at the international level. Figure 2.7A and 2.7B plot the country-level relation between the management fraction of employment versus mean firm size (the two plots show different occupation classification regimes). Empirical data is shown in black, while model predictions are shown in the background with the span of control indicated by color. Different mean firm sizes are produced by varying the exponent of the firm size power law distribution (for a technical discussion of this model, see Appendix A.8).

The model nicely reproduces the observed relation between mean firm size and the management fraction of employment. However, this fit is achieved by freely manipulating the span of control parameter. Thus, it is important to check that the modelled span of control range is consistent with the span range for *real* firms.

Ideally we would be able compare the span range of the model to the span distribution of a large, global sample of firms. Unfortunately, data constraints make this impossible. Due to the proprietary nature of firm personnel data, only a handful of studies have analyzed firm hierarchies. Figure 2.7C shows data from 12 such studies that together sample firms from 7 different nations (Denmark, Japan, Netherlands, Portugal, the United Kingdom, the United States, and Sweden). The resulting firm sample gives relatively good coverage of wealthy nations, but unfortunately does not include any firms from developing countries (due to the lack of available studies). For a summary of the data sources, see Appendix A.1.

Boxplots in Figure 2.7C correspond to the span of control range found by each study. Note that the data is a mixture of case studies of single firms and aggregate studies that analyze the structure of many different firms. While these aggregate studies give better scope than the case studies, many focus only on the upper levels of the hierarchy (where data is more easily obtained). The important finding in Fig. 2.7C is that the model's fitted span of control range is consistent with the available empirical data.

To summarize these findings, a simple hierarchical firm model of society is

able to replicate the observed relation between mean firm size and the management share of employment. The changes in mean firm size are achieved by varying the exponent of a firm size power law distribution, while the management fraction of employment is fitted by 'tuning' the span of control range (assumed to be the same both within and between all modelled firms). Importantly, the resulting fitted span range is consistent with the existing empirical data on the internal structure of the firm. The success of this model gives support to hypothesis E, and suggests that increases in mean firm size are characteristic of a generalized increase in social hierarchy.

2.4.4 Causality

I have proposed hypotheses A-E as a chain of reasoning connecting energy consumption to institution size. But which way does causation run? Do increases in energy consumption *cause* institutions to become larger, or is the reverse true? As I discuss below, it seems likely that causation runs in *both* directions.

Although hypotheses A-E are framed in terms of *increases* in energy use (and institution size), I think that a discussion of causation is clearer when framed in terms of *constraints* and *decline*. For instance, I think it *must* be the case that energy constraints place limits on institution size. This is for the simple reason that energy conversion technology is *useless* without an energy input. I have proposed that large institutions provide the social coordination necessary to build and operate large technologies. But without sufficient energy input, these technologies cannot be operated, and the institution's raison d'être ceases to exist. Imagine how long a large steel firm would stay in business if there was not enough coke to fuel its large blast furnaces. This line of thinking implies that a decline in energy consumption (due to scarcity) can *cause* a decline in institution size.

However, recent history (the collapse of the Soviet Union) suggests that causality can operate in the *reverse* direction. Figure 2.8 shows energy and government employment share trends in six nation-states that emerged after the dissolution of the USSR. In the aftermath of the Soviet collapse, these six countries experienced *drastic* reductions in both government size and energy use. During this period, there was no global energy shortage, meaning biophysical energy constraints can likely be ruled out as a causal factor. Instead, it seems likely that institutional collapse is the driving factor here.

This case is illustrative because the Soviet economy relied on an unusually high degree of government control of production, placing an enormous amount
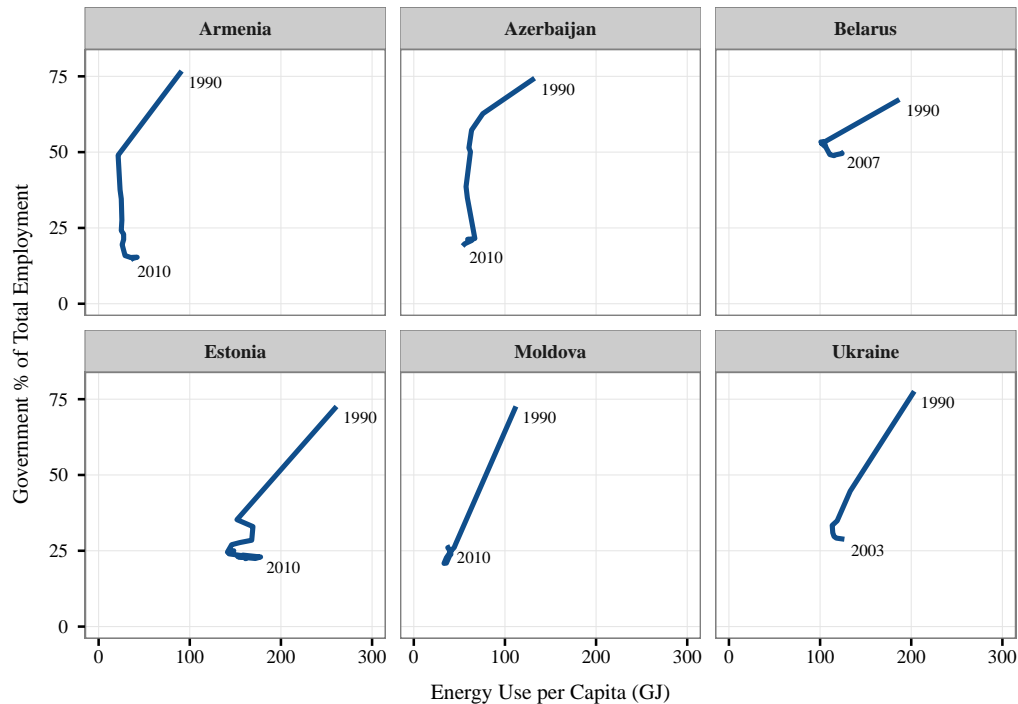
Figure 2.8: A Case Study in Causality: The Collapse of the Soviet Union

This figure tracks the path through time of six nations that emerged after the collapse of the Soviet Union (in 1990-91). As the collapse unfolded, the fraction of people employed by the government shrank rapidly, as did energy use per capita. Since the USSR collapse was an institutional crisis (not an energy crisis), this suggests that at least in this case, causality runs from institution size to energy consumption.

---

of power in the hands of a single institution. Not surprisingly, the collapse of this institution led to social chaos and widespread economic decline. I think this shows quite clearly that institutional collapse can cause a decline in energy consumption.

The argument that causation can operate in *both* directions suggests that energy use and institution size exhibit a *feedback* relation (rather than linear causality). One possible avenue for furthering this research is to use systems modelling. Ugo Bardi has shown that a simple adaptation of the Lotka–Volterra equations can be used to model the relation between energy extraction and a technological stock [54]. A plausible line of future research would be to add institution size to this type of model.

It is also important to note that changes in energy use and institution size occur alongside other social changes, the two most obvious being urbanization and changes in sector composition [35]. It seems likely that these phenomena

are all interrelated — part of a complex process of social change accompanying changes in energy consumption. In Appendix A.9, I use an adaptation of the hierarchical firm model (used in Fig. 2.7) to explore the institution size constraints that are inherent in the sectoral composition of agrarian societies. The results offer a promising way of broadening our understanding of why energy use is related to institution size.

2.5   Conclusions

All life on earth is united by a common struggle — a "struggle for free energy available for work" [55]. The ability to harness energy places key constraints on the structure of life, from the level of the cell [56], to the organism [57, 58], to the ecosystem [59]. Within this unifying context, it seems plausible that the structure of human society ought to be related to the ability to harness energy.

Based on this line of reasoning, a branch of scholarship has emerged that studies the role of energy in human societies [36, 60–65]. However, to my knowledge, this paper is the first to explicitly connect energy use with institution size. This connection is important because it is not easily explained by existing institution size theories, which focus mostly on the monetary incentives for institution growth.

I have offered a new theory of institution size that is rooted in human biology, and the theorized limitations of our ability to maintain social relations. I have proposed that institutions (firms and governments) are social hierarchies that serve to increase the scale of social coordination beyond that which is possible through egalitarian relations. I have argued that increases in energy consumption require a general increase in the scale of social coordination, and that increases in technological scale are a plausible reason for this connection. There is, of course, no need for increases in technological scale to be the *only* reason why social coordination increases with energy use — it is simply the easiest to study.

An important prediction of this theory is that increases in energy consumption are associated with a general increase in social hierarchy, meaning power is concentrated in the hands of fewer and fewer people. Although this starkly contradicts neoclassical economic theory, it is consistent with the power-based approach to political economy offered by Nitzan and Bichler [17]. If concentrations of power are at the heart of increases in energy consumption, then the theory developed here may be useful for studying a broad range of modern political economic phenomena.

References

1. Cottrell F. Energy & Society (Revised): The Relation Between Energy, Social Change, and Economic Development. Bloomington: AuthorHouse; 2009.

2. Hall C, Tharakan P, Hallock J, Cleveland C, Jefferson M. Hydrocarbons and the evolution of human culture. Nature. 2003;426(6964):318–322.

3. Soddy F. Virtual Wealth and Debt: the Solution of the Economic Paradox. London: George Allen & Unwin; 1926.

4. White LA. Energy and the evolution of culture. American Anthropologist. 1943;45(3):335–356.

5. Kondepudi DK, Prigogine I. Modern Thermodynamics: from Heat Engines to Dissipative Structures. Chichester: John Wiley & Sons; 1998.

6. Adams RN. Man, Energy, and Anthropology: I Can Feel the Heat, But Where's the Light? American Anthropologist. 1978;80(2):297–309.

7. Dunbar RI. Neocortex size as a constraint on group size in primates. Journal of Human Evolution. 1992;22(6):469–493.

8. Turchin P, Gavrilets S. Evolution of complex hierarchical societies. Social Evolution and History. 2009;8(2):167–198.

9. Gibrat R. Les inegalites economiques. Recueil Sirey; 1931.

10. Hart PE. The size and growth of firms. Economica. 1962;29(113):29–39.

11. Hymer S, Pashigian P. Firm size and rate of growth. The Journal of Political Economy. 1962;70(6):556–569.

12. Singh A, Whittington G. The size and growth of firms. The Review of Economic Studies. 1975;42(1):15–26.

13. De Wit G. Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models. International Journal of Industrial Organization. 2005;23(5):423–450.

14. Sutton J. Gibrat's legacy. Journal of Economic Literature. 1997;35(1):40–59.

15. Coase RH. The nature of the firm. Economica. 1937;4(16):386–405.

16. Geroski PA. The growth of firms in theory and in practice. Center for Economic Policy Research. 1999;2092.

17. Nitzan J, Bichler S. Capital as Power: A Study of Order and Creorder. New York: Routledge; 2009.

18. Lucas Jr RE. On the size distribution of business firms. The Bell Journal of Economics. 1978;9(2):508–523.

19. Gomes PM, Kuehn Z. Human capital and the size distribution of firms. IZA Discussion Papers Series. 2014;(8268).

20. Poschke M. The firm size distribution across countries and skill-biased change in entrepreneurial technology. IZA Discussion Papers Series. 2014;7991.

21. Barney J. Firm resources and sustained competitive advantage. Journal of management. 1991;17(1):99–120.

22. Peteraf MA. The cornerstones of competitive advantage: A resource-based view. Strategic Management Journal. 1993;14(3):179–191.

23. Priem RL, Butler JE. Tautology in the resource-based view and the implications of externally determined resource value: further comments. Academy of Management Review. 2001;26(1):57–66.

24. Meltzer AH, Richard SF. A rational theory of the size of government. The Journal of Political Economy. 1981;89(5):914–927.

25. Peltzman S. The growth of government. Journal of Law and Economics. 1980;23(2):209–287.

26. Brennan G, Buchanan JM. The power to tax: Analytic foundations of a fiscal constitution. New York: Cambridge University Press; 1980.

27. Niskanen WA. Bureaucracy and Representative Government. Piscataway, NJ: Transaction Publishers; 1974.

28. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, et al. In search of homo economicus: behavioral experiments in 15 small-scale societies. American Economic Review. 2001;91(2):73–78.

29. Beck T, Demirgus-Kunt A, Maksimovic V. Financial and legal institutions and firm size. The World Bank Development Research Group. 2003;2997.

30. Biggs T, Oppenheim J. What drives the size distribution of firms in developing countries. Employment and Enterprise Policy Analysis, Discussion paper. 1986;(6).

31. Gollin D. Nobody's business but my own: Self-employment and small enterprise in economic development. Journal of Monetary Economics. 2008;55(2):219–233.

32. Georgescu-Roegen N. The entropy law and the economic process. Cambridge, MA: Harvard University Press; 1971.

33. Ayres R, Warr B. Exergy, power and work in the US economy, 1900-1998. Energy. 2003;28(3):219–273.

34. Cleveland C, Costanza R, Hall C, Kaufmann R. Energy and the US economy: a biophysical perspective. Science. 1984;225(4665):890–897.

35. Fix B. Rethinking Economic Growth Theory from a Biophysical Perspective. Hall C, editor. New York: Springer; 2015.

36. Hall C, Klitgaard K. Energy and the Wealth of Nations: Understanding the Biophysical Economy. New York: Springer; 2012.

37. Simon HA, Bonini CP. The size distribution of business firms. The American Economic Review. 1958;48(4):607–617.

38. Biham O, Malcai O, Levy M, Solomon S. Generic emergence of power law distributions and Levy-Stable intermittent fluctuations in discrete logistic systems. Physical Review E. 1998;58(2):1352.

39. Gabaix X. Zipf's Law for Cities: An Explanation. The Quarterly Journal of Economics. 1999;114(3):739–767. doi:10.1162/003355399556133.

40. Kesten H. Random difference equations and Renewal theory for products of random matrices. Acta Mathematica. 1973;131(1):207–248. doi:10.1007/BF02392040.

41. Axtell RL. Zipf distribution of US firm sizes. Science. 2001;293:1818–1820.

42. Gaffeo E, Gallegati M, Palestrini A. On the size distribution of firms: additional evidence from the G7 countries. Physica A: Statistical Mechanics and its Applications. 2003;324(1–2):117–123. doi:10.1016/S0378-4371(02)01890-3.

43. Angelini P, Generale A. On the evolution of firm size distributions. The American Economic Review. 2008;98(1):426–438.

44. Cabral LM, Mata J. On the evolution of the firm size distribution: Facts and theory. American economic review. 2003;93(4):1075–1090.

45. Bottazzi G, Secchi A. Explaining the distribution of firm growth rates. The RAND Journal of Economics. 2006;37(2):235–256.

46. Stanley MH, Amaral LA, Buldyrev SV, Havlin S, Leschhorn H, Maass P, et al. Scaling behaviour in the growth of companies. Nature. 1996;379(6568):804–806.

47. Bottazzi G, Secchi A. Why are distributions of firm growth rates tent-shaped? Economics Letters. 2003;80(3):415–420. doi:10.1016/S0165-1765(03)00142-3.

48. Fu D, Pammolli F, Buldyrev SV, Riccaboni M, Matia K, Yamasaki K, et al. The growth of business firms: Theoretical framework and empirical evidence. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(52):18801–18806. doi:10.1073/pnas.0509543102.

49. Dunbar RI. The social brain hypothesis and its implications for social evolution. Annals of human biology. 2009;36(5):562–572.

50. Jamnagar. About Jamnagar; 2011. Available from: http://web.archive.org/web/20111027062343/http://www.ril.com/downloads/pdf/about_jamnagar.pdf.

51. Dunbar RI. Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences. 1993;16(04):681–694.

52. Hamilton MJ, Milne BT, Walker RS, Burger O, Brown JH. The complex structure of hunter–gatherer social networks. Proceedings of the Royal Society of London B: Biological Sciences. 2007;274(1622):2195–2203. doi:10.1098/rspb.2007.0564.

53. Strauss G. The Changing Role of the Working Supervisor. The Journal of Business. 1957;30(3):202–211.

54. Bardi U, Lavacchi A. A simple interpretation of Hubbert's model of resource exploitation. Energies. 2009;2(3):646–661.

55. Boltzmann L. Populare Schriften. SEVERUS Verlag; 2011.

56. Lane N, Martin W. The energetics of genome complexity. Nature. 2010;467(7318):929–934.

57. Ahlborn BK. Thermodynamic Limits of Body Dimension of Warm Blooded Animals. Journal of Non-Equilibrium Thermodynamics. 2005;25(1):87–102. doi:10.1515/JNETDY.2000.006.

58. McNab BK. Resources and energetics determined dinosaur maximal size. Proceedings of the National Academy of Sciences. 2009;106(29):12184–12188. doi:10.1073/pnas.0904000106.

59. Odum HT. Systems Ecology; an introduction. New York: Wiley; 1983.

60. Ayres RU, Warr B. The economic growth engine: how energy and work drive material prosperity. Cheltenham, UK: Edward Elgar Publishing; 2010.

61. Chaisson E. Cosmic Evolution: The Rise of Complexity in Nature. Cambridge, Mass.: Harvard University Press; 2002.

62. Giampietro M, Mayumi K, Sorman A. The Metabolic Pattern of Societies: Where Economists Fall Short. New York: Routledge; 2012.

63. Kummel R, Lindenberger D, Eichhorn W.   The productive power of energy and economic evolution.   Indian Journal of Applied Economics. 2000;8(2):1–26.

64. Pimentel D, Pimentel MH. Food, Energy, and Society, Third Edition. New York: CRC Press; 2007.

65. Smil V. Energy in nature and society: general energetics of complex systems. Cambridge, MA: MIT press; 2008.

Chapter 3

Evidence for a Power Theory of Personal Income Distribution

Abstract

This paper proposes a new 'power theory' of personal income distribution. I hypothesize that income is most strongly determined by *hierarchical power* — which I define as the number of subordinates under one's control. Using this definition, I find that relative income within firms scales strongly with hierarchical power. I also find that hierarchical power has a stronger effect on income than any other factor for which data is available. I conclude that this is evidence for a power theory of personal income distribution.

3.1 Introduction

Over the last decade, concerns about income inequality have risen to the forefront of public attention. As testament to this interest, Thomas Piketty's expansive treatise on inequality, *Capital in the Twenty-First Century*, became an unlikely best seller when it was published in 2014. Due in no small part to the work of Piketty and colleagues [1–5], *empirical* study of income inequality has flourished. But this plethora of new data has not led to a corresponding *theoretical* revolution.

The problem, I believe, is an unwillingness to question and test the basic assumptions on which current theory rests. Most theories of personal income distribution are deeply wedded to the assumption that income is proportional to *productivity*. However, this approach has a simple, but little discussed problem: income is distributed *far* more unequally than documented differentials in human labor productivity. But if not productivity, then what explains differentials in income?

I hypothesize that personal income is explained most strongly by *hierarchical*

power, as manifested by one's rank in an *institutional hierarchy*. Using the common definition of power as the 'ability to influence or control others', I measure hierarchical power in terms of the number of subordinates under an individual's control. From this definition, it follows that power, unlike productivity, tends to be *very* unequally distributed within hierarchies — a natural consequence of the tree-like chain of command that concentrates control at the top.

I test the power-income hypothesis in two ways. First, using the available firm case study data, I look for correlation between income and my metric for hierarchical power. I find that relative income within firms is strongly correlated with hierarchical power. I also find a strong correlation between *changes* in income and changes in hierarchical power. Second, I test the strength of the power-income effect against a wide range of other income-affecting factors. I find that grouping individuals by hierarchical rank has the strongest effect on income. I conclude that this is evidence for a power theory of personal income distribution.

The paper is organized into the following parts. In section 3.2, I review and critique existing theories of personal income distribution, and summarize the key failings of the dominant 'productivist' approach. In section 3.3, I outline the principles and motivations behind my proposed power theory of income distribution. In section 3.4, I test the power-income hypothesis against empirical evidence. All methods and sources are documented in the Appendix.

3.2   Theories of Personal Income Distribution

My reading of the history of personal income distribution theory is that the field has struggled to meet the following two *mutually contradictory* goals:

1. Address and explain the 'Galton-Pareto' paradox;
2. Maintain consistency with prevailing theories of functional income distribution.

The 'Galton-Pareto paradox' refers to the large discrepancy between the observed distribution of *human abilities* and the observed distribution of *income*. The former was first documented by Francis Galton [6], who found that human abilities were normally distributed, and hence quite equal. The latter was first documented by Vilfredo Pareto [7], who found that income distributions were highly skewed and unequal.

Following the findings of Galton and Pareto, political economists have spent a century struggling to reconcile these two facts [8]. The process has been

made difficult primarily because the two dominant theories of *functional* (class-based) income distribution assume a connection between individual productivity (hence ability) and income.

At the present time, two main approaches to personal income distribution theory exist: the *stochastic* and the *productivist* approach. The *stochastic* school solves the Galton-Pareto paradox by ignoring prevailing theories of function income distribution. In contrast, the *productivist* school purports to both resolve the Galton-Pareto paradox *and* maintain consistency with the rest of economic theory. However, a closer look reveals that this 'success' relies on untestable assumptions and circular logic. I review both theories below.

3.2.1   Stochastic Theories

The discrepancy between Galton and Pareto's findings is a *paradox* only if one expects that income should be somehow related to ability. Clearly the simplest resolution is to assume that ability plays a *negligible* role in determining income. This is precisely the road taken by *stochastic* models, which explain income distribution in terms of random events that have little (if anything) to do with the characteristics of individuals.

In 1953, David Champernowne demonstrated that a simple statistical process could be used to explain the 'Pareto' (or power law) distribution [9]. In this model, individuals are subjected to a series of random, exogenous 'shocks' that perturb their income. Over time, this process leads to an equilibrium power law distribution. Champernowne's model was later recognized to be part of a general class of interrelated models in which 'multiplicative' randomness is the generative mechanism for a skewed distribution [10–14].

More recently, econophysicists have used this stochastic line of thinking to draw explicit parallels between the distribution of income and the distribution of kinetic energy in gases. These *kinetic exchange* models explain income distributions in terms of the random exchange of money between individuals [15–18]. Under the assumption that money is conserved, kinetic exchange models generate distributions of income that closely resemble those in the real world.

Despite their successes, stochastic models have been mostly ignored by the economics profession. One reason is that the assumptions underlying this type of theory (especially kinetic exchange models) are often unrealistic [19]. Kinetic exchange models imply a world in which money is conserved for all time, nothing is ever produced, there are no groups, institutions or classes of people, and the world exists in static equilibrium.

However, a more insidious reason that stochastic models have been ignored is that they are inconsistent with the prevailing theories of *functional* income distribution, and the latter form the 'hard core' of political economic theory.

3.2.2   Productivist Theories

The discipline of political economy essentially arose in response to questions about class-based (or *functional*) income distribution. As David Ricardo saw it, the role of political economy was to "determine the laws" that regulate the distribution of income between the "classes of the community" [20].

Out of the 19th century debate over these laws, two great schools of thought merged — *Marxist* and *neoclassical*. Over the following century, virtually all economic theory was built on top of either Marxist or neoclassical assumptions about income distribution. The result is that if a new theory of *personal* income distribution contradicts these prevailing theories of *functional* income distribution, accepting the new theory logically requires discarding not only the functional income distribution theory, but a large part of political economic theory as well. Perhaps understandably, economists have hesitated to take this road. Instead, they have largely opted for personal income distribution theories that prioritize consistency with the rest of economic thought.

Although Marxist and neoclassical schools are usually positioned in opposition to one another, they both posit a similar link between *productivity* and income [21]. In neoclassical theory, income is attributed to *marginal productivity* — the incremental increase in output caused by the incremental increase in inputs of capital/labor [22,23]. Thus, if a capitalist makes more than a worker, it is because an additional unit of his 'capital' adds more to output than an additional unit of the worker's labor.

The logical implication of this theory is that income differences between *workers* — who all earn *labor* income — must be due to differences in *individual* productivity. Out of this line of reasoning came *human capital* theory, which attributes workers' productivity to some internal stock of 'human capital' [24–26].

Unlike neoclassical theory, Marxist theory posits that labor is the *sole* producer of value [27]. Therefore, both labor and capitalist income ultimately stem from workers' productivity. The Marxist twist is to treat capitalist income as *parasitic* – the result of the expropriation of surplus value created by workers. The relative balance between labor and capitalist income is then a function of the 'degree of exploitation' of workers. But when it comes to income distribution *among* workers, Marxists come to conclusions that are very similar to their neo-

classical counterparts. Since labor is the sole source of value, skilled workers who earn more than unskilled workers must somehow be more *productive* [28].

This productivity-income hypothesis has made it difficult for neoclassical and Marxist theories to address the Galton-Pareto paradox. Since individual productivity is presumably related to *ability*, one cannot take the easy road and simply negate any relation between *ability* and *income*. Instead, one must explain why productivity is as unequally distributed as income, but ability is not.

The most common resolution to the Galton-Pareto paradox is to assume that different abilities, each normally distributed, somehow interact to have a *multiplicative* effect on productivity [29, 30]. This multiplicative effect can be expressed as a production function in which a worker's output (Y) is an exponential function of the sum of different abilities (a_i): $Y = e^{a_1 + a_2 + \dots + a_i}$. This hypothesis is central to human capital theory, which proposes that investments in human capital yield multiplicative returns to productivity [24, 25].

But is this actually the case? Is productivity as unequally distributed as income? Unfortunately, this question is not as easily answered as it might seem. The problem is this: how do we compare the productivity of different workers who have *qualitatively different outputs*? For instance, how can we determine if a farmer, who produces potatoes, is more productive than a composer, who produces music? Any such comparison of qualitatively different outputs inevitably requires choosing a common unit of analysis. But the choice of this unit is *subjective*, and different units will lead to different results. The logical implication is that there are no *objective* grounds for comparing the productivity of workers with qualitatively different types of output. The same problem occurs when attempting to measure the productivity of *capital*: one can only compare capitalists with exactly the same output. There are other measurement problems inherent in marginal productivity theory. These include the inability to objectively measure capital [21, 32, 33], as well as the inability to isolate the effect on output caused by changes in capital versus changes in labor (see Pullen [34] for a good review).

Taking these measurement problems seriously means that one can compare productivity only between workers who have *exactly the same output*. [31] have compiled data that does exactly that — they report differences in productivity among workers doing the *same task*. In Figure 3.1, I take this data and convert it into a Gini index of 'productivity inequality' so that it is directly comparable to income inequality within nation states. This evidence indicates that differences in productivity are *systematically* too small to account for observed levels of inequality.
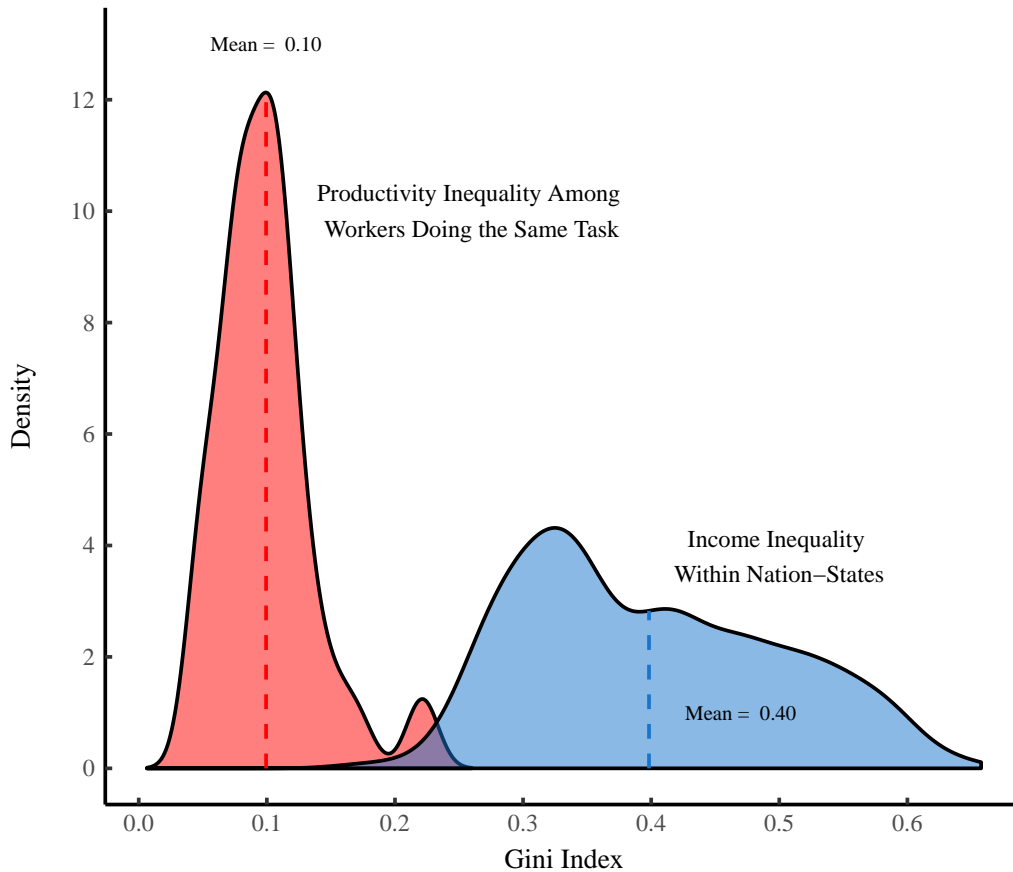
Figure 3.1: Labor Productivity Inequality vs. Income Inequality

Using a Gini index, this figure compares the inequality of worker productivity to income inequality within nation-states. Data for the former comes from Hunter et al. [31], who report the coefficient of variation of productivity among workers conducting the *same task*. Data plotted here shows the distribution of productivity inequality for 55 different tasks. I convert Hunter's data to a Gini index by assuming that worker productivity is lognormally distributed. The Gini index (G) of a lognormal distribution with a coefficient of variation c_v is $G = \mathrm{erf}(\frac{1}{2}\sqrt{\log(c_v^2 + 1)})$. I plot the resulting distribution against the distribution of Gini indexes of income inequality for all country-year observations in the World Bank database (series SI.POV.GINI).

However, the link between productivity and income is not typically measured in such restrictive terms. Instead, the standard practice is to adopt *monetary value* as a common unit of comparison for measuring different outputs. Thus, labor productivity is generally measured in terms of sales or value-added per worker [35–41]. The problem with this approach is that it relies on circular logic. According to theory, income is *explained* by productivity. But when the theory is tested, productivity is *measured* in terms of income. And based purely on accounting principles, we expect wages to be correlated with sales/value-added per worker.

Double entry accounting principles dictate that the value-added (Y) of a firm is equivalent to the sum of all wages/salaries (W) and capitalist income (K). If we divide by the number of workers (L), we find that value-added per worker is equivalent to the average wage ($w = W/L$) plus K/L:

$$\frac{Y}{L} = \frac{W + K}{L} = w + \frac{K}{L} \qquad (3.1)$$

Sales (S) are similar, but include an additional non-labor cost term (C):

$$\frac{S}{L} = \frac{W + K + C}{L} = w + \frac{K + C}{L} \qquad (3.2)$$

Thus, if we look for correlation between average wage (w) and value-added/sales per worker (Y/L or S/L), we will surely find it, since simple accounting definitions dictate that the former is a major component of the later.

To summarize, existing theories of personal income distribution are plagued by fundamental problems. The two main schools reviewed here — stochastic and productivists — both have major shortcomings. The stochastic approach, while interesting from a mathematical standpoint, makes assumptions that are unrealistic and have little to do with the real world. The productivist school, on the other hand, has waged an uphill battle with empirical evidence. Its successes have been achieved by basing empirical tests on circular logic.

I argue that a new approach is needed. Rather than focus on productivity (or stochastic interactions) I propose that personal income is best explained by the *hierarchical power structure of institutions*.

3.3   A Hierarchical Power Theory of Personal Income Distribution

The premise of this paper is that income distribution can be explained primarily in terms of differentials in hierarchical power. But before diving into the specifics of this theory, I want to provide a rationale based on the big picture of human history. Why? There is nothing like looking at the past to gain fresh insight into the present. Let's ask a simple question: what aspects of human history suggest that hierarchical power might affect how we distribute resources (which is what income distribution is all about).

Let's begin with our deep history — the evolutionary backdrop of the human species. Humans are but one of a wide variety of social mammals, virtually all of which form dominance hierarchies, or 'pecking orders' [42–47]. A key characteristic of these dominance hierarchies is that high social rank is associated with preferential access to resources, particularly sexual mates [48–52].

Of course, human behavior is far more complex than even the most intelligent (non-human) primates. Just because we evolved from hierarchy-forming animals does not necessarily mean that hierarchical rank *still* plays a role in how we divide up the pie. However, there is good evidence that humans *do* have an instinctual behavior towards hierarchy formation. Several studies have shown that children and adolescents spontaneously form dominance hierarchies when placed into small groups [53–55]. Other studies have shown that, like other social mammals, human reproductive success increases with social status [56,57]. There is even evidence that social status at birth is epigenetically imprinted on human DNA [58] — something that also occurs in Rhesus monkeys [59]. Given our evolutionary heritage, it seems plausible that hierarchy plays a role in the way humans distribute resources.

Another reason to suspect that resource distribution has to do with hierarchy and power is the ubiquity of *inherited status* in human history. It is hard to justify the wealth of a hereditary aristocracy as stemming from anything but power and privilege. Interestingly, inherited status has surprisingly deep historical roots. There is tentative archaeological evidence for inherited status beginning in the neolithic era [60–62], and widespread evidence beginning in the bronze age around 5000 years ago [63–67]. It is around this time that the first Egyptian dynasty formed [68], followed later by dynasties in Mesopotamia [69] and China [70].

Since then, as Gaetano Mosca observes, the existence of a hereditary ruling class has been the norm:

> There is practically no country of longstanding civilization that has not had a hereditary aristocracy at one period or another in its history. We find hereditary nobilities during certain periods in China and ancient Egypt, in India, in Greece before the wars with the Medes, in ancient Rome, among the Slavs, among the Latins and Germans of the Middle Ages, in Mexico at the time of the Discovery and in Japan down to a few years ago. [71]

But while history may be sordid, there is always the possibility that modern societies have made a clean break with the past. Power may have played a central role in the distribution of resources in past societies, but in modern societies *reciprocal exchange* is what matters most. This is the story that emerged in the writings of Adam Smith [72] and was codified into neoclassical theory by Jevons, Menger and Walras [73–75]. To paraphrase George Orwell [76], this is now the prevailing orthodoxy that most right-thinking economists accept without question.

But what if there has *not* been a clean break with the past? What if power still plays an important role in shaping resource distribution? A wide variety of scholars have argued that this is the case. A non-exhaustive list would include [21, 77–89]. These scholars argue that power plays a central role in shaping income distribution.

If there is to be a power-based theory of income distribution, what should it look like? According to Christopher Brown:

> ... [A] theory of distribution should be indistinguishable from a theory of power. A satisfactory theory of power would, beyond defining what power is, elucidate principles to explain how power is established, enlarged or diminished, protected and perpetuated, redistributed, exercised, and rendered legitimate or illegitimate. [90]

A full-fledged theory of power is a tall order. In this paper, I narrow the focus to look only at *hierarchical power* in the context of personal income distribution. My ideas stem from the work of Simon [91] and Lydall [92], who independently proposed income distribution models based on the hierarchical structure of firms.

The focus of Simon and Lydall's work is the *branching* nature of institutional hierarchies, in which each superior controls multiple subordinates. This structure is unique to humans. All other animals form *linear* hierarchies — an ordinal ranking from top to bottom. The most important feature of a branching hierarchy is that it *concentrates* power in the hands of the few. I propose that this concentration of hierarchical power can be used to explain income inequality.

The main theoretical contribution of this paper is to offer a quantifiable def-

inition of hierarchical power that allows power differentials to be directly compared to income differentials. I test the following hypothesis:

Hypothesis: Income is most strongly determined by *hierarchical power*, as measured by the number of subordinates under one's control.

3.3.1   Measuring Hierarchical Power

What is *hierarchical power*? I define it as the ability to control subordinates within a hierarchical chain of command. The link between hierarchy and power is implicit in the etymology of the word 'hierarchy' itself, which derives from the Greek term *hierarkhēs*, meaning 'sacred ruler' [93]. In essence, an institutional hierarchy is a nested set of power relations between a superior (a ruler) and subordinates (the ruled). It is a control structure that concentrates power at the top [94]. I propose that one's power within a social hierarchy is proportional to the *number of subordinates under one's control*. I put this in formula form as:

$$\text{hierarchical power} = \text{number of subordinates} + 1 \qquad (3.3)$$

The logic of this equation is that all individuals start at a baseline power of 1, indicating that they have control over themselves. Power then increases linearly with the number of subordinates.

If we had access to the exact chain of command structure of an institution, we could use this definition to measure the power of each individual within a hierarchy. Unfortunately, chain of command information is rarely available. Instead, existing case studies report *aggregate* hierarchical structure only — total employment by hierarchical level. While we cannot calculate the power of *specific* individuals, we can use this data to calculate the *average* power of all individuals in a specific hierarchical level:

$$\bar{P}_h = \bar{S}_h + 1 \qquad (3.4)$$

Here $\bar{P}_h$ is the average power of individuals in hierarchical level h, and $\bar{S}_h$ is the average number of subordinates below these individuals. The average number of subordinates $\bar{S}_h$ is equal to the sum of employment (E) in all subordinate levels, divided by employment in the level in question. Figure 3.2 shows a sample calculation, where red individuals occupy the level in question, and blue individuals are subordinates. Each red individual has 2 direct subordinates, and 4 indirect subordinates, for a total of 6 subordinates. The average hierarchical power in level three is therefore 7.

$$\bar{S}_3 = \frac{E_1 + E_2}{E_3} = \frac{16 + 8}{4} = 6$$
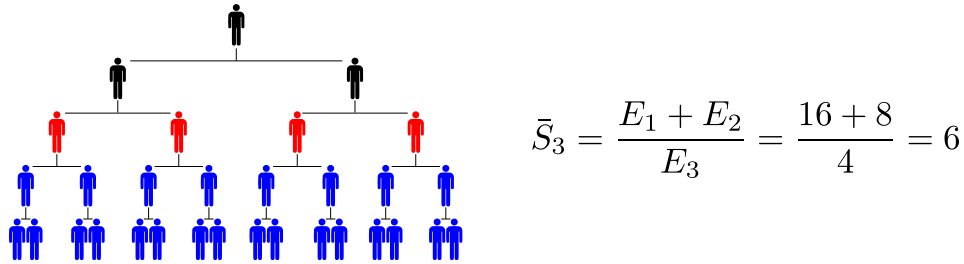
Figure 3.2: Calculating the Average Number of Subordinates

Using summation notation, we can write the following general equation for the average number of subordinates in hierarchical level h (here $h = 1$ is the bottom hierarchical level 1):

$$\bar{S}_h = \sum_{i=1}^{h-1} \frac{E_i}{E_h} \tag{3.5}$$

Together, equations 3.4 and 3.5 allow us to define and measure the average power of individuals in an institutional hierarchy.

3.3.2 The Concentration of Power Within Firms

The focus of this paper is the hierarchical structure of business firms, which are the dominant institutions in capitalist societies. I treat firms as 'dedicated hierarchies' [95]. The premise of my theory is that firm hierarchies concentrate power, which causes concentrations of income.

But before we look at the relation between power and income, there is a prior question: how concentrated is hierarchical power within firms? Just as we can with income inequality, we can use the Gini index to quantify the concentration of hierarchical power within firms. Figure 3.3 shows the distribution of hierarchical power in a hypothetical firm. In this firm, the Gini index of power inequality is 0.58. To put this in perspective, if income within this firm was *exactly* proportional to power, the firm would have income inequality on par with South Africa (according to World Bank data).

Of course, this is a contrived example. What we really want to know is — how concentrated is power in real-world firms? I have identified six firm case studies that provide adequate data to calculate average power by hierarchical level. These studies (discussed in detail in Appendix B.2) offer a sample of firms from the United States, Britain, the Netherlands, and Portugal. While a larger firm sample would be better, the proprietary nature of firm payroll data has
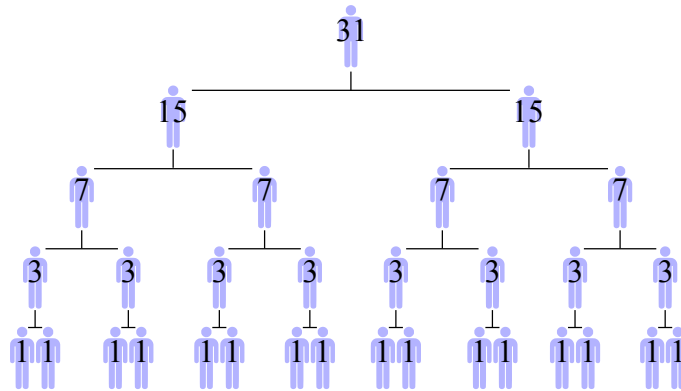
Figure 3.3: The Distribution of Power Within A Firm

This figure shows how power is distributed within a hypothetical hierarchical firm. The number above each individual indicates their power, as defined by Eq. 3.4 and 3.5.

---

proved a major obstacle to empirical research. As a result, data on firm hierarchical structure is quite limited.

To calculate power inequality in these case study firms, I first use equations 3.4 and 3.5 to quantify average power by hierarchical level in each firm. I then assign each member of the firm the average power in their respective hierarchical level. This results in a distribution of hierarchical power, from which we can calculate the Gini index. Figure 3.4 shows the resulting distribution of Gini indexes of power inequality within these case study firms (each firm-year observation gets one Gini index).

To compare this concentration of power to concentrations of income, Figure 3.4 also shows the distribution of income inequality within nation-states (the same distribution as in Figure 3.1). Interestingly, hierarchical power within these firms is much more concentrated than income within nation-states. This suggests that hierarchical power is a good starting point for a theory of income distribution. If nothing else, it means that a power-based theory will not suffer from the under-explanation problem that plagues productivist theory.

3.4   Testing the Power-Income Hypothesis

To reiterate, the power-income hypothesis proposes that income is most strongly determined by *hierarchical power*, as measured by the number of subordinates under one's control. To test this hypothesis, it is helpful to break it down into two parts:
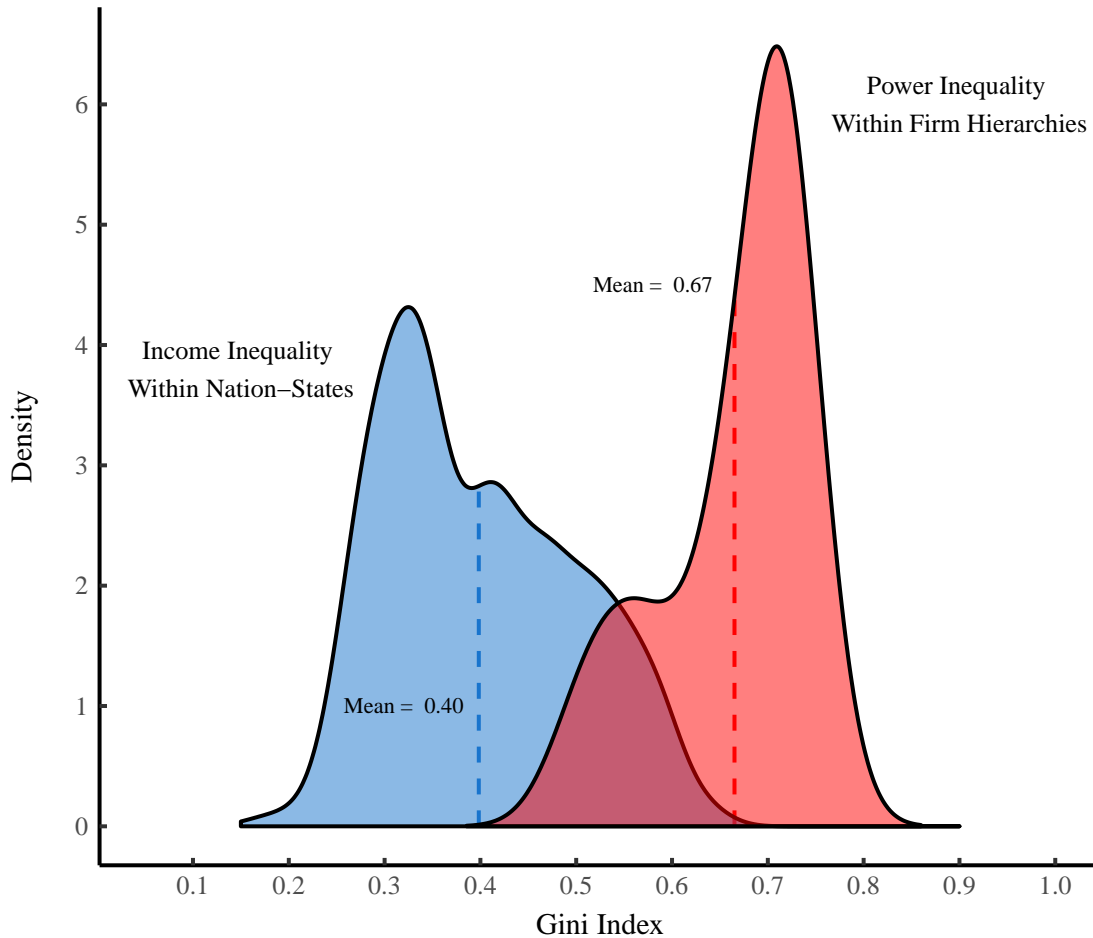
Figure 3.4: Income Inequality vs. Power Inequality within Firms

This compares the distribution of income inequality within nation states to the distribution of hierarchical power inequality within six case study firms. To calculate the distribution of power, I use equations 3.4-3.5. Firm case study data comes from [96–101] (for details, see Appendix B.2). The distribution of power inequality is calculated using Gini indexes for all firm-year observations. The distribution of income inequality within nation-states is calculated using all countries-year observations in the World Bank database series SI.POV.GINI.

Hypothesis A: Relative income within a hierarchy is proportional to hierarchical power.

Hypothesis B: Hierarchical power affects income more strongly than *any* other factor.

Hypothesis A is an important initial test of the power-income hypothesis. If income within a hierarchy is not significantly correlated with our metric for hierarchical power, then the power-income hypothesis is *false*. However, even a substantial correlation is only partial evidence, since many factors other than power are well-known to strongly affect income (education is the most widely recognized). Thus, we must go one step further and test if the power-income effect is *stronger than all others*. If we find empirical support for both hypotheses A and B, then we conclude that there is evidence for the power-income hypothesis.

3.4.1   Power-Income Correlation

To test hypothesis A, I look for both a *static* and a *dynamic* correlation between income and hierarchical power. I begin with the static test.

I analyze the correlation between power and income in six case study firms — the same firms used in section 3.3.2 (for a detailed discussion of these studies, see Appendix B.2). For each firm in each observation year, I use equations 3.4 and 3.5 to calculate average power by hierarchical level. I then compare average power to average relative income by hierarchical level. In order to make comparisons across firms (and across time), I normalize all income data so that the mean income in the bottom hierarchical level is always equal to 1.

The results of this analysis are shown in Figure 3.5. Each point represents a single firm-year observation, with the different case-study firms indicated by color. Although the firm sample is small, the evidence is conclusive: there is a strong correlation between relative income in our case-study firm hierarchies and our metric for power.

While Figure 3.5 shows a correlation between *static* levels of hierarchical power and pay, it is also important to test for a *dynamic* correlation. That is, we want to know if *changes* in power are related to *changes* in income when individuals are promoted/demoted within a firm. I conduct such a test using the data published by Baker, Gibbs, and Homstrom [97] — the 'BGH dataset'. This dataset contains raw personnel data for a large US firm over the years 1969-1985.

I define a promotion/demotion as any change in an individual's hierarchical level. For each such event, we define the fractional change in power ($\Delta \bar{P}$) as
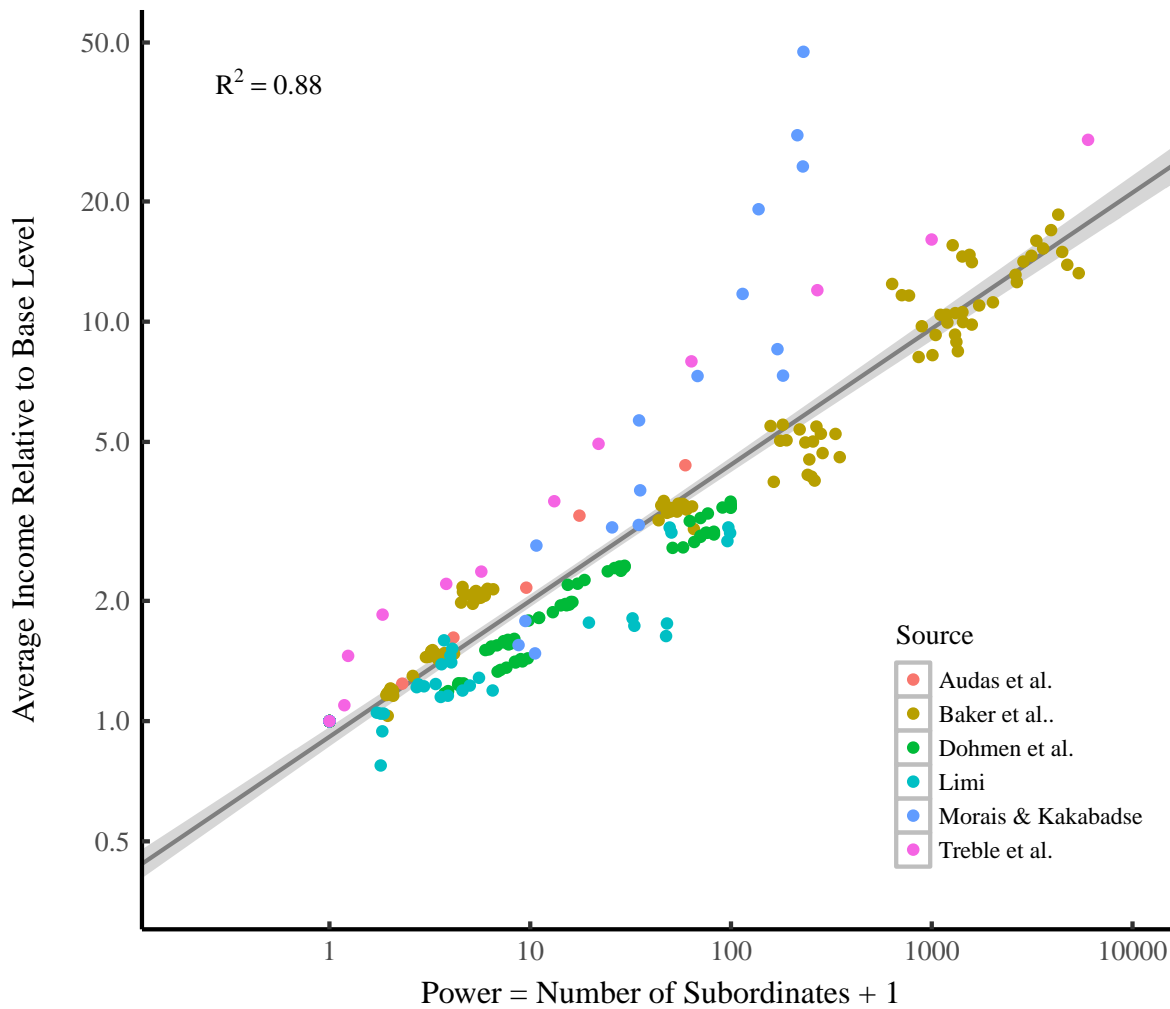
Figure 3.5: Average Income vs. Hierarchical Power Within Case-Study Firms

This figure shows data from six firm case studies [96–101]. The vertical axis shows average income within each hierarchical level of the firm (relative to the base level), while the horizontal axis shows our metric for average power, which is equal to one plus the average number of subordinates below a given hierarchical level (see Eq. 3.4 and 3.5). Each point represents a single firm-year observation, and color indicates the particular case study. Grey regions around the regression indicate the 95% confidence region.
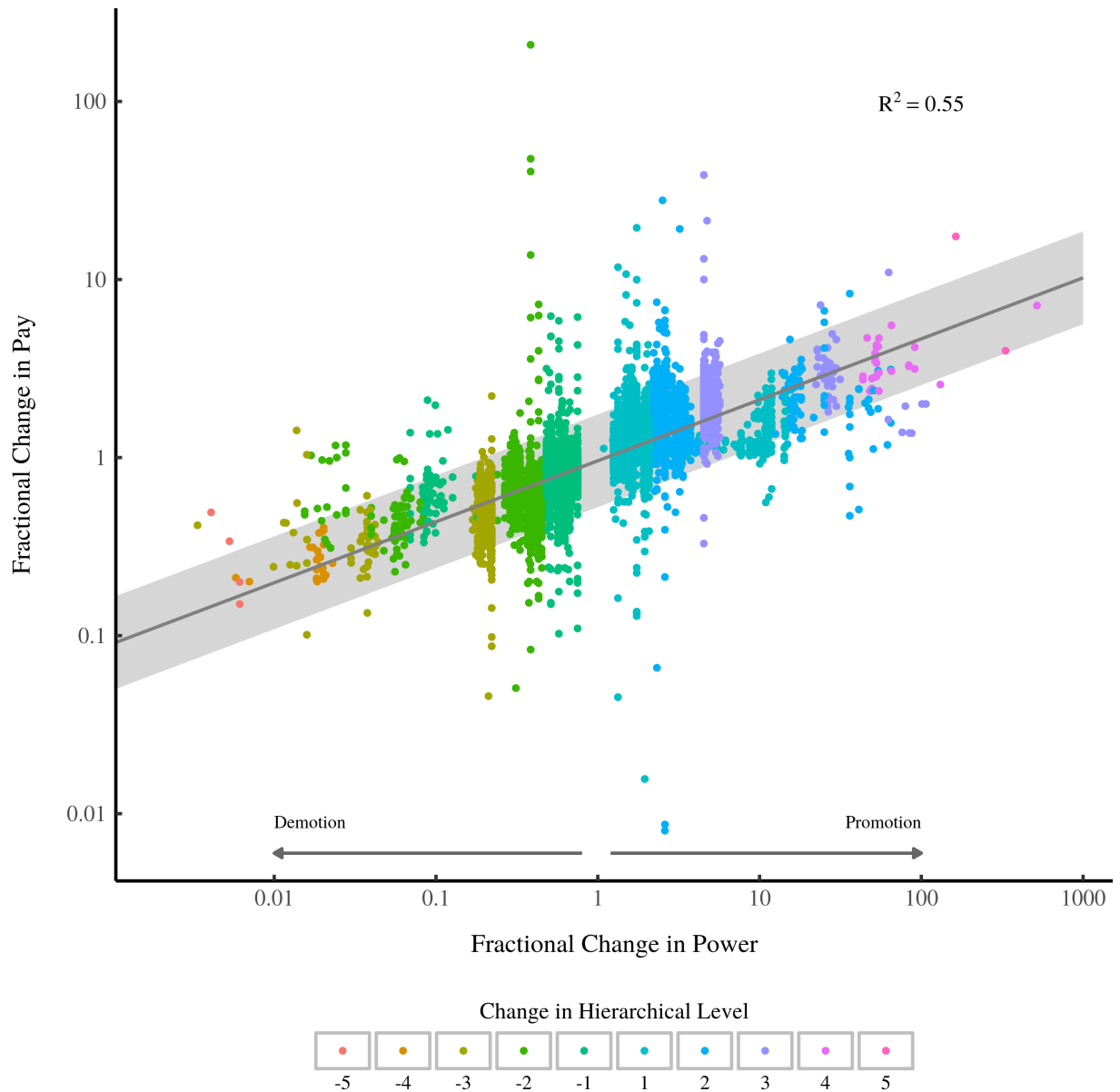
Figure 3.6: Changes in Hierarchical Power and Pay During Intra-Firm Promotions

This figure plots the fractional change in pay (Eq. 3.7) versus the fractional change in hierarchical power (Eq. 3.6) for individual promotions/demotions in the Baker, Gibbs, and Holmstrom (BGH) dataset [97]. Each point represents the resulting change in pay and power of a *single* individual. Over 16,000 promotion/demotion events are plotted here. Change in hierarchical level is indicated by color. The grey region indicates the 95% prediction interval of a log-log regression. The BGH data comes from an anonymous US firm over the period 1969-1985. The dataset is available at http://faculty.chicagobooth.edu/michael.gibbs/research/index.html

the ratio of power *after* versus power *before* the promotion/demotion (Eq. 3.6). An individual's power is defined by Eq. 3.4. Since we do not know the exact chain of command, I assign all individuals the *average* power of their respective hierarchical level.

$$\Delta \bar{P} = \frac{\bar{P}_{\text{after}}}{\bar{P}_{\text{before}}} \tag{3.6}$$

For each promotion/demotion, I define the fractional change in income (ΔI) as the ratio of income *after* versus income *before* the event (Eq. 3.7). In order to isolate the effect of the promotion from the exogenous effects of inflation and/or general wage increases, I measure all incomes *relative* to the firm mean income ($\bar{I}$) in the appropriate year.

$$\Delta I = \frac{I_{\text{after}}/\bar{I}_{\text{after}}}{I_{\text{before}}/\bar{I}_{\text{before}}} \tag{3.7}$$

Figure 3.6 show the results of this dynamic analysis. Here each plotted point represents the fractional change in pay and power for the promotion/demotion of a *single individual*. For the over 16,000 promotions/demotion events analyzed here, a highly significant correlation exists between changes in power and changes in individual income.

Interestingly, the correlation holds both for promotions and for *demotions*, the latter occurring when an individual drops hierarchical levels. The relative pay reductions accompanying these demotions are difficult to understand from a productivist approach. Do these individuals suddenly experience a drastic reduction in ability/productivity? The evidence in Figure 3.6 suggests a better explanation: within the BGH firm, pay is largely a function of the power of a specific hierarchical *position*, irrespective of the person holding this position.

To conclude, the available evidence is consistent with hypothesis A. Relative income within firms is both *statically* and *dynamically* correlated with hierarchal power. Having survived this first hurdle, we now move on to test the power-income effect in the more stringent form of hypothesis B.

3.4.2 The Strength of the Power-Income Effect

Hypothesis B states that hierarchical power affects income more strongly than *any other factor*. To test this hypothesis, I use an analysis of variance method to quantify the income effect of from wide variety of different factors. In order

A. Small Effect on Income   **B. Large Effect on Income**

$$
\begin{aligned}
\text{Between} - \text{Group}: \quad & G_B = 0.07 \\
\text{Within} - \text{Group}: \quad & \bar{G}_W = 0.45 \\
\text{Indicator}: \quad & G_{BW} = 0.17
\end{aligned}
$$

$$
\begin{aligned}
\text{Between} - \text{Group}: \quad & G_B = 0.57 \\
\text{Within} - \text{Group}: \quad & \bar{G}_W = 0.14 \\
\text{Indicator}: \quad & G_{BW} = 4.07
\end{aligned}
$$
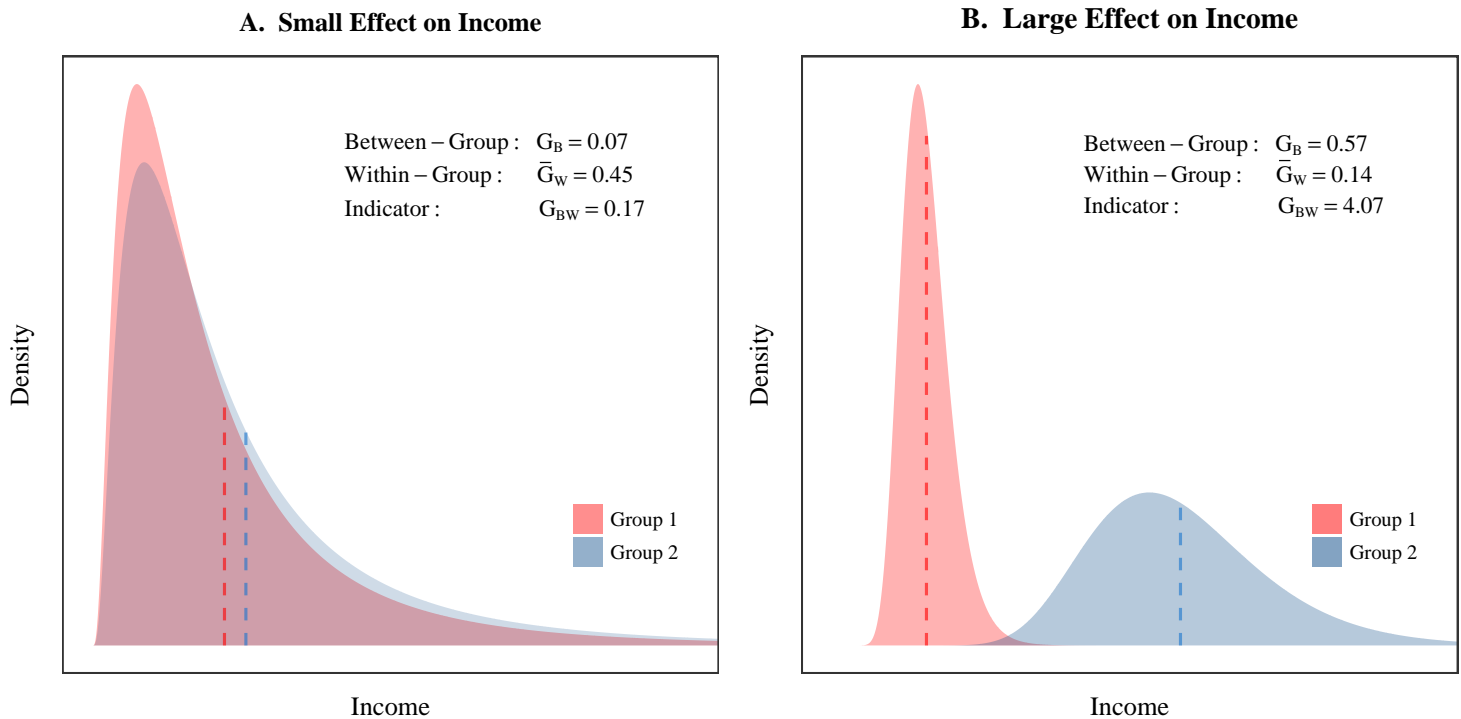
Figure 3.7: Analysis of Variance Using the Gini Index

This figure shows an example of the analysis of variance method that uses the Gini index. A hypothetical 2-group variable (like 'sex') is illustrated to have a small effect on income in panel A and a large effect in panel B. The means of each distribution are indicated by a dashed line. We use equation 3.8 to define the between-within Gini indicator, G_{BW}. A small effect on income is indicated by a G_{BW} that is close to zero, while a large effect is indicated by a G_{BW} greater than one.

---

to make the test as thorough as possible (given data constraints), some data is *model dependent* (see the Appendix for a detailed discussion).

Method for Measuring Effect Size

While there are many conceivable ways that hypothesis B could be tested, the format of available data makes the *analysis of variance* method the most appropriate. This is because many factors that affect income (such as 'sex' or 'race') are *qualitative* variables. Even factors like 'education' and 'age' that could conceivably be quantitatively measured (in units of time) are typically reported in qualitative groups such as 'college graduate' or ages '50-59'. The analysis of variance (ANOVA) method provides a simple way of determining how strongly qualitative variables affect income. The essence of this approach is to compare

between-group income dispersion to within-group income dispersion for a given factor. The larger the between-group dispersion is relative to the within-group dispersion, the larger the effect on income.

This approach is most easily understood by way of an example. Figure 3.7 shows a hypothetical example of how a two-group variable like 'sex' might affect income. When the separate income distributions of the two groups are plotted together, we can clearly see a *small* effect in Fig. 3.7A and a *large* effect in Fig. 3.7B. How do we quantify the size of this effect? Most people likely judge the difference in group means against the dispersion within each group. We might call this a *signal-to-noise* ratio, where the 'signal' is the difference in group means and the 'noise' is the within-group dispersion. The larger the signal is relative to the noise, the larger the effect.

The ANOVA method allows us to generalize this concept of effect to more than two groups. The corresponding signal-to-noise ratio is often called Cohen's f^2. For this metric, the 'signal' is the dispersion *between* group means, while the 'noise' is the dispersion *within* groups, where dispersion is measured as the sum of squared differences from the mean [102, 103]. While Cohen's f^2 is a common measure of effect size, its calculation requires either raw data on individual income, or data for within-group *variance* (or *standard deviation*). Unfortunately, this type of data is difficult to obtain. Instead, what *is* readily available are aggregate statistics reporting within-group *Gini indexes*. Because of the ubiquity of the Gini index, I use it to measure effect size.

Similar to Cohen's f^2, my effect size metric is a signal-to-noise ratio (Eq. 3.8). However, rather than the sum of squares, I use the Gini index to measure both within-group and between-group dispersion. I call this metric the *between-within Gini ratio* (G_{BW}).

$$G_{BW} = \frac{G_B}{\overline{G_W}} \tag{3.8}$$

Here G_B is the between-group Gini index (the Gini index of group mean incomes), while ($\overline{G_W}$) is the average of all within-group Gini indexes. For a detailed discussion of the relation between G_{BW} and f^2 see Appendix B.8.

The value of G_{BW} can range from 0 to infinity, with larger values indicating a larger effect on income (see the example in Fig. 3.7). Of particular interest is the value $G_{BW} = 1$, which occurs when between-group dispersion is equal to within-group dispersion. Any factor that produces $G_{BW} > 1$ can be considered to have a significant impact on income, since inequality *between* groups is larger than inequality *within* groups. However the primary use of the G_{BW} metric is not

its *absolute* value, but its *relative* value when different income-affecting factors are compared.

A well-known shortcoming of the Gini index is that it has a *downward bias* for small sample sizes. If the sample size is n, the maximum possible Gini index is:

$$G_n^{max} = \frac{n-1}{n} \tag{3.9}$$

Thus a sample size of $n = 2$ has a maximum Gini index of $G_2^{max} = 0.5$. This bias presents a problem for the calculation of the between-group Gini index G_B because the number of groups (n) is often extremely small (i.e. $n = 2$ for the factor 'sex'). While this small n is not really a *sample* (it is the *actual* number of groups), it still causes a bias in the Gini index. The result is that we cannot safely compare G_B between two income-affecting factors with different numbers of internal groups.

To correct for this bias, I use the method proposed by George Deltas [104]. The bias-adjusted Gini index (G^{adj}) is defined by dividing the unadjusted Gini (G) by the maximum possible Gini (G_n^{max}), given the number of internal groups n:

$$G^{adj} = \frac{G}{G_n^{max}} \tag{3.10}$$

All between-group Gini calculations in this paper use the adjusted Gini index, G^{adj}. However, for notational simplicity I refer to this adjusted between-group Gini as G_B for the remainder of the paper.

Some Clarifications on 'Effect Size'

It is important to clarify that my empirical method measures the effect on *income*, not the effect on *inequality*. There is a subtle, but important difference. In measuring the effect on *income*, group size is *irrelevant*. Any factor that has a strong effect on income (like education) necessarily involves zeroing in on a small, elite group of people (for instance, a small fraction of the population has a graduate degree). But the effect on *inequality* takes group size into consideration. Consider the scenario where all people with a graduate degree are millionaires, but there are only 10 such people (out of millions). Having/lacking a graduate degree will have a strong effect on income, but not on inequality. I hope this gives an intuitive understanding of the difference between the two types of effect size. For a technical discussion, see Appendix B.8.
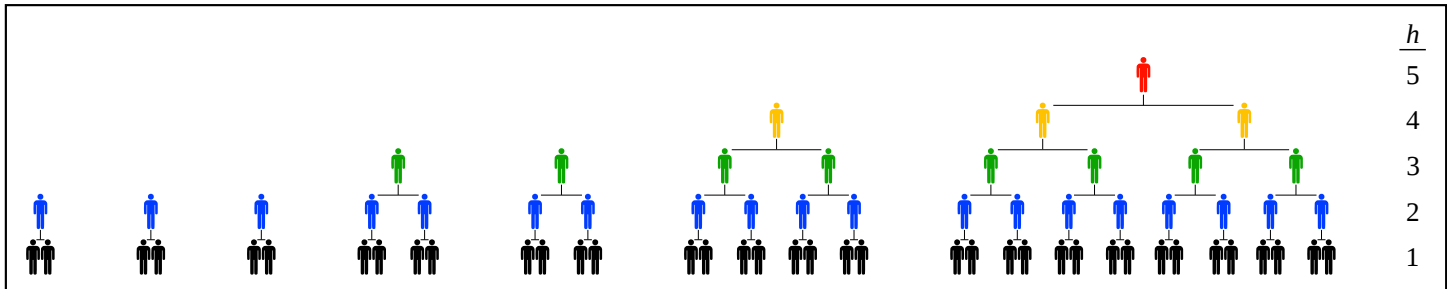
Figure 3.8: Grouping Power By Hierarchical Level

This figure shows my method for grouping individuals by their power. In this figure, each hierarchy represents a different firm. My proposed groups consist of all individuals (regardless of firm) that share the same hierarchical level. Groups are indicated by color.

On a different note, readers trained in econometrics will (correctly) observe that my method for measuring effect-size does not *isolate* the income-effects of a given factor. It does not show that, when *all* other factors are held constant, a change in factor *A* by amount *x* affects income by amount *y*. I make no attempt to do this because I think it is the wrong approach. As Keynes long ago argued, the only conceivable way that an econometric model can *isolate* an effect is if the model includes a *complete* list of causal factors [105]. But since we can never be sure that our causal list is complete, we can never know if our econometric model is wrong [106].

My thinking is more pragmatic. Given the complexities of human behavior, we can likely never isolate a factor to find its 'true' effect on income. But we can *rank* effect-size with the full understanding that when we measure one factor's effect on income, enumerable other factors are included in this measurement. In the face of enumerable confounding variables, Occam's razor would suggest that we simply chose the factor with the largest effect on income and use it to build a theory.

Grouping Individuals By Hierarchical Level

To test hypothesis B (that hierarchical power affects income more strongly than any other factor) using an analysis of variance method, we must group individuals into different categories/classes of social power. My method is to group individuals by *hierarchical level* across all firms, as illustrated in Figure 3.8.

This method is theoretically attractive because hierarchical level is the principle determinant of power. If a firm has a constant 'span of control' (the number of

subordinates below each superior), then power will increase *exponentially* with hierarchical level. In Figure 3.8, the span of control is constant both *within* and *between* firms. The result is that all individuals in each hierarchical level have the *same* power. In the real-world, we would expect this *not* to be the case. Evidence from firm case-study data suggests that the span of control varies both *within* and *between* firms (see Fig. 4 and 5 in Appendix B.2). As a result, we still expect that average power will increase exponentially with hierarchical level, but each hierarchical level will contain individuals with a *range* of different power.

While there are other conceivable ways of grouping individuals by power, this method is both theoretically attractive and *practical* for empirical analysis. The available data on firm hierarchies is limited, and the most commonly reported metric is the distribution of income by hierarchical level.

The Data

To test hypothesis B, I use the 19 different income-affecting factors shown in Table 3.1. With two exceptions (discussed below), data comes from the United States. Data sources as well as details about each category are discussed in Appendix B.1.

Before proceeding with a discussion of the data sources used for income by hierarchical level, it is worth reviewing why I do *not* use the same case study data that was used to test hypothesis A. Testing hypothesis B requires grouping

Table 3.1: Income-Affecting Factors Used to Test Hypothesis B

Geographic	Physical Attribute	Socioeconomic
Census Block Group	Age	Education
Census Tract	Cognitive Score*	Employee vs. Self-Employed
County	Race	Firm*
Urban vs. Rural	Sex	Full vs. Part Time
		Hierarchical Level*
		Home Owner vs. Renter
		Occupation
		Parents' Income Percentile
		Public vs. Private Sector
		Religion
		Type of Income (Labor/Property)

* Indicates variables that use model-dependent data (at least in part)

individuals by hierarchical level across a *large* number of firms. To be consistent, the firms should all be in the *same* country (ideally the United States), and the observations (that are compared) should be in the *same* year. The case study data does not meet these requirements: it is a *small* sample, with firms from many *different* countries with many *non-overlapping* years. As a result, the case study data is not useful for testing hypothesis B.

Instead, I use three different sources for estimating income distribution by hierarchical level. The first source is a seminal study by Mueller, Ouimet, and Simintzi [107] that reports income distribution by hierarchical level for 880 United Kingdom firms over the period 2004-2013. The second source is a study by Fredrik Heyman [108] that analyzes the pay distribution of the top 4 levels of management in 560 Swedish firms in the year 1995. Heyman's data comes with the caveat that it does not represent all hierarchical levels — just the top four. For this reason, I mark Heyman's results with an asterisk.

I use this non-US data because I am not aware of any equivalent US study that reports income distribution by hierarchical level over a large number of firms. While comparing US to UK/Swedish studies is not ideal, I proceed because of the lack of alternative data. If anything, the UK and Swedish data should lead to an *under*-estimate of the power-income effect in the United States. Why? Both the UK and Sweden have significantly less income inequality than the US (according to the World Bank, the most recent UK and Swedish Gini indexs are 0.33 and 0.27, while the most recent US Gini index is 0.46). If there is less *total* inequality, the potential for *between*-group inequality is diminished, resulting in a lower G_{BW} metric (see Eq. 3.8).

My third source for hierarchical level data is a model that uses the insights from firm case study data to estimate the hierarchical pay structure of 713 US firms in the *Compustat* database (covering the years 1992-2015). This 'Compustat Model' is discussed in detail in the Appendix, but I review its core components here.

The idea of the Compustat Model is that firm case-study data can be used to make generalizations about the hierarchical employment and pay structure of firms. Although different firms have differently shaped hierarchies (see Fig. 4 in Appendix B.2), there are underlying regularities shared by all firms. The following regularities are shown in Fig. 5 in Appendix B.2:

1. The span of control tends to *increase* with hierarchical level.
2. The ratio of average pay between adjacent hierarchical levels *increases* by level.
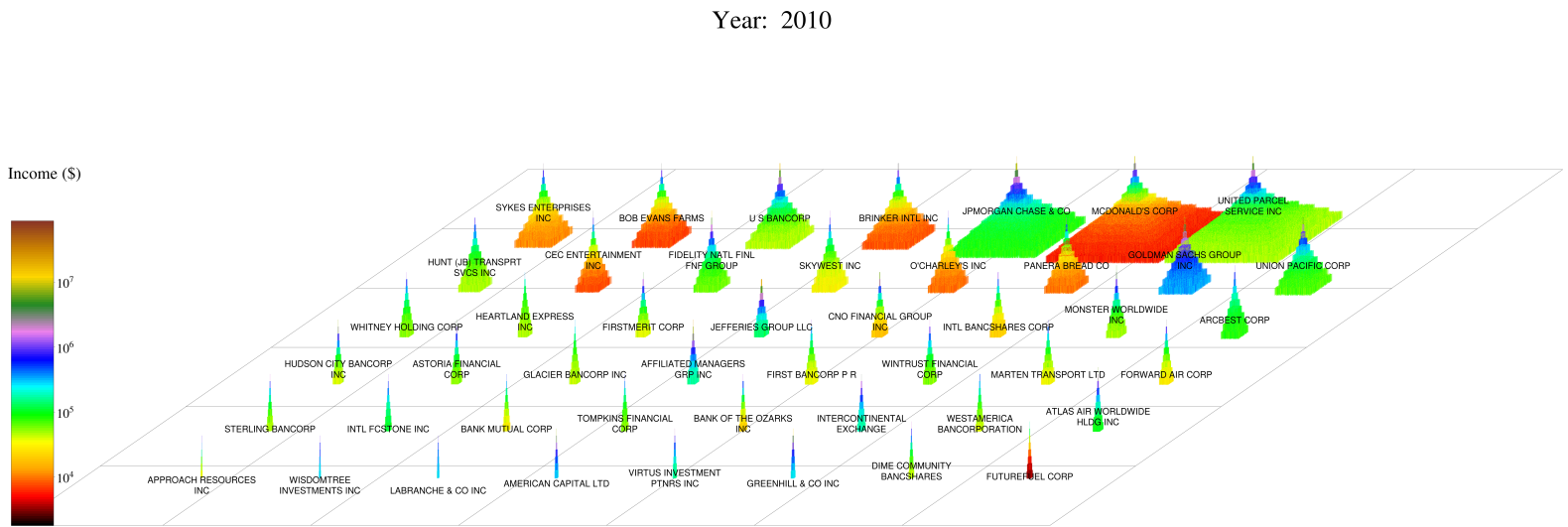
Year: 2010

Figure 3.9: Visualizing the Compustat Model

This figure visualizes the results of the Compustat Model for selected US firms in the year 2010. The data and method underlying this model are discussed in detail in the Appendix. Each pyramid represents a separate firm with volume proportional to total employment. The vertical axis corresponds to hierarchical level. Income is indicated by color.

3. Intra-level inequality tends to be *constant* across all hierarchical levels.

I use these regularities to construct a hierarchical model of the firm (see Appendix B.3). Given appropriate input data, this model can be used to estimate income inequality by hierarchical level (across firms). To make this estimate, I use the Compustat database, which provides the following data for 713 US firms over the period 1992–2015:

1. Number of Employees
2. Total Staff Expenses
3. CEO Pay

In conjunction with case-study regressions, this Compustat data can be used to estimate the hierarchical pay structure of individual US firms (see Appendix B.5 and F). While the details of the model are complex, the core idea is simple: since the CEO sits at the top of the corporate hierarchy, his/her *relative* pay (when compared to the average pay of all employees) gives an indication of the rate at which income increases by hierarchical level. When paired with assumptions about the 'shape' of the firm (derived from case-study regressions), the model gives an unambiguous prediction about firm internal pay structure. Results of the model are visualized in Figure 3.9 for selected firms in 2010.

The skeptical reader may be wondering why, after dismissing the case study data as not useful for testing hypothesis B, I nonetheless construct a model that hinges on this very data. The model is useful because the Compustat data (to which the model is fitted) adds a great deal of new information that is not contained within the case study data itself. The Compustat data adds a large number of US firms that exist over a continuous time-series, each having a different size, different mean pay, and different CEO pay ratio. While the case study data determines the hierarchical shape of all firms, the Compustat data determines everything else. In Appendix B.7 I analyze the sensitivity of this model to the case study data. I find that the key metric — the G_{BW} metric for income grouped by hierarchical level — is relatively robust to changes in case study data.

In addition to income distribution by hierarchical level, I also use the Compustat model to estimate the strength of the *firm*-income effect (how much working for different firms affects income). In this case, the Compustat database can be used to directly measure income inequality *between* firms, and the model is used to estimate inequality *within* each firm. I use this model-dependent data because I am not aware of any studies that directly measure internal income distributions of a large sample of firms.

Results

The results of the analysis of variance test of hypothesis B are shown in Figures 3.10 and 3.11. Figure 3.10 shows the between-within Gini ratio (G_{BW}) for our 19 different income-affecting factors. For all factors except religion and cognitive score, the boxplots indicate the variation of G_{BW} over time (typically the last 20 years). For religion, the boxplot range indicates uncertainty in the G_{BW} estimate, while for cognitive score, it indicates variation between different studies.

Figure 3.11 shows the same data, but in a slightly different format. The G_{BW} metric consists of a *ratio* of between-to-within group income dispersion (Eq. 3.8). Figure 3.11 decomposes this ratio and shows the individual components of the metric — between-group inequality (G_B) and within-group inequality ($\bar{G}_W$). Aside from religion, density plots indicate the distribution of these values over time (for religion, density plots indicate uncertainty). The important information here is the relative position of between-group inequality relative to within-group inequality.

This test of hypothesis B yields conclusive results: of the 19 different income-affecting factors tested, hierarchical level has the strongest effect on income. We can conclude that the available evidence supports hypothesis B: hierarchical power appears to affect income more strongly than any other factor. Interestingly, the Compustat model and the data from Mueller et al. and Heyman give G_{BW} ratios that are similar (although the underlying values of G_B and $\bar{G}_W$ are quite different). This may indicate that the strength of the hierarchy-income effect is consistent across countries that have different levels of inequality.

In addition to the support for the power-income hypothesis, Figures 3.10 and 3.11 reveal a few other notable findings. Firstly, physical attributes (age, cognitive score, race, and sex) have a relatively insignificant effect on income. Geographic effects are also quite small, although they become larger as the geographic area *decreases* (geographic factors ranked from largest to smallest area are: county, tract, block group).

Besides hierarchical level, only two other factors have G_{BW} ratios that are significantly greater than 1: labor vs. property income and full vs. part time. The latter is easily understandable: part-time individuals work significantly fewer hours than full-time individuals, so we would expect significant income differentials between the two groups. Added to this effect is the fact that part-time jobs are often in sectors such as retail that have lower wages than in sectors (like mining) where full-time employment is the norm.

But what should we make about the significant effect of *functional income*
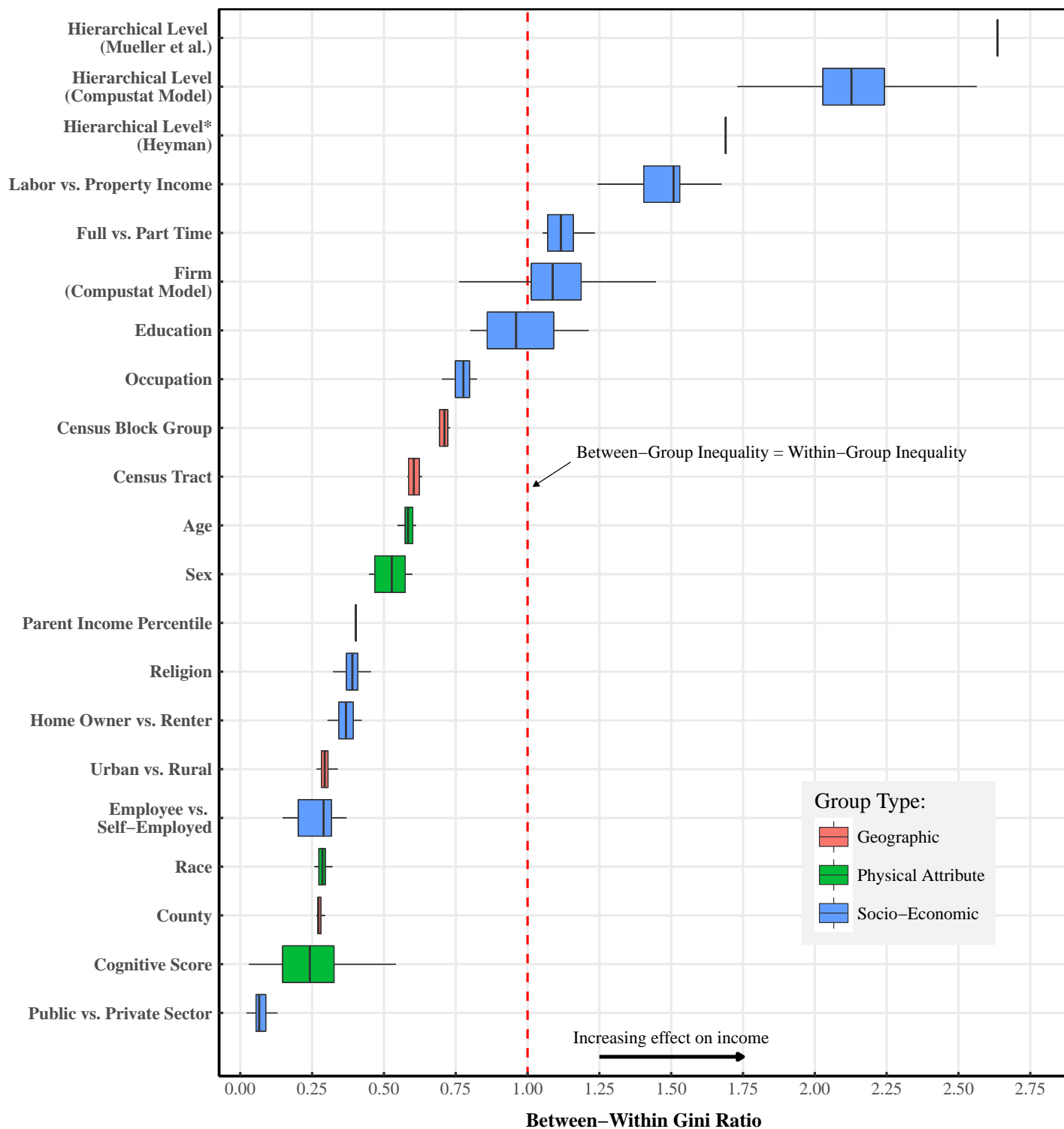
Figure 3.10: The G$_{BW}$ Ratio for Different Income-Affecting Factors

This figure shows the results of an analysis of variance test of hypothesis B using the method outlined in Sec. 3.4.2. According to this hypothesis, hierarchical power should affect income more strongly than any other factor. The horizontal axis shows the between-within Gini ratio (G_{BW}) defined by Eq. 3.8 (G_B is adjusted for bias using Eq. 3.10). A larger G_{BW} indicates a greater effect on income. The box plots indicate the total range (horizontal line), 25th to 75th percentile range (the box), and the median (vertical line). With the exception of hierarchical level data from Mueller et al. [107] and Heyman [108], all data is from the United States. For sources and methods, see Appendix B.1.

∗ Includes only top 4 hierarchical levels
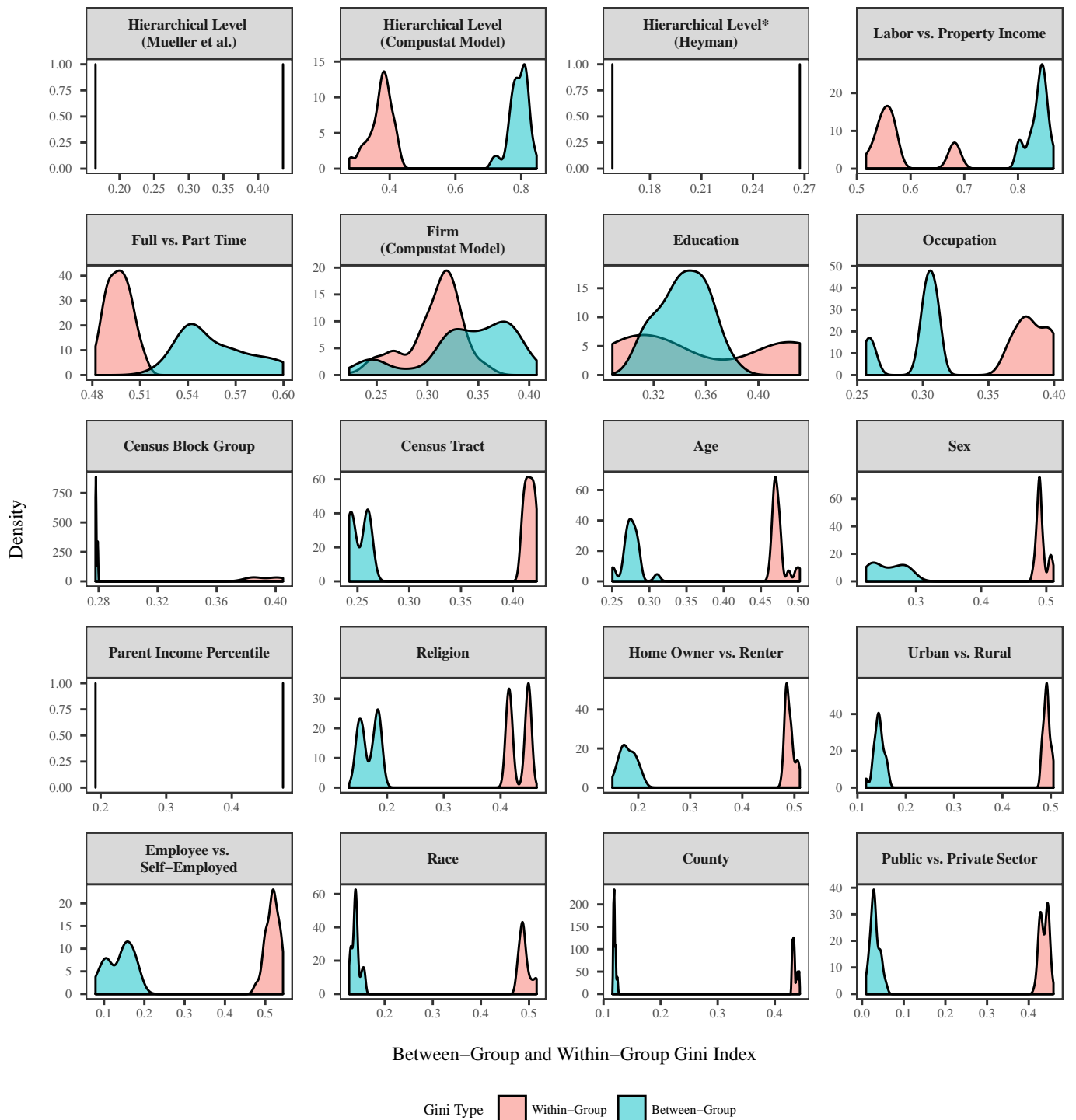
Figure 3.11: The G$_B$ and G$_W$ Index for Different Income-Effecting Factors
This figure shows the distribution of *between*-group Gini indexes (G_B, the Gini index of group mean incomes, shown in blue) in relation to the distribution of *within*-group Gini indexes ($\bar{G}_W$, average within-group inequality, shown in red). Each panel plots the results for a different income-affecting factor. With the exception of 'parent income percentile' and 'religion', the density curves represent the distribution of data over different years. Panels are sorted by effect size, declining (column-wise, then row-wise) from the top left to the bottom right. With the exception of hierarchical level data from Mueller et al. [107] and Heyman [108], all data is from the United States. For sources and methods, see Appendix B.1.

∗ Includes only top 4 hierarchical levels

*typ*e (property vs. labor)? At first glance, this may seem to support many political economists' (especially Marxists) deeply held convictions about functional income distribution: capitalists tend to be much wealthier than workers. While this may be true, the results shown here indicate something much different — that property income is on average *much less* than labor income.

This result is best thought of as an artifact of the US Census accounting method. In the Census data, 'property income' includes *anyone* with some form of dividend, interest, or rental income. The result is that the average property income is trivially small — about 8% of the average income from wages/salaries. This is because many people earn small amounts of property income in the form of interest on savings or dividends from small investments. Since these people likely earn income from other sources, a direct comparison of Census data for labor and property income has little meaning. However, I include it here for the sake of completeness.

To compare the income-effect of functional income type, what we really need to do is group individuals by the *proportion* of income coming from property sources. Based on the work of Piketty [3], it is reasonable to expect that this would strongly affect income. Piketty shows how the proportion of capitalist income increases with income fractile in the United States. But this grouping is the reverse of what would be required to apply my analysis of variance method. Piketty groups individuals by income size, while the method used here would require grouping individuals by the proportion of capitalist income. At present, I am not aware of the data sources that would allow such a grouping.

The evidence presented here demonstrates that grouping individuals by hierarchical level affects income more strongly than any of the other 18 factors tested. Of course we cannot rule out some as yet *unmeasured* factor that has a stronger effect on income, but this is the uncertain nature of empirical analysis.

3.5   Discussion

While science involves reductionism, income distribution theory has tended towards *greedy reductionism* — a term coined by Daniel Dennet. He writes: "in their eagerness for a bargain, in their zeal to explain too much too fast, [greedy reductionists] ... underestimate the complexities, trying to skip whole layers or levels of theory in their rush to fasten everything securely and neatly to the foundation" [109]. The two dominant theories of income distribution — neoclassical and Marxist — are *greedy reductionist*. They offer extremely simple principles that, on the face of it, are supposed to explain *everyone's income, all*

the time. Both posit a fundamental connection between income and productivity. The problem is that this productivist hypothesis seems to fail at the gate. When measured objectively, differences in productivity seem far too small to account for differences in income. Therefore, I believe it is time for a new hypothesis.

A natural tendency is to fight fire with fire — to reject one greedy reductionist theory and substitute another. If we followed this route, we might hypothesize that *power*, not productivity, explains everyone's income all the time. While this may well be true, it is not a useful way to do empirical science. In this paper, I have intentionally proposed a power theory of income distribution with *limited* scope. I focused only on power within firm hierarchies, and offered a very rigid definition of hierarchical power. My hypothesis was not that hierarchical power explains *all* variation in income, but rather, that hierarchical power explains *more* than any other factor.

With this narrower framework, the empirical evidence presented here is quite clear. Relative income within firms is strongly correlated with hierarchical power, and grouping individuals by hierarchical rank affects income more strongly than any other examined factor. This, I have claimed, is evidence for a power theory of income distribution. Of course, these results are contingent on the limited firm hierarchy data that is available. When better data comes along, results may change.

I want now to discuss the wider implications of a power theory of income distribution. An important question to ask is – what are the mechanisms that cause income to be correlated with hierarchical power? I doubt there is a simple answer, simply because there are is dizzying number of 'pathways to power' [110]. A hierarchical chain of command can function because each subordinate sincerely believes that their superior's power is *legitimate*. In this case, subordinates might simply agree that their superior should earn x amount more than them. But, as the history of slavery indicates, a hierarchy can also function through brutal repression. In this case, the income of superiors is an outcome of the judicious use of force. It seems plausible that the greater the inequality within a hierarchy, the more the chain of command functions via intimidation and fear rather than beliefs of legitimacy.

It is also plausible that belief in the legitimacy of the hierarchical system increases with hierarchical rank [111]. If this is true, it should manifest in opinions about income. Interestingly, a recent survey reveals that a majority of Americans question the legitimacy of CEO income [112]. Only 16% of the general public agree that CEOs are "paid the correct amount relative to the average worker". Yet a majority of Fortune 500 CEOs (64%) thought that CEO pay was 'correct'. It

would be fascinating to expand this type of survey to see if there is a gradient of opinion by hierarchical rank, and if this opinion changes as inequality increases.

Another complexity to the power-income relation is that firms are not islands unto themselves — there are power relations *between* institutions as well as within them [113]. Government regulation, for instance, can have a significant impact on CEO pay. CEOs in the US utility sector (which is highly regulated) have significantly lower pay than CEOs in other sectors (see Appendix B.6, as well as [114]). There is also evidence that CEO pay has a class-like cohesiveness. The average compensation of top US CEOs moves coherently with the capitalization of large firms [115]. This raises interesting implications for integrating the concept of hierarchical power with Nitzan and Bichler's [21] 'capital as power' hypothesis, in which capital is conceived as a symbolic representation of power. Because our state of knowledge on these matters is so dismal, the avenues for empirical research are quite expansive.

3.6 Conclusions

I want to conclude by reflecting on the philosophical underpinnings of income distribution theory itself. Discussion about income typically involves questions of *value*. What *value* does person *x* contribute to society? Is person *y* paid what they are *worth*? Conventional economic theory takes these questions seriously by proposing that people's contributions have *intrinsic* worth [90]. According to productivist theories, this worth corresponds to productivity. "*To each according to what he and the instruments he owns produces*", as Milton Friedman [116] famously wrote. For reasons that should be studied, most humans would likely agree that this is a *just* ethos.

But how strange this discussion looks when viewed from the outside. Would any serious biologists ever ask: "does an alpha male gorilla get the number of mates that he *deserves*?" I doubt it. To echo Alan Turing, the question is too meaningless to deserve discussion [117]. It is like asking if a species 'deserves' to go extinct. To put the matter harshly, the philosophical basis of a power theory of income distribution should be this: no one is *ever* 'paid what they are worth' because *intrinsic worth* is a scientifically meaningless concept. (But the fact that we like to frame the discussion of income distribution in terms of 'worth' and 'value' is an interesting indication of our belief system).

What matters for human resource distribution is agency and *beliefs* about agency. The range of possible human behavior is simply astounding, likely because there is a complex feedback relation between our beliefs and our ac-

tions. Beliefs affect behavior, and behavior affects beliefs. And this is as true of individuals as it is of entire societies. Thus, humans can live as egalitarian hunter-gatherers, as feudal caste societies, as brutal slave-owning societies, or as capitalist oligarchies. What changes is our beliefs about agency, and these beliefs change how we act. Certain beliefs encourage equality of agency (think of Boehm's reverse dominance hierarchy [118]), while others can allow agency to accumulate in the hands of the few (think of the doctrine of the Divine right of Kings). In the latter sense, agency becomes collective and institutional. It becomes concentrated *power*.

If we had to put a power theory of income distribution into mantra form, it would be this: '*To each according to his/her power to take*'. This ethos is profoundly unsettling, and for most people, likely *reviling*. But just because we find it vile, does not mean that it is *false*. Darwinian survival of the fittest is quite ethically vile, but nonetheless scientifically valid. This is not a doctrine of what our beliefs *should be*; rather, it is a hypothesis about the way that resources are actually distributed.

The corollary of this thinking is that the distribution of income has no natural state, and there are no natural laws that govern it. Distribution is an outcome of our collective belief structure, which can change over time. Moreover, if a power theory of income distribution is shown to be correct, then acts of income redistribution can be considered merely as *checks* on power — no different than the checks and balances that form the governmental basis of most liberal democracies.

References

1. Piketty T, Saez E. Income Inequality in the United States, 1913-1998 (series updated to 2000 available). National Bureau of Economic Research; 2001.

2. Piketty T, Saez E. The evolution of top incomes: a historical and international perspective. National Bureau of Economic Research; 2006.

3. Piketty T. Capital in the Twenty-first Century. Cambridge: Harvard University Press; 2014.

4. Atkinson AB, Piketty T. Top incomes: A global perspective. Oxford University Press; 2010.

5. Alvaredo F, Atkinson AB, Piketty T, Saez E. The top 1 percent in international and historical perspective. The Journal of Economic Perspectives. 2013;27(3):3–20.

6. Galton F. Hereditary genius: An inquiry into its laws and consequences. vol. 27. Macmillan; 1869.

7. Pareto V. Cours d'economie politique. vol. 1. Librairie Droz; 1897.

8. Sahota GS. Theories of personal income distribution: a survey. Journal of Economic Literature. 1978;16(1):1–55.

9. Champernowne DG. A model of income distribution. The Economic Journal. 1953;63(250):318–351.

10. Gibrat R. Les inegalites economiques. Recueil Sirey; 1931.

11. Rutherford RSG. Income distributions: a new model. Econometrica: Journal of the Econometric Society. 1955;23(3):277–294.

12. Simon HA. On a class of skew distribution functions. Biometrika. 1955;42(3/4):425–440.

13. Simon HA, Bonini CP. The size distribution of business firms. The American Economic Review. 1958;48(4):607–617.

14. Zipf GK. Human behavior and the principle of least effort. Cambridge: Addison-Wesley; 1949.

15. Angle J. The surplus theory of social stratification and the size distribution of personal wealth. Social Forces. 1986;65(2):293–326.

16. Chatterjee A, Chakrabarti BK. Kinetic exchange models for income and wealth distributions. The European Physical Journal B. 2007;60(2):135–149.

17. Chatterjee A, Chakrabarti BK, Chakraborti A. Econophysics and sociophysics: trends and perspectives. Milan: John Wiley & Sons; 2007.

18. Pianegonda S, Iglesias JR, Abramson G, Vega JL. Wealth redistribution with conservative exchanges. Physica A: Statistical Mechanics and its Applications. 2003;322:667–675.

19. Gallegati M, Keen S, Lux T, Ormerod P. Worrying trends in econophysics. Physica A: Statistical Mechanics and its Applications. 2006;370(1):1–6.

20. Ricardo D. Principles of political economy and taxation. London: G. Bell and Sons; 1817.

21. Nitzan J, Bichler S. Capital as Power: A Study of Order and Creorder. New York: Routledge; 2009.

22. Clark JB. The Distribution of Wealth. New York: Macmillan; 1899.

23. Wicksteed PH. An Essay on the Co-ordination of the Laws of Distribution (1932 Edition). London: London School of Economics; 1894.

24. Becker GS. Investment in human capital: A theoretical analysis. Journal of political economy. 1962;70(5, Part 2):9–49.

25. Mincer J. Investment in human capital and personal income distribution. The journal of political economy. 1958;66(4):281–302.

26. Schultz TW. Investment in human capital. The American Economic Review. 1961;51(1):1–17.

27. Marx K. Capital, Volume I. Harmondsworth: Penguin/New Left Review; 1867.

28. Rubin II. Essays on Marx's theory of value. Black Rose Books Ltd.; 1973.

29. Boissevain CH. Distribution of abilities depending upon two or more independent factors. Metron. 1939;13(4):49–58.

30. Mandelbrot B. The Pareto-Levy law and the distribution of income. International Economic Review. 1960;1(2):79–106.

31. Hunter JE, Schmidt FL, Judiesch MK. Individual differences in output variability as a function of job complexity. Journal of Applied Psychology. 1990;75(1):28.

32. Cohen AJ, Harcourt GC. Retrospectives: Whatever happened to the Cambridge capital theory controversies? Journal of Economic Perspectives. 2003;17(1):199–214.

33. Robinson J. The production function and the theory of capital. The Review of Economic Studies. 1953;21(2):81–106.

34. Pullen J. The marginal productivity theory of distribution: a critical history. London: Routledge; 2009.

35. Abowd JM, Kramarz F, Margolis DN. High wage workers and high wage firms. Econometrica. 1999;67(2):251–333.

36. Haltiwanger JC, Lane JI, Spletzer JR. Productivity differences across employers: The roles of employer size, age, and human capital. The American Economic Review. 1999;89(2):94–98.

37. Haskel J, Hawkes DD, Pereira SC. Skills, human capital and the plant productivity gap: UK evidence from matched plant, worker and workforce data. CEPR Discussion Paper. 2005;(5334).

38. Hellerstein JK, Neumark D, Troske KR. Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. National Bureau of Economic Research; 1996. Working Paper No. 5626.

39. Hoegeland T. Do higher wages reflect higher productivity? Education, gender and experience premiums in a matched plant-worker data set. Contributions to Economic Analysis. 1999;241:231–260.

40. Iranzo S, Schivardi F, Tosetti E. Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data. Journal of Labor Economics. 2008;26(2):247–285.

41. Oulton N. Competition and the dispersion of labour productivity amongst UK companies. Oxford Economic Papers. 1998;50(1):23–38.

42. Barroso FG, Alados CL, Boza J. Social hierarchy in the domestic goat: effect on food habits and production. Applied Animal Behaviour Science. 2000;69(1):35–53.

43. Guhl AM, Collias NE, Allee WC. Mating behavior and the social hierarchy in small flocks of white leghorns. Physiological Zoology. 1945;18(4):365–390.

44. Kondo S, Hurnik JF. Stabilization of social hierarchy in dairy cows. Applied Animal Behaviour Science. 1990;27(4):287–297.

45. Meese GB, Ewbank R. The establishment and nature of the dominance hierarchy in the domesticated pig. Animal Behaviour. 1973;21(2):326–334.

46. Sapolsky RM. The influence of social hierarchy on primate health. Science. 2005;308(5722):648–652.

47. Uhrich J. The social hierarchy in albino mice. Journal of Comparative Psychology. 1938;25(2):373.

48. Bradley BJ, Robbins MM, Williamson EA, Steklis HD, Steklis NG, Eckhardt N, et al. Mountain gorilla tug-of-war: silverbacks have limited control over reproduction in multimale groups. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(26):9418–9423.

49. Haley MP, Deutsch CJ, Le Boeuf BJ. Size, dominance and copulatory success in male northern elephant seals, Mirounga angustirostris. Animal Behaviour. 1994;48(6):1249–1260.

50. Girman DJ, Mills MGL, Geffen E, Wayne RK. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (Lycaon pictus). Behavioral Ecology and Sociobiology. 1997;40(3):187–198.

51. Gerloff U, Hartung B, Fruth B, Hohmann G, Tautz D. Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (Pan paniscus) determined from DNA analysis of faecal samples. Proceedings of the Royal Society of London B: Biological Sciences. 1999;266(1424):1189–1195.

52. Wroblewski EE, Murray CM, Keele BF, Schumacher-Stankey JC, Hahn BH, Pusey AE. Male dominance rank and reproductive success in chimpanzees, Pan troglodytes schweinfurthii. Animal Behaviour. 2009;77(4):873–885.

53. Frankel DG, Arbel T. Group formation by two-year olds. International Journal of Behavioral Development. 1980;3(3):287–298.

54. Savin-Williams RC. Dominance hierarchies in groups of middle to late adolescent males. Journal of Youth and Adolescence. 1980;9(1):75–85.

55. Strayer FF, Trudel M. Developmental changes in the nature and function of social dominance among young children. Ethology and Sociobiology. 1984;5(4):279–295.

56. Hopcroft RL. Sex, status, and reproductive success in the contemporary United States. Evolution and Human Behavior. 2006;27(2):104–120.

57. Betzig L. Means, variances, and ranges in reproductive success: comparative evidence. Evolution and Human Behavior. 2012;33(4):309–317.

58. Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M, et al. Associations with early-life socio-economic position in adult DNA methylation. International journal of epidemiology. 2012;41(1):62–74.

59. Massart R, Suderman MJ, Nemoda Z, Sutti S, Ruggiero AM, Dettmer AM, et al. The Signature of Maternal Social Rank in Placenta Deoxyribonucleic Acid Methylation Profiles in Rhesus Monkeys. Child Development. 2016;00(0):1–19.

60. Boric D. Social dimensions of mortuary practices in the Neolithic: a case study. Starinar. 1996;(47):67–83.

61. Halstead P. Spondylus shell ornaments from late Neolithic Dimini, Greece: specialized manufacture or unequal accumulation? Antiquity. 1993;67(256):603–609.

62. Van der Velde P. Banderamik social inequality—a case study. Germania. 1990;68(1):19–38.

63. Aranda G, Molina F. Wealth and Power in the Bronze Age of the South-East of the Iberian Peninsula: The Funerary Record of Cerro De La Encina. Oxford Journal of Archaeology. 2006;25(1):47–59. doi:10.1111/j.1468-0092.2006.00248.x.

64. Aranda-Jiménez G, Montón-Subías S, Jiménez-Brobeil S. Conflicting evidence? Weapons and skeletons in the Bronze Age of south-east Iberia. Antiquity. 2009;83(322):1038–1051.

65. Graziadio G. The process of social stratification at Mycenae in the Shaft Grave period: a comparative examination of the evidence. American journal of archaeology. 1991;95(3):403–440.

66. Kristiansen K. Europe before history. New York: Cambridge University Press; 2000.

67. Harding AF. European societies in the Bronze Age. Cambridge: Cambridge University Press; 2000.

68. Dee M, Wengrow D, Shortland A, Stevenson A, Brock F, Flink LG, et al. An absolute chronology for early Egypt using radiocarbon dating and Bayesian statistical modelling. Proc R Soc A. 2013;469:20130395.

69. Reade J. Assyrian king-lists, the royal tombs of Ur, and Indus origins. Journal of Near Eastern Studies. 2001;60(1):1–29.

70. Guo Z, Liu K, Lu X, Ma H, Li K, Yuan S, et al. The use of AMS radiocarbon dating for Xia–Shang–Zhou chronology. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms. 2000;172(1):724–731.

71. Mosca G. On the Ruling Class. New York: McGraw Hill; 1939.

72. Smith A. An Inquiry into the Nature and Causes of the Wealth of Nations. Edinburgh: A. and C. Black; 1776.

73. Jevons WS. The theory of political economy. London: Macmillan and co.; 1879.

74. Menger C. Principles of Economics. New York: University Press; 1871.

75. Walras L. Elements d'economie politique pure, ou, Theorie de la richesse sociale. F. Rouge; 1896.

76. Orwell G. The freedom of the press. Animal Farm: The Centennial Edition. 1972; p. 98–99.

77. Berle AA, Means GGC. The modern corporation and private property. Transaction Books; 1932.

78. Brown C. Income Distribution in an Institutional World. In: Mangum G, Philips P, editors. Three Worlds of Labor Economics. vol. 51; 1988. p. 515–530.

79. Commons JR. Legal foundations of capitalism. Transaction Publishers; 1924.

80. Galbraith JK. The anatomy of power. Boston, Mass.: Houghton Mifflin Co.; 1985.

81. Huber E, Huo J, Stephens JD. Power, policy, and top income shares. Socio-Economic Review. 2017;0(0):1–23. doi:10.1093/ser/mwx027.

82. Lenski GE. Power and privilege: A theory of social stratification. UNC Press Books; 1966.

83. Mills CW. The power elite. Oxford University Press; 1956.

84. Peach JT. Distribution and economic progress. Journal of Economic Issues. 1987;21(4):1495–1529.

85. Sidanius J, Pratto F. Social dominance: An intergroup theory of social hierarchy and oppression. Cambridge University Press; 2001.

86. Veblen T. The Theory of Business Enterprise. New York: Martino Fine Books; 1904.

87. Veblen T. Absentee ownership: business enterprise in recent times: the case of America. Transaction Pub; 1923.

88. Weber M. Economy and society: An outline of interpretive sociology. Univ of California Press; 1978.

89. Wright EO. Class structure and income determination. vol. 2. New York: Academic Press; 1979.

90. Brown C. Is there an institutional theory of distribution? Journal of Economic Issues. 2005;39(4):915–931.

91. Simon HA. The compensation of executives. Sociometry. 1957;20(1):32–35.

92. Lydall HF. The distribution of employment incomes. Econometrica: Journal of the Econometric Society. 1959;27(1):110–115.

93. Verdier N. Hierarchy: a short history of a word in Western thought. In: Hierarchy in natural and social sciences. Springer; 2006. p. 13–37.

94. Mones E, Vicsek L, Vicsek T. Hierarchy measure for complex networks. PloS one. 2012;7(3):e33799.

95. Rajan Raghuram G, Luigi Z, Rajan Raghuram G, Luigi Z. The firm as a dedicated hierarchy: A theory of the origins and growth of firms. The Quarterly Journal of Economics. 2001;116(3):805–851.

96. Audas R, Barmby T, Treble J. Luck, effort, and reward in an organizational hierarchy. Journal of Labor Economics. 2004;22(2):379–395.

97. Baker G, Gibbs M, Holmstrom B. Hierarchies and compensation: A case study. European Economic Review. 1993;37(2-3):366–378.

98. Dohmen TJ, Kriechel B, Pfann GA. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. Journal of Population Economics. 2004;17(2):193–228.

99. Lima F. Internal Labor Markets: A Case Study. FEUNL Working Paper. 2000;378.

100. Morais F, Kakabadse NK. The Corporate Gini Index (CGI) determinants and advantages: Lessons from a multinational retail company case study. International Journal of Disclosure and Governance. 2014;11(4):380–397.

101. Treble J, Van Gameren E, Bridges S, Barmby T. The internal economics of the firm: further evidence from personnel data. Labour Economics. 2001;8(5):531–552.

102. Fleishman AI. Confidence intervals for correlation ratios. Educational and Psychological Measurement. 1980;40(3):659–670.

103. Steiger JH. Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. Psychological methods. 2004;9(2):164–182.

104. Deltas G. The small-sample bias of the Gini coefficient: results and implications for empirical research. Review of economics and statistics. 2003;85(1):226–234.

105. Keynes JM. Professor Tinbergen's method. The Economic Journal. 1939;49(195):558–577.

106. Nitzan J. Inflation as restructuring. A theoretical and empirical account of the US experience. McGill University; 1992.

107. Mueller HM, Ouimet PP, Simintzi E. Within-Firm Pay Inequality. SSRN Working Paper. 2016;.

108. Heyman F. Pay inequality and firm performance: evidence from matched employer–employee data. Applied Economics. 2005;37(11):1313–1327.

109. Dennett DC. Darwin's Dangerous Idea: Evolution and the Meanings of Life. London: Penguin Books; 1995.

110. Price TD, Feinman GM, editors. Pathways to power: New Perspectives on the Emergence of Social Inequality. New York: Springer; 2010.

111. Schmidt J. Disciplined minds: A critical look at salaried professionals and the soul-battering system that shapes their lives. Rowman & Littlefield; 2001.

112. Larcker D, Thompson JT, Donatiello H, Tayan B. CEOs and Directors on Pay: 2016 Survey on CEO Compensation. Stanford University: Rock Center for Corporate Governance; 2016.

113. Bichler S, Nitzan J. Growing Through Sabotage: Energizing Hierarchical Power. Capital as Power Working Papers. 2017;2017(02).

114. Joskow P, Rose N, Shepard A, Meyer JR, Peltzman S. Regulatory constraints on CEO compensation. Brookings Papers on Economic Activity Microeconomics. 1993;1993(1):1–72.

115. Mishel L, Davis A. CEO pay continues to rise as typical workers are paid less. Issue Brief. 2014;380.

116. Friedman M. Capitalism and Freedom. Chicago: University of Chicago press; 1962.

117. Turing AM. Computing machinery and intelligence. Mind. 1950;59(236):433.

118. Boehm C, Barclay HB, Dentan RK, Dupre MC, Hill JD, Kent S, et al. Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. Current Anthropology. 1993;34(3):227–254.

Chapter 4

A Hierarchy Model of Income Distribution

Abstract

Based on worldly experience, most people would agree that firms are hierarchically organized, and that pay tends to increase as one moves up the hierarchy. But how this hierarchical structure affects income distribution has not been widely studied. To remedy this situation, this paper presents a new model of income distribution that explores the effects of social hierarchy. This 'hierarchy model' takes the limited available evidence on the structure of firm hierarchies, and generalizes it to create a large-scale simulation of the hierarchical structure of the United States economy. Using this model, I conduct the first quantitative investigation of hierarchy's effect on income distribution. I find that hierarchy plays a dominant role in shaping the tail of US income distribution. The model suggests that hierarchy is responsible for generating the power-law scaling of top incomes. Moreover, I find that hierarchy can be used to unify the study of personal and functional income distribution, as well as to understand historical trends in income inequality.

4.1 Introduction

The field of income distribution modeling is in need of new ideas. Ever since Pareto [1] discovered the power law scaling of top incomes and wealth, theorists have sought generative models for creating income distributions. In this regard, the field has been wildly successful. An impressive array of models now exist that can generate, from simple principles, observed distributions of income [2–21]. The problem is that this outward empirical success masks underlying assumptions that often have little to do with reality. To echo Leontief, "what is really needed, in most cases, is a very difficult and seldom very neat assessment and verification of these assumptions in terms of observed facts" [22].

This paper seeks to take the field of income distribution modeling in a new direction. In my view, the goal of an income distribution model should *not* be to generate income distributions from first principles. Such models are best left to physics, where there are actual 'first principles' (the laws of physics). Instead, a good income distribution model should be a tool for *dissecting* income distributions. A good model should be a tool for making generalizations from scattered and piecemeal observations of the real-world. A good model should be a tool for understanding connections to other branches of theory. A good model should be a tool for unifying ideas, and for understanding history.

A Focus on Hierarchy

In this paper, I build and test a model with the explicit purpose of understanding how *social hierarchy affects income distribution*. I am interested in hierarchy for a number of reasons. First, the use of hierarchy to distribute resources is ubiquitous among social animals. Virtually all social animals form dominance hierarchies, or 'pecking orders' [23–28]. Among such animals, hierarchical rank plays a key role in gaining access to resources, particularly sexual mates [29–33]. Given our evolutionary heritage, it seems quite reasonable to hypothesize that hierarchy plays a role in shaping resource distribution among humans.[1]

A second reason for my interest in hierarchy is that it offers a simple way of studying the class structure of society. Many social scientists have proposed that income distribution is connected to class structure [37–47]. However, there is no consensus on what, exactly, a 'class' is. Nor is there agreement on which classes are important for shaping income distribution. Hierarchy is useful for studying class structure because it is abstract and generalizable. A hierarchy is really just a particular form of *network* — one that has a tree-like structure [48]. Because human hierarchies represent a chain of command, they offer a natural way of grouping individuals by authority (or what I call *hierarchical power*). Do individuals with more hierarchical power earn more money? How does this affect income distribution? These are questions that I seek to answer.

---

[1] Lewontin and Levins note that "struggles for legitimacy between political ideologies eventually come down to struggles over what constitutes human nature" [34]. Although I think the human proclivity for forming hierarchies likely has an evolutionary basis, this does not mean that I think that hierarchy represents a 'natural order'. Rather, the evolutionary evidence simply suggests that we have an *instinct* for forming hierarchy. This says nothing about how society ought to be. Much of what we consider social progress consists of *suppressing* instinctual behavior. The decline of human violence from the evolutionary background rate is perhaps the best example of such progress [35, 36].

Lastly, I am interested in hierarchy because it is conspicuously *absent* from mainstream theory, and thus its role in shaping income distribution is poorly understood. The vast majority of income distribution models are atomistic — they focus solely on individuals. I believe this approach is misguided. While it would be a triumph of science if we could explain complex social structure in terms of the actions of individuals, we are very far from this goal. This 'bottom-up' approach requires a highly accurate model of human behavior — something that we are hopelessly far from having.

The Model

The hierarchy model that I construct in this paper is a different sort of beast than the typical economic model. The hierarchy model is *not* built on micro principles. It is *not* dynamic, and it is *not* agent based. Instead, it is a tool of necessity. There is simply too little empirical evidence about how hierarchy shapes income to draw conclusions directly from the data. I use the hierarchy model as a tool for making generalizations from the scattered evidence that does exist. It is essentially an *extrapolation* (albeit a complex one). The model fits trends to a small sample of firm-level data, and then generalizes these trends to create a large-scale simulation of the hierarchical structure of the United States economy. This simulated data can then be used to study how hierarchy affects income.

Goals

I use the hierarchy model to pursue two goals — one that is quite modest and one that is admittedly bold. The first (modest) goal, is to quantify the role that hierarchy plays in shaping income distribution in the United States. The second, admittedly bold, goal is to use hierarchy as a *unification mechanism*. I investigate how hierarchy can be used to unify both the study of personal and functional income, and our understanding of historical trends in income inequality.

Summary of Findings

The general finding in this paper is that the hierarchy model provides a rich framework (no pun intended) for understanding the behavior of top incomes. The hierarchy model explains why US income distribution has a power law tail, and it provides a tantalizing way of linking personal and functional income dis-

tribution. Lastly, hierarchical redistribution seems to be a fruitful way to understand historical changes in top income shares.

Key Results

1. Hierarchy links top incomes (and wealth) to large institutions: Top earning US executives, as well as the wealthiest Americans, work for (or own) firms that are much larger than those of the general population. The hierarchy model reproduces this effect.

2. Hierarchy shapes top incomes.   The model demonstrates a clear division between the *body* and the *tail* of the income distribution. The *body* of the distribution is primarily determined by between-firm income dispersion. However, the tail of the distribution is almost completely determined by hierarchy. The model reproduces the power law scaling of the top 1% of US incomes. I show that this is purely an effect of hierarchy.

3. Hierarchy links personal and functional income: Building on the work of Nitzan and Bichler [49], I test the hypothesis that being a 'capitalist' is a function of hierarchical power. Specifically, I propose that the fraction of income coming from capitalist sources scales with hierarchical power. A model implementing this hypothesis accurately predicts how capitalist income share increases with income size in the United States. The same model also reproduces the size distribution of US capitalist income, as well as the capitalist share of national income.

4. Changes in hierarchical pay explain historical changes in inequality.   I test the hypothesis that the recent explosion in US top income shares can be explained in terms of differential gains to hierarchical rank and power. By varying the rate at which income scales with hierarchical rank, I am able to use the model to reproduce historical trends. The model is able to replicate not only the increasing share of the top 1%, but also the increasing pay of top CEOs. The same model, when used in tandem with the capitalist gradient hypothesis, is able replicate (with 75% accuracy) the observed relation between US top income share and the dividend share of national income.

Paper Layout

The remainder of the paper is divided into three sections. In section 4.2, I review the basic characteristics of the hierarchy model. (A detailed, technical discussion of the model's algorithm can be found in the appendix). I then test the model against various aspects of US income distribution. Having confirmed that the hierarchy model gives sound results, I use it to estimate how hierarchy affects US income distribution. In section 4.3, I investigate if hierarchy can be used to unify the study of personal and functional income distribution. In section 4.4, I investigate if hierarchy can be used to unify our understanding of historical trends in income inequality.

4.2 A Hierarchy Model

The hierarchy model is based on the hypothesis that human institutions are hierarchically organized, and that hierarchical power (authority over subordinates) plays a key role in determining income. While this hypothesis is quite radical by the standards of neoclassical economics, I am certainly not the first scholar to suspect that power plays a role in income distribution (see, for instance [37, 42, 45, 49–61]).

The starting point for my approach is the seminal work of Herbert Simon [62] and H.F Lydall [63]. In the late 1950s, Simon and Lydall both developed simple models that focused on the branching structure of firm hierarchies. The distinguishing feature of a branching hierarchy is that each superior has control over *multiple* subordinates (see Fig. 4.1). This feature is important because it distinguishes human hierarchies from the linear dominance hierarchies (pecking orders) seen in animals. Within a linear hierarchy, there are as many ranks as there are individuals. Consequently, there is no class structure. However, a branching hierarchy naturally leads to a pyramid-shaped class system based on hierarchical rank.

Simon and Lydall both showed how branching hierarchical structure could explain regularities in income distribution. Simon used a simple hierarchical model of the firm to explain the observed scaling between CEO pay and firm sales [64]. Lydall showed how firm hierarchy could lead to a power law distribution of top incomes. Although promising, it seems that this work was largely ignored by the economics profession. Soon after these papers were published, human capital theory became the prevailing orthodoxy in personal income distribution
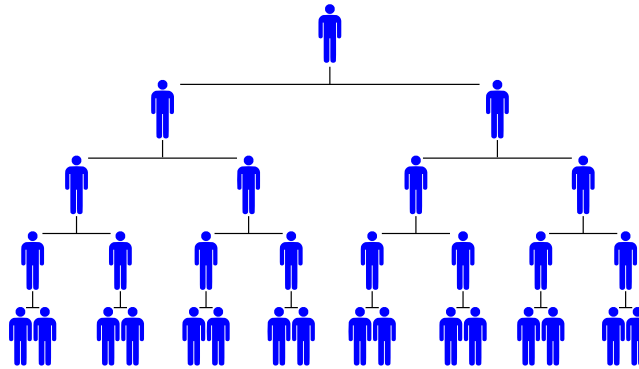
Figure 4.1: A Branching Hierarchy

This figure shows an idealized branching hierarchy in which each superior has two subordinates. This superior/subordinate ratio — often called the span of control — can be used to mathematically describe the hierarchy. Starting from the bottom rank, each consecutive rank decreases in size by a factor of the span of control. Evidence from real-world firms suggests that the span of control is not constant by rank, but instead tends to increase as one moves up the hierarchy (see Appendix C.2).

theory [65–67]. As a result, little work was done to explore the consequences of hierarchical organization.

This paper draws on the work of Simon and Lydall, but updates their model in light of recent empirical work. Both Simon and Lydall assumed a constant span of control. (The span of control is the number of subordinates per superior). Case study evidence (discussed in Appendix C.2) indicates that the span of control is *not* constant. Rather, it tends to increase as one moves up the hierarchy. Simon and Lydall also assumed a constant ratio of average income between adjacent hierarchical ranks. Again, case study evidence suggests that this is not quite true. Like the span of control, the pay ratio between ranks also tends to increase as one moves up the hierarchy.

Another key feature of my approach is that I take full advantage of modern computational power to build a large-scale, stochastic simulation. In contrast, Simon and Lydall used simple analytic methods. Simulation allows investigation that would otherwise be impossible with a purely analytic approach.

4.2.1   Modeling Goals and Methods

As I stated in the introduction, the goal of my modeling effort is not to generate a distribution of income from first principles. Instead, the model is designed to be a *surrogate* for data that does not exist. What do I mean by this? After scouring the scientific literature, I have been able to find only a handful of studies that document, with sufficient detail, the hierarchical structure of real-world firms (see Appendix C.2). This paucity of data likely owes to two things. Firstly, the discipline of economics is generally disinterested in hierarchy and power, so there is little incentive to do empirical work on this topic. Secondly, firm employment and payroll data is largely proprietary, meaning it is simply not available to researchers unless they have an inside connection.

This lack of data means that it is virtually impossible to study the general effects of hierarchy on income distribution solely by using the available case-study evidence. The hierarchy model is designed to generate data that I *wish* was available directly. The model takes the scant data that does exist, and fits trends (and parameterized distributions) to it. I then use the model to extrapolate these trends to a large-scale simulation of the economy. The resulting model is *entirely* dependent on the input, firm-level data. I do not tune the model to reproduce macro level results. The model output is purely what is implied by generalizing the trends found in input data.

The model is built on a tripartite income classification scheme that allows for three sources of income dispersion (see Fig. 4.2) :

Source 1: Income dispersion *between* hierarchical levels of each firm

(*inter*-hierarchical dispersion);

Source 2: Income dispersion *within* hierarchical levels of each firm

(*intra*-hierarchical dispersion);

Source 3: Income dispersion *between* different firms

(*inter*-firm dispersion).

Inter-firm and intra-hierarchical level dispersion are not explained by the model. (In the jargon of economic modeling, these dispersion sources are *exogenous*). In contrast, inter-hierarchical dispersion is *partially* explained by the model. It is explained in the sense that it is not *ex nihilo* — this dispersion does not come from nowhere. The model contains firms that have a specific hierarchical structure of employment and pay. However, the reason for this hierarchical structure is *not* explained by the model. Rather, hierarchical structure is determined
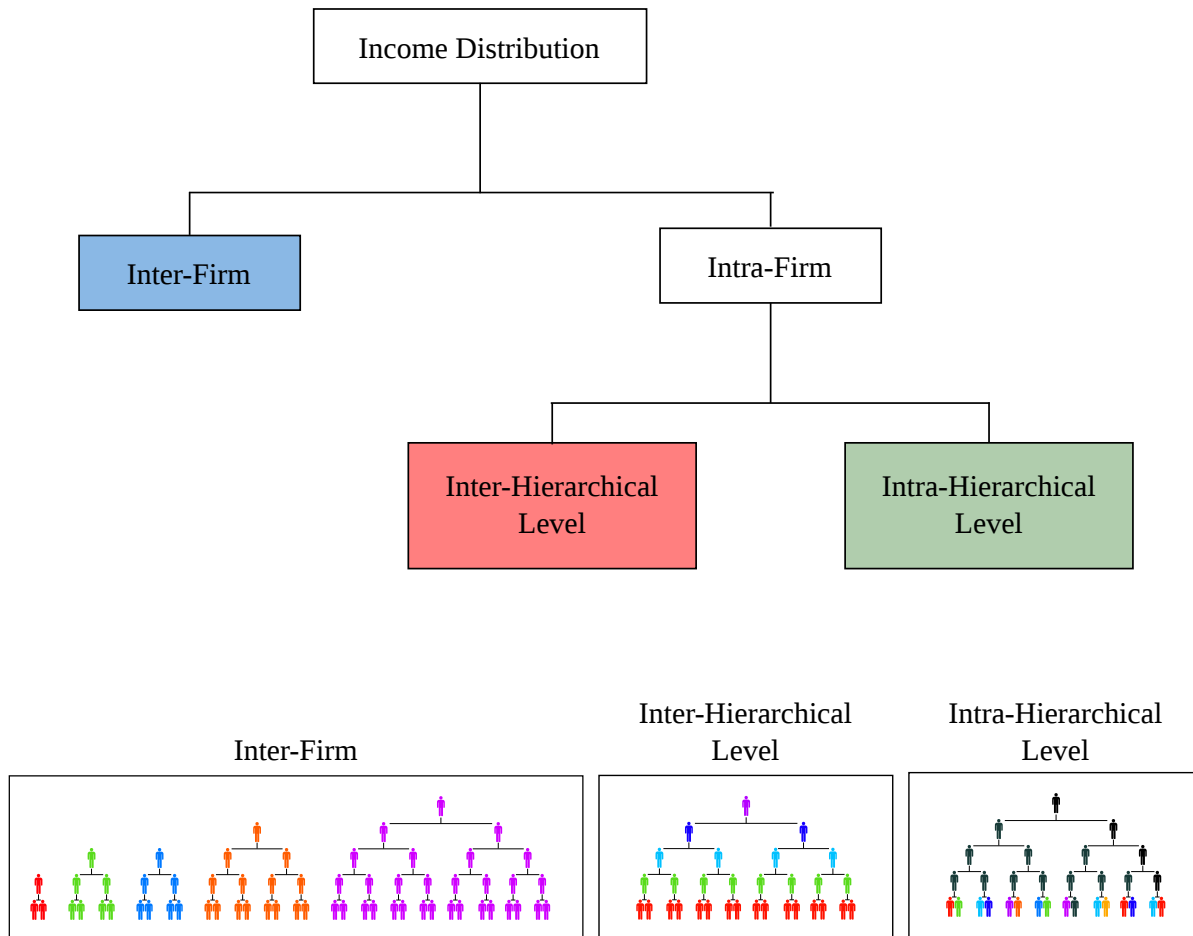
Figure 4.2: A Tripartite Division of Income Distribution

This figure illustrates the income distribution grouping scheme used by the hierarchy model. The model allows for three sources of income dispersion. *Inter-firm* dispersion consists of differences in (average) pay between firms. Within each firm, there are two further sources of dispersion. *Inter-hierarchical* level dispersion consists of differences in (average) pay between hierarchical levels, while *intra-hierarchical* level dispersion consists of differences in pay within each hierarchical level.

---

from regressions on case study data, in conjunction with firm-level data from the Compustat and Execucomp databases.

Modeling the United States

The model is designed to study the hierarchical structure of the US economy as it was (on average) over the years 1992-2015. At the highest level of abstraction, the model has three parts. First, the model creates a firm size distribution

that dictates how many firms of a given size will exist. Second, for each firm in this distribution, the model creates a hierarchical structure. This means the model determines how many ranks will exist, and how many individuals will occupy each hierarchical rank. Lastly, the model uses each of the three dispersion sources (outlined above) to stochastically generate an income for every individual in every firm. In a sense, everything else amounts to details about how each of these steps is carried out. I review here the most important elements of each step. A technical discussion can be found in the Appendix.

Step 1: Create a Firm Size Distribution. The first step of the model is to generate a distribution of firm sizes. The available evidence suggests that national firm size distributions can be modeled by a power law [68–70]. Under this assumption, the probability of finding a firm of size x is proportional to $x^{-\alpha}$, where α is a constant. I model the United States firm size distribution with 1 million firms distributed according to a discrete power law distribution with exponent $\alpha = 2.01$ (see Appendix C.5).

Step 2: Endow Firms with Hierarchical Structure. The hierarchy model captures only the *aggregate* hierarchical structure of firms. That is, I model the number of employees in each hierarchical level, not the exact chain of command. I base the model on a number of recent case studies that have documented the aggregate hierarchical structure of firms in various developed countries (see Appendix C.2). From this data, I make generalizations about the hierarchical structure of firms. The evidence suggests that the span of control (the ratio between adjacent hierarchical levels) increases with rank. I model this increase with an exponential function.

For simplicity, all firms in the model have the same hierarchical structure — that is, they are governed by the same span of control function. However, since there is a great deal of uncertainty in this function, I run the model many times. Each different model run uses a slightly different span of control function, determined by resampling from case study data. The result is that the hierarchical structure of firms varies stochastically between different model runs, allowing us to capture uncertainty in the underlying empirical data. For more details, see Appendix C.4 and C.5.

Step 3: Endow Individuals with Income After each firm has a hierarchical structure, we begin the most important part of the model, which is to assign ev-

ery individual an income. Because the model has three dispersion mechanisms, this last step has three components, outlined below.

Step 3A: Generate Inter-Hierarchical Level Dispersion.   In the model, hierarchical pay is constructed from the bottom up. Starting from the bottom rank, I define a function that determines the rate at which pay increases by hierarchical rank. This function is informed by case study data (see Appendix C.2). Unlike hierarchical employment structure, each modeled firm is given a *different* hierarchical pay structure. The process of assigning different hierarchical pay structure to each firm is heavily informed by firm-level data in the Compustat database. (See Appendix C.3 for a detailed discussion of the Compustat data).

The basic idea is this: before running the full simulation, I fit the hierarchy model to Compustat data for real-world American firms. Compustat (in conjunction with Execucomp) provides data on CEO pay, average pay, and firm employment. Assuming the CEO occupies the top hierarchical level, we can use this information to model the hierarchical pay structure of each Compustat firm. Once this is complete, we have an indication of how hierarchical pay should vary across firms. The model's main simulation is then informed by this variation. The result is a unique hierarchical pay structure for each firm. For more details, see Appendix C.4 and C.5.

Step 3B: Generate Inter-Firm Dispersion.   I create inter-firm income dispersion by varying (average) pay in the bottom hierarchical level of each firm. This variation is informed by firm-level data in the Compustat database. As discussed in Step 3A, prior to running a full-scale simulation, I fit the model to firms in the Compustat database. After having fit hierarchical pay, I use this information to estimate how base-level pay varies across these firms. This variation then informs the model's main simulation. For more details, see Appendix C.4 and C.5.

Step 3C: Generate Intra-Hierarchical Level Dispersion.   The last step is to model the income dispersion *within* the hierarchical levels of each firm. The available case study evidence suggests that income dispersion within hierarchical levels is roughly constant across all hierarchical levels (see Appendix C.2). To simplify the model, I further assume that intra-hierarchical level dispersion is constant across all firms. Informed by case study data, I use a single parameterized distribution to randomly generate income dispersion within all hierarchical levels of every firm. For more details, see Appendix C.4 and C.5.
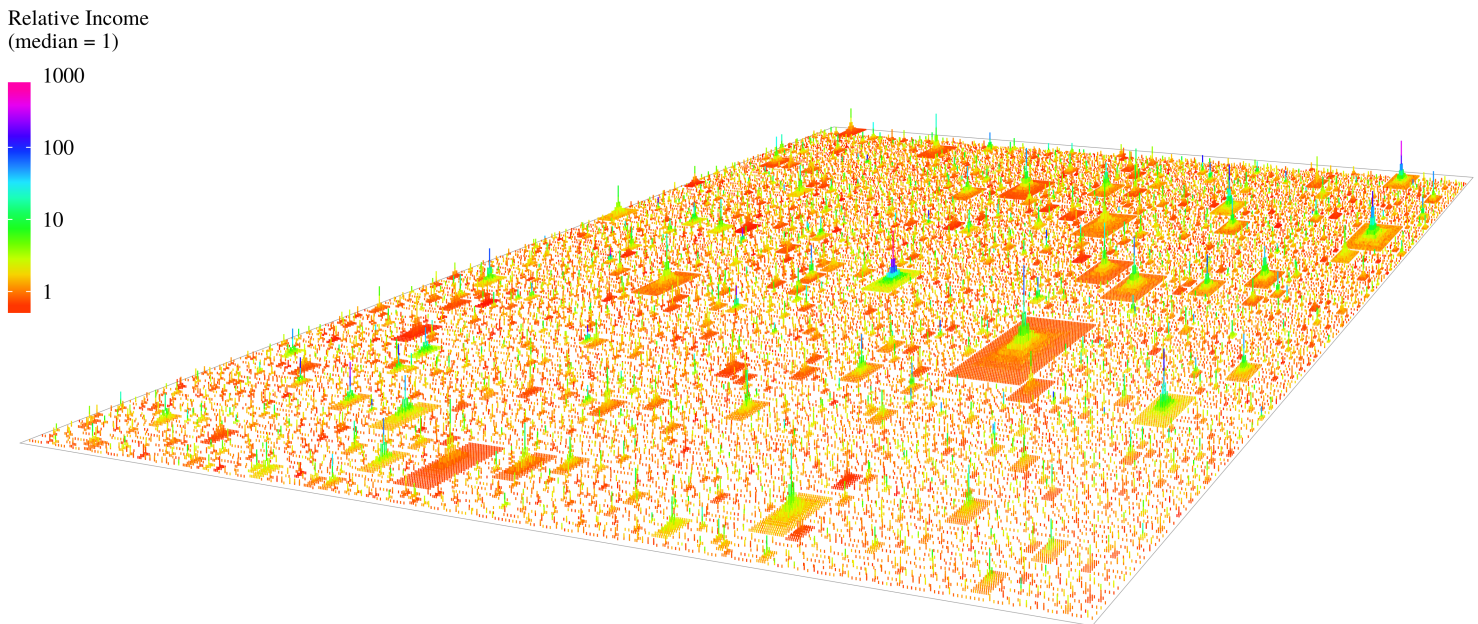
Figure 4.3: A Landscape View of the Hierarchy Model

This figure visualizes the US hierarchy model as a landscape of three dimensional firms. Each pyramid represents a single firm, with size indicating the number of employees and height corresponding to the number of hierarchical levels. If you look closely, you will see vertical lines corresponding to individuals. Income (relative to the median) is indicated by color. This visualization has 20,000 firms — a small sample of the actual model, which uses 1 million firms.

Visualizing the US Hierarchy Model

From the brief discussion here (or even from the technical discussion in the Appendix), it is not easy to gain an intuitive understanding of what the model 'looks like'. To aid with such an intuitive understanding, Figure 4.3 shows a 'landscape' view of the model's structure. Each pyramid represents a different hierarchically organized firm. The size of each pyramid corresponds to the number of employees, height represents hierarchical level, and color represents relative income.

Figure 4.3 nicely highlights the main characteristics of the model. The firm power law distribution is clearly visible. The vast majority of firms are small, but there are a few behemoths. Inter-firm income dispersion and inter-hierarchical level income dispersion are also visible, while intra-hierarchical level income

dispersion appears negligible. Lastly, top incomes are concentrated in upper hierarchical levels, and consequently occur mostly in larger firms. These facts, which are qualitatively visible here, become even more clear as we analyze the model results in quantitative terms.

4.2.2 Testing the Hierarchy Model (Part 1)

The purpose of the hierarchy model is to study the hierarchical structure of the United States economy. The first step, then, is to make sure that the model produces realistic results. To that end, Figure 4.4 compares the model's aggregate structure to US empirical data. Even though the model is an extrapolation from a limited set of data, it does a reasonably accurate job of reproducing US distribution of income.

A few things are obvious from this comparison. Firstly, the model *underestimates* US income inequality, both in terms of the Gini index (Fig. 4.4A) and the income share of the top 1% (Fig. 4.4B). What is the source of this discrepancy? Looking at the income probability density in Figure 4.4D, it appears that the US income distribution is more 'bottom heavy' than the model. That is, the model produces too few extremely small incomes, relative to the US. This tendency is also evident in the cumulative distribution (Fig. 4.4F).

Why does this discrepancy occur? I demonstrate in Appendix C.6 that the discrepancy can be removed by increasing the model's inter-firm income dispersion. This suggests that the model's under-estimate of US inequality is due to an under-estimate of inter-firm income dispersion. My guess is that this occurs because the model is based on Compustat firm data, which is not a representative sample of the US firm population. Compustat contains data for public firms only, and as a result, is biased towards large firms. I suspect that a more representative firm sample would give greater inter-firm income dispersion. (It is also possible that the model-empirical discrepancy results from some factor that is not included in the model. The most plausible would be unemployment, but many others are possible).

I include adjusted results in the Appendix to show that the model is *capable* of closely reproducing the important features of US income distribution (as any well-parameterized model should be). I do *not*, however, use this adjusted data for any of the proceeding analysis. The purpose of the model is to extrapolate empirical data, warts and all.

While the model slightly misrepresents the 'body' of US income distribution, it accurately reproduces the *tail*. This is evident in the complementary cumula-
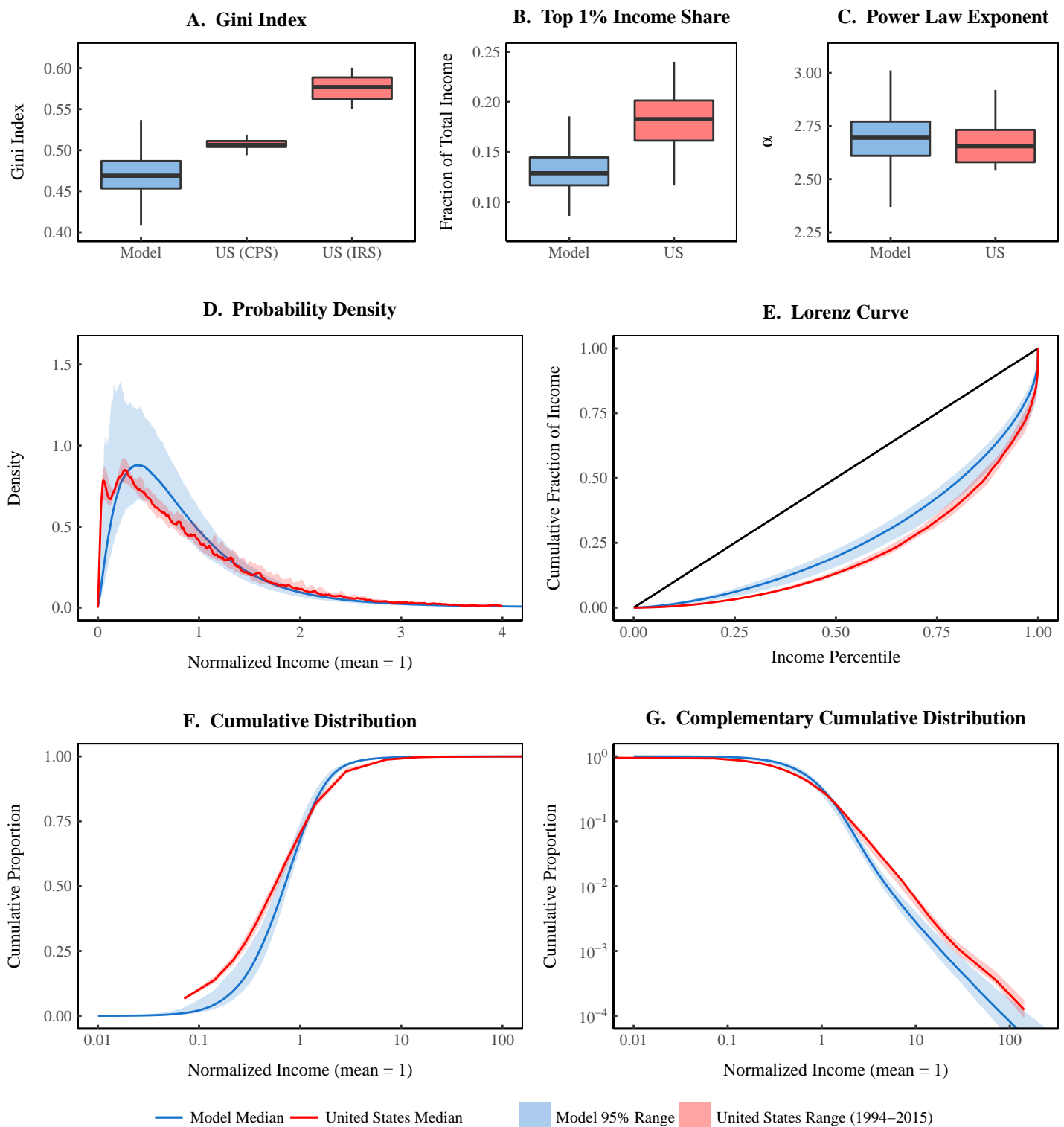
Figure 4.4: Modeled Income Distribution vs. US Data

This figure compares various aspects of the model's income distribution to US data over the years 1992-2015. Panel A shows the Gini index, with two different US sources — the Current Population Survey (CPS) and the Internal Revenue Service (IRS). Panel B shows the top 1% income share, using data from 17 different time series. Panel C shows the results of fitting a power law distribution to the top 1% of incomes (where α is the scaling exponent). Panel D plots the income density curve with mean income normalized to 1 (using data from the CPS). Panels E, F, and G use IRS data to construct the Lorenz curve, cumulative distribution, and complementary cumulative distribution (respectively). The cumulative distribution shows the proportion of individuals with income *less* than the given x value. The complementary cumulative distribution shows the proportion of individuals with income *greater* than the given x value. Note the log scale on the x-axis for these last two plots. For sources and methods, see Appendix C.1.

tive distribution (Fig. 4.4F) in the form of virtually identical model and empirical slopes in the right tail. How can these slopes be quantified? One way is to fit the tail of the income distribution to a power law — a method that dates back to the work of Pareto [1]. This approach provides a way of analyzing the tail of the income distribution *independently* from the body.

Under a power law distribution, the probability of finding someone with income x is proportional to $x^{-\alpha}$, where α is a constant (the power law exponent). The approximate power law scaling of top incomes is visible as the straight line in the tail of the complementary cumulative distribution (when plotted on a log-log scale). The choice of where the distribution 'tail' begins is arbitrary. I define the tail as the top 1% of incomes — a threshold that has been popularized by Piketty [71]. Figure 4.4C shows the results of fitting a power law to the top 1% of incomes (for methods, see Appendix C.1). The model produces power law exponents that are statistically indistinguishable from those found in the US data. Both a Kolmogorov–Smirnov test and a t-test indicate no significant differences (at the 5% level) between the model and empirical results.

To conclude, the model produces an income distribution that is roughly consistent with the US distribution of income. In particular the model closely reproduces the tail of the US distribution.

4.2.3   Testing the Hierarchy Model (Part 2)

When discussing the model visualization shown in Figure 4.3, I noted that large incomes appear to be clustered at the tops of *large* firms. This is a defining feature of the hierarchy model. It occurs because income scales strongly with hierarchical rank. As a result, top earners are found at the tops of large firms, because these firms have the most hierarchical levels. This prediction is not made by any other model of income distribution (to my knowledge). It is important, therefore, that we put it to the test.

To test this prediction, I look at the distribution of firm sizes associated with top earning individuals. What does this mean? I take a sample of Americans with top incomes, and then record the firms with which these individuals are associated. I then look at the size distribution of these firms. I do the same with the model, and compare the results.

I conduct this test using data from the Forbes 400 and Execucomp. The Forbes 400 list is useful because it is a definitive ranking of the 400 richest Americans, and it provides the institutional source of each individual's wealth. The caveat is that this list is a ranking by *wealth*, not income. I use the Forbes 400 as

a proxy for top US incomes, under the assumption that wealth and income are strongly related. I supplement the Forbes 400 data with the 'Execucomp 500'. The latter is composed of the 500 top paid US executives (in each year between 1992-2015) in the Execucomp database. The advantage of the Execucomp 500 is that it is a ranking explicitly by income. The disadvantage is that we do not know if these 500 executives are actually the top paid US individuals.

Before discussing the results of this test, it is instructive to know what a *null-effect* would look like. If there is absolutely no relation between income and firm membership, what sort of firm size distribution should be associated with top incomes? It turns out that for the United States, we should expect a null-effect to return a roughly *log-uniform* distribution (see Appendix C.7 for a derivation).

Results for the Fortune 400 and Execucomp 500 firm size distributions are shown in the main panel of Figure 4.5. To be clear, these density plots represent the size distribution of firms associated with the richest 400 Americans and the 500 top paid executives in the Execucomp database (respectively). To better visualize the distribution, I plot the density of the *logarithm* of firm size. Under this transformation, the null-effect result will appear as a *uniform* distribution. From the evidence shown in Figure 4.5, we can immediately conclude that the null-effect is false. There is definitely a relation between top incomes (wealth) and firm size. But is it the relation that is predicted by the hierarchy model?

To find out, I conduct the same analysis on the model. I select the model's 500 top paid individuals and record the size distribution of associated firms. The results are shown in Figure 4.5 as the 'Model 500'. The model predicts a relation between top incomes and firm size that is very similar to the US empirical data. To be sure, the model results are not identical to either the Forbes 400 or the Execucomp 500 distributions. But, given the paucity of data on which the model is based (as well as the general uncertainty in the empirical analysis of top incomes), I count this result as a success. The model produces results that are roughly consistent with the US data.

Since the model has three sources of income dispersion, we naturally want to know which of these sources is responsible for producing the results in Figure 4.5. To answer this question, I use a counterfactual analysis. I create three different counterfactual models to supplement the original (Model A). Each counterfactual model isolates a single source of dispersion as it appears in the original model. Model B has intra-hierarchical dispersion only, Model C has inter-firm dispersion only, and Model D has intra-hierarchical level dispersion only.

The results of this counterfactual analysis are shown in the right-hand panels in Figure 4.5. This analysis indicates that it is *exclusively* inter-hierarchical in-
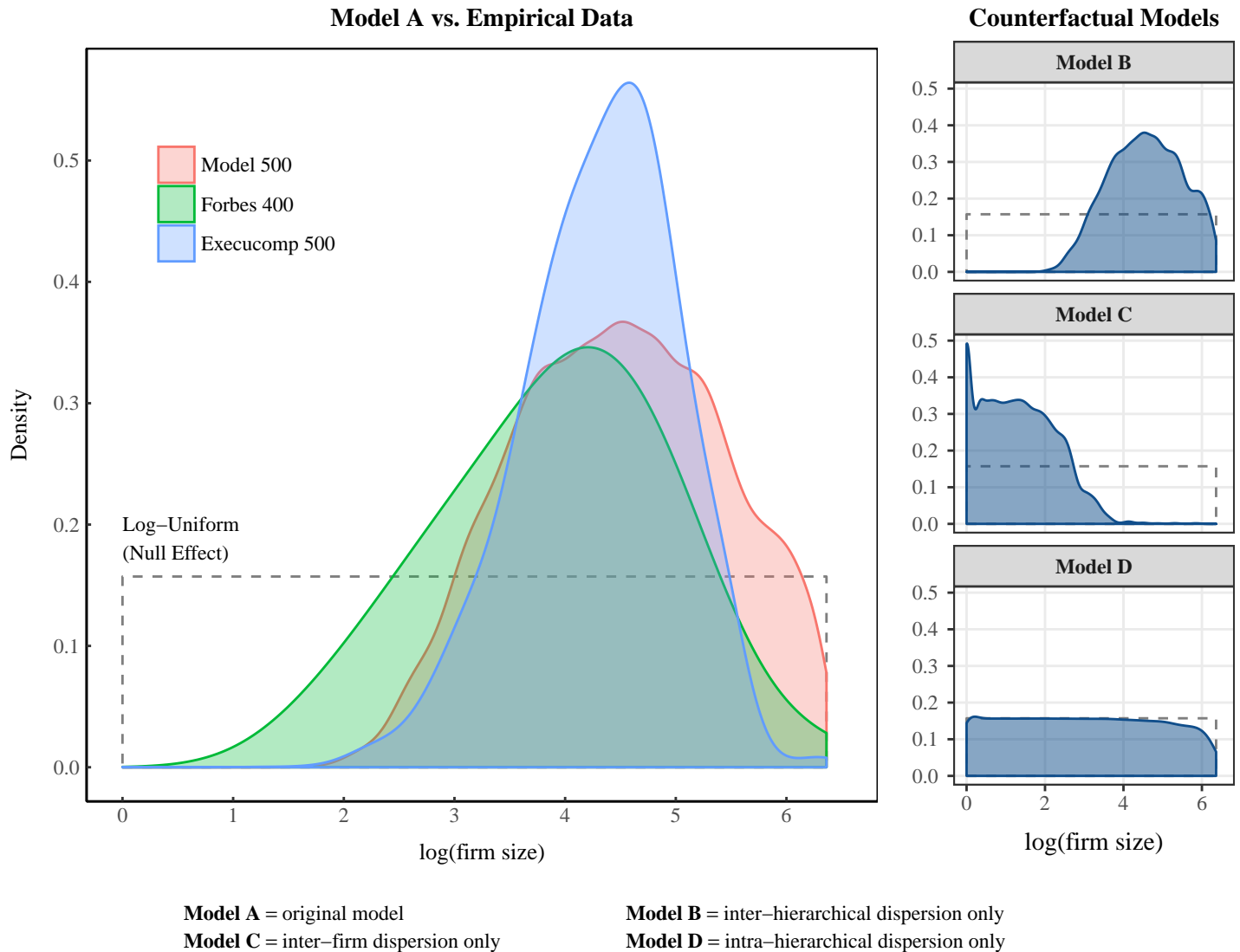
Figure 4.5: Firm Size Distributions Associated With Top Incomes and Wealth

This figure shows the size distribution of firms associated with top earning individuals in the US and in the hierarchy model (of the US). The 'Forbes 400' represents the size distribution of firms associated with (owned by) the wealthiest 400 Americans in the year 2014. The 'Execucomp 500' represents the size distribution of firms associated with the 500 top earning American executives (in each year from 1992-2015) in the Execucomp database. The 'Model 500' represents the size distribution of firms associated with the 500 top earning individuals in the hierarchy model. Results for counterfactual models are shown on the right. Each counterfactual model isolates a single source of income dispersion. Model B shows inter-hierarchical dispersion only, Model C shows inter-firm dispersion only, and Model D shows intra-hierarchical level dispersion only. In all plots, I also show the log-uniform distribution (dotted line), which is predicted if there is no relation between firm membership and income. For sources and methods, see Appendix C.1.

come dispersion (Model B) that is responsible for associating top incomes with large institutions. How do we know this? The inter-hierarchical dispersion model (B) produces results that are virtually identical to the original model. At the same time, inter-firm dispersion only (Model C) and intra-hierarchical level dispersion only (Model D) produce *drastically* different results.

Note that with intra-hierarchical dispersion only (Model D), we recover the null-effect (a log-uniform distribution). Why? In this model, firms play no part in determining income. (Income for all individuals is determined by a single stochastic function). Interestingly, this is a world that is implied by many models of income that focus solely on interactions between individuals [2, 3, 5, 6, 10–13, 15, 16, 20]. In these models, *there are no firms*. The implicit assumption must be that firms play no role in the distribution of income. Given the evidence in Figure 4.5, it would seem that these models need rethinking.

To conclude, the hierarchy model correctly predicts that top paid individuals should be associated with firms that are far larger than those of the general population. Moreover, the model indicates that this effect is purely a result of inter-hierarchical pay dispersion.

4.2.4 Quantifying Hierarchy's Effect on Income Distribution

Having established that the hierarchy model gives credible results, I now use it to investigate how hierarchy affects US income distribution. I isolate the effects of hierarchy by creating three different counterfactual version of the United States. Each version contains only *one* of the three sources of income dispersion used in the original model. By comparing these counterfactual models to the original model, we can determine how each dispersion source affects income distribution.

Let's begin with a seemingly simple question: how does hierarchy affect income inequality? The results in Figure 4.6 indicate that this question is not so simple. The affect seems to depend on *how* we measure inequality. Let's begin by using the the Gini index (Figure 4.6A). Here we see that the model with inter-firm dispersion has a Gini index that is closest to the original model. (The model with inter-hierarchical dispersion comes a distant second). This result suggests that hierarchy does not have a particularly strong effect on inequality.

However, things change drastically when we switch to measuring inequality in terms of the income share of the top 1% (Fig. 4.6B). Now we find that the model with inter-hierarchical dispersion has inequality that is nearly identical
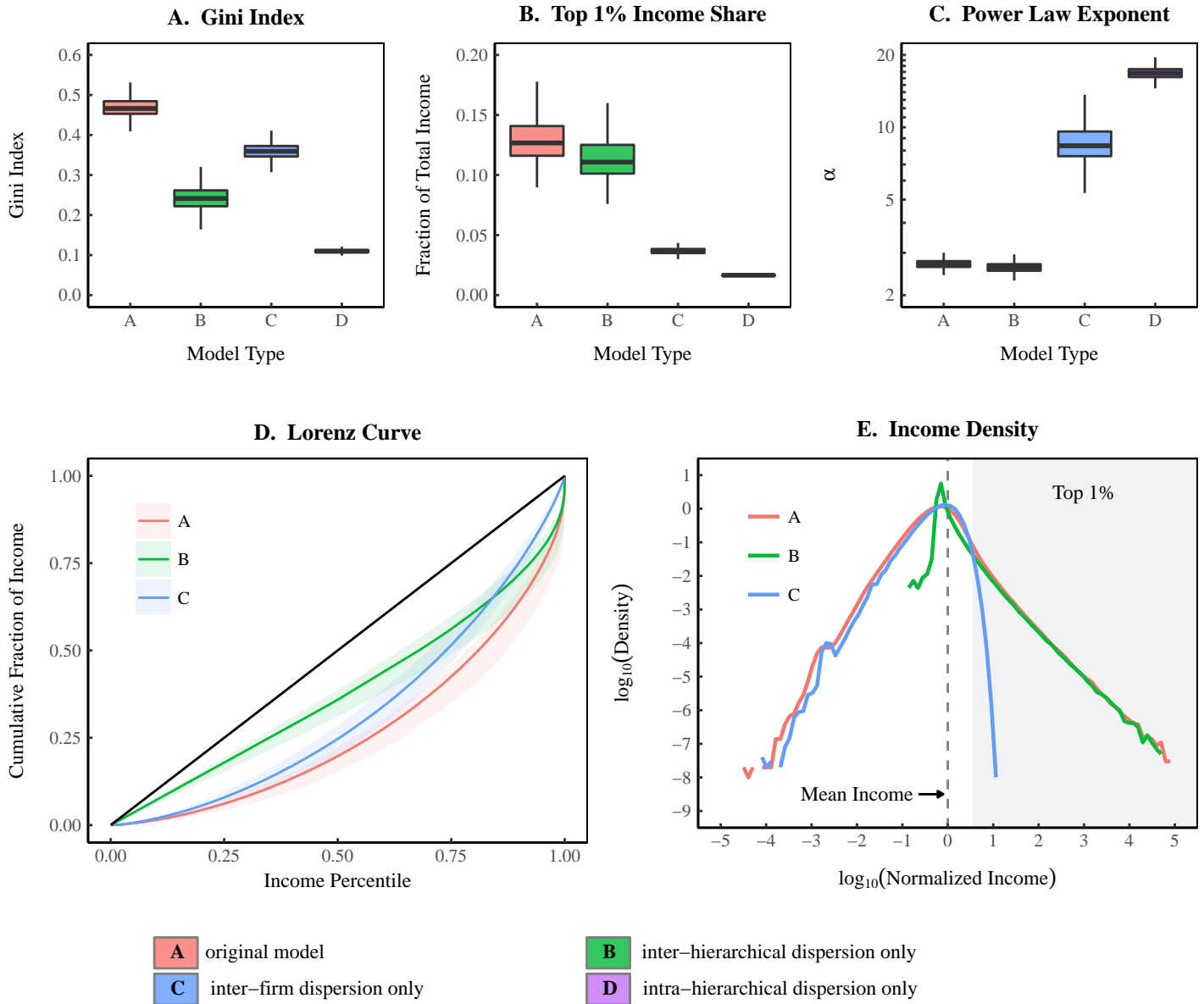
Figure 4.6: A Counterfactual Analysis of Model Properties

This figure compares the original hierarchy model of the United States to three different counterfactual models. Each counterfactual model contains only one of the three sources of income dispersion. Panel A compares the Gini index of each model, while panel B compares the top 1% income share. Note that since both of these inequality metrics are not additive, the inequality in the counterfactual models will not sum to the inequality in the original model. Panel C shows power law exponents fitted to the top 1% of incomes in each distribution. Panel D shows the Lorenz curve for each model, with shaded regions indicating the 95% range. Panel E shows the income density of each model, plotted on a log-log scale. The shaded region indicates the top 1% of incomes. For clarity (and because it plays a negligible role determining income distribution), the intra-hierarchical dispersion model is not shown in panels D and E.

to the original model. The other two sources of dispersion are inconsequential. How can this be?[2]

To understand this apparent contradiction, let's look at the Lorenz curves for each model (Fig. 4.6D). The Lorenz curve offers a convenient way to visualize the 'shape' of inequality. The curve traces the cumulative fraction of income held by all individuals below a given income percentile. The Gini index and the top 1% income share are both intimately related to the Lorenz curve. The Gini index is proportional to the area between the Lorenz curve and the line of perfect equality (the black line in Fig. 4.6D). The income share of the top 1% is equal to the vertical distance between the Lorenz curve and $y = 1$ (at the point $x = 0.99$).

The apparent contradiction between the Gini and top 1% results is now easy to understand. It is caused by an *intersection* between the inter-firm Lorenz curve and the inter-hierarchical level Lorenz curve. For incomes *below* this intersection, inter-firm dispersion plays the most important role in shaping inequality. However, for incomes *above* the intersection, hierarchy plays the most important role in shaping inequality. This nicely illustrates the pitfalls of quantifying inequality with a single metric: it is never possible to capture *all* of the information present in a Lorenz curve.

The counterfactual models indicate that inter-firm dispersion plays a very different role in shaping income inequality than does inter-hierarchical dispersion. This is made even more clear by Figure 4.6E. Here I plot the income density (in log-log form) of the original model. I then compare this to the density of the inter-firm and inter-hierarchical counterfactual models. This allows us to see how each factor contributes to the original model's distribution of income. To

---

[2] Some readers may note that I am using non-decomposable metrics to measure inequality. Since neither the Gini index nor the top 1% income share are decomposable, the inequality of the counterfactual models will *not* sum to the inequality of the original model. Thus we cannot quantify exactly 'how much' each factor contributes to income inequality. Although there are inequality metrics that are decomposable (such as the Theil index, or simply the variance), I choose not to use them here. For starters, such measures are generally far less intuitive than the Gini index or top income shares. Second, decomposable measures merely give *a* decomposition of inequality — not *the* decomposition. Decomposition requires deciding how to weight the number of incomes of a given size against the size of the income. Since there are many ways to do this, there are many equally valid decompositions of inequality. Anthony Shorrocks summarizes the problem nicely: "Inequality comparisons are invariably sensitive to the choice of inequality index used since alternative measures tend to emphasize inequality at different points in the distribution. Replacing one index by another will therefore almost always change the relative significance of the between- and with-group terms" [72].

interpret this plot, look at how closely the distribution of a specific counterfactual model comes to that of the original model. The closer it is, the more that factor influences income at the point in question. The results are unambiguous. A clear division exists between the *body* and *tail* of the distribution. The body of the distribution is almost completely determined by *inter-firm* dispersion. However, the tail of the distribution is almost completely determined by *inter-hierarchical* dispersion. (I do not include intra-hierarchical level dispersion in this plot because it plays a negligible role in shaping income distribution).

Figure 4.6C further attests to the importance of hierarchy for determining the tail of the distribution. This figure shows the fitted power law exponent for the top 1% of incomes in each counterfactual model. The power law exponent generated by the inter-hierarchical model is virtually identical to the exponent generated by the original model. The other counterfactual models produce wildly different results. This indicates that it is solely inter-hierarchical dispersion that is responsible for generating top incomes. (For a discussion of how hierarchical class structure works to create the power law tail, see Appendix C.8.)

To be clear, fitting a power law to a distribution does *not* indicate that the underlying distribution is actually a power law. We know *a priori* that neither inter-firm nor intra-hierarchy models actually produce power law tails, since dispersion within these models is generated with gamma and lognormal distributions, respectively (see Appendix C.5). In this case, the fitted power law exponent is purely descriptive. It allows us to quantify the *heaviness* of the distribution tail, independently from the body of the distribution. A heavier tail is indicated by a *smaller* power law exponent. The large exponents for inter-firm and intra-hierarchy models indicate that these distributions have tails that are far *less* heavy than the inter-hierarchical model.

To summarize, I have used the hierarchy model to gain insight into how hierarchy affects the US distribution of income. I find that hierarchy plays a decisive role in shaping the *tail* of the distribution of income. In contrast, the body of the distribution appears to be mostly determined by differences in pay between firms. This suggests that hierarchical class structure is primarily useful for understanding top incomes.

4.3   A Capitalist Gradient Hypothesis

Having established that the hierarchy model gives decent results, I now put it to a bold use. As I stated in the introduction, I believe that hierarchy shapes our social world in enumerable ways. As such, I want to know if hierarchy can

be used as a mechanism for unifying income distribution theory. I devote the remainder of the paper to this question. In this section, I investigate if hierarchy can be used to connect personal and functional income distribution.

4.3.1   Capitalists and the 1%

Long before the Occupy movement decried the separation between "the 1% and the rest of us" [73], the labor movement decried the separation between capitalists and the rest of us (workers). Are the two types of class division connected? I think so. And I think that hierarchy lies at the root of this connection. The hierarchy model suggests that top earners are *hierarchical elites*. I think the same is true of capitalists.

But for this hypothesis to make any sense, we must radically shift our ideas about what 'capital' is, and what it means to be a 'capitalist'. Building on Nitzan and Bichler's *capital as power* hypothesis [49], I propose that capitalist income is derived from power — hierarchical power. By owning firms, capitalists earn the legal right to helm firm hierarchies. From this position of power, capitalists can partition firm income streams as they see fit [74]. This hierarchical power, I suggest, is the source of capitalist income.

But a hierarchy does not have a *single* position of power. Rather, there is a *gradient* of power from top to bottom. Perhaps, along with this gradient of power, there is a gradient of ownership and a gradient of capitalist income? I call this the 'capitalist gradient' hypothesis. The idea is that the proportion of income individuals earn from capitalist sources tends to increase with hierarchical power. In other words, we can predict (in statistical terms) someone's capitalist income fraction simply by knowing their position within a firm hierarchy.

This is a bold and very much exploratory idea, but one worth testing. Surprisingly (from a mainstream perspective), I find that the capitalist gradient hypothesis has empirical support. Evidence suggests that the capitalist income fraction of US CEOs scales with hierarchical power (as I measure it). Using the hierarchy model, I generalize this CEO relation to test if it applies to the general US population. The model suggests that it does. This capitalist gradient model reproduces the US distribution of capitalist income as well as the scaling relation between income size and capitalist income fraction.

4.3.2   The Source of Capitalist Income

To begin our investigation of capitalist income, let's start with what all political economists can agree on. Capitalist income stems from owning *capital*. Beyond this trivial statement, opinions diverge rapidly. The sticking point is capital itself. True, capitalists earn income from capital — but *what is capital*?

Let's begin with the neoclassical vision of capital. In neoclassical theory, capital is a 'factor of production'. Capital consists of all the tools, technology, and infrastructure that are used to produce economic output.[3] Capitalists earn income because their capital is *productive* — it contributes to economic output [75, 76]. This thinking is illustrated in Figure 4.7.

Marxists start with a similar physical understanding of capital. According to Marx, capital is the 'means of production' — the tools, technology, and infrastructure that are used by society to create economic output [39]. The Marxist twist is to assert that capitalist income is *parasitic*. Marxists believe that labor is the source of all value. Because capitalists own the means of production, they are able to extract a surplus from labor. This thinking is illustrated in Figure 4.8.

Both neoclassical and Marxist theories of capital keep their eyes firmly on the 'real' sphere of production — on the ownership of *things*. The 'capital as power' approach, proposed by Nitzan and Bichler, is quite different. This approach focuses on ownership as an *institutional* act. What is the difference? Focusing on the *act* of ownership (and not what is owned) puts the focus on power. Nitzan and Bichler summarize: "ownership is wholly and only an institution of exclusion, and institutional exclusion is a matter of organized power" [49]. According to the capital as power hypothesis, capital is not a thing, but an *act*. It is a commodification of property rights — a vendible form of power.

In the context of studying hierarchy, the capital as power approach is useful because it puts the focus on the ownership of *institutions* (not things). Consider what it means to purchase all the shares in a company. What is it that you are buying? You are essentially purchasing *legal control* over the company. From this position of power, you have legal authority to divide up the firm's income stream as you see fit. You could slash wages and pay yourself a magnificent profit, or raise wages and earn no profit at all. From this perspective, capitalist income stems from one's power as owner.

---

[3] Admittedly, neoclassical economists have significantly broadened their definition of 'capital' over the years. For instance, there is now 'human capital' [65–67], 'knowledge capital' [77, 78], and even 'cultural capital' [79]. However, what has *not* changed is the insistence that 'capital' (in all its forms) is *productive*.
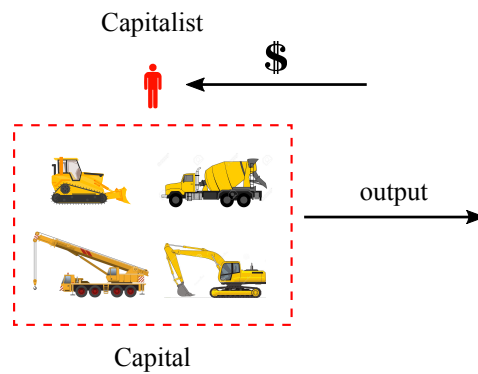
Figure 4.7: The Neoclassical Vision of Capitalist Income

In neoclassical theory, capitalists earn income because their capital is inherently productive. Capitalists earn the 'marginal product' of their capital — the incremental increase in output per incremental increase in capital input.
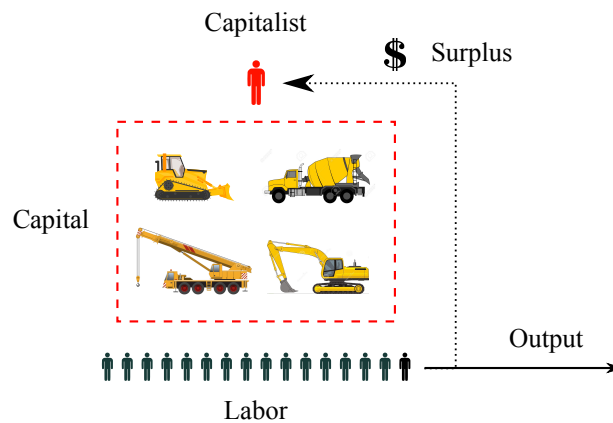
Figure 4.8: The Marxist Vision of Capitalist Income

In Marxist theory, capitalists earn income because they own the 'means of production'. Unlike neoclassical theory, Marxists see labor as the source of all value, and capitalists as parasites. Because capitalists control the means of production (capital), they are able to extract a surplus from labor.

For the present argument regarding the basis of capitalist income, I set aside the question of how the firm's income stream is derived. Instead, I am interested in how an owner wields power to partition a firm's income stream. The central hypothesis in this paper is that firms are hierarchically organized. This hypothesis implies that ownership confers the right to sit at the top of the firm hierarchy. From this position of hierarchical power (as owner), the capitalist has the authority to divide up the firm's income stream. This suggests that capitalist income stems from *hierarchical power*. This vision is illustrated in Figure 4.9.

While this vision is intuitive (at least to me), it is almost certainly too simplistic. The problem is that it is based on a 19th century, all-or-nothing concept of ownership. In this vision, a capitalist is *the* owner of a firm. Unfortunately, the rise of joint-stock companies muddies this tidy theory. Joint-stock companies allow ownership to by divided among many people. In the modern world, *partial* ownership is the rule. This realization led to the famous 'separation thesis' posited by Berle and Means [80]. The idea is that ownership has become so diffuse that capitalists no longer control the corporate hierarchy. Instead, control is ceded to managers, who are *employees*.

The problem with the separation thesis is that it acknowledges the rise of partial ownership, but insists on a traditional dichotomy between capitalists and laborers. The truth is that the line between being a capitalist and being a laborer has been blurred. Top managers often earn a large portion of their income from stock options. Conversely, owners of firms often pay themselves some form of salary. Instead of a capitalist-laborer *dichotomy*, what we need is a capitalist-laborer *gradient*. This implies that there is a steady range between being purely a capitalist and being purely a laborer. Figure 4.10 shows what this might look like when applied to a hierarchy. As one moves up the hierarchy, individuals become increasingly more capitalistic.

This capitalist gradient hypothesis can be interpreted a number of ways. The simplest interpretation is to assume a gradient of ownership within a *single* firm. However, this is realistic only for firms that are 100% employee owned. While such firms do exist (and can become quite large), they are not the norm. It is more common for a firm to have *partial* employee ownership via an employee stock ownership plan. In 2017, about 14 million Americans were enrolled in employee stock ownership plans (ESOP) [81]. This represents about 9% of the workforce. It is quite plausible that these employee stock options are preferentially rewarded to the top tiers of the hierarchy. However, ESOP assets constitute a small minority (roughly 4%) of total US market capitalization.[4] This means
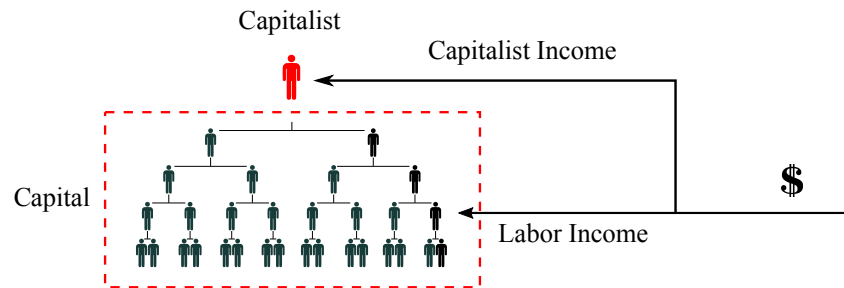
Figure 4.9: A Hierarchical Power Vision of Capitalist Income

This figure shows my interpretation of the capital as power framework, when applied to a hierarchically organized firm. Unlike in neoclassical and Marxist visions of capital (Fig. 4.7 and 4.8, respectively) I do not show physical capital. This is not to say that physical capital does not exist — we simply do not focus on it. Rather, we focus on ownership of *institutions*. Capital is conceived solely in terms of property rights. By purchasing a firm, a capitalist purchases the legal right to sit at the top of the firm hierarchy. From this position of power, the capitalist has the right to divide up the firm's income stream as he sees fit. Under this vision, hierarchical power is the source of capitalist income.
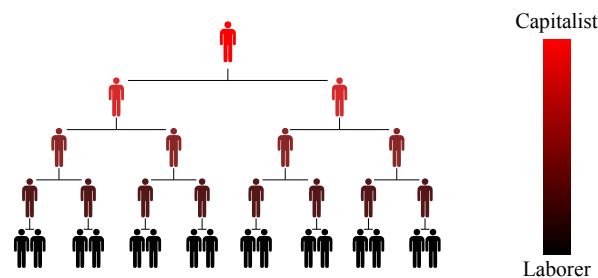
Figure 4.10: A Gradient Vision of Capitalist Income

This figure shows a more nuanced (than Fig. 4.9) interpretation of the relation between capitalist income and firm hierarchy. In this model, there is a smooth gradient between being 100% capitalist (earning all your income from capitalist sources) and being 100% laborer (earning all your income from labor sources). I hypothesize that the capitalist share of individual income tends to increase with hierarchical power.

they are probably not the main source of capitalist income.

Therefore, it is most realistic to interpret the gradient model as a statistical phenomenon that occurs at the *societal* level. We admit that the ownership structure of any given firm is likely complex. Similarly, we admit that individuals who earn capitalist income may receive it from a variety of firms. But at the *aggregate* level, we hypothesize that earning capitalist income is related to hierarchical class structure. This is the hypothesis that I test.

4.3.3 Measuring Hierarchical Power

To test the capitalist gradient hypothesis, we need to measure hierarchical power. What is *hierarchical power*? I define it as the ability to control subordinates within a hierarchical chain of command. Unlike the more general concept of 'social power', hierarchical power is easier to pin down and quantify. This is because the chain of command structure of a hierarchy clearly delineates *who has control over whom*. A hierarchy is nothing but a nested set of power relations between superior and subordinates (*ruler* and *ruled*). It is a control structure that concentrates power at the top [48].

I propose that one's power within a social hierarchy is proportional to the *number of subordinates under one's control*. I put this in formula form as:

$$\text{hierarchical power} = \text{number of subordinates} + 1 \qquad (4.1)$$

The logic of this equation is that all individuals start at a baseline power of 1, indicating that they have control over themselves. Power then increases linearly with the number of subordinates.
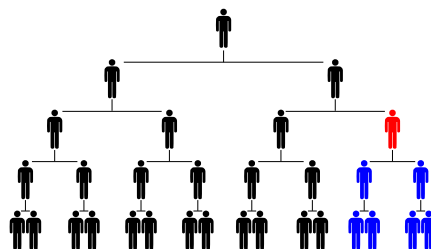
Figure 4.11: Measuring Hierarchical Power

---

[4] In 2017, ESOPs had total assets of roughly $1.3 trillion [81], while total US market capitalization was roughly $30 trillion, according to the Russel 3000 index.

As an example, suppose we want to find the hierarchical power of the red individual in Figure 4.11. This person has two direct subordinates, each of whom have 2 subordinates. Thus the red individual has control over 6 subordinates in total, mean his/her hierarchical power is 7. The general form of a branching hierarchy means that hierarchical power increases *exponentially* with rank.

4.3.4   Testing the Capitalist Gradient Hypothesis (Part 1)

If the capitalist gradient hypothesis is correct, we should be able to find evidence that capitalist income fraction increases with hierarchical power. I test the gradient hypothesis using CEO income data. This data is convenient for two reasons. First, CEO income data is easy to obtain. US regulation requires that public companies disclose CEO compensation. Second, we can estimate a CEO's hierarchical power without any knowledge of the firm's hierarchical structure. Under the assumption that the CEO holds the top hierarchical position in a firm, it follows that their hierarchal power is equivalent to the *number of employees in the firm*.

This thinking is visualized in Figure 4.12. If a firm has x employees, $x-1$ of them will be subordinate to the CEO. Since hierarchal power is defined as the number of subordinates plus one, the CEO's hierarchical power is simply firm size x. Thus, if we have data for firm size, we automatically have data for CEO hierarchical power.

So how do we calculate the 'capitalist' component of CEO income? I define the CEO capitalist income fraction as the portion of total income received from stock options:

$$\text{CEO Capitalist Income Fraction} = \frac{\text{Income from Stock Options}}{\text{Total Compensation}} \qquad (4.2)$$

Unlike cash compensation, there are many different ways to value stock options [82–84]. This means that CEO capitalist income fraction has some inherent ambiguity. However, the *nuances* of stock option valuation do not concern me here. Instead, I am interested in *general trends* in CEO compensation. For this task, the standard methods for stock option valuation will do just fine. I use CEO income data from the Execucomp database. The data series and their underlying methods are discussed in Appendix C.3.

Figure 4.13 shows the resulting relation between capitalist income fraction and firm size for roughly 40,000 American CEOs over the years 1992-2015. Two important findings emerge. Firstly, the capitalist fraction of CEO income tends
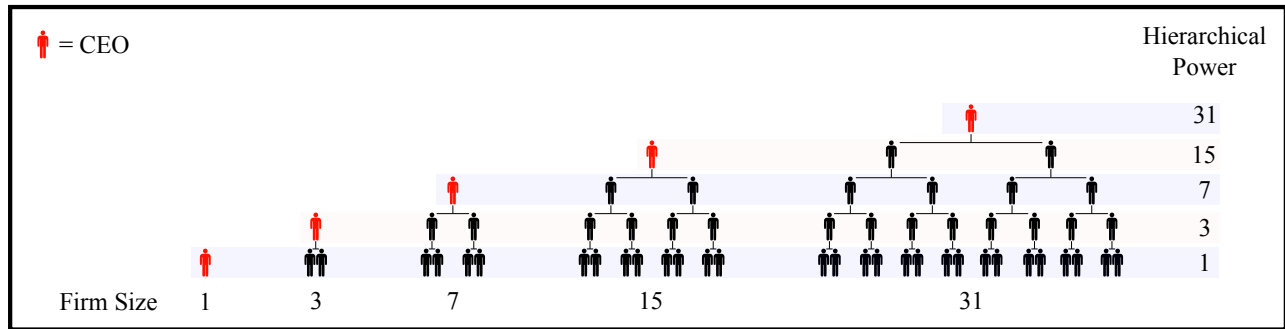
Figure 4.12: CEO Hierarchical Power

This figure shows the relation between firm size and CEO hierarchical power. Each hierarchy represents a different firm, with the CEO at the top (red). If hierarchical power is defined as the number of subordinates + 1 (Eq. 4.1), CEOs have hierarchical power equal to firm size.
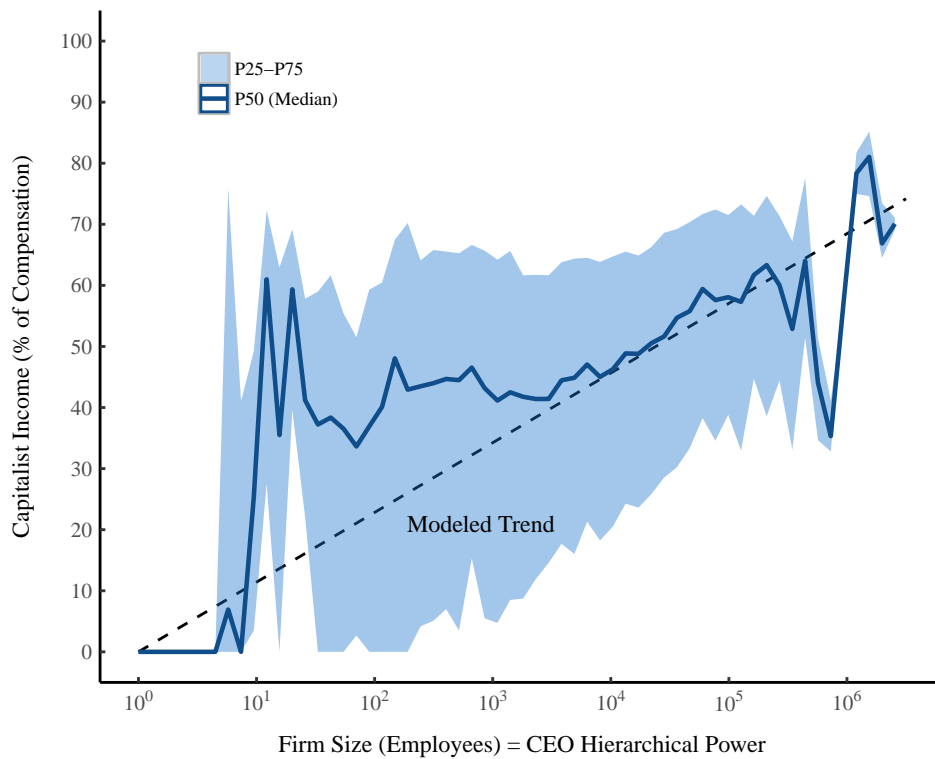
Figure 4.13: Capitalist Income Fraction of US CEOs

This figure plots the relation between capitalist income fraction and firm size for roughly 40,000 American CEOs over the years 1992-2015. Assuming that CEOs sit at the top of the corporate hierarchy, firm size is a direct indicator of CEO hierarchical power. The median (P50) and interquartile range (P25-P50) for capitalist income fraction are calculated using logarithmically spaced firm-size bins. The dashed line indicates the linear regression used for modeling purposes. Data comes from Execucomp and Compustat. For methods, see Appendix C.3.

to *increase* with firm size (and hence hierarchical power). Secondly, capitalist income fraction *tends towards zero* for CEOs in very small firms (fewer than 10 employees). These results are consistent with the capitalist gradient hypothesis — they support the idea that earning capitalist income is a gradient function of hierarchical power.

4.3.5 Testing the Capitalist Gradient Hypothesis (Part 2)

The evidence from US CEOs begs a question: does the relation between CEO capitalist income fraction and hierarchical power generalize to the broader US population? While data constraints stop us from answering this question directly (which is why we turned to CEO data in the first place), we can answer it *indirectly* by using the hierarchy model.

I do this by using the CEO data to create a simple function relating capitalist income fraction to hierarchical power. Once I have this function, I plug it into the hierarchy model and endow each individual with a capitalist income. I then check the model's results against US data. If the model produces results that are way off the mark, we know that the CEO results *do not* generalize to the whole population. However, if the model produces results that are consistent with US data, this is *indirect* evidence that capitalist income fraction increases with hierarchical power in the wider US population.

The first step is to idealize the Figure 4.13 trend between CEO capitalist income fraction and hierarchical power. The simplest interpretation of this trend is that CEO income fraction increases linearly with the logarithm of hierarchical power. I fit the CEO data with a one-parameter logarithmic function, resulting in the 'Modeled Trend' line shown in Figure 4.13. This gives the following function relating capitalist income fraction (K_{frac}) to hierarchical power (P):[5]

$$K_{\text{frac}} = 0.05 \ln(P) \tag{4.3}$$

This function is naive in the sense that it implies a *deterministic* relation between hierarchical power and capitalist income fraction — something that certainly does not exist in the real world. However, models are always simplifications, and it is often useful to simplify a noisy (stochastic) trend with a

---

[5] The discerning reader may note that, since a logarithmic function is uniformly increasing, Eq. 4.3 permits capitalist income fraction greater than 1. In practice, such results do not occur because the model does not permit firm sizes greater than 2.3 million — the largest US firm that has ever existed (Walmart, circa 2015). For this maximum hierarchical power of 2.3 million, Eq. 4.3 yields a capitalist income fraction of about 0.7.
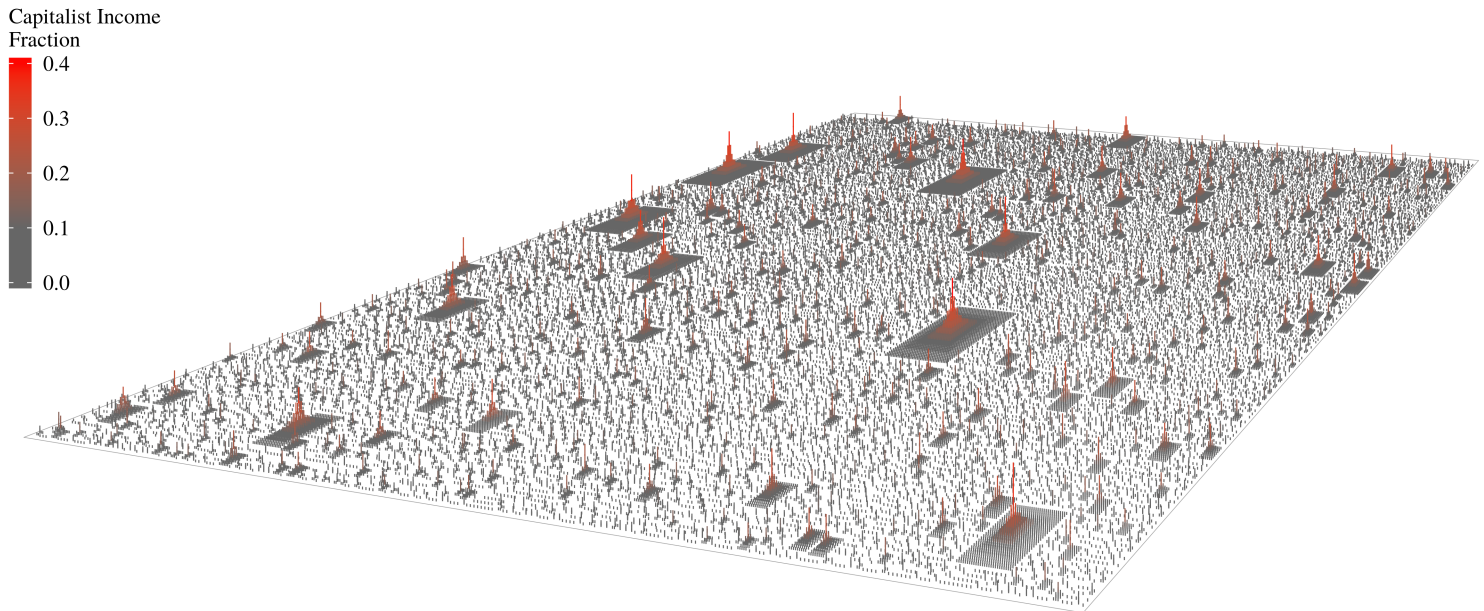
Figure 4.14: A Landscape View of the Capitalist Gradient Model

This figure visualizes the capitalist gradient model as a landscape of firms. Each pyramid represents a firm, with size indicating the number of employees. Hierarchical rank is indicated by height, and capitalist income fraction by color.

---

deterministic one. If the results are good, we can add more realism later. If the results are bad we throw away the model.

The next step is to plug this equation into the hierarchy model. We calculate the hierarchical power of each individual in the model (see Appendix C.4) and then use Eq. 4.3 to calculate the capitalist fraction of their income. The resulting capitalist gradient model is visualized in landscape form in Figure 4.14. As expected, capitalist income fraction is tightly related to hierarchical rank.

If the CEO capitalist income trend is generalizable, the capitalist gradient model should produces results that match US data. So does it? Figure 4.15 compares the model to the United States. Let's begin with the relation between capitalist income and total income size. This is effectively the relation between personal and functional income distribution — something that I have proposed that hierarchy can unify. Figure 4.15A plots Thomas Piketty's data showing how US capitalist income fraction increases with income percentile [71]. As illustrated by the inset plot (which uses a linear x-axis scale), there is an *explosion*

A. Capitalist Income Fraction vs. Income Percentile

B. Size Distribution of Capitalist Income

C. Gini Index

D. Top 1%

E. Capitalist Share of Total Income
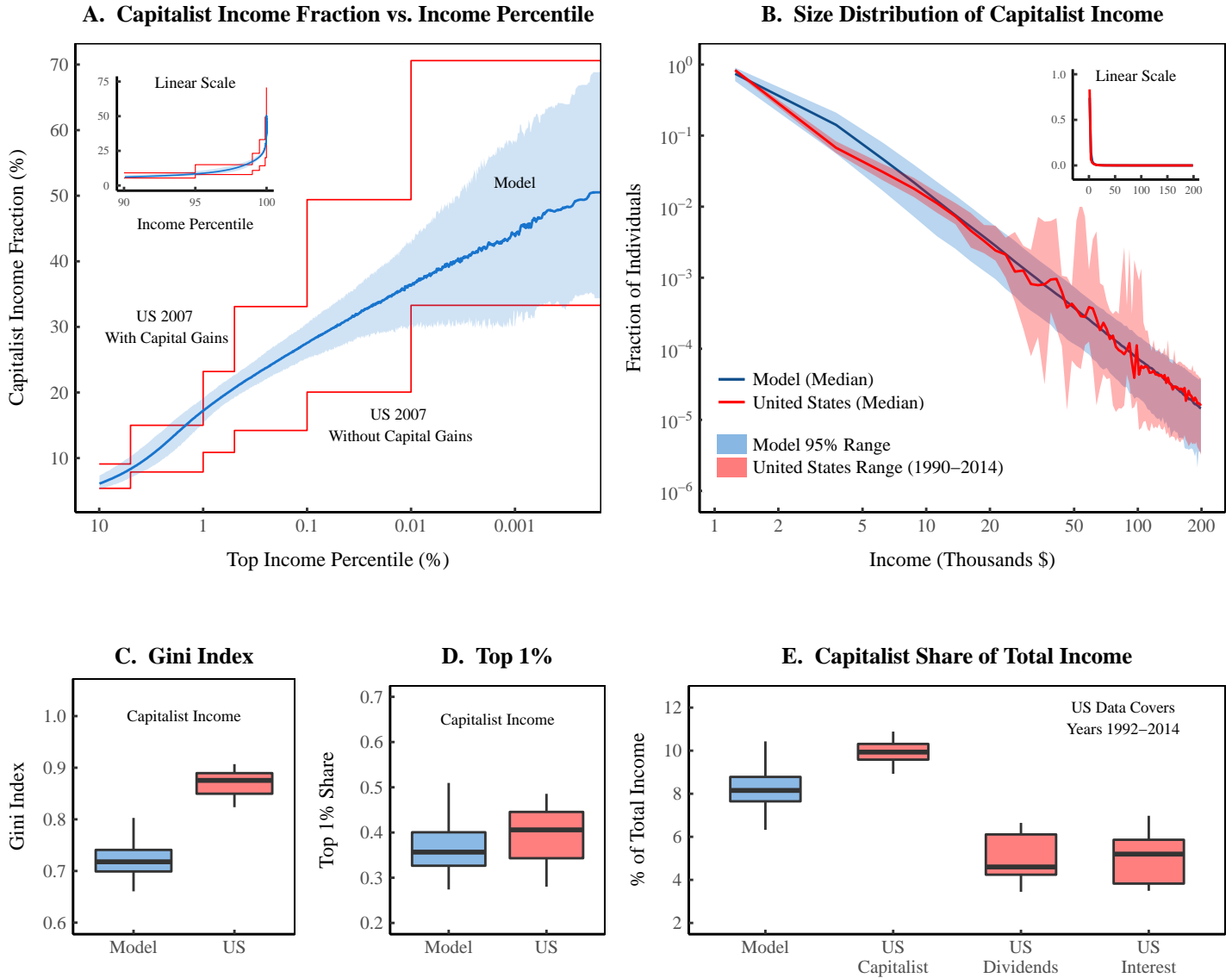
Figure 4.15: Comparing the Capitalist Gradient Model to US Data

This figure compares the income distribution generated by the capitalist gradient model to US data. Panel A shows how capitalist income fraction increases with income percentile (ranked by *total* income). The inset plot uses a linear *x*-axis scale, while the main plot uses an inverted logarithmic scale of top incomes. Note that US empirical data has 'steps' that correspond to the bins in the source data. The blue line and shaded regions indicate the model's median and 95% range, respectively. For panels B, C and D, US capitalist income is defined as the sum of income from dividends and interest. Data covers the years 1990 - 2014. Panel B shows the size distribution of capitalist income. The model data is normalized to have mean income in the same range as the US data. Panel C shows the inequality of capitalist income, as measured by the Gini index, while Panel D shows capitalist income inequality as measured by the income share of the top 1%. Panel E shows the capitalist share of total (national) income. For comparison, I also show the dividend and net interest share of US income. For sources and methods, see Appendix C.1.

of capitalist income that occurs in the topmost income percentiles. Evidently, those who earn very large incomes are overwhelmingly capitalists (and vice versa). The main panel spreads out this explosion by using an inverted logarithmic *x*-axis scale. Two different US trend-lines are shown. The upper line includes *capital gains* in the calculation of capitalist income, while the lower line does not. (The step-wise nature of these curves reflects Piketty's income bins.) Like the US data, the capitalist gradient model predicts an explosion in capitalist income amongst top earners.

Moving on, Figure 4.15B shows the size distribution of US capitalist income. For this graph (as well as Fig. 4.15C, D and E), I define capitalist income as the sum of income from dividends and interest. Although many people do earn *some* capitalist income, the amount is usually *inconsequentially* small. This fact is reflected in the inset panel, which plots the capitalist income distribution on a *linear* scale. Nearly all reported capitalist incomes are lower than $5000. In order to see the tail of the distribution, the main plot uses a log-log scale. Again, the model is consistent with US data. To get these results, I do nothing but index the model data so it has the same mean as US data. Without tuning it to do so, the model effectively reproduces the tail of US capitalist income distribution.

How about capitalist income inequality? Figure 4.15C and D show the Gini index and top 1% share of capitalist income, respectively. Just to be clear, the latter metric captures the share of total *capitalist* income held by the top 1% of reported *capitalist* incomes. First off, note how *unequal* US capitalist income is. The Gini index hovers around 0.9 (the maximum is 1), while the top 1% of capitalists earn about 40% of total capitalist income. The model reproduces this staggering income share of the top 1%, but falls short with the Gini index. Why? Part of the problem can be seen in Figure 4.15B — the model produces slightly too many capitalist incomes between $2000 to $5000.

However, the primary problem has to do with the function used to determine capitalist income (Eq. 4.3). Capitalist income is assumed to increase linearly with the logarithm of hierarchical power. Since $\log(1) = 0$, all individuals with a hierarchical power of 1 (the lowest amount possible) will have *exactly* zero capitalist income. When calculating inequality, these null incomes are (by convention) *excluded*. If we adjust the model slightly so that instead of having *no* income, these individuals have a *tiny* capitalist income, we get Gini index results that match US data. See Appendix C.6 for more details of this adjustment.

Lastly, Figure 4.15E shows the capitalist share of total (national) income. The model produces a capitalist income share that is slightly lower than (but in a similar range as) the US data (from 1992-2014). For future reference, I also

include the individual components of US capitalist income. (In section 4.4, I model historic trends in the dividend share of national income).

To summarize, the capitalist gradient model produces results that closely match US empirical data. This is indirect evidence suggesting that capitalist income fraction scales with hierarchical power in the general US population.

4.3.6 Property, Power, and Income

The results shown in Figure 4.13 and 4.15 are preliminary, and should be treated with appropriate uncertainty. That being said, I want to reflect on their potential significance. In effect, the capitalist gradient model connects three things. It suggests that hierarchical class structure, ownership class structure, and personal income distribution are all related. Put another way, hierarchical elites, capitalists, and top earners *are all the same people*.

What are we to make of this hypothesized relation between authority, property rights, and income? One interpretation is that it is nothing new. Suppose, when speaking about a feudal society, I stated that hierarchical elites, aristocrats, and the very rich are all the same people. This would be nothing particularly controversial. We are quite comfortable concluding that historical societies had a ruling class [85]. But many would bristle at that thought in our own society. Yet consider Reinhard Bendix's description of the relation between authority, property rights, and income in German feudal society. He writes:

> governmental functions were usable rights which could be sold or leased at will. For example, judicial authority was a type of property. The person who bought or leased that property was entitled to adjudicate disputes and receive the fees and penalties incident to such adjudication. [86] (p. 149)

If we paraphrase Bendix, we arrive at the same reasoning that I used to derive the capitalist gradient hypothesis. Building on the work of Nitzan and Bichler, I suggested that 'capitalist authority' is a 'type of property'. The person who buys this property is 'entitled' to wield hierarchical power and 'receive income' in return. From this reasoning came the hypothesis that capitalist income should be related to hierarchical rank and power.

From the perspective of mainstream economic theory, this hypothesis is quite radical. It undermines the ubiquitous assumption that capitalists earn income from a productive asset. But given Bendix's comments on feudal society, the capitalist gradient hypothesis may be quite conservative. Why? Conservatism implies a *lack of change* — a maintenance of the same order. The capitalist gradient hypothesis may be conservative because it suggests that income distribution

in modern capitalist societies might not be as different from past feudal societies as we would like to think.

4.4   A Hierarchical Redistribution Hypothesis

I turn now from modeling the static distribution of income, to modeling inequality *dynamics*. Over the last three decades, there has been an explosion in inequality in the United States (with less pronounced increases in other countries). I set aside the difficult 'why' question, and instead focus on the 'how' question. How did this increase occur? Does it have any relation to firm hierarchy? I think that it does. There is good evidence suggesting that the US has undergone a *hierarchical redistribution* of income — a transfer of income from the bottom to the top of firm hierarchies. I call this the 'hierarchical redistribution' hypothesis, and I test it using the hierarchy model.

4.4.1   The Evidence

Let's look at some evidence that hints at hierarchical redistribution. One trend that slaps us in the face is the post-1980 explosion in the CEO pay ratio. As shown in Figure 4.16A, this explosion corresponds closely with increases in the top 1% income share (Fig. 4.16B). Assuming that CEOs sit at the top of the corporate hierarchy, the increasing CEO pay ratio suggests that hierarchical redistribution has occurred.[6]

Figure 4.16D gives more evidence hinting at hierarchical redistribution. Here I show trends in the power law exponent of the top 1% of US incomes. This exponent quantifies the 'fatness' of the distribution tail (a smaller exponent means a fatter tail). This analysis demonstrates that rising top income inequality is associated with a fattening of the tail of the income distribution. What does this have to do with hierarchy? According to our model, hierarchy plays a dominant role in shaping the *tail* of US income distribution (section 4.2). Therefore, it is plausible that a fattening tail might be caused by hierarchical redistribution.

The connection between CEO pay and income inequality has been widely discussed [90–94], as has the fattening of the income distribution tail [8, 16, 95]. Less recognized, however, is the relation between rising inequality and the

---

[6] Mishel and Davis [87] note a strong correlation between CEO compensation and stock market returns. This raises the possibility of connecting income redistribution to capital accumulation, something that has been theorized by Bichler and Nitzan [88]. However, such an investigation is beyond the scope of this paper.
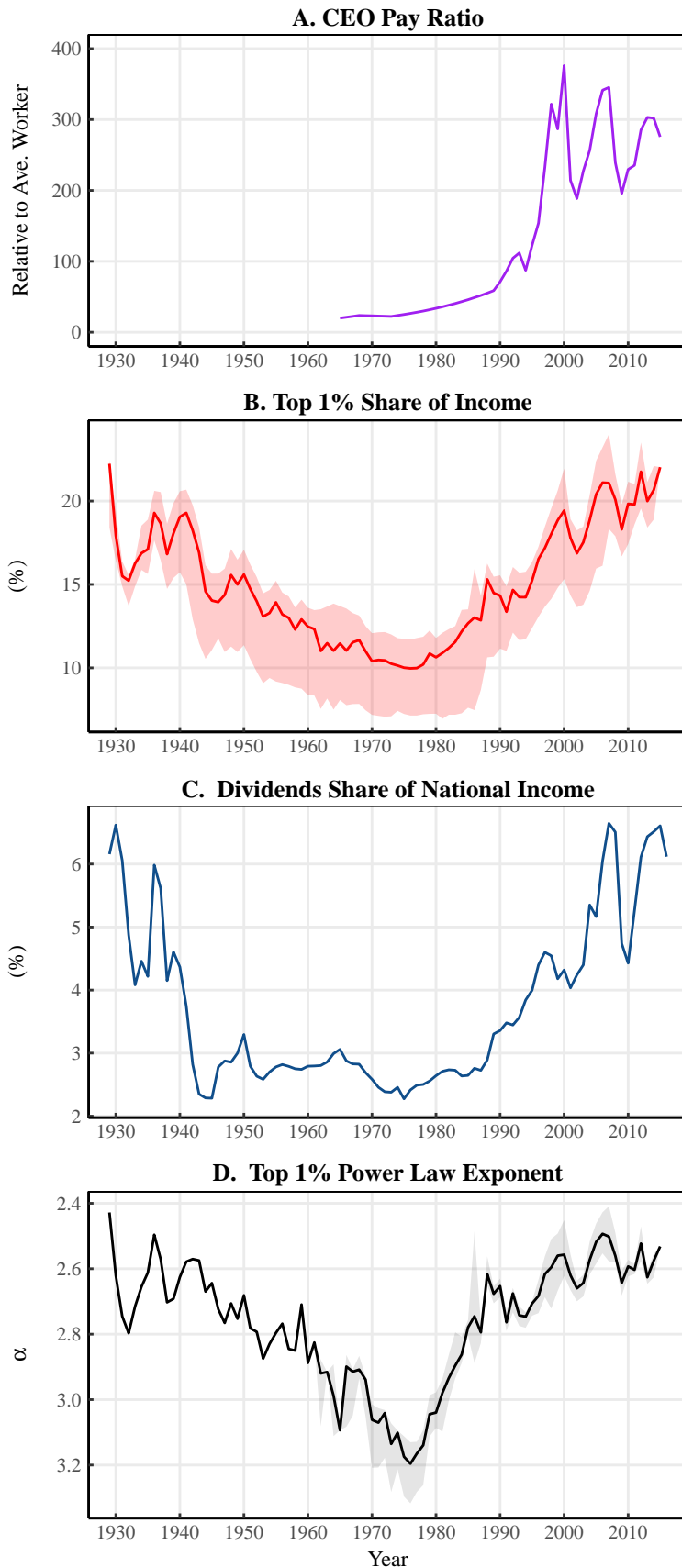
Figure 4.16: Historical Income Distribution Trends in the United States

This figure shows four trends in US income distribution that hint at hierarchical redistribution. Panel A shows the trend in the CEO pay ratio [89]. This ratio is calculated using CEO income in the 350 largest US firms (ranked by sales), compared to the average income of workers in the firms' respective industry. Panel B shows the trend in US inequality, as measured by the top 1% income share. The shaded region indicates the range of 17 different estimates for the top 1% income share. The line represents the median of these estimates. Panel C shows the trend in the dividend share of national income. Panel D shows the fitted power-law exponent for the top 1% of incomes. A smaller exponent indicates a 'fatter' tail. The grey region indicates the range of estimates (when different series are available). The line indicates the median estimate. For sources and methods see Appendix C.1.

redistribution of *functional* income. As shown in Figure 4.16C, changes in the US dividend share of national income are strongly correlated with changes in the top 1% income share. The correlation coefficient ranges between 0.82 and 0.90, depending on the choice of data. Is this trend also related to hierarchical redistribution? The capitalist gradient model suggests that it might be.

The capitalist gradient model proposes that individuals become more *capitalistic* as hierarchical rank increases (section 4.3). This model implies that a bottom-to-top redistribution of pay within firm hierarchies should correspond with an increase in the capitalist share of total income. Why? Top-ranked individuals are hypothesized to have a greater proportion of capitalist income relative to bottom-ranked individuals (regardless of the relative size of top and bottom incomes). If top-ranked individuals increase their share of the pie, the capitalist share of total income should increase as well.

To summarize, the trends in Figure 4.16 suggest a hierarchical redistribution of income within firms. To test this hypothesis, I use the hierarchy model.

4.4.2   Methods

The idea behind my test of the hierarchical redistribution hypothesis is quite simple. If the trends in Figure 4.16 are caused by a hierarchical redistribution of income, we ought to be able to replicate them with the hierarchy model. I attempt this replication by varying the rate at which modeled pay increases with hierarchical rank. I call the parameter that controls this rate the 'hierarchical pay-scaling parameter'. See Appendix C.5 for a technical discussion about what this parameter does.

Varying the hierarchical pay-scaling parameter changes the returns to hierarchical rank, and by extension, the returns to hierarchical power. This effect is illustrated in Figure 4.17. When the pay-scaling parameter is small (indicated by the color red), relative pay increases very slowly with hierarchical rank and power. But when the pay-scaling parameter is large (indicated by blue), there is an extremely rapid increase in pay with hierarchical rank and power.

To test the hierarchical redistribution hypothesis using the hierarchy model, I restrict the scope of analysis to the years 1965 onward. I do this for two reasons. Firstly, the CEO pay ratio data begins in 1965. Secondly, the model assumes an unchanging firm size distribution. From the late 1960s onward this assumption is valid — the US firm size distribution changed very little. However, prior to the 1960s the US firm size distribution changed rapidly [69], violating the model's assumptions.

A. Relative Income by Hierarchical Rank

B. Relative Income by Hierarchical Power
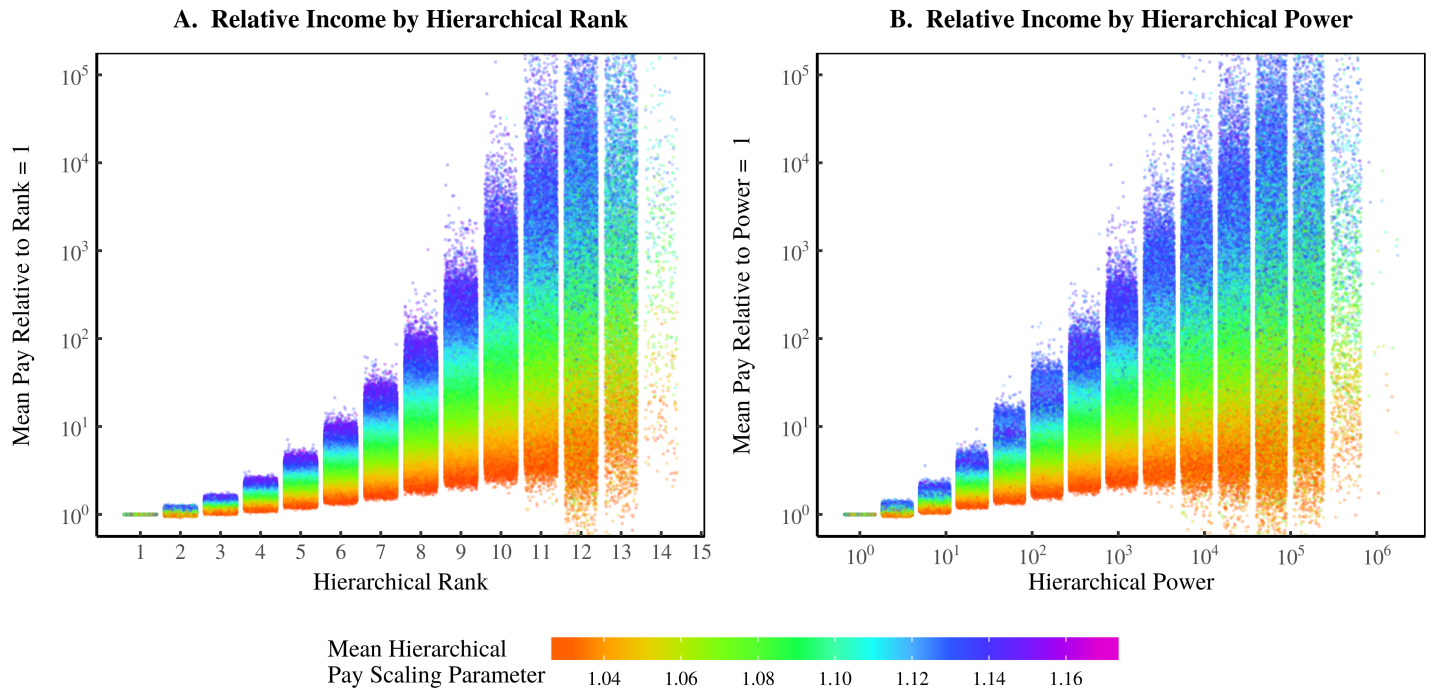
Figure 4.17: Changing How Income Scales with Hierarchical Rank and Power

This figure shows the results of the hierarchy model when the hierarchical pay-scaling parameter is allowed to vary (over different model iterations). Panel A shows how mean pay (relative to the bottom hierarchical level) increases by hierarchical rank. Different pay-scaling parameters are indicated by color. Panel B shows the same effect, but with hierarchical power (where hierarchical power is defined as the number of subordinates + 1). Individuals are grouped into log-spaced bins by hierarchical power. Note that the trends in both panels become increasingly noisy for the very top hierarchical ranks and very large hierarchical power. This is because individuals with very high rank are extremely rare, so the mean encompasses relatively few individuals. In both plots, horizontal 'jitter' is added to increase the visibility of all data points.

In this test, I vary *only* the hierarchical pay-scaling parameter. Inter-firm dispersion and intra-hierarchical level dispersion remain at the levels implied by modern case study and Compustat data. I continue to use the capitalist gradient model to decompose income into capitalist and labor components. Importantly, I do not vary the function that determines capitalist income fraction (Eq. 4.3).

To model the data in Figure 4.16, I add two assumptions to the hierarchy model. The first assumption is used to model the dividend share of national income. The capitalist gradient model predicts *total* capitalist income only, and

does not differentiate between interest and dividends. To model the dividend share of income, I assume that dividends constitute *exactly half* of capitalist income. This 50-50 split between interest and dividends is what the US has averaged over the last century (see Fig. 4.15E for the post-1990 split). Although there have been important historical variations in the composition of capitalist income [49], these are not included in the model.

The second assumption has to do with modeling the CEO pay ratio. The empirical CEO pay ratio in Figure 4.16 is calculated using CEO pay in the top 350 US firms, ranked by *sales*. The average pay of employees is calculated using average pay in each firm's respective *sector* [89]. The model has *neither* sales, nor sectors, nor explicit job titles. I assume that CEOs are the top-ranked individual in each firm hierarchy. I calculate the model's CEO pay ratio using CEO income in the top 350 firms, ranked by total *payroll*. I use payroll as a proxy for sales, since the two metrics are highly correlated (see Appendix C.3). Because the model has no sectors, I use the average pay in the whole model to calculate average worker pay.

4.4.3   Results

Results of the hierarchical redistribution model are shown in Figure 4.18. Because the model has no time element, I compare only the relation between trends. (Note that the top 1% share is the common x-axis in all panels). Each panel shows both US empirical and model relations. As in Figure 4.17, variation in the hierarchical pay-scaling parameter is indicated by color. The take-home message here is that, by varying hierarchical pay, the hierarchy model is able to reproduce the general form of the empirical trends identified in Figure 4.16.

To be sure, the model's results are not perfect. In general, the model tends to underestimate the top 1% income share. This causes a leftward shift in the modeled relations (relative to the empirical ones). The hierarchy model is heavily dependent on the Compustat database, which is biased towards large firms. I have hypothesized that this bias causes the model to underestimate inter-firm income dispersion. In Appendix C.6, I show that increasing inter-firm dispersion (so that the model almost perfectly reproduces the US distribution of income) improves the accuracy of the hierarchical redistribution model.

Another problem is that the model's dividends versus top 1% slope is not quite correct. This slope turns out to be heavily dependent on the particular functional relation between hierarchical power and capitalist income fraction. While the function that I use is based on empirical data (see section 4.3), there
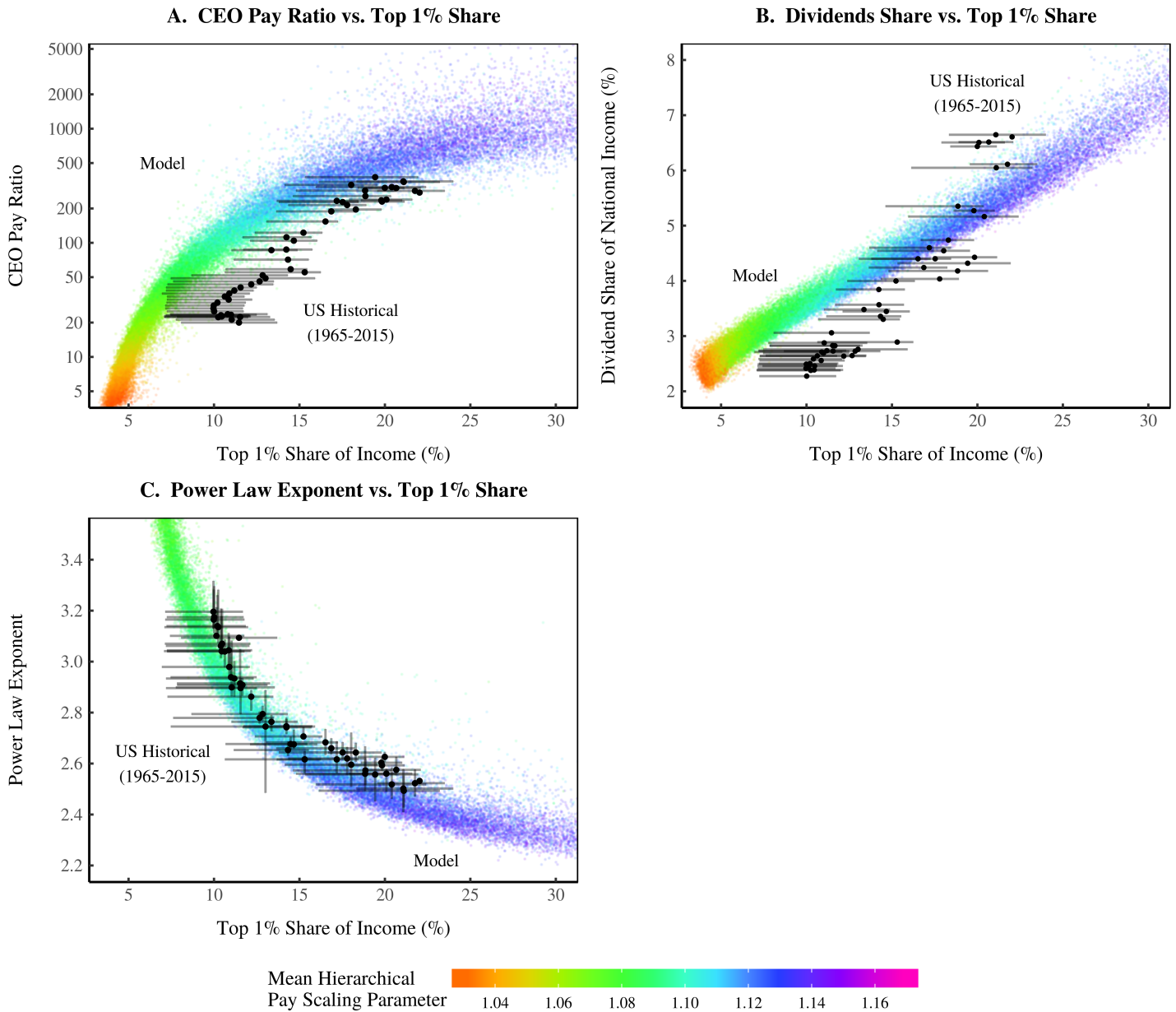
Figure 4.18: The Hierarchical Redistribution Model vs. US Data

This figure compares model results to historical trends in US income distribution. Model results are produced by varying the hierarchical pay-scaling parameter, indicated by color. Each colored point represents a single model iteration. US empirical data is shown in black, with horizontal error lines indicating the range of 17 different estimates for the top 1% income share. The point indicates the median of these estimates. Panel A plots the CEO pay ratio against the top 1% share, while panel B plots the dividend share of national income against the top 1% share. Panel C plots the fitted power law exponent of the top 1% of incomes against the top 1% income share. For sources and methods, see Appendix C.1.

is tremendous uncertainty in this relation. More empirical research is needed to understand the source of this model discrepancy.

4.4.4 Discussion

While we should always be cautious about drawing conclusions from a model, I want to offer my thoughts on the significance of these results. There has been a tendency, in political economy, to explain human income distribution in terms of 'natural law'. For instance, John Bates Clark began his foundational text on marginal productivity by declaring: "It is the purpose of this work to show that the distribution of the income of society is controlled by a *natural law*" [75]. This tendency was only strengthened when Pareto discovered the ubiquitous power law scaling of top incomes [1].

But what is curious about 'natural law' theories is that they are almost always atomistic. Thus, Clark showed that perfectly competitive markets distribute income according to 'natural law'. But leviathan governments are mysteriously absent from this picture. In a sense, the term 'natural law' is used as a euphemism for '*in the absence of concentrated power*'. Thus, 'natural law' explanations of skewed income distribution tails are typically based on atomistic premises, in which there are isolated individuals but no institutions [5, 18]. From this perspective, power is a *distortion*.

But what if concentrated power is the *reason* that income distribution has a power law tail? This is the story told by the hierarchy model. This model suggests that hierarchy — a form of concentrated power — is responsible for producing the fat tail of US income distribution. The same model suggests that *changes* in the tail are a result of a hierarchical redistribution of pay. Thus, hierarchy provides a potentially potent tool for understanding both the *regularities* of income distribution over time and space, but also the *variation*. I propose that the *regularity* of power-law income distribution tails owes to the ubiquity of social hierarchy. Conversely, I propose that *variation* in the tail owes to hierarchical redistribution.

I conclude by visualizing the hierarchical redistribution that has occurred in the United States (as suggested by the hierarchy model). Figure 4.19 shows two modeled versions of the United States. On top is the 1965 version. On the bottom is the 2015 version. The difference between the two is subtle — it is almost completely isolated to the tops of large firms. Here we see a massive, order of magnitude increase in relative pay — a clear redistribution of income to top-ranked individuals. If the model is correct, we can conclude that the US has
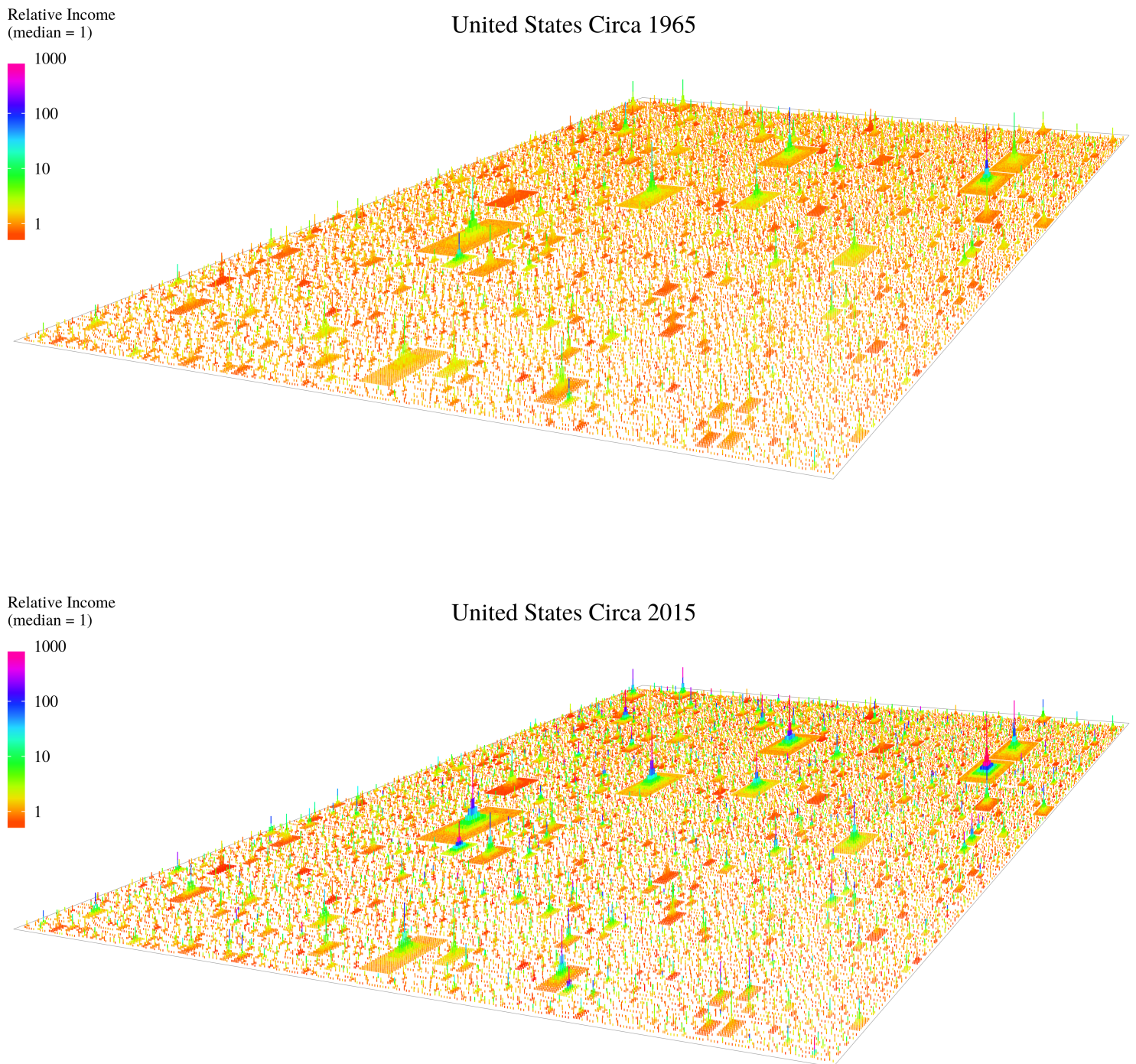
Figure 4.19: A Visualization of US Hierarchical Income Redistribution

This figure shows the model's representation of historical hierarchical income redistribution in the United States. The top model represents the US in 1965 while the bottom represent the US in 2015. I create these models by choosing the hierarchical pay-scaling parameter that best matches the US CEO pay ratio, top 1% and dividend share data in the year in question. The difference between the two model's is mostly visible at the tops of large firms as an order of magnitude increase in the pay of top-ranked individuals.

undergone a massive hierarchical redistribution of income in the last 30 years.

4.5   Conclusions: Modeling from the Top Down

Many economists have an understandable desire to model human society from from the 'bottom up' [7]. This means that they seek to explain complex social structures solely in terms of the interaction of individuals. The bottom up strategy is a noble one, *in principle*. It would be a triumph of science if we could explain macro-level income distribution based purely on the interactions of individuals. In the same way, it would be a triumph of science if we could understand the emergence of consciousness based purely on the interactions of atoms and molecules. This is a noble pursuit in principle. In *practice*, however, it is misguided.

The problem is two-fold. The first problem is computational feasibility. Suppose we had a highly accurate model of the human psyche, comparable to the accuracy of quantum mechanics. If we did, it's highly likely that meaningful questions would be computationally unfeasible. Even though it is the general scientific consensus that consciousness emerges from matter alone (i.e. there is no mind-body dualism) I know of no attempt to simulate consciousness using the laws of physics. The problem is simply too difficult. Quantum physics is so computationally complex that it is difficult to simulate large *molecules*, let alone brains.

The second problem is that to build a model from the bottom up, we need a highly accurate model of the 'fundamental particles'. We have a pretty good model of atoms. Do we have a good model of the human psyche? Hardly. I believe we should be humble and admit that we know very little about human behavior. As a consequence, when we model from the bottom up, we are essentially groping in the dark. We must make blind assumptions about how agents behave. The problem is that the *entirety* of the modeling effort depends on these assumptions. The model may very well give good results — it may seem to 'explain' the social phenomena in question. But if the underlying assumptions are incorrect, the entire model is wrong.

The dream of explaining income distribution from the bottom up is a noble one. The problem is that we are hopelessly far from being able to do this the right way. The bottom up models that do exist make extremely naive assumptions about how humans behave. While these models give good results, it is a fallacy to think that this validates their underlying assumptions.

The alternative to the bottom-up approach is to model from the *top down*.

What does this mean? Instead of having social structure emerge from the bottom-up actions of individuals, we (the modelers) *impose* structure from the top down. In essence, we impose structure on society and then explore the consequences. The *origin* of this structure is left unexplored. The top-down approach is useful because it allows realism and ignorance to coexist. A realistic model of income distribution *must* have institutions — they are simply too important to ignore. But we know very little about how and why institutions form. The top-down approach allows us to model institutions without having any idea of why they exist.

This is the philosophy that underlies the hierarchy model. The model is based on two observations of the real-world: (1) firms are the dominant institution for organizing paid human activity (in capitalist societies); and (2) firms are hierarchically organized. The model takes these facts as given, and explores their consequences.

The central finding of the hierarchy model is that hierarchy shapes the *tail* of the income distribution. According to our model, it is hierarchy that causes the distinctive power-law scaling of top incomes. This is important because explaining the power-law distribution of top incomes has been one of the primary concerns of income distribution modelers. The over-whelming majority of power law generating models are based on atomistic premises. As far as I am aware, the hierarchy model is the only power law generating model that *includes* institutions.

But this is not all. The hierarchy model is, to my knowledge, the only power law generating model that is *completely* empirically grounded. As I have stated many times, the hierarchy model amounts to an *extrapolation* of real-world evidence. The model takes the little information of firm hierarchy that does exist, and extrapolates it to create a large-scale simulation of the US economy. To risk overstating this, there is nothing in the model that is not implied by empirical data.

The story that the hierarchy model tells is this: the power-law distribution of top incomes arises from *concentrations of power*. The model suggests that without large, hierarchically organized firms, there would be no power law distribution of top incomes. This finding is significant in its own right, but made more so by its stark contrast with mainstream, neoclassical economic theory. James T. Peach summarizes the neoclassical approach: "Individual productivity and exogenously determined shifts in supply and/or demand curves determine distributive shares. ... [T]here is no power and there is no income distribution problem" [57]. If the hierarchy model is correct, concentrated power is not an
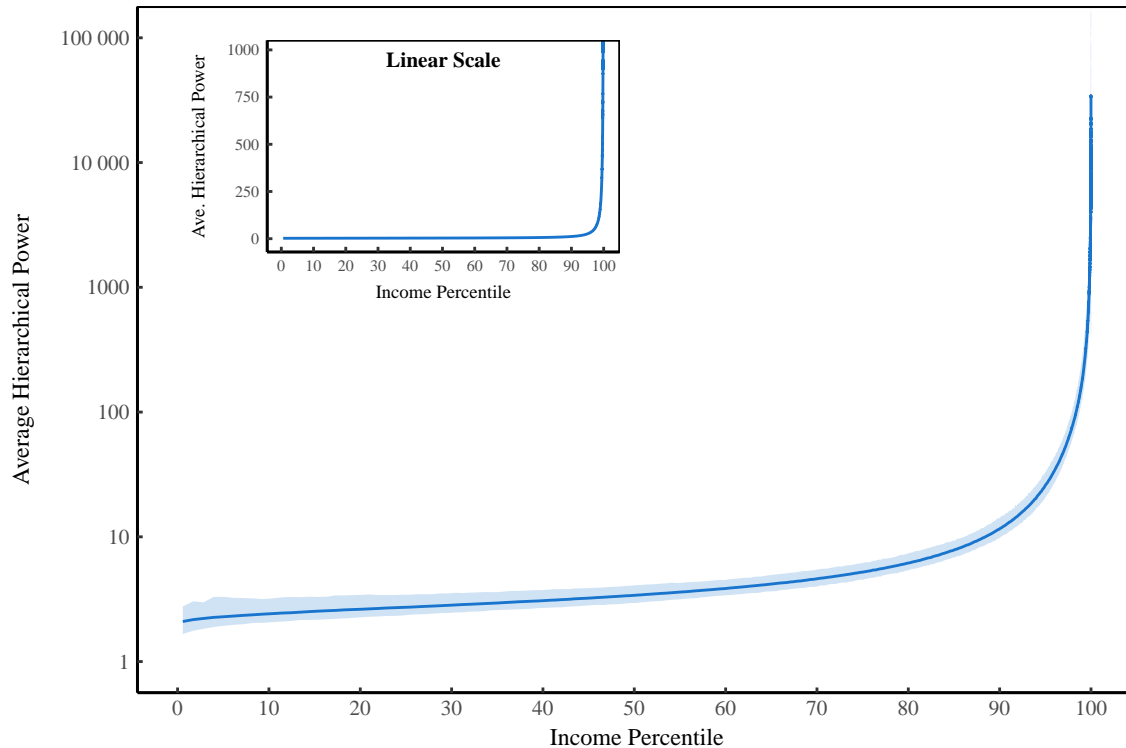
Figure 4.20: The Rich and Powerful — Hierarchical Power and Top Incomes
This figure plots average hierarchical power (number of subordinates + 1) against income percentile for individuals in the hierarchy model of the United States. The shaded regions indicates the 95% range, while the line indicates the median. In order to show the entire range of data, the main panel uses a logarithmic scale on the y-axis. The inset panel uses a linear y-axis to illustrate how rapidly hierarchical power increases in the top 1% of incomes.

———————————————

aberration — it is the norm. Based on the model results, I have suggested that power-law scaling of top incomes is ubiquitous because concentrated power (in the form of hierarchical institutions) is also ubiquitous.

To put matters simply, the hierarchy model gives new meaning to the phrase 'rich and powerful'. This is made clear by Figure 4.20. Here I plot average hierarchical power against income percentile for the hierarchy model of the United States. Two completely different populations emerge — those with power and those without. The vast majority of people have very little hierarchical power. But things change drastically for the small minority in the upper income percentiles. Here there is an explosion of hierarchical power. This power, I believe, is the origin of the great inequalities that plague human society (now and in the past). Hierarchical power gives preferential access to resources, plain and

simple.

That being said, there is no fixed relation between income and hierarchical power. Gerhard Lenski [55] gives the curious example of Robert McNamara's move from the Ford Motor Company to the position of US Secretary of Defense. McNamara's new position had far more power, and yet his income did not increase. Instead, it *decreased* by an order of magnitude. Why? These are questions we must ask. Unlike Clark's theory of marginal productivity, a theory of income distribution based on hierarchy and power has no 'laws'. Things can and do change.

To conclude, the hierarchy model is a first attempt at quantitatively studying the distributional consequences of hierarchical organization. If nothing else, the model suggest that hierarchy must be taken seriously — it is a grave mistake to 'assume' hierarchy away when building income distribution models. If we want to alleviate income inequality, we need to understand it. This understanding will undoubtedly require models, but these models must be rooted in the real world — a world in which concentrated power appears to be the norm.

References

1. Pareto V. Cours d'economie politique. vol. 1. Librairie Droz; 1897.

2. Angle J. The surplus theory of social stratification and the size distribution of personal wealth. Social Forces. 1986;65(2):293–326.

3. Angle J. The inequality process as a wealth maximizing process. Physica A: Statistical Mechanics and Its Applications. 2006;367(15):388–414.

4. Champernowne DG. A model of income distribution. The Economic Journal. 1953;63(250):318–351.

5. Chatterjee A, Chakrabarti BK, Chakraborti A. Econophysics and sociophysics: trends and perspectives. Milan: John Wiley & Sons; 2007.

6. Chatterjee A, Chakrabarti BK. Kinetic exchange models for income and wealth distributions. The European Physical Journal B. 2007;60(2):135–149.

7. Epstein JM, Axtell R. Growing artificial societies: social science from the bottom up. Washington, D.C.: Brookings Institution Press; 1996.

8. Gabaix X, Lasry JM, Lions PL, Moll B. The dynamics of inequality. Econometrica. 2016;84(6):2071–2111.

9. Gibrat R. Les inegalites economiques. Recueil Sirey; 1931.

10. Hegyi G, Neda Z, Santos MA. Wealth distribution and Pareto's law in the Hungarian medieval society. Physica A: Statistical Mechanics and its Applications. 2007;380(1):271–277.

11. Kitov IO. Mechanical model of personal income distribution. arXiv preprint arXiv:09030203. 2009;.

12. Nirei M, Souma W. A two factor model of income distribution dynamics. Review of Income and Wealth. 2007;53(3):440–459.

13. Nirei M. Pareto distributions in economic growth models. IIR Working Paper WP#09-05. 2009;.

14. Mandelbrot B. The Pareto-Levy law and the distribution of income. International Economic Review. 1960;1(2):79–106.

15. Pianegonda S, Iglesias JR, Abramson G, Vega JL. Wealth redistribution with conservative exchanges. Physica A: Statistical Mechanics and its Applications. 2003;322:667–675.

16. Silva AC, Yakovenko VM. Temporal evolution of the "thermal" and "superthermal" income classes in the USA during 1983–2001. EPL (Europhysics Letters). 2004;69(2):304.

17. Simon HA. On a class of skew distribution functions. Biometrika. 1955;42(3/4):425–440.

18. Scheffer M, Bavel Bv, Leemput IAvd, Nes EHv. Inequality in nature and society. Proceedings of the National Academy of Sciences. 2017; p. 201706412. doi:10.1073/pnas.1706412114.

19. Rutherford RSG. Income distributions: a new model. Econometrica: Journal of the Econometric Society. 1955;23(3):277–294.

20. Yakovenko VM, Rosser Jr JB. Colloquium: Statistical mechanics of money, wealth, and income. Reviews of Modern Physics. 2009;81(4):1703.

21. Wold HO, Whittle P. A model explaining the Pareto distribution of wealth. Econometrica, Journal of the Econometric Society. 1957;25(4):591–595.

22. Leontief W. Theoretical assumptions and nonobserved facts. American Economic Review. 1971;61(1):1–7.

23. Barroso FG, Alados CL, Boza J. Social hierarchy in the domestic goat: effect on food habits and production. Applied Animal Behaviour Science. 2000;69(1):35–53.

24. Guhl AM, Collias NE, Allee WC. Mating behavior and the social hierarchy in small flocks of white leghorns. Physiological Zoology. 1945;18(4):365–390.

25. Kondo S, Hurnik JF. Stabilization of social hierarchy in dairy cows. Applied Animal Behaviour Science. 1990;27(4):287–297.

26. Meese GB, Ewbank R. The establishment and nature of the dominance hierarchy in the domesticated pig. Animal Behaviour. 1973;21(2):326–334.

27. Sapolsky RM. The influence of social hierarchy on primate health. Science. 2005;308(5722):648–652.

28. Uhrich J. The social hierarchy in albino mice. Journal of Comparative Psychology. 1938;25(2):373.

29. Bradley BJ, Robbins MM, Williamson EA, Steklis HD, Steklis NG, Eckhardt N, et al. Mountain gorilla tug-of-war: silverbacks have limited control over reproduction in multimale groups. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(26):9418–9423.

30. Haley MP, Deutsch CJ, Le Boeuf BJ. Size, dominance and copulatory success in male northern elephant seals, Mirounga angustirostris. Animal Behaviour. 1994;48(6):1249–1260.

31. Girman DJ, Mills MGL, Geffen E, Wayne RK. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (Lycaon pictus). Behavioral Ecology and Sociobiology. 1997;40(3):187–198.

32. Gerloff U, Hartung B, Fruth B, Hohmann G, Tautz D. Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (Pan paniscus) determined from DNA analysis of faecal samples. Proceedings of the Royal Society of London B: Biological Sciences. 1999;266(1424):1189–1195.

33. Wroblewski EE, Murray CM, Keele BF, Schumacher-Stankey JC, Hahn BH, Pusey AE. Male dominance rank and reproductive success in chimpanzees, Pan troglodytes schweinfurthii. Animal Behaviour. 2009;77(4):873–885.

34. Lewontin R, Levins R. Biology under the influence: Dialectical essays on the coevolution of nature and society. New York: Monthly Review Press; 2007.

35. Pinker S. The better angels of our nature: Why violence has declined. New York: Penguin Books; 2011.

36. Gómez JM, Verdú M, González-Megías A, Méndez M. The phylogenetic roots of human lethal violence. Nature. 2016;538(7624):233.

37. Dahrendorf R. Class and class conflict in industrial society. Stanford: Stanford University Press; 1959.

38. Giddens A. The class structure of the advanced societies. Hutchinson London; 1973.

39. Marx K. Capital, Volume I. Harmondsworth: Penguin/New Left Review; 1867.

40. Reardon SF, Bischoff K. Income Inequality and Income Segregation. American Journal of Sociology. 2011;116(4):1092–1153. doi:10.1086/657114.

41. Robinson RV, Kelley J. Class as conceived by Marx and Dahrendorf: Effects on income inequality and politics in the United States and Great Britain. American Sociological Review. 1979; p. 38–58.

42. Weber M. Economy and society: An outline of interpretive sociology. Univ of California Press; 1978.

43. Wodtke GT. Social Class and Income Inequality in the United States: Ownership, Authority, and Personal Income Distribution from 1980 to 2010. American journal of sociology. 2016;121(5):1375–1415.

44. Wright EO. Approaches to class analysis. Cambridge University Press; 2005.

45. Wright EO. Class structure and income determination. vol. 2. New York: Academic Press; 1979.

46. Wright EO, Perrone L. Marxist class categories and income inequality. American Sociological Review. 1977;42(1):32–55.

47. Wright EO. Race, class, and income inequality. American Journal of Sociology. 1978;83(6):1368–1397.

48. Mones E, Vicsek L, Vicsek T. Hierarchy measure for complex networks. PloS one. 2012;7(3):e33799.

49. Nitzan J, Bichler S. Capital as Power: A Study of Order and Creorder. New York: Routledge; 2009.

50. Brown C. Income Distribution in an Institutional World. In: Mangum G, Philips P, editors. Three Worlds of Labor Economics. vol. 51; 1988. p. 515–530.

51. Brown C. Is there an institutional theory of distribution? Journal of Economic Issues. 2005;39(4):915–931.

52. Commons JR. Legal foundations of capitalism. Transaction Publishers; 1924.

53. Galbraith JK. The anatomy of power. Boston, Mass.: Houghton Mifflin Co.; 1985.

54. Huber E, Huo J, Stephens JD. Power, policy, and top income shares. Socio-Economic Review. 2017;0(0):1–23. doi:10.1093/ser/mwx027.

55. Lenski GE. Power and privilege: A theory of social stratification. UNC Press Books; 1966.

56. Mills CW. The power elite. Oxford University Press; 1956.

57. Peach JT. Distribution and economic progress. Journal of Economic Issues. 1987;21(4):1495–1529.

58. Price TD, Feinman GM. Social inequality and the evolution of human social organization. In: Pathways to power. Springer; 2010. p. 1–14.

59. Sidanius J, Pratto F. Social dominance: An intergroup theory of social hierarchy and oppression. Cambridge University Press; 2001.

60. Veblen T. The Theory of Business Enterprise. New York: Martino Fine Books; 1904.

61. Veblen T. Absentee ownership: business enterprise in recent times: the case of America. Transaction Pub; 1923.

62. Simon HA. The compensation of executives. Sociometry. 1957;20(1):32–35.

63. Lydall HF. The distribution of employment incomes. Econometrica: Journal of the Econometric Society. 1959;27(1):110–115.

64. Roberts DR. A general theory of executive compensation based on statistically tested propositions. The Quarterly Journal of Economics. 1956;70(2):270–294.

65. Becker GS. Investment in human capital: A theoretical analysis. Journal of political economy. 1962;70(5, Part 2):9–49.

66. Mincer J. Investment in human capital and personal income distribution. The journal of political economy. 1958;66(4):281–302.

67. Schultz TW. Investment in human capital. The American Economic Review. 1961;51(1):1–17.

68. Axtell RL. Zipf distribution of US firm sizes. Science. 2001;293:1818–1820.

69. Fix B. Energy and Institution Size. PLOS ONE. 2017;12(2):e0171823. doi:doi:10.1371/journal. pone.0171823.

70. Gaffeo E, Gallegati M, Palestrini A. On the size distribution of firms: additional evidence from the G7 countries. Physica A: Statistical Mechanics and its Applications. 2003;324(1–2):117–123. doi:10.1016/S0378-4371(02)01890-3.

71. Piketty T. Capital in the Twenty-first Century. Cambridge: Harvard University Press; 2014.

72. Shorrocks AF. The class of additively decomposable inequality measures. Econometrica: Journal of the Econometric Society. 1980;48(3):613–625.

73. Di Muzio T. The 1% and the rest of us: A political economy of dominant ownership. Zed Books Ltd.; 2015.

74. Fix B. Putting Power Back into Growth Theory. Review of Capital as Power. 2015;1(2):1–37.

75. Clark JB. The Distribution of Wealth. New York: Macmillan; 1899.

76. Wicksteed PH. An Essay on the Co-ordination of the Laws of Distribution (1932 Edition). London: London School of Economics; 1894.

77. Carr DL, Markusen JR, Maskus KE. Estimating the knowledge-capital model of the multinational enterprise. The American Economic Review. 2001;91(3):693–708.

78. Lööf H, Heshmati A. Knowledge capital and performance heterogeneity:: A firm-level innovation study. International Journal of Production Economics. 2002;76(1):61–85.

79. Throsby D. Cultural capital. Journal of cultural economics. 1999;23(1):3–12.

80. Berle AA, Means GGC. The modern corporation and private property. Transaction Books; 1932.

81. NCEO. ESOPs by the Numbers; 2017. Available from: https://www.nceo.org/articles/esops-by-the-numbers.

82. Jarrow R, Rudd A. Approximate option valuation for arbitrary stochastic processes. Journal of financial Economics. 1982;10(3):347–369.

83. Carr P, Madan D. Option valuation using the fast Fourier transform. Journal of computational finance. 1999;2(4):61–73.

84. Core J, Guay W. Estimating the value of employee stock option portfolios and their sensitivities to price and volatility. Journal of Accounting research. 2002;40(3):613–630.

85. Mosca G. On the Ruling Class. New York: McGraw Hill; 1939.

86. Bendix R. Kings or people: Power and the mandate to rule. Berkeley: University of California Press; 1980.

87. Mishel L, Davis A. CEO pay continues to rise as typical workers are paid less. Issue Brief. 2014;380.

88. Bichler S, Nitzan J. A CasP model of the stock market. Real-World Economics Review. 2016;(77):119–154.

89. Mishel L, Schieder J. Stock market headwinds meant less generous year for some CEOs: CEO pay remains up 46.5% since 2009. Washington, D.C.: Economic Policy Institute; 2016. 109799. Available from: http://www.epi. org/publication/ceo-and-worker-pay-in-2015/.

90. Conyon MJ, Murphy KJ. The prince and the pauper? CEO pay in the United States and United Kingdom. The Economic Journal. 2000;110(467):640–671.

91. Gabaix X, Landier A. Why has CEO pay increased so much? The Quarterly Journal of Economics. 2008;123(1):49–100.

92. Krugman P. For richer. Critical social issues in American education: Democracy and meaning in a globalizing world. 2005;3.

93. Mishel L, Sabadish N. CEO pay and the top 1%. Economic Policy Institute Issue Brief. 2012;331:1–7.

94. Piketty T, Saez E. The evolution of top incomes: a historical and international perspective. National Bureau of Economic Research; 2006.

95. Atkinson AB. Pareto and the upper tail of the income distribution in the UK: 1799 to the present. Economica. 2017;84(334):129–156.

Chapter 5

Conclusion: A Glimpse of a Synthesis?

The great success of the natural sciences has been achieved substantially by the reduction of each physical phenomenon to its constituent elements, followed by the use of the elements to reconstitute the holistic properties of the phenomenon.

— E.O. Wilson [1]

As Wilson observes, science has two parts — reduction and synthesis. To understand a complex phenomenon, it is necessary to break it down into smaller parts. Once we understand these simple parts, the hope is that we can synthesize previously disparate branches of knowledge. In this dissertation, I have used social hierarchy as the 'constituent element' of social structure. I have shown that hierarchy plays an important role in many aspects of human society, from institution size, to energy consumption, to income distribution. This can be considered a 'reduction' (albeit a messy and incomplete one). Is there a corresponding synthesis? I think that there is. I believe that hierarchy offers the glimpse of a synthesis between economic growth (understood in biophysical terms) and income inequality. I conclude by discussing this synthesis. But before doing so, I want to reflect on the epistemology of 'economics from the top down'.

5.1   What is the Unit of Analysis in Economics?

When pursuing a reduction, the difficult question is this: how far down do we go? How do we know when we have isolated the 'constituent elements' of a system? This is a far more difficult question than many scientists would like to admit. Why? Because the answer depends both on what we are trying to understand, and on the limitations of our present knowledge.

As an example of this difficulty, consider how we might seek to understand cellular metabolism. If we chose a unit of analysis that is too large, we will reach

a dead end. It is obviously impossible to understand cellular metabolism without breaking the cell into smaller components (such as organelles and metabolic chemicals). But if we go too far down, we can't see the forest for the trees. Thus, it is not useful (at the present time) to reduce cellular metabolism to the level of sub-nuclear particles (i.e quarks). The art of doing science involves finding the happy medium — the unit of analysis that is small enough but not too small. And this happy medium should be flexible — it should change as science progresses. Fifty years ago our understanding of cellular metabolism stopped at the molecular level. But current evidence suggests that quantum effects (electron tunneling) play a fundamental role in metabolism [2–4] (for a non-technical overview, see [5]).

Because the 'how far down' question is so difficult, scientists have developed 'good tricks' [6] to make life easier . The principal 'good trick' is the academic discipline — a partitioned realm of investigation with an agreed upon unit of analysis. Disciplines rely on rules of thumb — something like: "if we want to understand phenomenon x, we agree that it is useful to focus on unit y". This trick is useful, because it allows science to proceed without the burden of constantly answering difficult epistemic questions.

While useful, we should remember that this good trick is precisely that — a *trick*. The problem comes when rules of thumb are treated as laws of the land. This is what has happened in economics. Consider Brennan and Tullock's dogmatic assertion used as the epigraph for this dissertation: "in modern economics ... the ultimate unit of analysis is *always* the individual; more aggregative analysis must be regarded as only provisionally legitimate [7] (emphasis added). In economics, a rule of thumb — focusing on the individual — has become a dogma.

Of course, the problem is not unique to economics. It is instructive to look at ongoing debates in other disciplines to get a sense for this type of epistemological problem (and to dispel the fog of disciplinary myopia that often pervades economics [8]). In Chapter 1, I quoted E.O Wilson's withering critique of modeling practices in biology [1] . Recently, Wilson has been involved in a heated debate about the proper unit of analysis in evolutionary biology [9]. The debate is over the validity of 'inclusive fitness' theory, which seeks to explain the origin of sociality and altruism at the level of the individual. I highlight this controversy in evolutionary biology because it parallels the problem in modern economics. Criticizing inclusive fitness theory, Allen, Nowak, and Wilson observe:

> The concept of inclusive fitness[1] arises when one attempts to explain the evo-
> lution of social behavior at the *level of the individual*. For example, inclusive

> fitness theory seeks to explain the existence of sterile ant workers in terms of the behaviors of the workers themselves. The proposed explanation is that workers *maximize their inclusive fitness* by helping the queen rather than producing their own offspring. ... Inclusive fitness theory attempts to find a universal design principle for evolution that applies at the level of the individual. The result is an *unobservable* quantity that ... has *no predictive or explanatory value*. [10] (emphasis added)

If we paraphrase Allen, Nowak, and Wilson, we arrive at a very cogent critique of standard economic theory. Like inclusive fitness theory, standard economic theory attempts to explain social phenomena at the level of the individual. To do so, the theory assumes that individuals maximize utility. The goal is to find a universal design principle for social structure that applies at the level of the individual. As critics have observed [11–14], the result is a theory based on an unobservable quantity that has no predictive or explanatory value.[2]

The concept of individual maximizing behavior has proved seductive for biologists and economists alike. Its great advantage is that it reduces the complexities of individual behavior to a single mathematical function that is to be optimized. This has the convenient effect of making models analytically tractable. And yet no one has ever observed this internal maximizing function — it is conveniently unobservable. And attempts to measure it externally in humans (by measuring the acceptance of monetary pay offs) have failed. Humans, it seems, do not maximize external pay offs [21]. Moreover, recent work in evolutionary biology suggests that, in general, natural selection does not lead individuals to act 'as if' maximizing *any* quantity [22].

Given the failures of the individual maximizing model, what are we to do? There are two choices: (1) revise our model of individual behavior; or (2) change the unit of analysis. The problem with the first solution is that when we abandon the maximizing model, we are confronted with a bewildering sea of

---

[1] In evolutionary biology, an organism's 'fitness' is defined as its expected number of offspring. Inclusive fitness adds to this concept the expected number of offspring of an individual's close relatives. The idea is that, by helping close relatives reproduce (instead of themselves), organisms can maximize the spread of their genes. The problem, according to Nowak et al., is that other than in contrived situations, inclusive fitness cannot be calculated [9].

[2] While there is a strong parallel between the critique of inclusive fitness theory and the critique of neoclassical marginal utility theory, the criticism has played out quite differently in each discipline. In biology, the critique of inclusive fitness [9] was published in a prestigious scientific journal (Nature), with much ensuing discussion [15–19]. In contrast, critiques of neoclassical economics are more or less excluded from mainstream economic journals. Seminal analysis by Joe Francis shows that economics has become a discipline largely devoid of debate [20].

complexity. Humans (and other organisms) likely have numerous goals. Some of them are conscious, and many of them are not. Moreover, these goals can be mutually conflicting, and they can change over time. Lastly, in humans, goals are influenced by culture, and vice versa. It is hubristic to think that with our present state of knowledge, we can capture this complexity in a model.

If revising our model of the individual is too difficult, that leaves changing the unit of analysis as the only viable option. In evolutionary biology, Allen, Nowak and Wilson argue that the way forward is to take the focus off individuals, and put it onto the gene [10]. Obviously this is not a fruitful way forward for economic theory — human cultural evolution occurs orders of magnitude faster than genetic evolution. Richard Dawkins has offered the *meme* as the cultural equivalent of the gene [23]. A meme is a piece of information (an idea) transmitted between humans. The problem with this sub-individual approach is that it still requires an understanding of the individual. To understand the transmission of ideas, one must have a theory of the mind.

Rather than move to a *sub*-individual unit of analysis, I have proposed that we move the focus to a *super*-individual unit — hierarchical organization. My contention is that social hierarchy is a useful (but not the ultimate) unit of analysis for understanding social phenomena. In analytical terms, hierarchy is convenient because its structure is simple — it is easy to model mathematically. But unlike the model of the maximizing individual, the structure of a hierarchy is observable.

Of course, mathematical simplicity is no guarantee that a unit of analysis is useful. This is an *empirical* question. The principal aim of this dissertation has been to show that focusing on hierarchy is a fruitful way to do empirical research. Of course we should absolutely seek to understand how and why hierarchies form, by appealing to a lower unit of analysis. But in the mean time, we can get on with the investigation by taking hierarchical structure as an empirical given. I have called this approach *economics from the top down*.

5.2   The Reduction

What I have done throughout this dissertation is to show that a wide variety of social phenomena are connected to social hierarchy. In essence, this constitutes a *reduction* — albeit a rough and incomplete one. I have reduced these phenomena to the 'constituent element' of social hierarchy. Let's review.

In Chapter 2, I found that increases in energy consumption (across both space and time) are associated with a systematic increase in institution size. I showed

Biophysical
Economic Growth   *Synthesis?*   Income Distribution

Evidence for a Power Theory of
Personal Income Distribution

Energy and Institution Size

A Hierarchy Model of Income Distribution
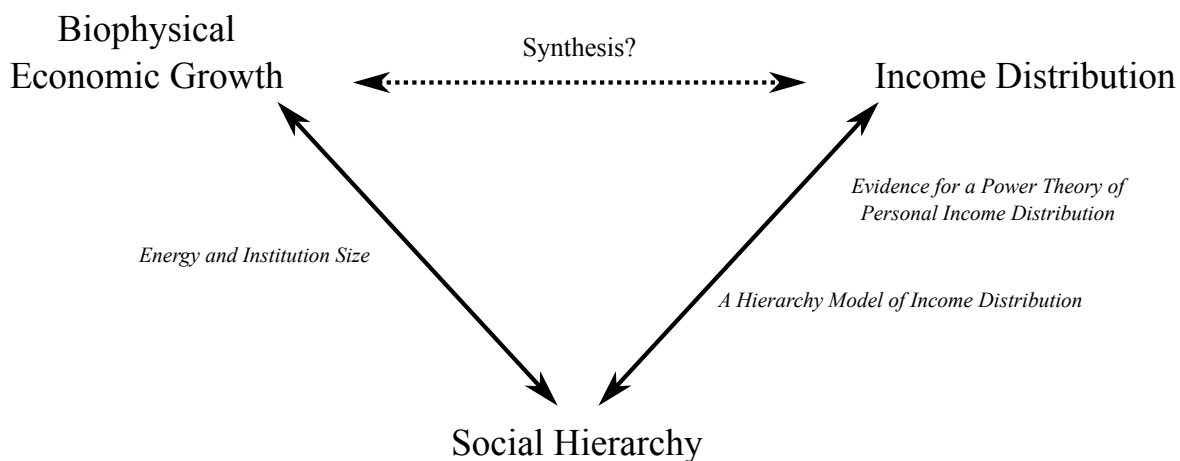
Social Hierarchy

Figure 5.1: A 'Reduction' Looking for a Synthesis

This figure shows how I conceive the big-picture structure of this dissertation. Each of the three papers (Ch. 2-4) connect either biophysical economic growth or income distribution to social hierarchy (this constitutes a sort of 'reduction'). But this connection begs a question: are growth and income distribution also related? I explore this possibility in Chapter 5.

---

that these changes can be understood as an overall increase in the hierarchical structure of society. I gave a possible reason why this occurs: technological change requires increasing social coordination which (due to human biological limitations) is achieved through hierarchical organization.[3]

In Chapter 3, I explored the relation between hierarchical power and personal income. Measuring hierarchical power in terms of the number of subordinates under one's control, I found that income within firms scales strongly (in both static and dynamic terms) with hierarchical power. Moreover, I found that grouping individuals by hierarchical power affected income more strongly than any other factor tested.

In Chapter 4, I explored the wider connection between hierarchy and in-

---

[3] Commenting on my work, Bichler and Nitzan have offered an alternative explanation for why hierarchy relates to energy consumption [24]. They propose that hierarchy is an unfortunate byproduct of increasing energy consumption. They believe hierarchy formation results from an innate human drive to accumulate power — a drive that is separate from the social process of technological development. I welcome this alternative hypothesis. Since the energy-hierarchy link is a very new empirical finding, it is wise to explore many possible explanations for its existence. However, for the arguments in this chapter, the reason for the energy-hierarchy connection is less important than the connection itself.

come distribution. Using an empirically informed model, I found that hierarchy plays a dominant role in shaping top incomes. The model suggested that the power-law distribution of top incomes — a celebrated empirical regularity — is a consequence of hierarchical organization. Moreover, I found that hierarchy seems to play a role in shaping the composition of top incomes, and the dynamics of income inequality.

To summarize, I have shown that both biophysical economic growth and income distribution can be reduced (again roughly and incompletely) to the constituent element of social hierarchy (see Fig. 5.1). This hints at a tantalizing link between growth and income distribution themselves.

5.3   A Synthesis of Growth and Distribution?

The cumulative findings in this dissertation suggest a relation between biophysical growth and income distribution. The basic idea is that increasing energy consumption is associated with increasing hierarchical organization, which causes the concentration of power and the concentration of income. Of course, the idea that growth should lead to increasing inequality is not new. Henry George thought as much a century ago:

> Where the conditions to which material progress everywhere tends are most fully realized — that is to say, where population is densest, wealth greatest, and the machinery of production and exchange most highly developed — we find the deepest poverty, the sharpest struggle for existence, and the most enforced idleness. [25]

The growth-inequality link is also implicit in the surplus theory of social stratification that is popular in anthropology and sociology [26–36].[4]   While the growth-inequality hypothesis is not new, the results in this dissertation allow us to move from a qualitative to a quantitative discussion. In particular, the evidence in Chapter 2 effectively provides an *empirical* relation between hierarchy and energy consumption. With a few assumptions, we can use the hierarchy model (developed in Ch. 4) to make concrete predictions about how the concentration of hierarchical power (and with it income) might change with energy consumption. To my knowledge, this prediction is the first of its kind.

---

[4]In the surplus theory of social stratification, the thinking goes something like this. The majority of the population are 'producers' who produces more than they consume. This economic surplus is controlled by a small group of elites. Therefore, as the surplus increases (via economic growth), so does inequality. Unfortunately this approach has many problems. One is that it requires differentiating between 'producers' and 'non-producers' (or productive and unproductive

An Energy-Hierarchy Model

To explore the relation between energy and hierarchy, I take the hierarchy model used in Chapter 4, and tack onto it the empirical relation between energy and institution size discovered in Chapter 2. I call the result the 'energy-hierarchy' model. This model creates a simulated economy of hierarchical institutions exactly like the original hierarchy model. However, this time I allow the institution size distribution to *vary* (by changing the power exponent).

When institutions are very small, little hierarchical structure will exist. Conversely, when institution size is larger more hierarchical structure will exist. For each different model iteration, I calculate mean institution size and then use the energy-firm-size regression (shown in Fig. 2.1C) to predict the level of energy consumption. Like the original hierarchy model, the energy-hierarchy model is essentially an extrapolation. It takes the available evidence and extrapolates it to make a prediction about how energy consumption should relate to hierarchical organization. Using this model requires the following assumptions:

1. Institution size is distributed according to a power law. Changes in the institution size distribution correspond to a change in the power law exponent.

2. The hierarchical structure of institutions is constant across time and equivalent to that found in modern firm case studies (Appendix C.2).

3. The modern trend between energy use per capita and firm size is applicable to non-capitalist societies (see Fig. 2.1 in Ch. 2 for this trend).

These assumptions are meant to allow exploratory analysis — they justify projecting modern trends into the past. Are they realistic? Regarding assumption 1, there is evidence that pre-capitalist societies had a power law distribution of institution size. Obviously, what constitutes an 'institution' will change in different societies. In feudal societies, we might imagine that the feudal manor is

---

labor). But this cannot be done objectively [13]. Furthermore, the concept of surplus is impossibly fuzzy [37]. Should it be measured in *absolute* terms — anything above bare subsistence needs? If so, Milanovic shows that we should expect economic growth to be associated with runaway inequality [38, 39]. But this is not what has happened historically. Perhaps surplus should be measured in *relative* terms? This would allow for changes in the standards of living of the 'producer' population. Unfortunately this definition means that the size of the surplus is dependent on the *distribution of income*. Needless to say, this relative view of surplus cannot be used to make predictions about income distribution, since the measurement of surplus depends on distribution in the first place.

Subsistence Society
Energy Use per Capita ~ 5 GJ / Year
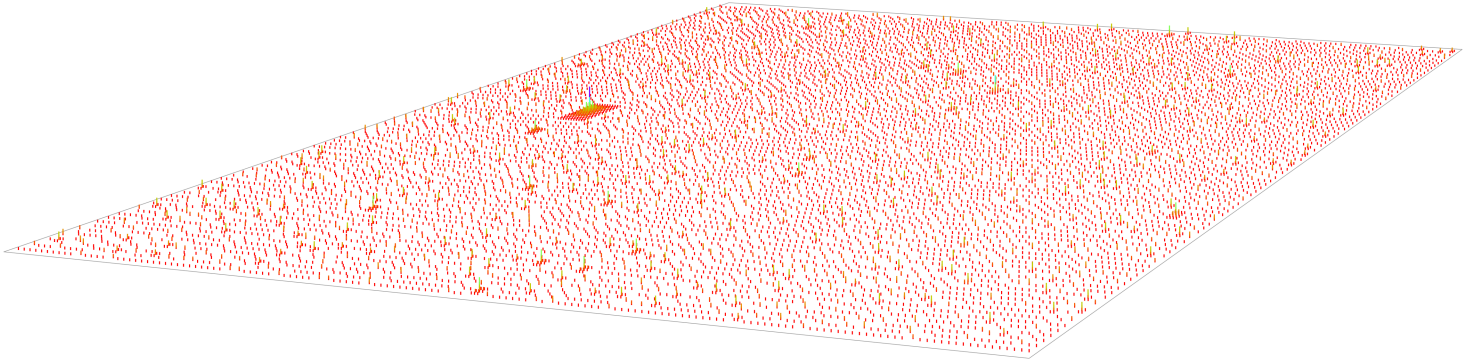
Industrial Society
Energy Use per Capita ~ 300 GJ / Year
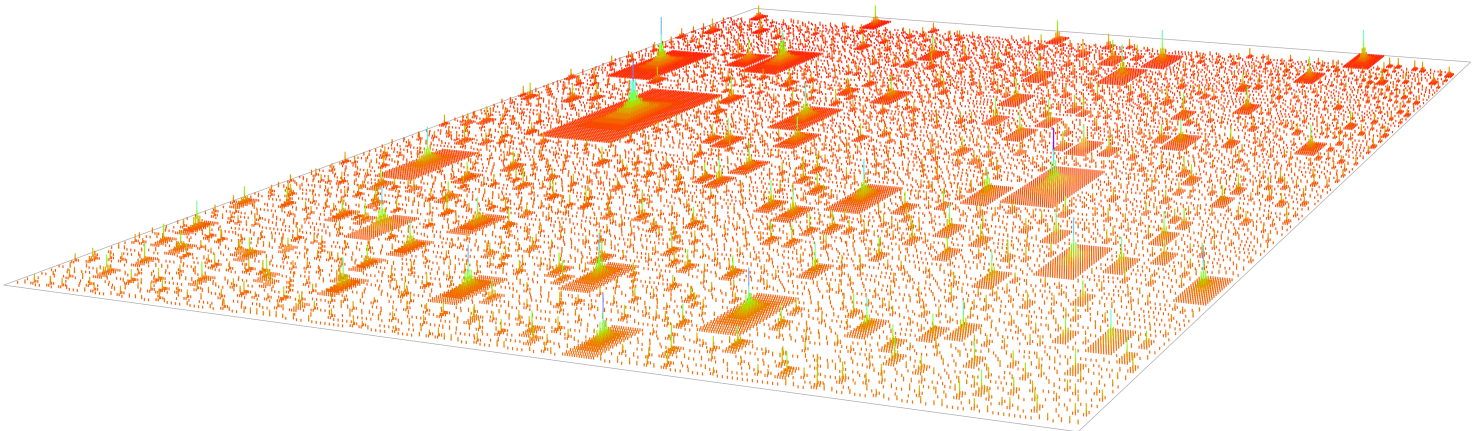
Figure 5.2: Visualizing the Energy-Hierarchy Model

This figure shows the hierarchical structure of two hypothetical societies. Hierarchies are visualized as pyramids (color indicates hierarchical rank). The top panel shows a low-energy society that consumes 5 GJ of energy per capita per year —- not much more than daily caloric food requirements. In such a society, the model predicts very little hierarchical organization. In contrast, the bottom panel shows a high-energy industrial society with energy use on par with the modern United States. Such a society is predicted to have a considerable amount of hierarchical organization.

the dominant institution. There is evidence that feudal manors were power-law distributed. For instance Hegyi et al. find an approximate power law distribution of serf ownership by nobles/aristocrats in 16th century Hungary [40]. Similarly, Kahan finds a highly skewed distribution of serf ownership in 18th century Russia [41] (although this distribution is better fit with a lognormal function). In hunter-gatherers societies, we might imagine that institutions consist of families, clans, and tribes. Because hunter-gatherer societies are largely prehistoric, the archaeological evidence is all that remains. On this front, recent evidence suggests that hunter-gatherer settlement sizes had a power law distribution (in the tail) [42]. While the evidence is limited, assumption 1 (a power-law institution size distribution) seems reasonable.

What about assumption 2 (the hierarchical structure of institutions is constant across time)? Unfortunately, we know very little about the 'shape' of hierarchal institutions in pre-capitalist societies. What we do know is that in agrarian and hunter-gatherer societies, there is a reliable scaling relation between population size and the number of hierarchical levels of socio-political organization [43,44]. This is evidence that these societies were hierarchical. However, I am not aware of any work on the micro-structure of hierarchy in pre-capitalist societies. As such, assumption 2 is speculative.

Lastly, the assumption 3 (the modern energy-institution-size relation applies to non-capitalist societies) is purely speculative at the present time. But given an empirical trend, why not extrapolate it and see where it takes us?

To get an intuitive sense for what the energy-hierarchy model looks like, Figure 5.2 visualizes it in landscape form. I show the modeled hierarchical structure of two very different societies. The top image shows a hypothetical subsistence society that consumes 5GJ of energy per capita per year. (This is equivalent to 3200 Kcal per day — not much above metabolic needs of an average human). In this subsistence society, the energy-hierarchy model predicts that very little hierarchical organization should exist. In contrast, the bottom image shows a hypothetical industrial society that consumes 300GJ of energy per capita per year. This rate is on par with the modern United States. The model predicts that such a society should have significant hierarchical organization.

A Prediction: Energy and the Concentration of Hierarchical Power

The most important feature of the energy-hierarchy model is that it makes a quantitative prediction about how hierarchical power concentration should vary with energy consumption. To make this prediction, I run the model many times,
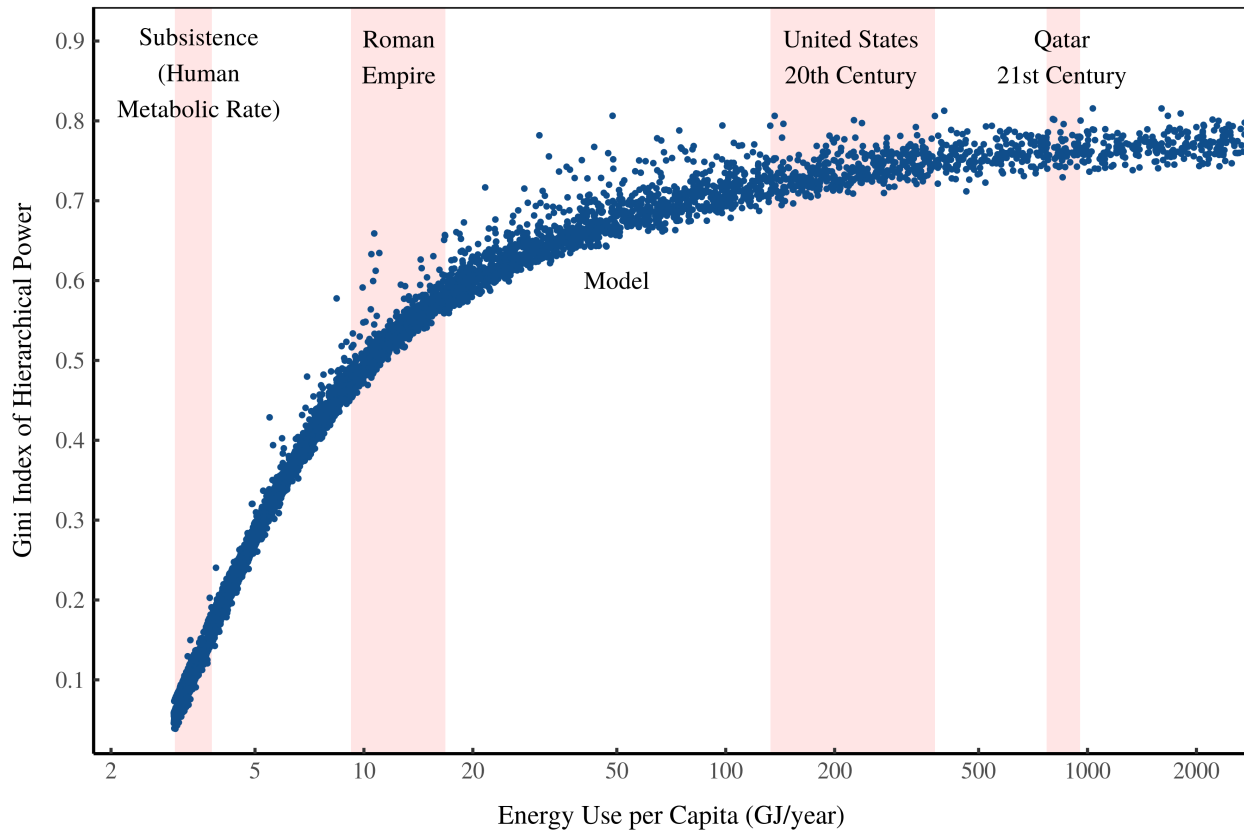
Figure 5.3: Hierarchical Power Concentration and Energy Use per Capita

This figure show the results of the energy-hierarchy model, produced by stochastically varying the institution size distribution. Each dot indicates a different model iteration. Shaded regions show the energy consumption range for various types of real-world societies. Sources: Qatar data comes from the World Bank (series EG.USE. PCAP.KG.OE). US total energy consumption is from HSUS, Tables Db164-171 (1890-1948) and EIA Table 1.3 (1949-2012). US population is from Maddison [45] (1890-2009) and World Bank series SP.POP.TOTL (2010-2012). Roman Empire data comes from Malanima [46]. Human metabolic needs are assumed to range from 2000 Kcal to 2500Kcal per day.

each with a different institution size distribution. For each iteration, I calculate the Gini index of hierarchical power concentration (as I did in Fig. 3.4). Over many iterations, the model produces the relation shown in Figure 5.3.

The results are interesting. Virtually all of the increases in hierarchical power concentration are predicted to occur during the transition from subsistence to agrarian levels of energy consumption. (Here the Roman Empire serves as the example of an agrarian society). Counterintuitively (to me, at least), the model predicts that further increases of energy consumption to industrial levels should have little impact on the concentration of hierarchical power. Note also that the model predicts a collapse of hierarchical power concentration when energy consumption is at subsistence levels. (Given the uncertainty in the evidence underlying the energy-hierarchy model, this result is serendipitous).

What we are ultimately interested in is the relation between inequality and growth. In Chapter 3, we found that income scales strongly with hierarchical power. Therefore, all other things being equal, we expect that greater concentrations of hierarchical power should lead to greater concentrations of income. In general terms, then, the energy-hierarchy model predicts that the transition from hunter-gatherer to agrarian levels of energy consumption should be associated with a significant increase in inequality. (The size of this increase will depend on how strongly income scales with hierarchical power). After this transition, we should expect more or less no relation between growth and inequality.

Is this prediction correct? The limited available evidence suggests that it is on the right track. Because there are few estimates of energy consumption for ancient societies, it is difficult to create an inequality-vs-energy plot that would allow a direct empirical comparison to Figure 5.3. However, what we can do is divide societies into different modes of energy capture. Figure 5.4 shows inequality estimates based on a division into four modes: hunter-gatherer, horticultural, agrarian, and industrial. I plot these modes in the (likely) order of increasing energy consumption. The caveat here is that the inequality data for pre-industrial societies is measured in terms of *house size* [47]. This is not strictly comparable to the *income* inequality data used for industrial societies.

Caveats aside, this evidence supports the basic prediction of the energy-hierarchy model. The evidence suggests that inequality increased rapidly during the transformation from hunter-gatherer to agrarian levels of energy consumption, and was then more or less unaffected by further increases in energy consumption. These results are certainly promising. They suggest that the origin of inequality is a consequence of the increasing hierarchical organization associated with increasing energy consumption. In big-picture terms, the energy-
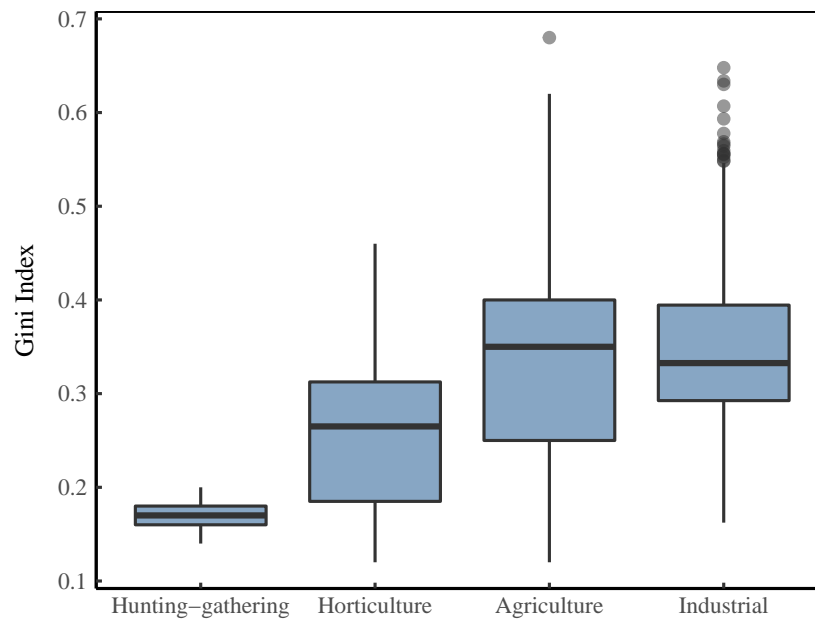
Figure 5.4: Inequality vs. Mode of Energy Capture

This figure shows how inequality relates to the mode of energy capture for four different types of societies. Inequality data for hunter-gatherers, horticulture, and agrarian societies comes from Kohler et al. [47] and is calculated using archaeological studies of house size. Inequality in industrial societies uses data from the World Bank, series SI.POV.GINI. I define an 'industrial' society as having an energy consumption above the (arbitrary) threshold of 50GJ per capita. The caveat here is that income inequality and household size inequality are not strictly comparable.

Notes: Kohler et al. conceive the house size Gini index as a metric of *wealth* inequality. Perhaps we should compare Kohler's data to modern levels of wealth inequality rather than income inequality (in Fig. 5.4)? I think this is unwise. House size inequality is most comparable to the inequality of equity on a principal residence. Evidence from Edward Wolff suggests that principal residence wealth inequality is far more equally distributed than total wealth inequality in the United States. In 2007, Wolff finds that the top 10% of households owned 39% of principal residence wealth, but 73% of total wealth [48]. In the same year, World Bank data (series SI.DST.10TH.10) indicates that the top 10% held 31% of total income. This suggests that house size inequality is far closer to income inequality than it is to wealth inequality.

hierarchy model offers a potential way to synthesize distribution and growth. In more narrow terms, the model unifies the cumulative results in this dissertation, and gives an example of how they can be applied.

5.4 Conclusion

It is common to present the history of science as a progressive march towards greater understanding. What is less discussed is the philosophical upheaval that accompanies this process. Noam Chomsky notes that scientific progress has consistently required lowering the philosophical bar of what constitutes 'understanding'. Early scientists like Galileo believed that 'understanding' came only if one could explain a natural phenomenon in *mechanical* terms:

> The mechanical philosophy provided the very criterion for intelligibility in the sciences. Galileo insisted that theories are intelligible, in his words, only if we can "duplicate [their posits] by means of appropriate artificial devices." The same conception, which became the reigning orthodoxy, was maintained and developed by the other leading figures of the scientific revolution: Descartes, Leibniz, Huygens, Newton, and others. [49]

While useful for kick starting the scientific revolution, the mechanical philosophy was slowly abandoned. Why? The fundamental forces (gravity, and later electromagnetism) seemed to require 'action at a distance'. Although necessary for his theory of gravitation, Newton himself regarded action at a distance as "so great an absurdity, that I believe no man who has in philosophical matters a competent faculty of thinking, can ever fall into it" [50]. Chomsky notes that the problem of action at a distance was not resolved by finding a mechanical explanation, but rather by tacitly lowering philosophical goals [49, 51]. Physicists abandoned the original goal of a mechanistic understanding of force, and instead became content with explaining forces in terms of the non-mechanical concept of a 'field'.

I believe that the social sciences have their own version of the mechanistic philosophy — namely the philosophy of methodological individualism. Consider Max Weber's standard for judging 'understanding':

> ... collectivities must be treated as solely the resultants and modes of organization of the particular acts of individual persons, since these alone can be treated as agents in a course of subjectively *understandable* action. [52] (emphasis added)

According to Weber, we do not really *understand* a social phenomenon unless we can explain it in terms of the purposeful action of individuals. If the model

of intelligibility for early scientists was "when mechanism fails, understanding fails" [51], the model of intelligibility for Weber (and many other social scientists) is 'when purposeful, individual action fails, understanding fails'.

I suspect that both of these philosophies are artifacts of our evolved minds. The mechanical philosophy arises because we intuitively understand the world in terms of material objects that 'touch' one another. Touch is one of our five senses, and is a key part of how we interpret world. Similarly, we intuitively assign rational intent to the agents (animals, people, and sometimes things) that we interact with. Daniel Dennet calls this the 'intentional stance', and he believes it is something that we evolved in order to predict the behavior of other agents. The intentional stance works as follows:

> ... first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent will do. [53]

Just like the sensation of 'touch', the intentional stance is highly functional — it gives us an intuition about how agents ought to behave. While it is crucial for our everyday lives as a social species, this does not mean that assigning rational intentionality is correct in a scientific sense. One could argue that scientific advance has involved a progressive *limitation* of the intentional stance. The first step was to remove intentionality from inanimate matter. For instance, weather is now understood in terms of the laws of physics, and not in terms of the action of a purposeful god. Next, intentionality was removed from the study of non-human life. In modern biology, purposeful action is not a criterion for understanding animal behavior. The social sciences appear to be the last hold out of the intentional stance — and for good reason. It is extremely difficult for us to consider the behavior of a fellow human without assigning rational intention. Doing so would mean questioning our own rationality — something most people would abhor. But if we look at the course of science, it seems plausible that a theory of human behavior will not involve intentionality.

To be sure, such a theory is a long way off (if not forever beyond our grasp). This implies that understanding social structure in terms of the purposeful action of individuals may be too lofty a goal. I believe that we (social scientists) should lower our goals for judging 'understanding'. This dissertation has attempted

to do so by investigating social structure not in terms of individual action, but in terms of *social hierarchy*. In this context, hierarchy represents an empirical regularity in our collective behavior. Hierarchical organization is simple enough that we can model it — and explore its implications — without an understanding of how or why it exists. In Chomsky's terms, this approach amounts to lowering our philosophical goals in order to advance science.

Lofty philosophical reasons aside, there are more immediate *political* reasons to focus on hierarchy and not individuals. Focusing on individuals makes it extremely difficult to study *social power*, which is an inherently collective phenomenon. This is no accident. Joseph Heath observes that methodological individualism was appealing to early 20th century thinkers such as Hayek [54–56] and Popper [57–59] because it *negated* the study of collective power:

> For both Hayek and Popper, the primary motivation for respecting the precepts of methodological individualism was to avoid "grand theory" in the style of Auguste Comte, G.W.F. Hegel and Karl Marx. Yet the motivation for avoiding this sort of grand theory was not so much that it promoted bad theory, but that it promoted habits of mind, such as "collectivism," "rationalism," or "historicism," that were thought to be conducive to totalitarianism. Thus the sins of "collectivism," and "collectivist" thought patterns, for both Hayek and Popper, were primarily political. [60]

Interestingly, the political unease with collective power was typically one-sided and based on a fear of worker revolution (and not capitalist power). For instance, John Bates Clark's motivation for developing marginal productivity theory was based explicitly on his fear that workers would take Marxist theory seriously. Clark noted that if workers believed that they "produce an ample amount and get only a part of it, many of them would become revolutionists, and all would have the right to do so" [61]. This fear of worker revolution led to a theory that actively denied the existence of power (in all its forms). This denial is evident as a basic asymmetry in neoclassical economics: consumers are treated as *individuals* but producers are treated as black-box *firms* (a collectivity). Why not open up the firm and treat it as a collection of individuals? Presumably because one cannot do so without seeing *concentrated power*. Abraham Zaleznik observes:

> Whatever else organizations may be (problem-solving instruments, sociotechnical systems, reward systems, and so on), they are *political structures*. This means that organizations operate by distributing authority and setting a stage for the *exercise of power*. [62] (emphasis added)

This dissertation has sought to investigate the effects of social power by opening up firms and studying their internal power structure. My central hypothesis has been that firms are hierarchically organized, and that this hierarchical organization has implications for higher-level social structure (namely income distribution).

It is important to recognize that this type of research is in its infancy, and thus based on less data than we might prefer. I have done my best to tease out the implications of firm hierarchy based on a handful of case studies. More detailed analysis requires more data, which requires that empirical researchers be motivated to study hierarchy. Unfortunately this involves a Catch-22 type situation. So long as power-blind neoclassical economics is the emperor of the social sciences, there is little academic incentive to study hierarchy. Better data thus depends on more researchers realizing that the emperor 'has no clothes' [63].

We should expect that better data will lead to results that differ from those discussed in this dissertation. This is the way that empirical science works. When Edwin Hubble discovered the expansion of the universe, he got the rate wrong by an order of magnitude [64]. But while wrong about the specific value, Hubble was correct that the universe is expanding. In a similar way, the regression coefficients and model parameters used in this dissertation are almost certainly open to revision. However, the key trends that I have documented seem fairly secure. These trends are as follows.

1. Social hierarchy is not constant throughout human history; rather, the evidence suggests that hierarchy has increased over time.

2. Increases in hierarchy are accompanied by an increase in energy consumption. This suggests that hierarchical organization has a biophysical basis.

3. Hierarchy plays a key role in shaping income and income distribution.

Having identified these trends/facts, what are some avenues for future research? The first, and most important, is replication. The hypotheses advanced in this dissertation need to be tested using better data. To that end, I have made all of my data and code available for researchers wishing to do replication research. Second, we need to study how and why hierarchies form, and place this hierarchy formation in a biophysical context. The work done in Chapter 2 is preliminary, and needs more investigation. Third, whenever we uncover generalities in social behavior, the door opens for *comparative* work that studies

departures from the general trend. It would be fascinating to conduct a comparative study of the different ways that hierarchy relates to income distribution in various societies (now and in the past). Lastly, we should look for ways to use the study of hierarchy to solve practical problems.

It is on this front that I conclude. I stated at the outset that 'economics from the top down' is not about prescriptive policy of any kind. I felt this was necessary to differentiate my approach from so-called 'top down' or 'supply side' economics — a policy first, facts second ideology. This may have given the impression that I am disinterested in policy, which I am not. Quite the opposite. My primary motivation for conducting this research was not an innate interest in hierarchy, but rather, a desire to find pathways to a better future. As I see it, the great challenge facing humanity is how to chart a path to sustainability in a way that is just and equitable.

This is a monumental task that will almost certainly involve great institutional change. The research in this dissertation gives hints at what this change should look like. If we want to reduce the scale of the economy (to lessen our impact on the biosphere) we will need to consume less energy. This means that we should seek smaller institutions and less hierarchy. But the paradox is this: unless the reduction of hierarchy is enormous, the evidence suggests that there is no guarantee that a future low-energy society will be more equitable. If greater equality is a goal (and I think it should be), we need to do one of two things: either we learn how to organize without hierarchy (a tall order), or we learn how to put checks on hierarchical power so that it does not lead to vast concentrations of income.

References

1. Wilson EO. Consilience: The unity of knowledge. New York: Random House; 1999.

2. Beratan DN, Onuchic JN, Winkler JR, Gray HB. Electron-tunneling pathways in proteins. Science. 1992;258(5089):1740–1742.

3. Chance B, Devault DC, Frauenfelder H. Tunneling in Biological Systems: A Colloquium of the Johnson Research Foundation. Academic Press; 2013.

4. Gray HB, Winkler JR. Electron tunneling through proteins. Quarterly reviews of biophysics. 2003;36(3):341–372.

5. Lane N. The vital question: energy, evolution, and the origins of complex life. New York: WW Norton & Company; 2015.

6. Dennett DC. Darwin's Dangerous Idea: Evolution and the Meanings of Life. London: Penguin Books; 1995.

7. Brennan G, Tullock G. An economic theory of military tactics: Methodological individualism at war. Journal of Economic Behavior & Organization. 1982;3(2-3):225–242.

8. Brown PG, Timmerman P. Ecological economics for the anthropocene: An emerging paradigm. New York: Columbia University Press; 2015.

9. Nowak MA, Tarnita CE, Wilson EO. The evolution of eusociality. Nature. 2010;466(7310):1057.

10. Allen B, Nowak MA, Wilson EO. Limitations of inclusive fitness. Proceedings of the National Academy of Sciences. 2013;110(50):20135–20139.

11. Hodgson GM. On the Limits of Rational Choice Theory. Economic Thought: History, Philosphy, and Methodology. 2012;1(1).

12. Lee FS, Keen S. The incoherent emperor: a heterodox critique of neoclassical microeconomic theory. Review of Social Economy. 2004;62(2):169–199.

13. Nitzan J, Bichler S. Capital as Power: A Study of Order and Creorder. New York: Routledge; 2009.

14. Robinson J. Economic philosophy. Chicago: Aldine Pub. Co.; 1962.

15. Abbot P, Abe J, Alcock J, Alizon S, Alpedrinha JAC, Andersson M, et al. Inclusive fitness theory and eusociality. Nature. 2011;471(7339):E1–E4. doi:10.1038/nature09831.

16. Boomsma JJ, Beekman M, Cornwallis CK, Griffin AS, Holman L, Hughes WOH, et al. Only full-sibling families evolved eusociality. Nature. 2011;471(7339):E4–E5. doi:10.1038/nature09832.

17. Nowak MA, Tarnita CE, Wilson EO. Nowak et al. reply. Nature. 2011;471(7339):E9–E10. doi:10.1038/nature09836.

18. Ferriere R, Michod RE. Inclusive fitness in evolution. Nature. 2011;471(7339):E6–E8. doi:10.1038/nature09834.

19. Birch J. The inclusive fitness controversy: finding a way forward. Royal Society Open Science. 2017;4(170335). doi:10.1098/rsos.170335.

20. Francis J. The Rise and Fall of Debate in Economics; 2014. Available from: http://www.joefrancis.info/economics-debate/.

21. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, et al. In search of homo economicus: behavioral experiments in 15 small-scale societies. American Economic Review. 2001;91(2):73–78.

22. Lehmann L, Alger I, Weibull J. Does evolution lead to maximizing behavior? Evolution. 2015;69(7):1858–1873.

23. Dawkins R. The Selfish Gene. Oxford: Oxford University Press; 1976.

24. Bichler S, Nitzan J. Growing Through Sabotage: Energizing Hierarchical Power. Capital as Power Working Papers. 2017;2017(02).

25. George H. Progress and poverty: An enquiry into the cause of industrial depressions, and of increase of want with increase of wealth. The Remedy. New York: K. Paul, Trench & Company; 1879.

26. Cancian F. Social stratification. Annual Review of Anthropology. 1976;5(1):227–248.

27. Childe V. Man makes himself. London: Watts; 1936.

28. Engels F, Hunt T. The origin of the family, private property and the state. Zurich: Verlag der Schweizerischen Volksbuchhandlung; 1884.

29. Gilman A. The development of social stratification in Bronze Age Europe. Current anthropology. 1981;22(1):1–23.

30. Henrich J, Boyd R. Division of labor, economic specialization, and the evolution of social stratification. Current Anthropology. 2008;49(4):715–724.

31. Lenski GE. Power and privilege: A theory of social stratification. UNC Press Books; 1966.

32. Lenski G. Ecological-evolutionary theory: Principles and applications. London: Paradigm Publishers; 2015.

33. Orans M. Surplus. Human Organization. 1966;25(1):24–32.

34. Sahlins MD. Social stratification in Polynesia. vol. 29. University of Washington Press; 1972.

35. Price TD, Feinman GM. Foundations of social inequality. vol. 1. New York: Springer Science & Business Media; 1995.

36. Price TD, Feinman GM, editors. Pathways to power: New Perspectives on the Emergence of Social Inequality. New York: Springer; 2010.

37. Pearson HW. The economy has no surplus: critique of a theory of development. In: Trade and market in the early empires. The Free Press; 1957. p. 320–41.

38. Milanovic B. Ancient inequality. NBER WORKING PAPER SERIES. 2008;Working Paper 13550.

39. Milanovic B. The inequality possibility frontier: extensions and new applications. World Bank Policy Research Working Paper. 2013;WPS6449.

40. Hegyi G, Neda Z, Santos MA. Wealth distribution and Pareto's law in the Hungarian medieval society. Physica A: Statistical Mechanics and its Applications. 2007;380(1):271–277.

41. Kahan A, Hellie R. The plow, the hammer, and the knout: An economic history of eighteenth-century Russia. Chicago: University of Chicago Press; 1985.

42. Haas Jr WR, Klink CJ, Maggard GJ, Aldenderfer MS. Settlement-size scaling among prehistoric hunter-gatherer settlement systems in the New World. PloS one. 2015;10(11):e0140127.

43. Hamilton MJ, Milne BT, Walker RS, Burger O, Brown JH. The complex structure of hunter–gatherer social networks. Proceedings of the Royal Society of London B: Biological Sciences. 2007;274(1622):2195–2203. doi:10.1098/rspb.2007.0564.

44. Turchin P, Gavrilets S. Evolution of complex hierarchical societies. Social Evolution and History. 2009;8(2):167–198.

45. Maddison A. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD; 2008. Available from: http://www.ggdc.net/maddison/Maddison.htm.

46. Malanima P. Energy Consumption in the Roman World. In: The Ancient Mediterranean Environment between Science and History. Boston: Brill; 2013. p. 13–36.

47. Kohler TA, Smith ME, Bogaard A, Feinman GM, Peterson CE, Betzenhauser A, et al. Greater post-Neolithic wealth disparities in Eurasia than in North America and Mesoamerica. Nature. 2017;551(7682):619–622. doi:10.1038/nature24646.

48. Wolff EN. Recent trends in household wealth in the United States: Rising debt and the middle-class squeeze-an update to 2007. Levy Economics Institute. 2010;Working Paper No. 589.

49. Chomsky N. Science, Mind, and Limits of Understanding; 2014. Available from: https://chomsky.info/201401__/.

50. Cohen IB, editor. Isaac Newton's papers and letters on natural philosophy and related documents. Cambridge, Mass.; 1961.

51. Chomsky N. Chomsky Notebook. Columbia University Press; 2010.

52. Weber M. Economy and society: An outline of interpretive sociology. Univ of California Press; 1978.

53. Dennett DC. The intentional stance. Cambridge: MIT press; 1989.

54. Hayek Fv. Scientism and the study of society. Part I. Economica. 1942;9(35):267–291.

55. Hayek Fv. Scientism and the study of society. Part II. Economica. 1943;10(37):34–63.

56. Hayek Fv. Scientism and the study of society. Part III. Economica. 1944;11(41):27–39.

57. Popper K. The Poverty of Historicism, I. Economica. 1944;11(42):86–103.

58. Popper K. The poverty of historicism, II. A Criticism of Historicist Methods. Economica. 1944;11(43):119–137.

59. Popper K. The poverty of historicism, III. Economica. 1945;12(46):69–89.

60. Heath J. Methodological Individualism. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. 2015th ed. Metaphysics Research Lab, Stanford University; 2015.Available from: https://plato.stanford.edu/archives/spr2015/entries/methodological-individualism/.

61. Clark JB. The Distribution of Wealth. New York: Macmillan; 1899.

62. Zaleznik A. Power and Politics in Organizational Life. Harvard Business Review. 1970;(May 1970).

63. Keen S. Debunking Economics: The Naked Emperor of the Social Sciences. New York: Zed Books; 2001.

64. Huchra JP. The Hubble Constant; 2008. Available from: https://www.cfa.harvard.edu/~dfabricant/huchra/hubble/.

Appendix A

Appendices for Energy and Institution Size

Supplementary materials for this paper are available at PLOS ONE, where the published version resides:

The supplementary materials include:

1. Data for all figures appearing in the paper;
2. Raw source data;
3. R code for all analysis and modeling;

Acronyms

BEA US Bureau of Economic Analysis

BLS US Bureau of Labor Statistics

EIA US Energy Information Agency

HSUS Historical Statistics of the United States

ILO International Labour Organization

GEM Global Entrepreneurship Monitor

WBES World Bank Enterprise Survey

A.1   Sources and Methodology

Electricity Use per Capita

US electricity use is from HSUS table Db228 (1920 - 1948) spliced to EIA table 7.1, *Electricity End Use, Total* (1949-2015). US population is from Maddison [1] (1920-2009) and World Bank series SP.POP.TOTL (2010-2015).

Energy Use per Capita – International

International energy use per capita data is from the World Bank (series EG.USE. PCAP.KG.OE).

Energy Use per Capita – United States

US total energy consumption is from HSUS, Tables Db164-171 (1890-1948) and EIA Table 1.3 (1949-2012). US population is from Maddison [1] (1890-2009) and World Bank series SP.POP.TOTL (2010-2012).

Energy Use per Capita – US Industry

US Industry energy use is from EIA Table 2.1 (Energy Consumption by Sector). Industry employment is from BEA Table 6.8B-D (Persons Engaged in Production by Industry), where 'Industry' is defined to include Mining, Manufacturing and Construction.

Energy Use per Capita – US Manufacturing Subsectors

US manufacturing sub-sector energy use is from EIA Manufacturing Energy Consumption Survey Table 1.1 (First Use of Energy for All Purposes) 2002, 2006, and 2010. Manufacturing subsector employment is from Statistics of U.S. Businesses (US 6 digit NAICS) for 2002, 2006, and 2010.

Firm Age Composition

The fraction of firms under 42 months old (3.5 years) is calculated from the GEM dataset aggregated over the years 2001-2011 (data series *babybuso*). This series gives true/false values for whether or not a given firm is under 42 months old. Uncertainty in this data is estimated using the bootstrap method [2].

Firm Age Model

In order to model firm age accurately, I use a time step interval of 0.5 years (this allows us to calculate firms under 3.5 years so that we can compare to GEM data). However, most empirical data on firm growth rates are reported with a time interval of 1 year. In order to facilitate comparison with empirical data, I convert model growth rate parameters (μ and σ) into the equivalent parameters for a time step of 1 year. Code for this conversion process is provided in the supplementary material.

Firm size – International

International mean firm size data is estimated using the Global Entrepreneurship Monitor (GEM) database, series *omnowjob*. Data is aggregated over the years 2000-2011. In order to account for the over-representation of large firms, *I remove firms with more than 1000 employees from the database* (see Appendix A.2 for a discussion).

This 'truncation' amounts to removing the top 0.2% of firms in the GEM database. The effects of this truncation on GEM country samples are shown in
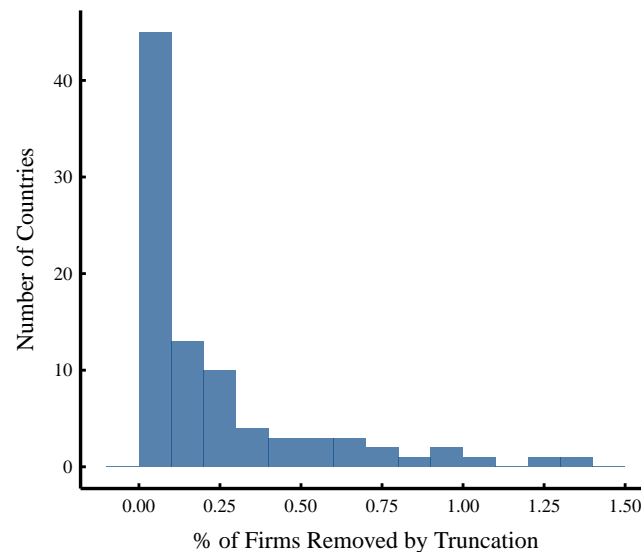
Figure A.1: The Effects of Truncating the GEM Database < 1000

This figure plots the country-level distribution of the percentage of firms removed by truncation (firms <1000). The x-axis shows the percentage of firms within each GEM country sample that are removed by truncation. The y-axis shows the number of countries with the given percentage range.

Figure A.1. For 35 out of 89 counties, this has *no effect*, since these country samples do not contain firms larger than 1000 employees. The median percentage of firms removed (by country sample) is 0.01%. For a small number of countries, this truncation removes more than 1% of firms.

Firms with zero employees are assigned a size of 1. This is an attempt to deal with the ambiguity associated with incorporation. The owner of an incorporated sole-proprietorship is usually treated as an employee (by most statistical agencies), but the owner of an unincorporated sole-proprietorship is not. Both types of firms have a single member.

To compare the resulting firm size observations with other time-based series, I use the average year of each country's aggregated data.

Uncertainty in mean firm size is estimated using the bootstrap method [2]. This involves resampling (numerous times, with replacement) the data for each country and calculating the mean of each resample. Confidence intervals are then calculated using the resampled mean distribution.

For comparison between firm size and energy consumption, Yemen and Trinidad are removed as outliers.

Firm size – United States

Average firm size data for 1977-2013 is calculated by dividing the number of persons engaged in production (BEA Table 6.8B-D) by the number of firms. The latter is calculated as the sum of all employer firms in US Census Business Dynamics Statistics plus the number of unincorporated self-employed individuals (BLS series LNU02032192 + LNU02032185).

Average firm size data for 1890-1976 uses firm counts from HSUS Ch408 (which excludes agriculture) and total private, non-farm employment from HSUS Ba471-473 (total employment less farm and government employment). To construct a continuous time-series, the two data sets are spliced together at US Census levels for 1977.

Firm size – US Industry

Mean firm size is calculated using data from Statistics of U.S. Businesses, US 6 digit NAICS and 4 digit SIC between 1992 and 2013. 'Industry' is defined to include Mining, Construction and Manufacturing.

Firm size – US Manufacturing Sub-sectors

Mean firm size is calculated using data from Statistics of U.S. Businesses, US 6 digit NAICS 2002, 2006, and 2010.

Government Employment Share – International

International government employment data is from ILO LABORSTA database (total public sector employment: level of government = Total, sex code = A, sub-classification = 06). Total employment in each country uses World Bank series SL.TLF.TOTL.IN.

Government Employment Share – United States

US government employment data is from HSUS Ba473 (1890-1928), Ba1002 (1929-40), and BEA 6.8A-D persons engaged in production (1940-2011). Total US employment is from HSUS Ba471 (1900-1928), Ba988 (1929-1940), and BEA tables 6.8A-D (1941-2011).

Large Firm Employment Share – International

The measurement of the large firm employment share is inspired by the work of Nitzan and Bichler [4]. Global data is from Compustat Global Fundamentals (series EMP). Total employment in each country uses World Bank series SL.TLF.TOTL.IN. In some countries, the Compustat data exhibits sharp discontinuities. In order to remove these discontinuities, I have removed the following data: Thailand (1999, 2008, 2010, 2011), Phillipines (2003), Croatia (2011, 2012), and Oman (2010).

Large Firm Employment Share – United States

Data for the largest firms in the United States (ranked by employment) is from Compustat North America, series DATA29 (Figure 2 uses the top 200 firms, while Figure 3 uses the top 25). Total US employment is from BEA tables 6.8A-D (Persons Engaged in Production).

Large Firm Employment Share – US Industry

The employment of the largest 25 firms in US Industry is calculated using the Compustat database, series DATA29. 'Industry' is defined to include Mining,
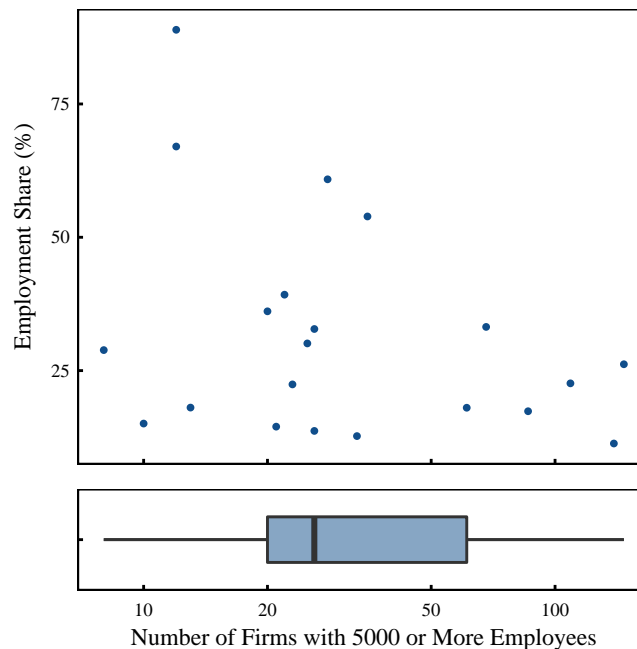
Figure A.2: Large Firms in Manufacturing Subsectors — Analyzing Bias Caused by Variations in the Number of Firms

The top panel plots the employment share of 'large firms' versus the number of firms that are defined as 'large' ($\geq$ 5000 employees). Each data point represents a single manufacturing subsector. The bottom panel shows the distribution of the number of 'large firms' per subsector.

Construction, and Manufacturing (all SIC codes between 1000 and 3999). Total Industry employment is from BEA tables 6.8A-D (Persons Engaged in Production).

Large Firm Employment – US Manufacturing Subsectors

Large firm employment share is calculated using data from Statistics of U.S. Businesses, US 6 digit NAICS 2010. 'Large firms' are defined here as those with 5000 or more employees. This differs from other data in Figure 2.3 of the main paper, where the 25 largest firms are used. Figure A.2 analyzes the bias in this method. As expected, the number of firms with 5000 or more employees varies significantly by manufacturing subsector. However the median value is 26 firms, meaning that this method should yield similar results to the 'top 25' method used elsewhere. There is also no significant correlation between the number of firms with 5000 or more employees, and the sectoral employment share of these

firms. Therefore, the variability in the sample size of 'large firms' does not cause a directional bias to the employment share of 'large firms'.

Management Employment Share

Management fraction = management employment / total employment. International management employment is from the ILO LABORSTA database using ISCO-88 (Legislators, senior officials and managers) and ISCO-1968 (Administrative and managerial workers). Total employment is from World Bank series SL.TLF.TOTL.IN. For ISCO-88, Argentina is removed as an outlier. For ISCO-1968, Syria is removed as an outlier.

US management employment is from BLS Occupational Employment Statistics (various tables, 1999-2014), ILO LABORSTA ISCO-88 (1970-1998) and HSUS Ba1037 (1860-1970). All series are spliced to BLS data. Total US employment is from HSUS Ba1033 (1860-1890), HSUS Ba471 (1900-1928), Ba988 (1929-1940), and BEA tables 6.8A-D (1941-2011). All series are spliced to BEA data.

Power Plants – Construction Labor Time vs. Capacity

Data is compiled by the author from numerous sources. Data and sources are provided in spreadsheet form in the Supplementary Material.

Power Plants — US Plant Mean Capacity

Plant nameplate capacity data comes from EIA 860 forms from 1990 to 2015. Mean plant capacity counts only power plants that are operational in the given year. Note that form 860 reports *generator* capacity. To calculate *plant* capacity, I aggregate all generators with the same Plant Code.

Self-Employment — International

International self-employment data is from the World Bank, series SL.EMP.SELF.ZS.

Self-Employment — United States

US self-employment data is from HSUS Ba910 (1900-1928), Ba988 (1929-1940) and BEA tables 6.7A–D (1941-2011). Total US employment is from HSUS Ba471 (1900-1928), Ba988 (1929-1940), and BEA tables 6.8A-D (1941-2011).

Self-Employment — US Industry

Industry self-employment data is from BEA tables 6.7A–D. Industry total employment is from BEA tables 6.8A-D (Persons Engaged in Production). Industry is defined to include Mining, Construction, and Manufacturing.

Small Firms — US Manufacturing Subsectors

Small firms are defined as those with 0–4 employees. Data is from Statistics of U.S. Businesses, US 6 digit NAICS 2002, 2006, and 2010.

Span of Control

The span of control is calculated as the employment ratio between adjacent hierarchical levels. Data sources are listed in Table A.1.

Technological Scale

Data for technological scale increases (shown in Table 2.1 of the main paper) is compiled by the author. Sources are available in spreadsheet form in the Supplementary Material.

Table A.1: Span of Control Data Sources

Source	Ref	Years	Type	N	Country	Firm Levels
Ariga	[5]	1981-1989	A	unknown	Japan	All
Audas	[6]	1992	C	1	Britain	All
Baker	[7]	1969-1988	C	1	United States	Management
Bell	[8]	2001-2010	A	552	United Kingdom	Top 3
Dohmen	[9]	1987-1996	C	1	Netherlands	All
Eriksson	[10]	1992-1995	A	210	Denmark	Management
Heyman	[11]	1991,1995	A	560	Sweden	Management
Lima	[12]	1991-1995	C	1	Portugal	All
Morais	[13]	2007-2010	C	1	Undisclosed	All
Mueller	[14]	2004-2013	A	880	United Kingdom	All
Rajan	[15]	1986-1998	A	∼300	United States	Top 2
Treble	[16]	1989-1994	C	1	Britain	All

Notation: Ref = Reference, N = number of firms A = Aggregate Study, C = Case Study

Notes: The 'Firm Levels' column indicates the coverage of the study. 'All' indicates that the study covered all hierarchical levels with the firm(s). 'Management' indicates that only managers were studied. 'Top 2' and 'Top 3' indicate that only the top 2 or 3 hierarchical levels were studied. Raw data from Baker (the BGH dataset) is available for download at http://faculty.chicagobooth.edu/michael.gibbs/.

In many cases, the above papers report results in a table of values, which were then used in this paper. However, some papers report their results only in graphical form. In these cases, I used the *Engauge Digitizer* program to extract data from the graphics.

A.2 Assessing Size Bias within Firm Databases

Like all scientific inquiry, the study of firm size distribution requires reliable data. Unfortunately, accurate firm-size data (with reasonable international coverage) is difficult to find. There are two primary data avenues available: government statistics (the *macro* level) and firm-level databases (the *micro* level). Each avenue has drawbacks.

The problem with relying on macro-level data is that it intrinsically limits the number of countries that can be studied. Apart from wealthy (OECD) nations, reliable macro statistics on firm size distribution are hard to find. This dearth of data often leads researchers to use micro-level databases instead.

The problem with using these micro-level databases to study firm size distribution is that they are rarely (if ever) designed to be *accurate* samples of the wider firm 'population'. As the analysis in this section demonstrates, firm-level databases typically under-represent small firms and over-represent large-firms. Thus, when using a micro database to study the firm size distribution, one must ask: is the database an accurate sample of the firm population? The question that immediately follows is: how do we know if the database is (or is not) biased?

In order to assess database bias, one must inevitably make comparisons to macro-level data. The key is to find macro data that is both relevant and *available* (the second criteria being the more difficult to fulfill). In the following sections I present and apply two methods for assessing firm-size bias within micro datasets.

Methods for Determining Firm-Size Bias within a Database

> **Method 1**: Compare *macro* and *micro*-level average firm-sizes.

> **Method 2**: Compare *micro*-level small-firm employment share to *macro*-level self-employment rates.

Method 1 is straightforward: it involves calculating the average firm-size within a *micro* database and comparing it to the average firm-size calculated from *macro* data. This approach is limited by the availability of macro data. For OECD countries, it is possible to directly compare firm-size averages between micro and macro data. I conduct such an analysis in Table A.2 (visualized in Figure A.6). Unfortunately, for most non-OECD countries, this approach is not
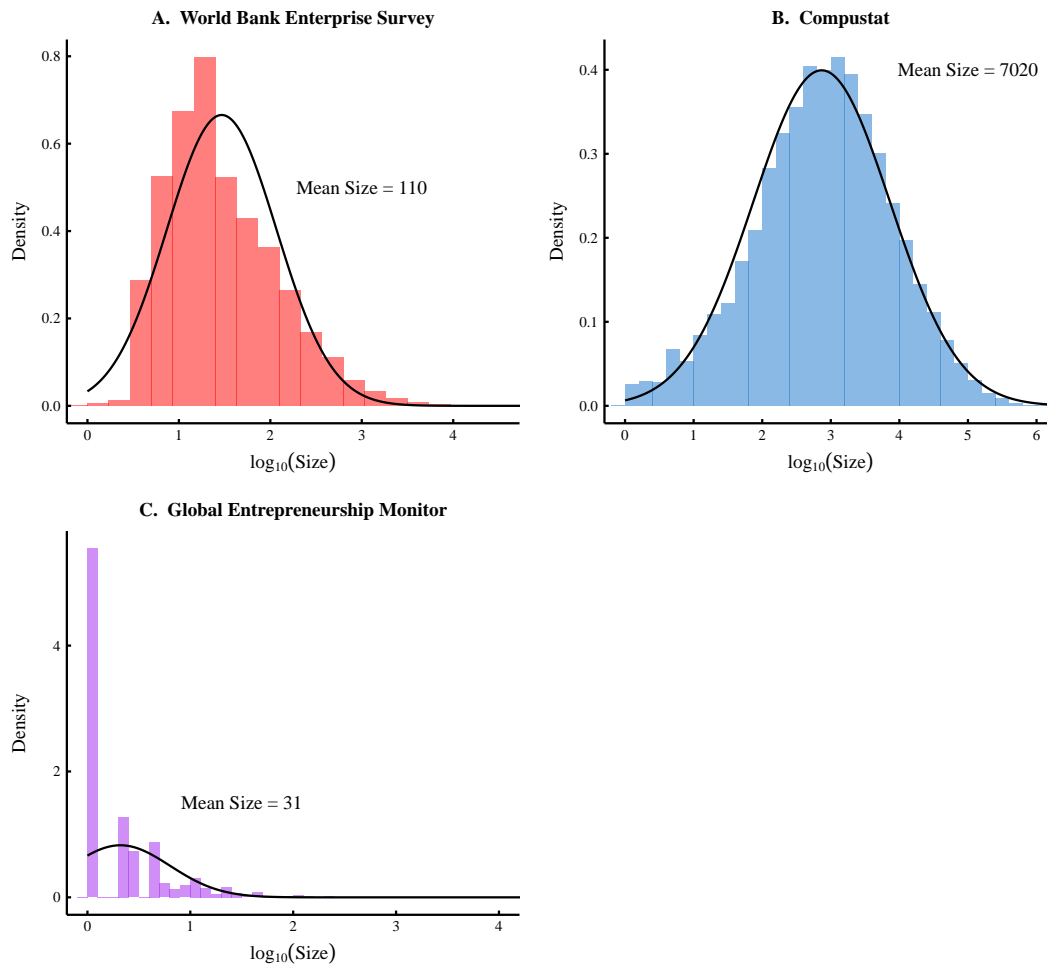
Figure A.3: Firm Size Distributions in Selected Micro Databases

Histograms show the firm size distribution within each database (firm size = number of employees). Note that data is log-transformed. Black curves show the best log-normal fit. Panel A shows the firm size distribution of the entire World Bank Enterprise Survey database (for all years). Panel B shows the firm size distribution within the Compustat database (Compustat North America merged with Compustat Global – all available years). Panel C shows the firm size distribution of the Global Entrepreneurship Monitor (GEM) database (from 2000-2011). Note that the log-normal distribution fits both World Bank and Compustat data fairly well, but fits the GEM data very poorly.

feasible because relevant macro-level data does not exist (hence our need for micro data in the first place).

Method 2 is more indirect (and is dependent on some assumptions); however, its advantage is that self-employment data is readily available for most countries. The basic logic of method 2 is as follows:

1. Self-employed individuals work in *small* firms.
2. We can think of the self-employment rate as an indicator of the share of employment held by the smallest firms.
3. By comparing the self-employment rate to the small-firm employment share within a particular database, we can infer the degree of database bias.

As a starting point, I believe method 2 is more useful, since relevant data is more widely available. In Section A.2 I apply method 2 to three databases: Compustat, the World Bank Enterprise Survey (WBES), and the Global Entrepreneurship Monitor (GEM). Figure A.3 shows the firm size distribution within these three databases. The distributions are log-transformed in order to show the log-normal character of two of the three databases (Compustat and WBES).

While all three databases are global in scope, their respective firm size distributions are quite different (note the disparities in mean firm-size). Which database gives the most accurate picture of the underlying population of firms? Analysis reveals that the GEM database is the most consistent with available macro data. Based on these results, in Section A.2 I then conduct a more detailed analysis of the GEM database (see Fig. A.6).

Small Firm Employment Share as a Database Bias Test

The basic methodology of this test is to use macro-level self-employment rates as an indicator of the share of employment held by small firms. By comparing this rate to the small-firm employment share within a micro database, we can assess the level of bias.

To begin, we define the small firm employment share as the share of employment held by firms with x or fewer employees (where x is an arbitrary number). We then vary x and see if we can match the resulting small-firm employment share with empirical self-employment rates. Figure A.4 conducts such an analysis on the Compustat, GEM, and WBES databases by comparing their respective small firm employment shares to the global self-employment rate.

First, we note that the small firm employment share in all three databases matches global self-employment rates only for a choice of x that is too large to be
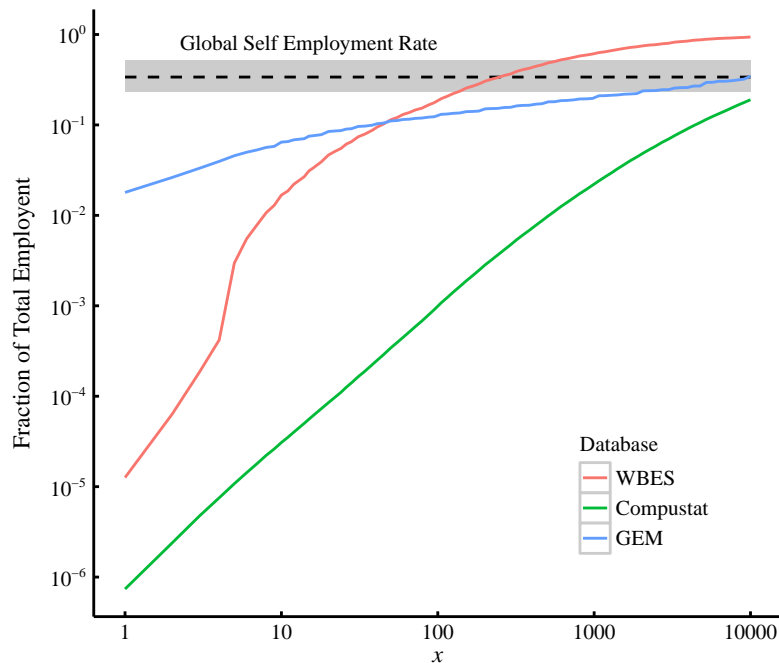
Figure A.4: Small Firm Employment Share in Selected Micro Databases

This figure assesses the relative bias within the World Bank Enterprise Survey (WBES), Compustat, and Global Entrepreneurship Monitor (GEM) databases. The share of employment held by firms with x or fewer employees (in each database) is compared to the global self-employment rate between 1990 and 2013 (the dotted line is the median, while the shaded region shows the interquartile range). Sources: Global self-employment data is for self-employed workers who are *non-employers*. This is calculated by subtracting *employer* rates (series SL.EMP.MPYR.ZS) from total self-employment rates (series SL.EMP.SELF.ZS).

---

believably related to 'self-employment'. For WBES, the small firm employment share is similar to the global self-employment rate when x is of order 100. For the GEM and Compustat databases this does not happen until x is of order 10000. This suggests that all three databases have a significant bias towards the under-representation of small firms.

Which database has the least bias? To decide this, we must settle on a believable range for the size of self-employer firms. In the real-world, the boundary x, separating *self-employer* from *employer*, does not exist. However, we can make an educated guess at the likely size range of self-employer firms.

Although a firm size of 1 typically comes to mind when we think of self-employment, the statistical definition of 'self-employment' (as defined by the World Bank) is quite broad. It consists of the following sub-categories:[1]

>   1. Own-account workers
>   2. Members of producers' cooperatives
>   3. Contributing family workers

The inclusion of contributing family workers is important, especially in developing countries where household production is still common. In this context, the size of a self-employer 'firm' will be similar to the size of a family. Since very few families are larger than 10, a believable range for which the small firm employment share should relate to self-employment rates is for $1 \leq x \leq 10$.

Over this range, the GEM small firm employment share is by far the closest to the actual rate of self-employment. While the WBES claims to be a "representative sample of an economy's private sector", this analysis suggests otherwise. The WBES small firm employment share is 2-4 orders of magnitude off the global self-employment rate for $1 \leq x \leq 10$. The Compustat database produces even worse results (off by 4-5 orders of magnitude), but this is expected. Compustat maintains records only for *public* corporations, giving it an inherent bias towards larger firms.

Note that the WBES and GEM small firm employment shares *cross* at a firm size of roughly 50. Why? The WBES contains very few small firms (size 1-10) and too many *medium* size firms (size 10-50). The GEM database, on the other hand, contains many small firms, but seems to contain too many large firms (size > 1000). This causes the crossing behaviour observed in Figure A.4.

This analysis indicates that the GEM database is the most consistent with observed global levels of self-employment. However, it still seems to contain some size bias. The problem, as I discuss in the next section, is that the GEM database contains too many extremely large firms.

Assessing Firm-Size Bias Within the GEM Database

While sufficient to weed out extremely biased databases, the method used in Figure A.4 ignores the internal distribution of data within each database. In general, micro databases with global coverage do not contain equal sized samples for each country. Thus, a large, biased sample from one country could potentially skew the entire database, even if other samples are relatively unbiased. To further test database bias, it is important to group data at the national level. In this section I investigate national-level bias within the GEM database.

---

[1]World Bank self-employment data also contains a fourth category called 'Employers'. This category is more aptly called 'owners'. Since firms of all size have owners, I have adjusted the self-employment rate by subtracting the 'Employer' rate.

A. Database Distributions

B. Country–Level Correlation
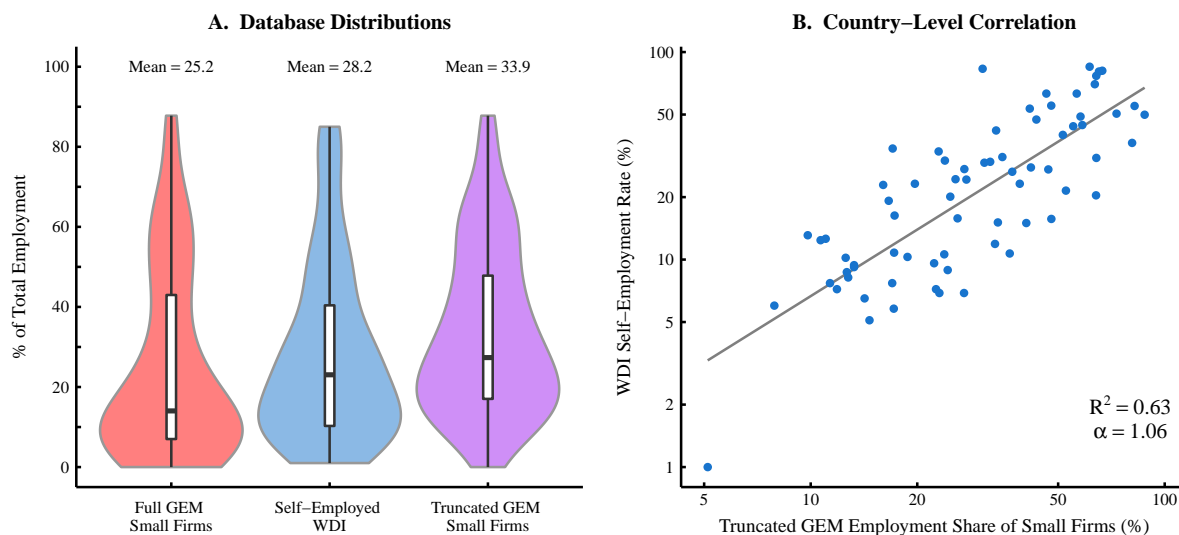
Figure A.5: Assessing Small-Firm Bias in the GEM Database

Notes: This figure compares the employment share of small firms (≤ 5 members) in the GEM database to the distribution of self-employment rates (non-employer firms only) within the WDI dataset. Only countries for which data is mutually available are shown (72 countries in total). Unlike Figure A.4 all data is aggregated at the national level (countries with small/large sample sizes are all weighted equally). Panel A shows how country-level data is distributed within each database. The 'violin' shows the distribution of data. The internal box plot shows the interquartile range (the 25th to 75th percentile), with the median marked as a horizontal line. Corresponding mean values are shown above. Panel B shows a scatter-plot of country-level data (each point is a country) for the self-employment rate vs. the small-firm employment share in the truncated GEM database. The line shows the best-fit power regression. Note that the regression exponent, α, is nearly 1. Thus, the relation between self-employment rates and small-firm employment share is roughly one-to-one. A similar regression for the non-truncated GEM database (not shown) gives $R^2 = 0.48$ and $\alpha = 0.54$, far from a one-to-one relation. This discrepancy between the full and truncated GEM dataset is the result of the over-representation of large firms within a handful of countries. This skews the small firm employment share downwards (note the low median for the full GEM database in Panel A). Thus, the truncated GEM database is more consistent with self-employment data, meaning we can infer that it has less of a firm-size bias.

Sources: Non-employer rates are calculated by subtracting employer rates (series SL.EMP.MPYR.ZS) from the total self-employment rate (series SL.EMP.SELF.ZS). WDI data is chosen for which the data year most closely matches the GEM year (which is calculated as the country-level mean year of all data entries from 2000-2011).

Table A.2: Mean Firm-Size in the GEM Dataset vs. Macro Data

Country	Macro	GEM Trunc	GEM Full
Austria	7.6	11.7	12
Belgium	5.6	6.3	6
Czech Republic	3.5	13.5	30
Denmark	9.2	8.5	26
Finland	6.8	5.3	13
France	7.5	5.3	22
Germany	10.4	11.9	151
Hungary	5.7	6.1	8
Italy	3.5	2.8	17
Netherlands	6.1	10	27
Poland	4.7	2.9	16
Portugal	3.6	8.9	9
Russian Federation	18	9	16
Slovakia	4.2	11.8	17
Slovenia	4.9	13	19
Spain	5.5	4.5	10
Sweden	5.8	5.7	15
Switzerland	10.8	6.5	180
Turkey	3	9.5	18
United Kingdom	7.7	7	26
United States of America	9.1	10	164
India	2.6	5.2	6
Ghana	1.5	2.2	2
Mean	6.4	7.7	35.2

Notes: This table compares mean firm sizes within the GEM database to macro-level data. Data is shown for both the full GEM database, and its truncated version, which removes all observations of firms with more than 1000 employees. The rational for truncation is that large firms are over-represented within the dataset, skewing mean firm-size.

Sources and Methodology: Macro-level mean firm-size is calculated by dividing total employment by the number of firms. The number of firms N_{total} is calculated using Eq. (A.1), where N_{gov} is government data for the number of firms, S_T is the self-employment rate, S_E the self-employed *employer* rate, U is the fraction of self-employed firms that are unincorporated (hence not counted in official statistics), and L is the size of the labor-force.

$$N_{total} = N_{gov} + (S_T - S_E) \cdot U \cdot L \tag{A.1}$$

Data for S_T, S_E and L come from World Development Indicators (WDI) series SL.EMP.SELF.ZS, SL.EMP.MPYR.ZS, and SL.TLF.TOTL.IN, respectively. Data for the official number of firms comes from OECD Entrepreneurship at a Glance 2013. Due to lack of data, U is assumed to be 0.7, the level observed in the US [17]. For Ghana, all data comes from Sandefur [18], Table 1 and 2. For India, all data comes from Hasan and Jandoc [19], Table 1 and Table 3 (using the sum of the ASI and NSSO datasets). For US data sources, see Appendix A.1.
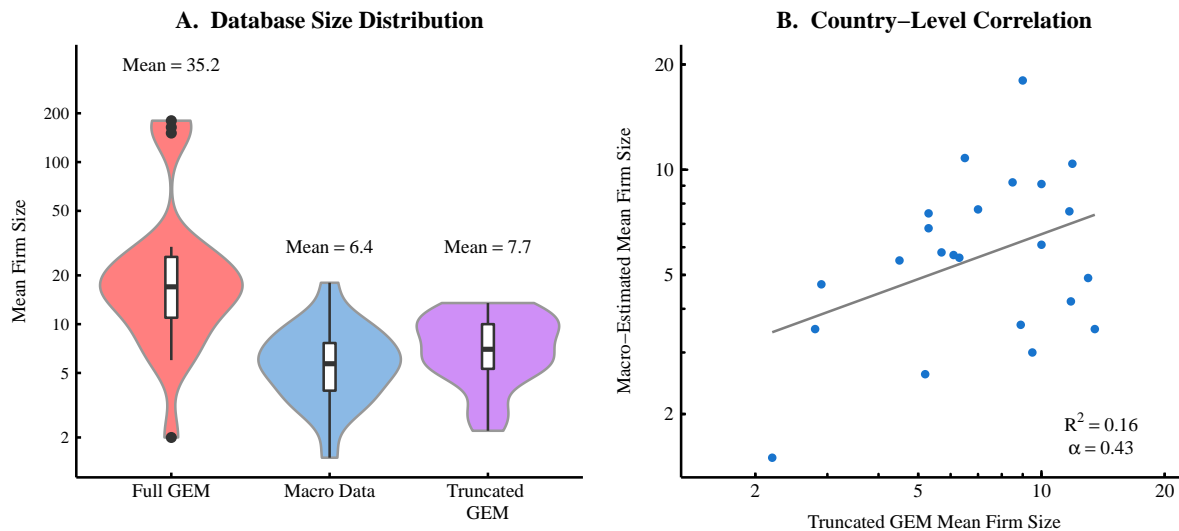
Figure A.6: GEM Mean firm size distribution vs. Macro Data

Notes: This figure visualizes the mean firm-size data for the countries shown in Table A.2. Panel A shows the mean firm-size distribution within each database. Relative to macro data, the full GEM database clearly over-represents large firms. The mean firm-size in the truncated GEM database is also slightly larger than the macro data, but given the small sample size, the difference is statistically insignificant ($p = 0.20$). Panel B shows the correlation between macro data and the truncated GEM data. A power regression gives an exponent $\alpha = 0.47$, below the desired one-to-one level that would indicate perfect agreement between the micro and macro data. Despite these shortcomings, the truncated GEM database appears to be a fairly accurate sample of the international firm-size distribution.

I begin with a continuation of the self-employment/small-firm method developed above. However, I now group all data at the national level. The GEM database contains firm samples from a total of 89 countries, 72 of which also have data available in the WDI database. For each country, the employment-share of firms with 5 or fewer employees is calculated (from GEM data) and compared to the WDI self-employed rate (non-employers only). This calculation is done for both the full GEM dataset, and a *truncated* version in which all firms with more than 1000 employees are excluded. This truncated version is tested on the hunch that the full GEM database still over-represents large firms (a hunch that is confirmed in Fig. A.6).

The results of this analysis are shown in Figure A.5. Both the full and truncated GEM databases have a small-firm employment-share distribution that is roughly equivalent to the WDI self-employment rate distribution. Of particular

interest is the fact that the small-firm employment share within the truncated GEM database gives a nearly one-to-one prediction of WDI self-employment rates (see Fig A.5A).

This analysis suggests that both the full and truncated GEM databases give a reasonably accurate sample of the international firm size distribution. In order to differentiate between the two, it is helpful to compare mean firm-size estimates with macro data. Due to macro data constraints, this must be done with a much smaller sample size than the 72 countries used above. Table A.2 shows the 23 countries for which data is available.

Note that macro-level mean-size estimates are predicated on a few assumptions. Government published statistics usually include firm-counts for *employer* firms only (i.e. firms with employees). Non-employer firms are excluded. Thus, unincorporated self-employed individuals are typically not counted as 'firms' (incorporated self-employed workers are technically counted as employees of their business, and are thus employer firms). As a result, calculations done using official firm-counts only will give a mean firm-size that is disproportionately large. To account for this bias in macro data, I adjust the official firm-count by adding an estimate for the number of self-employer firms (see the methodology in Table A.2).

The results of this investigation are visualized in Figure A.6. From this analysis, there is convincing evidence that the full GEM database over-represents large firms. For a few countries (Germany, Switzerland, and the US) this leads to a mean firm-size estimate that is a factor of 10 larger than macro estimates. Truncating the GEM database seems to effectively adjust for this bias.

Why is truncation effective (and is it justified)? The problem of firm-size bias is partially due to the extremely skewed nature of the firm size distribution. The presence of even a *single* extremely large firm can have a large effect on the mean of a sample. For instance, the GEM database contains roughly 170,000 observations. Suppose that the mean firm-size of these observations is 5. If we add a single observation of a Walmart-sized firm (2 million employees), the resulting average more than *triples* (to roughly 17). Of course, firms this large do exist, but the chance of observing one in a sample should be *extremely* small.

The fact that large firms are over-represented in the GEM database demonstrates a sampling bias. Discarding observations of very large firms is one method for dealing with this bias. Other methods are certainly possible, but I do not discuss them here.

Functional Form of the Firm Size Distribution

One of the first tasks for understanding an empirical distribution (of any kind) is to look for theoretical distributions that can be used to model it. Many observers have used the log-normal distribution to model firm size distributions [20–25]. As shown in Figure A.3, the log-normal distribution is a suitable model for the firm size distribution within the Compustat and WBES databases. However, the preceding analysis showed that these databases are rather poor representations of the actual global firm size distribution.

It may be that the use of the log-normal distribution is an artefact of researchers' reliance on biased micro databases [26]. For data that is more representative of the actual firm size distribution (i.e. the GEM dataset), a power law distribution is a much better fit. The characteristic feature of the log-normal distribution is that its logarithm is *normally distributed* (hence the reason for the log transformation in Fig. A.3). A power law distribution, however, will not *not* appear normally distributed under a log transformation. Instead, it will decline monotonically as the GEM database does.

Unlike Compustat and WBES, the GEM database is much better fitted with a power law than with a log-normal distribution (see Fig. A.7A). For firms under 10,000 employees, the GEM database is consistent with a power law with a scaling exponent $\alpha \approx 1.9$. Note that the tail of the GEM database is 'fatter' than expect for a power law (it is above the 99% confidence interval). This is consistent with our earlier conclusion that the GEM database over-represents large firms. Macro data from for the US firm size distribution is also consistent with a power law (Fig. A.7B).
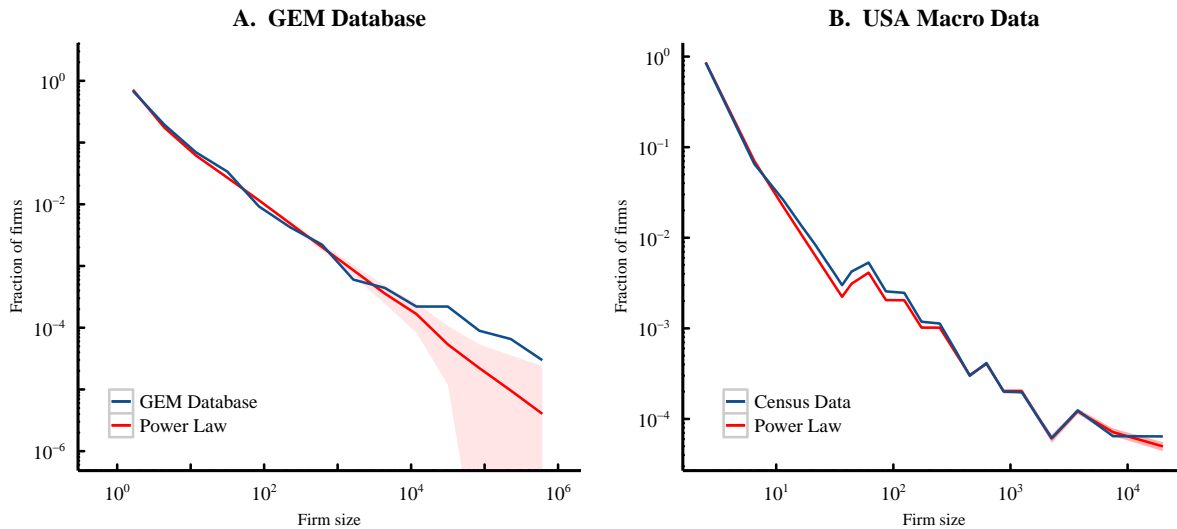
Figure A.7: GEM and US Census Data are Consistent with Power Laws

Notes: Panel A shows the firm size distribution of the Global Entrepreneurship Monitor database (all years). For firms with less than 10,000 employees, the database is consistent with a discrete power-law distribution with exponent $\alpha \approx 1.9$. Panel B shows the US firm size distribution, which is consistent with a discrete power-law distribution with exponent $\alpha \approx 2$. Shaded regions show the 99% confidence interval for a simulated power law distribution with a sample size similar to each dataset.

Sources and Methodology: US data for employer firms is from the US Census Bureau, Statistics of U.S. Businesses (using data for 2013). This data is augmented with Bureau of Labor Statistics data for unincorporated self-employed workers (series LNU02032185 and LNU02032192). The histogram preserves Census firm-size bins, with self-employed data added to the first bin. The last point on the histogram consists of all firms with more than 10,000 employees. Both power-law distributions are simulated using the R poweRlaw package, and plotted with the same histogram bins used to plot empirical data. The GEM simulation uses 170,000 observations while the US simulation use 10 million observations.

Note: many readers will expect power law distributions to appear linear when plotted on a log–log scale. Departures from linearity shown in Panel B are artefacts of US census bin sizes (which do not always grow proportionately).

A.3   The Firm Size Distribution as a Variable Power Law

Recent studies have found that firm size distributions in the United States [26] and other G7 countries [27] can be modelled accurately with a power law. Less is known about other countries. In this section, I test if country-level firm size

distributions in the GEM database are consistent with a power law. I find that a power law distribution is favored over other heavy-tail distributions in the vast majority of countries. I also find that international variations in 3 summary statistics (mean, self-employment, and large firm employment share) are mostly consistent with a power law distribution.

Power Laws in the GEM Database

The firm size distribution in the *entire* GEM database is roughly consistent with a power law, although the end of the tail is slightly too heavy (Fig. A.7A). In this section, I analyse the GEM firm size distribution at the *country* level to assess how well the data fit a power law distribution. I use the truncated GEM database, which contains only firms with fewer than 1000 employees. The rational is that the full GEM database slightly over-represents large firms (see Appendix A.2).

Historically, power law distributions have been fitted by using an ordinary least-squares (OLS) regression on the logarithm of the histogram. However, this approach is inaccurate, and it violates the assumptions that justify the use of OLS [28]. A more appropriate approach for fitting distributions is to use the *maximum likelihood method*. The likelihood function $\mathscr{L}$ assesses the probability that a set of data x came from a probability density function with the parameter(s) θ.

$$\mathscr{L}(\theta|x) = P(x|\theta) \tag{A.2}$$

The best fit parameter(s) θ_{mle} maximizes the likelihood function. Like any fitting method, the maximum likelihood indicates only the best fit parameters of the *specified* model, not the appropriateness of the model itself. To discriminate between two different models (1 and 2), we compare their respective maximum likelihoods in ratio form (Λ). The larger likelihood indicates the better fitting model.

$$\Lambda_{1,2} = \frac{\mathscr{L}_1(\theta_{mle}|x)}{\mathscr{L}_2(\theta_{mle}|x)} \tag{A.3}$$

It is often more convenient to use the log-likelihood ratio, $\log \Lambda$. The sign of $\log \Lambda$ indicates the preferred model (positive indicates that model 1 is better, negative indicates that model 2 is better). The magnitude of $\log \Lambda$ indicates the strength of this preference.

I use this method to assess if country-level firm size distributions in the GEM database are best modelled with a power law. I compare the likelihood
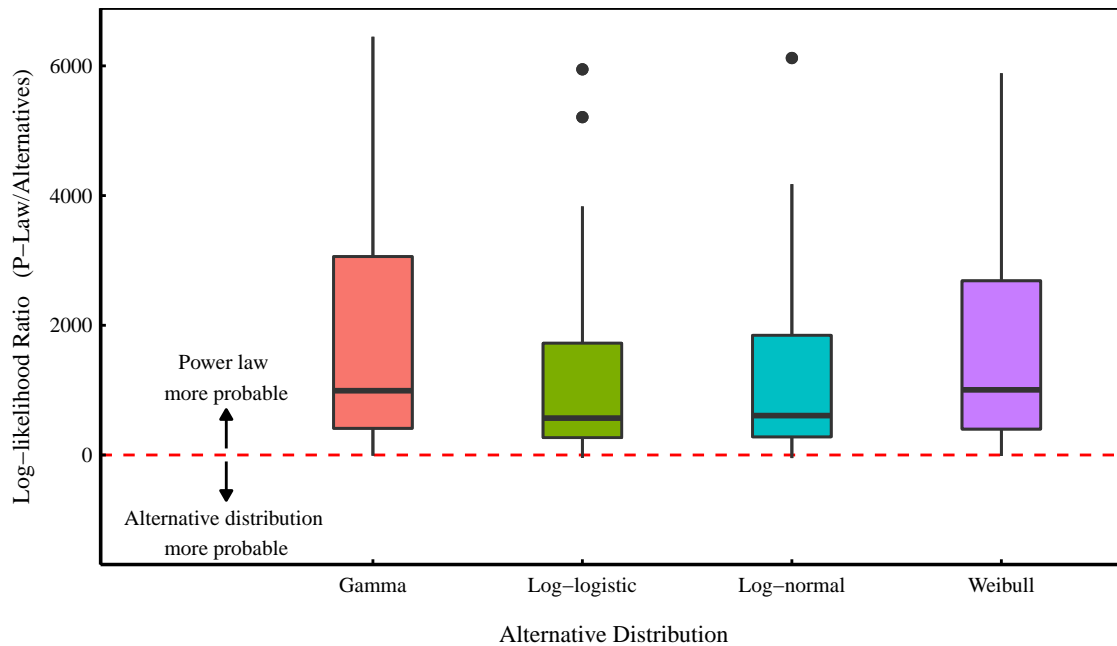
Figure A.8: Comparing the Power Law to Alternatives in the GEM Database

Using country-level firm size distributions from the GEM database, this figure assesses the goodness of fit of a power law relative to four other heavy-tail distributions. The firm size distribution in each country in the GEM database is fitted with a power law, gamma, log-logistic, log-normal, and Weibull distribution. For each country, the log-likelihood ratio is computed between the power law and the four alternative distributions. The box plots display the resulting range of ratios. A positive ratio indicates that the power law is more probable, while a negative ratio indicates that the alternative distribution is more probable. In order to better display the majority of data, several large outliers favoring a power law are not shown. For all but 3 countries, a power law distribution is the best fit.

Notes: This figure shows the mean log-likelihood ratios for 100 re-samples (with replacement) of each country. Maximum likelihoods are calculated using the R packages 'poweRlaw' (for a power law) and 'fitdistrplus' (for alternative distributions). Although empirical data is discrete, all models used here are continuous.

of a power law distribution to the likelihood of four other heavy-tail distributions: gamma, log-logistic, log-normal, and Weibull. The resulting range of log-likelihood ratios (one for each country in the GEM database) is shown in Figure A.8. A power law distribution is favored over other distributions in the vast majority of countries (97%).

International Summary Statistics

Firm size summary statistics can be used as another way to test if the firm size distribution is consistent with a power law. This has the advantage of broadening the evidence to include more data sources (I combine GEM, World Bank, and Compustat data). My method is to pair two statistics and test if the resulting empirical relation can be reproduced by simulated samples from a power law distribution. I look at two pairings: (1) the self-employment rate vs. mean firm size; (2) the large firm employment share vs. mean firm size.

Self-Employment vs. Mean Firm Size

The rational for looking at the self-employment rate is that it indicates the relative share of employment held by small firms. Figure A.9A shows the empirical relation between self-employment rates and mean firm size (black dots). The simulated relation is shown in the background, where the power law exponent α is indicated by color. Creating this simulation requires making assumptions about the size of self-employer firms. I assume that *all* firms below the size boundary L_s are considered self-employer firms. The simulated self-employment rate then consists of the fraction of employment held by firms with employment less than or equal to L_s.

To account for international variation in the size of self-employer firms, I let the boundary point vary randomly over the range $1 \leq L_s \leq 10$. In Figure A.9A, $L_s = 1$ corresponds to the bottom of the coloured region, and $L_s = 10$ to the top. Why choose the upper bound to be so large? My reasoning is based on the definition of 'self-employment', which consists of 3 sub-categories: own-account workers, cooperatives, and family workers.[2] Especially in developing countries, where household production is still common, a self-employer 'firm' is synonymous with a *family*. A size of 10 seems a reasonable upper limit on the size of family. Given this assumption, a majority of countries (75%), as well as
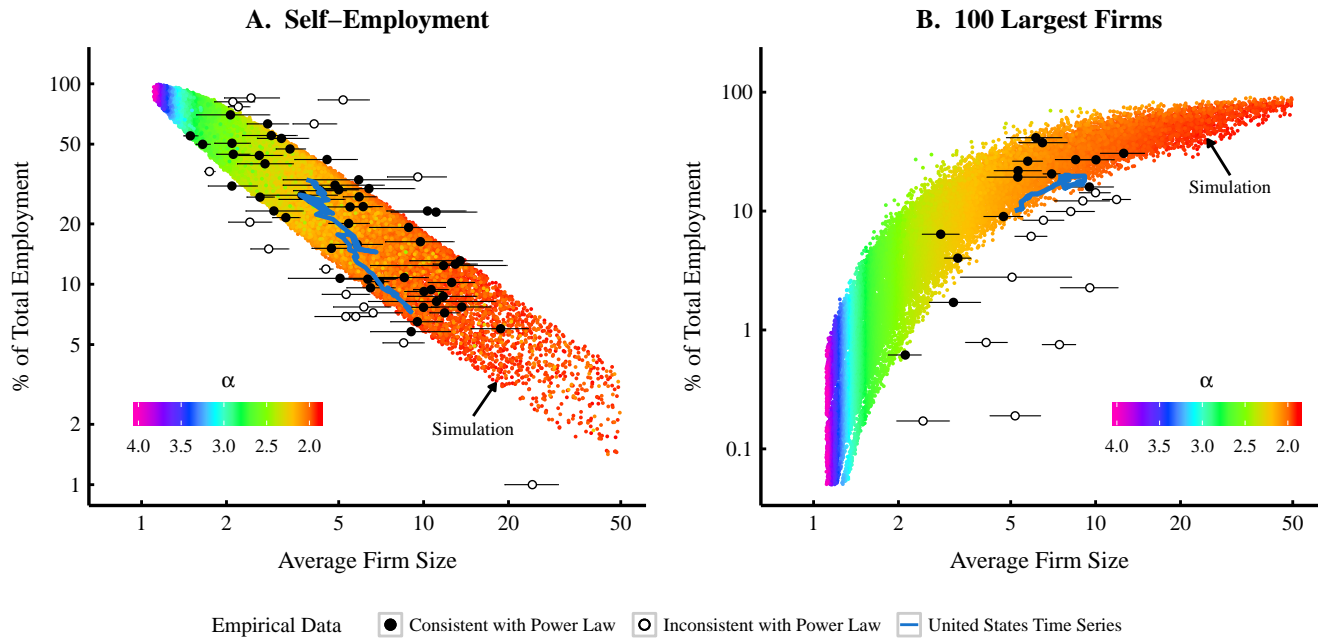
Figure A.9: International Summary Statistics, Empirical vs. Power Law

This figure compares pairings of summary statistics for empirical and simulated data. Empirical data is at the country level. Simulated data is randomly generated from a power law distribution (the exponent α is indicated by color). Panel A shows self-employment rates vs mean firm size while panel B shows large firm employment share vs. mean firm size. Self-employment rates are modelled as the employment share of all firms less than the size L_s, which varies randomly over the range $1 \leq L_s \leq 10$. Uncertainty in mean firm size (95% level confidence intervals) is indicated by horizontal lines. Empirical data is judged to be consistent with a power law when the error bar is within the 99% range of simulated data. For data sources, see Appendix A.1.

---

the entire time series for the United States, have a self-employment vs. mean firm size relation that is consistent with a power law.

Large Firm Employment Share vs. Mean Firm Size

To test if variations in the large firm employment share are consistent with a power law distribution, I use the same method as above: I plot the employment share of the 100 largest firms against mean firm size (Fig A.9C). I then compare this relation to the one predicted by simulated power law data. To allow for the

---

[2]Most statistical databases add a fourth category of 'employers' (i.e. capitalists). Because this category is not related to small firms, I remove it from analysis.

effects of differing country size, simulation sample sizes vary over the range of national firm populations (which are estimated by dividing the labor force by the mean firm size).

A slight majority of countries (56%), as well as the entire time-series for the United States, have a large firm employment share vs. mean firm size relation that is consistent with a power law distribution. Note that all data points that are *not* consistent with a power law lie *below* the simulation zone (rather than above). This could indicate that these countries have firm size distributions with a tail that is thinner than a power law, but it could also indicate a problem with the data. I have assumed that the 100 largest firms in the Compustat database are actually the largest firms in each nation. There is no guarantee that this assumption is true: the Compustat database may not give complete coverage of the largest firms, especially if a country has many large *private* companies. Further research is needed to determine if these findings indicate a departure from a power law distribution, or if they are artefacts of incomplete data.

A.4   Testing Gibrat's Law Using the Compustat Database

Gibrat's 'law' states that firm growth rates are *independent* of firm size. To what extent is this supported by empirical evidence? I investigate here using the Compustat US database. My results are consistent with previous analysis of the Compustat database: growth rates are approximately Laplace distributed, and volatility declines with firm size [29]. However, I show that this decline is of importance to only a small subset of firms.

Analysis

Rather than directly calculate the mean and variance of Compustat firm growth rates, I fit the growth rate distribution with a truncated Laplace density function (growth rates less than -100% are rounded to -100%). I then investigate how the parameters of this function change with firm size (Fig. A.10). The advantage of this approach is that it is not biased by large outliers, and it allows a direct comparison of empirical data to modelled data (where firm growth rates are drawn from a Laplace distribution).

To estimate the Laplace parameters, I fit the histogram of simulated data to the histogram of empirical data (using a Monte Carlo technique that minimizes the absolute value of the error). The results are displayed in Figure A.10C-D. The location parameter (μ) shows no significant relation to firm size. However, growth rate volatility (the scale parameter, b) declines monotonically with firm size.

Interestingly, the location parameter is always less than zero, meaning the most probable rate of growth is *negative*. This finding is consistent with the conditions predicted by a stochastic model with a reflective lower bound. Such a model will be stable only when there is a net *negative* drift to firm size (Appendix A.5). In Appendix A.6 I reproduce the US firm size distribution using a model with a location parameter of -1%, which is consistent with Compustat data.

Extrapolating to the Entire Economy

Because the Compustat database contains data only for publicly traded firms, it is not an accurate sample of the wider US firm population (see Appendix A.2). However, based on the assumption that the US firm size distribution is a power law, we can estimate how the volatility-percentile relation shown in Figure A.10D might look for the economy as a whole. The method for this process is shown in Table A.3.

A. Compustat 10th Decile

B. Compustat 2nd Decile

C. Fitted Location Parameter by Percentile

D. Fitted Scale Parameter by Percentile
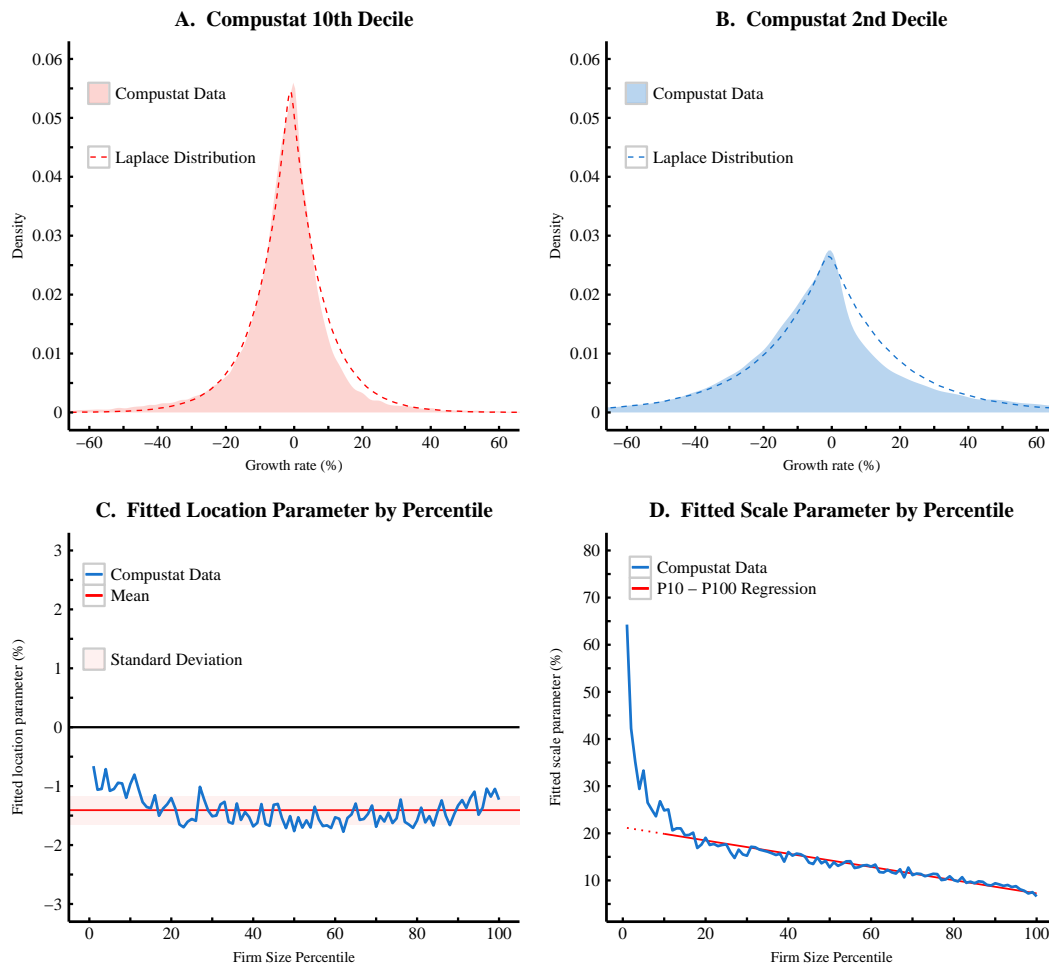
Figure A.10: Firm Growth Rate Distribution in the Compustat US Database

This figure analyses firm growth rates (by employment) within the Compustat US database from 1970 to 2013. Panel A shows the growth rate distribution for firms in the 10th (top) decile, while Panel B shows the distribution for firms in the 2nd decile. Dotted lines indicate the best-fit Laplace distribution. Panel C and D show the results of Laplace regressions at the percentile level. Panel C shows the estimated location parameter (μ), while Panel D shows the estimated scale parameter (b). Laplace distributions are fitted using a Monte Carlo method. This analysis indicates that growth rate volatility is a function of firm size, while the growth rate mode is not. Given the firm-size bias of the Compustat database, results for lower percentiles (i.e. P1-P10) should be treated with scepticism.

The first step is to generate a US firm sample from a power law distribution that best fits empirical data (I use $\alpha = 2.01$ here), and then compute size percentiles. Next, we select a particular percentile (the green cell) and note the corresponding firm size in the Compustat database (left pink cell). We than find all firms within the power-law sample that have the same size (right pink cells). The scale parameter for the selected Compustat percentile (left purple cell) is then mapped onto these firms, and their corresponding percentiles. The result (right purple cells) is a transformed relation between firm percentile and scale parameter that serves as our economy-wide estimate.

The results of this transformation are shown in Figure A.11. Two different estimates are shown. The blue curve shows results using the raw data shown in Figure A.10D, while the red dotted curve shows results using a linear regression for P10-100, extrapolated over all percentiles.

Why two different methods? The bias in the Compustat increases as firm size decreases: coverage for large firms is nearly complete, while coverage of small firms (under 10) is extremely limited. Thus, it is quite possible that the large increase in volatility for firm percentiles 1–10 may be an artefact of this bias. By using the linear regression of P10-P100, we remove this potential artefact. We can think of the two curves in Figure A.11 as representing a plausible range for the US economy. The stochastic model used to reproduce the US firm size distribution (Fig. A.12), has a location parameter of 34%, which is much nearer the lower bound of our Compustat estimates.

This analysis suggest that declines in growth rate volatility are important only to a small minority of firms.
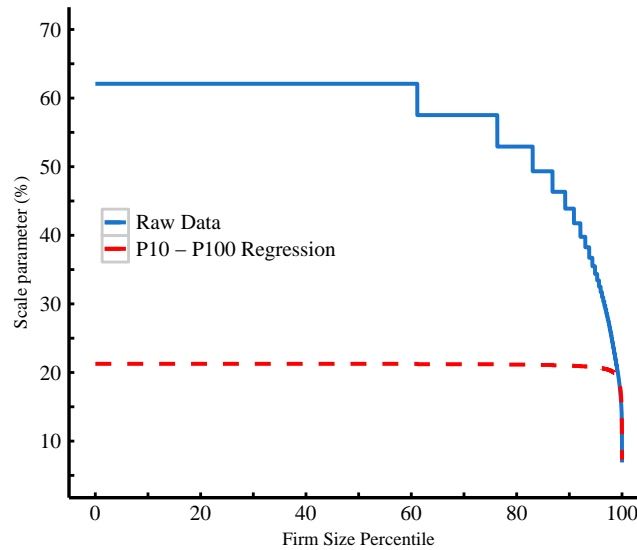
Figure A.11: Scale Parameter vs. Percentile, Economy-Wide Estimates

This figure shows a transformation of the Compustat scale-percentile regressions (Fig. A.10D) to a form that is consistent with the firm size distribution of the entire US economy. The US distribution is modelled with a power law ($\alpha = 2.01$). The blue curve shows the relation that would result from using the entire range of the Compustat regressions (P1-100). The step-wise pattern is a result of discrete data (steps correspond to a change in firm size by 1). The red dotted curve shows the relation resulting from using a linear regression of Compustat P10-100 (red line in Fig. A.10D), extrapolated over P1-10.

Table A.3: Method for Transforming Compustat Scale Parameter Regressions

Percentile	Compustat Firm Size	Scale	Power Law Firm Size	Transformed Scale
1	1	60	1	60
2	3	50	1	60
3	–	–	1	60
4	–	–	1	60
5	–	–	1	60
6	–	–	2	50

This table demonstrates the method for transforming the Compustat scale-percentile relation to an estimated relation for the whole economy. The first step is to select a percentile (the green cell P1 is selected here). We then match the Compustat firm size of this percentile to the equivalent power law firm size (pink cells). The Compustat scale parameter is then mapped onto all power law percentiles with matching firm sizes, resulting in a transformed scale function (purple cells).

A.5   Instability of the Gibrat Model

The Gibrat model assumes that firm growth is a stochastic, multiplicative process. If L_0 is the initial firm size and x_i the annual growth rate, then firm size at time t is given by:

$$L(t) = L_0 \cdot x_1 \cdot x_2 \cdot \ldots \cdot x_t \;\; = \;\; L_0 \prod_{i=1}^{t} x_i \tag{A.4}$$

The instability of this model was first noted by Kalecki [30]. It stems from the model's diffusive nature: the resulting firm size distribution tends to *spread* with time. This tendency can be understood by relating the model to the classic example of diffusion: the one-dimensional random walk.

In a random walk model, a particle is subjected to a series of random additive shocks (y_i) that cause its position to change over time. At any given time, the particle's displacement from the initial position $d(t)$ is simply the sum of all of these shocks:

$$d(t) = y_1 + y_2 + \ldots + y_t \;\; = \;\; \sum_{i=1}^{t} y_i \tag{A.5}$$

In order to intuitively understand how this leads to diffusion, let us suppose that the shocks y_i are drawn from the uniform distribution $\{-1, 1\}$. At any given

time, we can ask: what is the maximum possible displacement? In this case, it is exactly equal to t (the number of time intervals that have passed). When we introduce many randomly moving particles, some may attain this maximum displacement (however unlikely it is). Since the maximum grows with time, we can conclude that the displacement distribution must spread with time.[3]

The Gibrat model shares this property, except that the diffusion is *exponential*. To see this, we take the logarithm of Eq. A.4, which allows us to express the growth rate product as a *sum*.

$$\log(L(t)) = \log(L_0) + \log(x_1) + \log(x_2) + ... + \log(x_t)$$
$$= \log(L_0) + \sum_{i=1}^{t} \log(x_i) \tag{A.6}$$

We then exponentiate to get:

$$L(t) = L_0 e^{\sum_{i=1}^{t} \log(x_i)} \tag{A.7}$$

By setting $\log(x_i) = y_i$, we can see that Eq. A.7 is just Eq. A.5 in exponential form: our firm growth model is a one-dimensional, *exponential* random walk. The resulting firm size distribution will therefore spread rapidly with time – a fact that is inconsistent with available evidence. For instance, we know that the US firm size distribution has changed little since 1970 (see Fig. 2.2 in main article).

The second problem with this model is that it gives rise to a *log-normal* distribution, contradicting our finding that most firm size distributions are best described by a power law. The proof that this model leads to a log-normal distribution is straightforward. For a sufficiently large number of iterations, the Central Limit Theorem dictates that the sum of independent, random numbers will be *normally* distributed. Thus, for a large number of random walkers, the displacement $d(t)$ will be normally distributed (so long as the distribution of y_i satisfies certain conditions). Because Eq. A.7 is the exponential form of Eq. A.5, the *logarithm* of $L(t)$ will be normally distributed – the defining feature of the log-normal distribution.

---

[3]For a step size drawn from the uniform distribution $\{-1, 1\}$, the standard deviation of the displacement is equal to $\sqrt{t}$. For a good derivation, see Feynman [31] Ch. 6.

Adding a Reflective Lower Bound

One simple way to reform this model is to add a *reflective lower bound* that stops firms from shrinking below a certain size [32–34]). This slight change will cause the model to generate a power law, rather than a log-normal distribution. It also leads to model stability (under certain conditions).

Why does the introduction of a reflective boundary lead to a power law distribution? One way of understanding this is to relate back to the additive random walk. If a reflective barrier is added to a one-dimensional random walk, it will no longer tend towards normal distribution; rather, it will tend towards an *exponential* distribution (see [35], p 15 for a proof).

Recall that a multiplicative process can be transformed into an additive process by taking the logarithm. Therefore, for a multiplicative firm model with a lower bound, the logarithm of firm size (L) will be exponentially distributed. Thus, the firm size distribution $p(L)$ is given by Eq. A.8, which reduces to a power law (where C is the normalizing constant, and α is the scale parameter).

$$\begin{aligned}
p(L) &= Ce^{-\alpha \cdot \log(L)} \\
&= CL^{-\alpha}
\end{aligned} \tag{A.8}$$

For a firm size distribution, the obvious choice for a minimum lower bound is $L = 1$ (a sole-proprietor with no employees). In the proceeding model, I implement this reflection through the following conditional statement, which is evaluated at every time interval:

$$\text{if } L(t) < 1, \text{ then } L(t) = 1 \tag{A.9}$$

Introducing a reflective lower bound also solves the instability problem, but only when growth rates have a negative 'drift'. Why? Intuitively, we can state that a model will be stable if it is not possible for a firm to shrink or grow *forever*. Introducing a lower bound automatically stops firms from shrinking forever, but it does nothing to stop the possibility of unending growth.

However, if firm growth rates have a net *downward* drift, all firms will tend towards a size of 0, given enough time. This downward drift occurs when the *geometric* mean of the growth rate distribution is less than 1. We can draw an analogy with gas particles moving in a gravitational field on earth. The particles move randomly, but there must be a small net downward drift due to the force of gravity. The result is a stable distribution of particles. If we remove gravity,

the particles are free to diffuse forever. Similarly, if we remove the downward bias to firm growth rates, the distribution becomes unstable.

A.6   Properties of Stochastic Models

Despite their simplicity, stochastic models of firm growth are able to replicate many important properties of the real world. I review three such properties here. Stochastic models can be used to:

1. Generate a firm size distribution that is consistent with empirical data;
2. Reproduce the relation between firm size and firm age;
3. Simulate new firm survival rates over time.

Modelling the US Firm Size Distribution

The model used here assumes scale-free growth with a reflective lower bound at a firm size of one. Growth rates are drawn from a Laplace distribution that is truncated by rounding all (fractional form) growth rates less than 0 to 0. In order to maintain a discrete distribution, firms with non-integer size are rounded to the nearest integer (after the application of each growth rate).

This simple model can be used to replicate the US firm size distribution (Fig. A.12). In this case, model parameters $\mu = 0.99$ and $b = 0.34$ are used. The model shows the distribution of 1 million firms after 100 time iterations. In order to capture fluctuations around the equilibrium, the model is run 100 times, with the shaded region showing the resulting range of outcomes.

Firm Age vs. Firm Size

Firm age is calculated as the time since a firm's last 'reflection'. The model described above can be used to replicate the size-age relation of firms in the World Bank Enterprise Survey (WBES) database (Fig. A.13A). The fitted parameters are $\mu = 0.97$, $b = 0.55$. Note that the model diverges from WBES data for firms with fewer than 10 employees. Due to the size bias within the WBES database (see Appendix A.2), it is not clear if this divergence is significant, or an artefact of database bias.

Firm Survival Rates

The survival rate of new firms tends to decline exponentially over time (Fig. A.13B). To replicate this behavior, we give our stochastic model an initial firm size distribution and then track firm survival over time. A firm 'dies' when it is
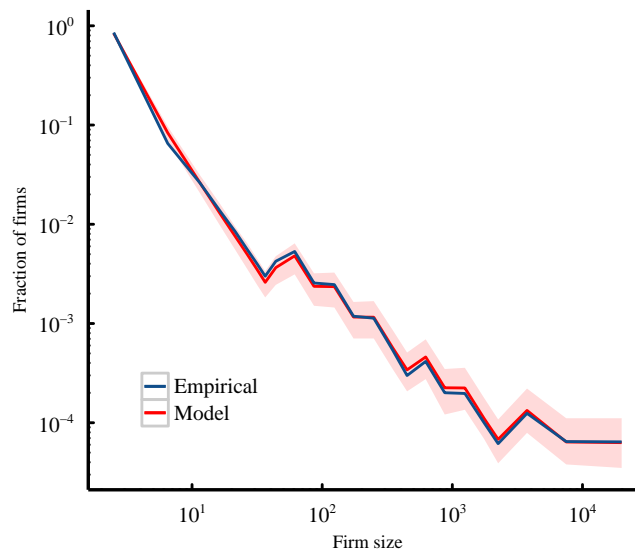
Figure A.12: A Stochastic Model of US Firm Size Distribution

The US firm size distribution is shown for the year 2013 (blue line), along with a stochastic model (red) of 1 million firms with growth rates drawn from a truncated Laplace distribution with parameters $\mu = 0.99$, $b = 0.34$. The shaded region indicates the 90% confidence region of the model. US Data for employer firms is from the US Census Bureau, Statistics of U.S. Businesses (using data for 2013). This data is augmented with Bureau of Labor Statistics data for unincorporated self-employed workers (series LNU02032185 and LNU02032192). The histogram preserves Census firm-size bins, with self-employed data added to the first bin. The last point on the histogram consists of all firms with more than 10,000 employees. The model histogram uses Census bins to allow direct comparison.

––––––––––––––––––––––

reflected for the first time. At any given time, the firm survival rate is given by the fraction of firms that have never been reflected.

In order to model firm survival rates, we must choose an initial distribution of firms. We can make guesses about this distribution based on BLS *establishment* data. In 1994 — the first year the BLS tracked survival rates — the average size of new establishments was 7.3. In the same year, the average size of *all* US establishments was 16.9 (using data from Census Business Dynamics Statistics. It seems reasonable to assume that the average size of new *firms* might also be about half the average for all firms. It also seems reasonable to assume that the distribution of new firms can be modelled with a power law. Using these assumptions, I model the initial firm size distribution with a power law of $\alpha = 2.1$. This gives a mean size of close to 5 (about half the US average).

The empirical data shown in Figure A.13 comes from the US Bureau of Labor
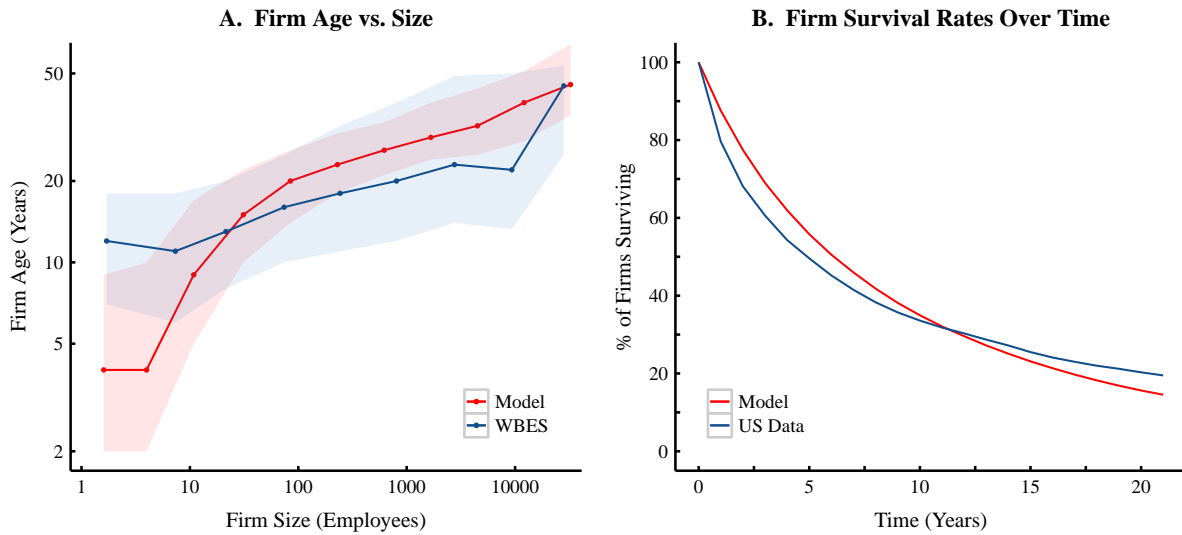
Figure A.13: Stochastic Models Can Reproduce Firm Age/Survival Data

Panel A shows the relation between firm size and firm age within the World Bank Enterprise Survey (WBES) database (blue). A stochastic model (red) with growth rates drawn from a truncated Laplace distribution with parameters $\mu = 0.97$, $b = 0.55$ produces a similar firm size-age relation. Lines indicate medians and shaded regions indicate the interquartile range. Logarithmic bin locations are indicated with points. Panel B shows the survival rates of new firms over a period of 21 years. Empirical data (blue) is from the BLS Business Employment Dynamics database, Table 7, Survival of private sector establishments by opening year. The model (red) draws growth rates from a truncated Laplace distribution with parameters $\mu = 0.99$, $b = 0.35$.

Statistics (BLS). A caveat is that this data is for *establishment* (not *firm*) survival rates. An establishment refers to a specific business location, while a firm is a legal entity that may contain multiple establishments. For modelling purposes, I ignore this distinction here and assume that establishments are equivalent to firms.

Empirical and modelled survival rates are shown in Figure A.13B). The survival rate model parameters ($\mu = 0.99$, $b = 0.35$) are nearly identical to the parameters ($\mu = 0.99$, $b = 0.34$) used to replicate the US firm size distribution (Fig. A.12). These parameters are also consistent with the range estimated from Compustat data (Appendix A.4).

A.7   Bias and Error in the GDP Labor Time Method

My method for estimating power plant construction time is to take total cost and divide by (nominal) GDP per capita in the country and year of construction (henceforth called 'the GDP method'). In this section, I estimate the *bias* in this method. To do so, we need to investigate in detail the assumptions made by this approach.

The total cost of construction (C) of a power plant can be attributed to direct labor costs (L_d), profits and interest (denoted as K, for *capitalist* income), and non-labor costs (N):

$$C = L_d + K_d + N \tag{A.10}$$

By the rules of double-entry accounting, all non-labor costs will eventual become the *income* of other firms. Thus, after a long digression, we can eventually attribute non-labor costs to either indirect labor costs (L_i) or indirect capitalist costs (K_i):

$$C = L_d + L_i + K_d + K_i \tag{A.11}$$

Since we are not interested in differentiating between direct and indirect costs, we define L as the sum of direct and indirect labor costs, and K as the sum of direct and indirect capitalist costs:

$$C = L + K \tag{A.12}$$

Next, we define w as the average wage of all of the workers who are directly and indirectly involved in the construction project. Total labor cost (L) is then the average wage times total labor time (t). Substituting $L = w \cdot t$ into Eq. A.12 gives:

$$C = w \cdot t + K \tag{A.13}$$

Solving for total labor time gives:

$$t = \frac{C - K}{w} \tag{A.14}$$

Equation A.14 gives an accurate estimate of the total labor time involved in construction. Unfortunately, it is difficult (if not impossible) to calculate K (direct and indirect capitalist expenses) and w (the average wage of all direct and

indirect workers). In order to get around this lack of data, I make the assumption that capitalist income can be neglected — that labor costs are approximately the same as total costs:

$$L = C - K \approx C \tag{A.15}$$

Furthermore, I assume that w is approximately the same as nominal GDP per capita (Y_{pc}).

$$w \approx Y_{pc} \tag{A.16}$$

Under these assumptions, Eq. A.14 is approximated by Eq. A.17:

$$t \approx \frac{C}{Y_{pc}} \tag{A.17}$$

By using GDP per capita as a measure of average income, we implicitly assume that all aspects of power plant construction occur within one country. For older plants, this is likely a good assumption. However, in the modern era of globalized production, this assumption is most likely violated to some degree, especially for key components of the plants like the generators and turbines. Unfortunately there is simply no way to disaggregate construction/manufacture costs to their various regions. However, we can correct for this bias to some degree by including power plants from as many nations as possible. The GDP method will then overestimate the labor time for plants constructed in developing countries (where GDP per capita is very low) and underestimate labor time for plants constructed in wealthy countries (where GDP per capita is very high). The hope is that these divergent biases will cancel themselves out.

How accurate is the GDP method? Unfortunately, we cannot compare GDP method estimates to the true labor time value (Eq. A.14) because this latter formula contains *unknowable* quantities (K and w). However, we can test Eq. A.14 against an alternative estimate for labor time that makes more accurate assumptions.

To proceed, let us first rewrite Eq. A.14 as follows by factoring out C in the numerator:

$$t = \frac{C\left(1 - \dfrac{K}{C}\right)}{w} \tag{A.18}$$

We then make the assumption that capitalists involved (indirectly and directly)

with the project earn profit and interest at approximately the *national average rate*. This means we assume that the capitalist share of total costs (K/C) is approximately the same as the capitalist share of national income (k_s).

$$\frac{K}{Y} \approx k_s \tag{A.19}$$

We also assume that workers involved (indirectly and directly) with the project earn the *national average wage* (w_n). Given these assumptions, Eq. A.18 can be rewritten as:

$$t \approx \frac{C\,(1-k_s)}{w_n} \tag{A.20}$$

We now have two way of estimating the labor time involved in the construction of a power plant (Eq. A.17 and Eq. A.20). Our expectation is that Eq. A.20 is the more accurate estimate. To quantify the discrepancy between the two estimates, we construct an error ratio, which is the ratio of the two labor time estimates (Eq. A.17 / Eq. A.20):

$$\text{error ratio} = \frac{w_n}{Y_{pc}\,(1-k_s)} \tag{A.21}$$

Figure A.14 shows this error ratio calculated using US data from 1929–2015. The results indicate that the GDP method (Eq. A.17) *overestimates* labor time by roughly 60%. Why? By neglecting capitalist income, our estimate *inflates* the numerator in Eq. A.14. Furthermore, GDP per capita is typically slightly lower than the average annual wage of a full-time worker, so the GDP method *deflates* the denominator in Eq. A.14. Of course, this error estimate is itself based on the assumptions contained in Eq. A.20. Still, it seems safe to conclude the following:

1. The GDP method likely *overestimates* the true labor time of power plant construction;

2. This overestimate is relatively stable over time.

Since our interest in this study is how labor time *scales* with plant capacity (and not with *absolute* labor time), this constant overestimate is of little concern. It will have no effect on the scaling of construction labor time with power plant size.

What is of more concern, however, are the *changes* in the error ratio that occur over time. How might this affect the estimation of power plant construction
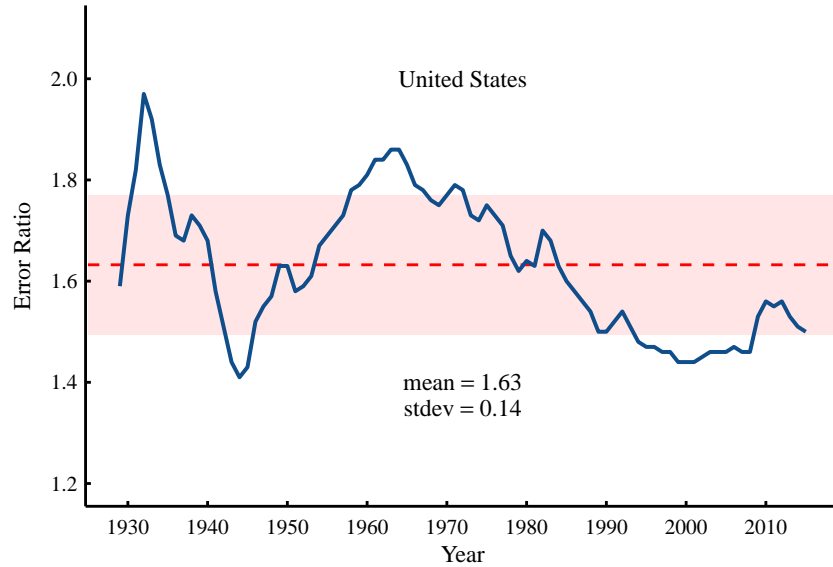
Figure A.14: Error estimate of the GDP method for calculating labor time
This figure shows calculations of the error ratio (Eq. A.21) using US data. All data is from the BEA. National income data is from Table 1.12, National Income by Type of Income. Capitalist share of national income is equal to profits (with CCA and IVA) and net interest divided by national income. The average wage is calculated by dividing the sum of the compensation of employees and proprietor income by the total persons engaged in production (Table 6.6B-D). US population data is from Maddison [1] and the World Bank series SP.POP.TOTL. Nominal GDP data is from the file *gdplev*.

---

time? It is actually quite simple to model the effect of measurement error on a scaling relation. We begin by assuming that two variables, x and y, exhibit perfect power law scaling *identical* to that found between power plant capacity and construction labor time:

$$y = x^{1.26} \tag{A.22}$$

To study the effect of measurement error, we introduce a 'noise factor' ϵ (drawn from a lognormal distribution), that perturbs the perfect scaling relation:

$$y = x^{1.26} \cdot \epsilon \tag{A.23}$$

The effect of larger/smaller error can be modelled by increasing/decreasing the relative dispersion of ϵ. Suppose, for argument's sake, that Figure A.14 *severely* underestimates the error associated with the GDP method. In reality, let us assume that the error is *10 times larger*. Since the relative standard deviation

of the Figure A.14 error ratio is 0.086, we can model the effect of a tenfold increase in error by setting ϵ to have a relative standard deviation of 0.86.

Figure A.15 shows how the effects of this error factor (on our power law scaling relation) change as the orders of magnitude spanned by the dependent variable (x) increase. The horizontal axis shows the orders of magnitude spanned by the variable x, while the vertical axis shows the R^2 value of a log-log regression on the relation $y = x^{1.26} \cdot \epsilon$. The important result is that even though the measurement error is quite large, it becomes increasingly *inconsequential* as the data span increases.

Why? The R^2 value indicates the proportion of the variance in the dependent variable (y) that is predictable from the independent variable (x). Since we are conducting a log-log regression, it is helpful to look at the log transformed relation:

$$\log(y) = 1.26 \cdot \log(x) + \log(\epsilon) \qquad \text{(A.24)}$$

Now, the variance in $\log(y)$ is affected both by the variance in $\log(x)$ and by the variance in $\log(\epsilon)$. But notice that the variance in both $\log(y)$ and $\log(x)$ will be proportional to the logarithm of the *range* of x. But this is equivalent to the orders of magnitude spanned by x (since orders of magnitude indicate scaling by factors of 10). Thus, the variance in $\log(x)$ and $\log(y)$ scales with the orders of magnitude spanned by x. However the variance in $\log(\epsilon)$ is *constant* — it does not change as the range of x increases. Because the variance in $\log(\epsilon)$ does not scale, its importance decreases as the range of x increases. That is, the fraction of variance in $\log(y)$ that is attributable to $\log(\epsilon)$ is inversely related to the orders of magnitude spanned by x.

So what does this result imply for the accuracy of the GDP method? Clearly, accuracy is a function of the orders or magnitude spanned by plant capacity. In our case study, plant capacity spanned *seven orders of magnitude*. According to Figure A.15, even if the GDP method had a *severe* error factor (i.e. only accurate to within a factor of 3), the resulting measurement error would still not have a significant effect on the observed scaling relation. Thus, despite the error that is implicit in the GDP method, it is likely that our results are robust.

Still, given that the GDP method has a bias, why not use the more accurate approach given in Eq. A.20? The problem with this formula is that it requires data on the capitalist share of national income as well as data on the average annual income of full time workers. This data is much more difficult to obtain
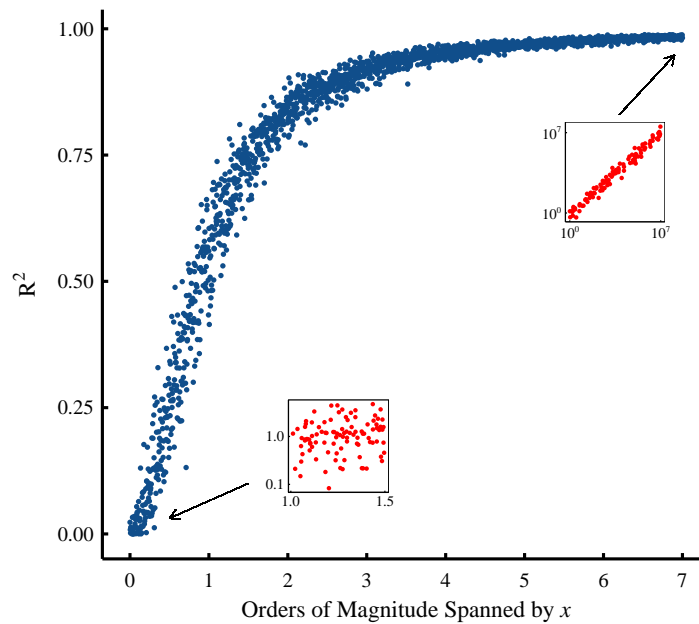
Figure A.15: Data span vs. the effect of measurement error on a scaling relation

This figure shows multiple log-log regressions on data defined by the relation $y = x^{1.26} \cdot \epsilon$. Here x is a random variable whose logarithm is uniformly distributed, and ϵ is a noise factor drawn from a lognormal distribution with mean 1 and standard deviation 0.86 (which is 10 times the relative standard deviation of the error ratio in Fig A.14). The horizontal axis shows the orders of magnitude spanned by the variable x, while the vertical axis shows the resulting R^2 value of the y vs. x regression. Each dot represents a *single* regression. Inset plots (red) show raw data underlying two different regressions — one with a small data span (bottom left) and one with a large data span (top right). For data that spans less than 2 orders of magnitude, the noise dominates the subsequent regression. However, once the span of x surpasses 4 orders of magnitude, the noise becomes inconsequential to the regression.

than GDP per capita (especially in developing countries). Thus my use of the GDP method is mostly one of convenience: it makes analysis easier.

A.8   A Hierarchical Model of the Firm

An 'ideal' hierarchy has a constant span of control throughout — meaning the employment ratio between each consecutive hierarchical level is *constant* (Fig. A.16). This property allows total employment to be expressed as a geometric series of the span of control s. If the number of individuals in the top hierarchical level is a, and h_t is the total number of hierarchical levels, then total employment L is given by the following series:

$$L = a\left(1 + s + s^2 + \ldots + s^{h_t - 1}\right) \tag{A.25}$$

Using the formula for the sum of a geometric series, Eq. A.25 can be rewritten as:

$$L = a\frac{1 - s^{h_t}}{1 - s} \tag{A.26}$$

We make the assumption that individuals in and above the hierarchical level h_m are considered *managers*. The number of managers M in a firm with h_t levels of hierarchy is equivalent to the employment of a firm with $h_t - h_m + 1$ levels of hierarchy:

$$M = a\frac{1 - s^{h_t - h_m + 1}}{1 - s} \tag{A.27}$$

We can use Eq. A.27 and Eq. A.26 to express management as a fraction of total employment (M/L):

$$\frac{M}{L} = \frac{1 - s^{h_t - h_m + 1}}{1 - s^h} \tag{A.28}$$

Asymptotic Behavior of the Management Fraction

The management fraction tends to grow with the number of hierarchical levels, but only to a certain point (Fig. A.17). For $h_t > 10$ the management fraction approaches an asymptotic limit that depends only on the span of control s. Finding the asymptotic behavior of M/L requires evaluating the following limit:

$$\lim_{h_t \to \infty} \frac{M}{L} = \lim_{h_t \to \infty} \frac{1 - s^{h_t - h_m + 1}}{1 - s^{h_t}} \tag{A.29}$$
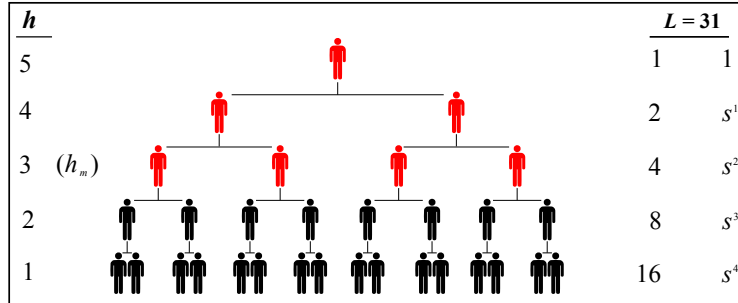
Figure A.16: A Perfectly Hierarchical Firm

Within a perfectly hierarchical firm, the number of individuals in adjacent hierarchical levels differs by a factor of the span of control s (in this diagram, $s = 2$). This characteristic allows total employment L to be expressed as a geometric series of s. Managers (red) are defined as all individuals in and above level h_m (which equals 3 here).
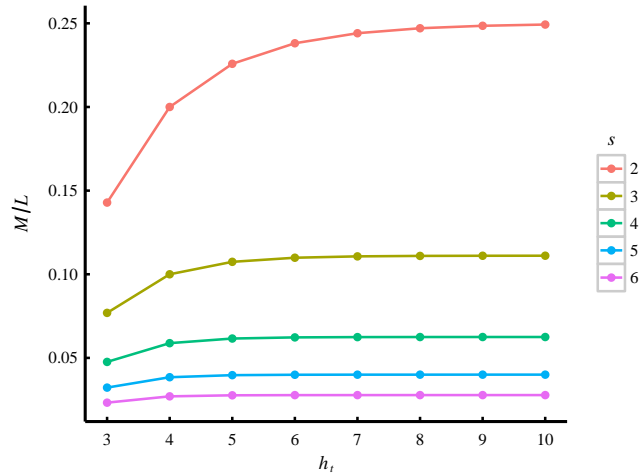
Figure A.17: Asymptotic Behavior of the Management Fraction

This figure shows a plot of Eq. A.28 for $h_m = 3$ and various s. As the total number of hierarchical levels (h_t) increases, the management fraction (M/L) within a firm grows rapidly, but soon reaches an asymptotic limit. This asymptote is a function of the span of control s, and the choice of h_m (the definition of where management begins).

To evaluate this limit, I use L'Hospital's Rule, which states that $\lim \frac{f(x)}{g(x)} = \lim \frac{f'(x)}{g'(x)}$. We first rewrite Eq. A.29 in a differentiable form, with a base e exponent:

$$\lim_{h_t \to \infty} \frac{M}{L} = \lim_{h_t \to \infty} \frac{1 - e^{\log(s) \cdot (h_t - h_m + 1)}}{1 - e^{\log(s) \cdot h_t}} \tag{A.30}$$

Applying L'Hospital's Rule, we take the derivative (with respect to h_t) of both the numerator and the denominator in Eq. A.30, giving:

$$\lim_{h_t \to \infty} \frac{M}{L} = \lim_{h_t \to \infty} \frac{-\log(s) \cdot e^{\log(s) \cdot (h_t - h_m + 1)}}{-\log(s) \cdot e^{\log(s) \cdot h_t}} \tag{A.31}$$

This simplifies to:

$$\lim_{h_t \to \infty} \frac{M}{L} = e^{\log(s)(-h_m + 1)} = s^{-h_m + 1} \tag{A.32}$$

Therefore, the asymptotic behavior of the management fraction depends only on the span of control, and our definition of management.

An Algorithm for Creating Hierarchies

The management model uses a power law simulated firm size distribution. In order to calculate the number of managers, each firm must be organized into hierarchical levels. I have developed the following algorithm to carry out this process.

Having selected a firm, we know its employment L and its span of control s; however, the total number of hierarchical levels h_t is unknown. To calculate h_t, we assume, for the moment, that the size of the top hierarchical level is one. Therefore, h_t must satisfy:

$$L = \frac{1 - s^{h_t}}{1 - s} \tag{A.33}$$

Solving for h_t gives:

$$h_t = \frac{\log[1 + L(s - 1)]}{\log(s)} \tag{A.34}$$

Since h_t must be discrete, we round the solution to the nearest integer. My

method is then to 'build' the hierarchy from the bottom up. If the bottom hierarchical level contains b workers, then L is defined by the series:

$$L = b \left(1 + \frac{1}{s} + \frac{1}{s^2} + ... + \frac{1}{s^{h_t - 1}} \right) \tag{A.35}$$

Using the formula for the sum of a geometric series, this becomes:

$$L = b \, \frac{1 - 1/s^{h_t}}{1 - 1/s} \tag{A.36}$$

At the moment, L is known but b is unknown. We therefore solve for b (and round the answer to the nearest integer):

$$b = L \, \frac{1 - 1/s}{1 - 1/s^{h_t}} \tag{A.37}$$

Once we have b, we can differentiate the firm into hierarchical levels by dividing b by powers of s (Eq. A.35). Due to rounding errors, the sum of the employment of all hierarchical levels may differ from the original firm size L. Any discrepancies are added (or subtracted) to the base level to give the correct firm size. The number of managers M is then simply the sum from hierarchical level h_m to h_t.

A.9   An Agrarian Model of Institution Size

In this section, I use an adaptation of the hierarchical firm model (used in Fig. 2.7 of the main paper and discussed in Appendix A.8) to explain the institution size limits posed by an agrarian economy. In agrarian societies, the *vast* majority of the population is directly engaged in agricultural activities — a direct result of low agricultural labor productivity. This model aims to demonstrate that the large size of the agricultural population places inherent constraints on agrarian institution size. The model makes the following assumptions:

1. All agrarian institutions are 'ideal' hierarchies with the same span of control.

2. The agricultural population forms the bottom hierarchical level of *all* institutions.

3. Agrarian institution sizes are distributed according to a power law.

The model is depicted graphically in Figure A.18. In formulating this model, I have in mind a feudal society in which the institutional unit can be loosely thought of as the feudal manor. These institutions are organized around the extraction of an agricultural surplus from peasants/serfs, and are defined by a rigid caste system (with serfs at the bottom). For the sake of simplicity, we assume that all peasants/serfs are engaged in agriculture.

There is evidence that feudal manors (like modern firms) were power-law distributed. For instance Hegyi et al. find an approximate power law distribution of serf ownership by nobles/aristocrats in 16th century Hungary [36]. Similarly, Kahan finds a highly skewed distribution of serf ownership in 18th century Russia [37] (although this distribution is better fit with a lognormal function).

Although the above assumptions may well be wrong (or oversimplifications), this model is intended mostly as a thought experiment. Figure A.19A shows the modelled relation between the agricultural portion of the total population and mean institution size (with the span of control varying between 2 and 3). The prediction is that the agricultural population should decline rapidly as mean institution size increases.

In this model, the fraction of the population engaged in agriculture places strict limitations on institution size. Estimates vary on the size of this agricultural fraction of the population in historical agrarian societies. In Figure A.19, I use Cottrell's estimate that 95% of the population in ancient Egypt was directly

Figure A.18: The Decline of Agricultural Workers as a Function of Institution Size in an Agrarian Economy

This figure shows an adaptation of the hierarchical firm model presented in Appendix A.8. In agrarian societies, we assume that the bottom hierarchical level of all institutions is constituted entirely of agricultural workers. As institution size increases, the relative size of the agrarian population declines. All institutions are assumed to be 'ideal' hierarchies with constant spans of control. In the model (not accurately represented here) the institution size is distributed according to a power law.

———————————

engaged in agricultural activity [38] (indicated by the red horizontal line in Fig. A.19A). According to the model, this limits mean institution size to between 1.2 and 1.32 people (indicated by the grey region).

If we further assume that the modern relation between mean firm size and energy use per capita is applicable to agrarian institutions, we can make predictions about rates of energy consumption. We input the estimated mean institution size range into the firm size versus energy regression from Figure 2.1C (main paper) to predict a range of energy use per capita for this model society (Fig. A.19B).

The predicted interval of roughly 10 to 30 GJ per capita is a surprisingly realistic range for a typical agrarian society. For instance Warde estimates that England used 20 GJ of energy per capita in 1560 [39]. Similarly, Malanima estimates that 1st and 2nd century Romans consumed between 9 and 17 GJ per capita [40].
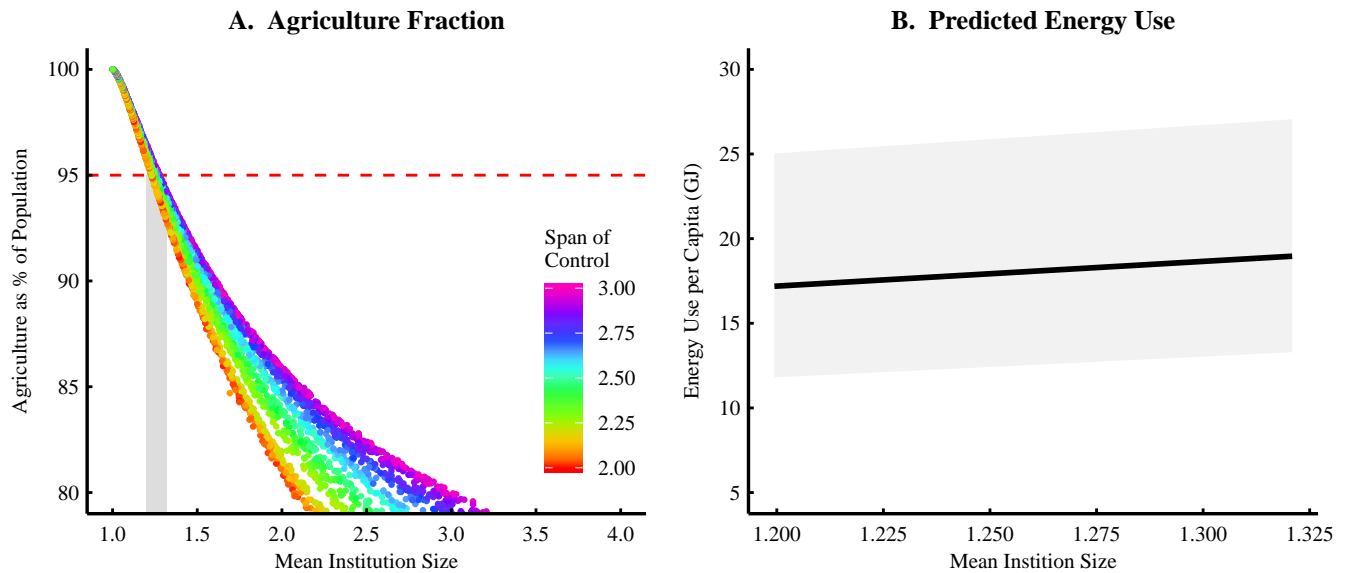
Figure A.19: Modelling Agricultural Constraints on Institution Size, and the Implication for Energy Use per Capita in Agrarian Societies

This figure shows how a hierarchical model of an agrarian society can be used to relate the size of the agricultural population to institution size and energy use per capita. Panel A shows the modelled relation between the agricultural portion of the population and mean institution size. Different mean institution sizes are generated by varying the exponent of the institution size distribution. Different spans of control are indicated by color. The red horizontal line corresponds to a society with 95% of the population in agriculture, and the shaded region shows the corresponding prediction for mean institution size. Panel B shows the energy use per capita predictions for this range of institution size. These predictions are made using the national mean firm size vs. energy use per capita regression shown in Fig. 2.1C of the main paper. The formula is $E_{pc} = 14.3\bar{L}^{1.02}$, where E_{pc} is energy per capita and $\bar{L}$ is mean firm size. The grey region indicates the 95% confidence interval of the prediction.

This model can be used to understand how energetic constraints place limits on institution size within agrarian societies. In all societies, the relative size of the agricultural population is a function of agricultural labor *productivity* [41]. The agriculture sector must produce a surplus of food in order to feed the non-agricultural population [42]. It follows that the fraction of workers in agriculture can decline only if their per person output of surplus food increases.

In agrarian societies, agricultural workers relied exclusively on human and animal labor, which meant that output per worker was extremely low compared to modern industrial agriculture. The result was that the agricultural surplus was very small, allowing only a small non-agricultural population to exist [43]. According to our model, this leads to inherent constraint on institution size.

Agricultural productivity, in turn, is directly related to energy use. Increasing agricultural labour productivity requires that each worker convert more energy into useful work. Historically, this meant first introducing more draft animals per worker, followed by the widespread adoption of fossil fuel powered equipment (tractors, combines, etc.). As agricultural workers increase their energy use, this will impact per capita energy use for society at large.

Unfortunately, this model cannot be used to study the transformation from an agrarian to an industrial society because its premise *breaks down* as this transition proceeds. The model is based on a feudal society organized around the expropriation of an agricultural surplus from a serf/peasant class. As feudal relations give way to market relations, this social structure ceases to exist. New institutions form that have nothing to do with agriculture, meaning assumption 2 (the bottom level of all institutions is entirely made up of agricultural workers) becomes *absurd*.

Despite its shortcomings, this model is useful for understanding the possible limitations placed on institution size by the energetic constraints of an agrarian economy.

References

1. Maddison A. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD; 2008. Available from: http://www.ggdc.net/maddison/Maddison.htm.

2. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: CRC press; 1994.

3. Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G. Historical statistics of the United States: millennial edition. vol. 3. Cambridge: Cambridge University Press; 2006.

4. Nitzan J, Bichler S. Capital as Power: A Study of Order and Creorder. New York: Routledge; 2009.

5. Ariga K, Brunello G, Ohkusa Y, Nishiyama Y. Corporate hierarchy, promotion, and firm growth: Japanese internal labor market in transition. Journal of the Japanese and International Economies. 1992;6(4):440–471.

6. Audas R, Barmby T, Treble J. Luck, effort, and reward in an organizational hierarchy. Journal of Labor Economics. 2004;22(2):379–395.

7. Baker G, Gibbs M, Holmstrom B. Hierarchies and compensation: A case study. European Economic Review. 1993;37(2-3):366–378.

8. Bell B, Van Reenen J. Firm performance and wages: evidence from across the corporate hierarchy. CEP Discussion paper. 2012;1088.

9. Dohmen TJ, Kriechel B, Pfann GA. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. Journal of Population Economics. 2004;17(2):193–228.

10. Eriksson T. Executive compensation and tournament theory: Empirical tests on Danish data. Journal of Labor Economics. 1999;17(2):262–280.

11. Heyman F. Pay inequality and firm performance: evidence from matched employer–employee data. Applied Economics. 2005;37(11):1313–1327.

12. Lima F. Internal Labor Markets: A Case Study. FEUNL Working Paper. 2000;378.

13. Morais F, Kakabadse NK. The Corporate Gini Index (CGI) determinants and advantages: Lessons from a multinational retail company case study. International Journal of Disclosure and Governance. 2014;11(4):380–397.

14. Mueller HM, Ouimet PP, Simintzi E. Within-Firm Pay Inequality. SSRN Working Paper. 2016;.

15. Rajan RG, Wulf J. The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. The Review of Economics and Statistics. 2006;88(4):759–773.

16. Treble J, Van Gameren E, Bridges S, Barmby T. The internal economics of the firm: further evidence from personnel data. Labour Economics. 2001;8(5):531–552.

17. Hipple S. Self-employment in the United States: an update. Monthly Lab Review. 2004;127:13.

18. Sandefur J. On the evolution of the firm size distribution in an African economy. Center for the Study of African Economies Working Paper Series. 2010;5.

19. Hasan R, Jandoc KR. The Distribution of Firm Size in India: What Can Survey Data Tell Us? Asian Development Bank Economics Working Paper Series. 2010;(213).

20. Angelini P, Generale A. On the evolution of firm size distributions. The American Economic Review. 2008;98(1):426–438.

21. Cabral LM, Mata J. On the evolution of the firm size distribution: Facts and theory. American economic review. 2003;93(4):1075–1090.

22. Gibrat R. Les inegalites economiques. Recueil Sirey; 1931.

23. Gupta HM, Campanha JR, de Aguiar DR, Queiroz GA, Raheja CG. Gradually truncated log-normal in USA publicly traded firm size distribution. Physica A: Statistical Mechanics and its Applications. 2007;375(2):643–650.

24. Hart PE, Prais SJ. The analysis of business concentration: a statistical approach. Journal of the Royal Statistical Society Series A (General). 1956;119(2):150–191.

25. Stanley MH, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA, Stanley HE. Zipf plots and the size distribution of firms. Economics letters. 1995;49(4):453–457.

26. Axtell RL. Zipf distribution of US firm sizes. Science. 2001;293:1818–1820.

27. Gaffeo E, Gallegati M, Palestrini A. On the size distribution of firms: additional evidence from the G7 countries. Physica A: Statistical Mechanics and its Applications. 2003;324(1–2):117–123. doi:10.1016/S0378-4371(02)01890-3.

28. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661–703.

29. Stanley MH, Amaral LA, Buldyrev SV, Havlin S, Leschhorn H, Maass P, et al. Scaling behaviour in the growth of companies. Nature. 1996;379(6568):804–806.

30. Kaldor N. A model of economic growth. The Economic Journal. 1957;67(268):591–624.

31. Feynman RP, Leighton RB, Sands M. The Feynman Lectures on Physics, Desktop Edition Volume I. vol. 1. Basic books; 2013.

32. Kesten H. Random difference equations and Renewal theory for products of random matrices. Acta Mathematica. 1973;131(1):207–248. doi:10.1007/BF02392040.

33. Biham O, Malcai O, Levy M, Solomon S. Generic emergence of power law distributions and Levy-Stable intermittent fluctuations in discrete logistic systems. Physical Review E. 1998;58(2):1352.

34. Gabaix X. Zipf's Law for Cities: An Explanation. The Quarterly Journal of Economics. 1999;114(3):739–767. doi:10.1162/003355399556133.

35. Harrison JM. Brownian motion and stochastic flow systems. New York: Wiley; 1985.

36. Hegyi G, Neda Z, Santos MA. Wealth distribution and Pareto's law in the Hungarian medieval society. Physica A: Statistical Mechanics and its Applications. 2007;380(1):271–277.

37. Kahan A, Hellie R. The plow, the hammer, and the knout: An economic history of eighteenth-century Russia. Chicago: University of Chicago Press; 1985.

38. Cottrell F. Energy & Society (Revised): The Relation Between Energy, Social Change, and Economic Development. Bloomington: AuthorHouse; 2009.

39. Warde P. Energy Consumption in England & Wales, 1560-2000. Consiglio nazionale delle ricerche, Istituto di studi sulle societa del Mediterraneo; 2007.

40. Malanima P. Energy Consumption in the Roman World. In: The Ancient Mediterranean Environment between Science and History. Boston: Brill; 2013. p. 13–36.

41. Fix B. Rethinking Economic Growth Theory from a Biophysical Perspective. Hall C, editor. New York: Springer; 2015.

42. Giampietro M, Mayumi K, Sorman A. The Metabolic Pattern of Societies: Where Economists Fall Short. New York: Routledge; 2012.

43. Pimentel D, Pimentel MH. Food, Energy, and Society, Third Edition. New York: CRC Press; 2007.

Appendix B

Appendices For Evidence for a Power Theory of Personal Income Distribution

Supplementary materials for this paper are available at the Open Science Framework repository:

The supplementary materials include:

1. Data for all figures appearing in the paper;
2. Raw source data;
3. R code for all analysis;
4. Compustat model code.

B.1   Data Sources

Age

Age mean income and within-group Gini index data is from US Census Tables PINC-02 over the years 1994-2015. Age is grouped into the following 4 categories: 18-24, 25-44, 45-64, 65 and older.

Census Blocks

Census blocks data comes from the US Census American Community Survey (ACS) over the years 2010-2014. This data is tabulated at the *household* (rather than individual) level. Neither mean household income nor household Gini index data is directly available from the ACS at the census block level. I calculate mean household income by dividing aggregate household income by the number of households.

Within group Gini indexes are estimated from binned income data using the R 'binequality' package. I construct two different estimates: one using a parametric method and the other using the midpoint method. For the parametric method, I fit either a lognormal or gamma distribution (whichever is best) to the binned data. Gini indexes are then calculated from this fitted distribution. The midpoint method uses midpoints of the bins to estimate the Gini index. The midpoint of the upper bin (which has an open upper bound) is estimated from a best-fit power law (again, implemented in the R binequality package). Both Gini estimates are used in Figures 10 and 11 in the main paper. The R code implementing this method is included in the Supplementary Material.

Census Tracts

Census tract data comes from the US Census American Community Survey (ACS) over the years 2010-2015. Mean income data comes from series S1902, while intra-tract Gini indexes come from series B19083.

Cognitive Score

The between-within indicator for cognitive score is estimated using data from Figure 6 in Bowles et al. [1]. Bowles' figure presents 65 different estimates (from 24 studies between 1963 and 1992) of the relation between individual income and cognitive score. The strength of this relation is quantified using the beta

coefficients (β) of a log-linear regression. This coefficient represents the slope of the regression equation shown in Eq. B.1, where the logarithm of income – $\log(I)$ – and cognitive score (S) have first been normalized to have a mean of 0 and standard deviation of one.

$$\log(I) = \alpha + \beta S \tag{B.1}$$

I use Engauge Digitizer to extract data from Bowles' graph. I then use a model to estimate the G_{BW} metric from Bowles' reported beta coefficients. The model creates a stochastic log-linear scaling relation between income and cognitive score. By adjusting the strength of this relation, we can create modeled data that has an equivalent beta coefficient to any of the points in Bowles' figure. I then use the model to calculate a G_{BW} for this beta coefficient.

The model assumes that cognitive score (S) is a normally distributed random variate with a mean of 100 and standard deviation of 15:

$$S \sim \mathcal{N}(100, 15) \tag{B.2}$$

We assume that the natural log of mean income ($\ln \bar{I}$) scales exponentially with cognitive score (Eq. B.3). Since there is no evidence that extreme IQs lead to extreme incomes (at either the bottom or top end), I do not include them in the model. I model only those individuals with scores that are within two standard deviations of the mean ($70 < S < 130$). The parameter a determines how strongly cognitive score affects average income.

$$\ln(\bar{I}) = a(S - 70) \quad \text{for} \quad 70 < S < 130 \tag{B.3}$$

We assume that individual income (I) is a stochastic variable that is distributed according to a lognormal distribution defined by the location parameter μ and scale parameter σ:

$$I \sim \ln \mathcal{N}(\mu, \sigma) \tag{B.4}$$

Equation B.5 shows how mean income $\bar{I}$ is related to μ and σ.

$$\bar{I} = e^{u + \frac{1}{2}\sigma^2} \tag{B.5}$$
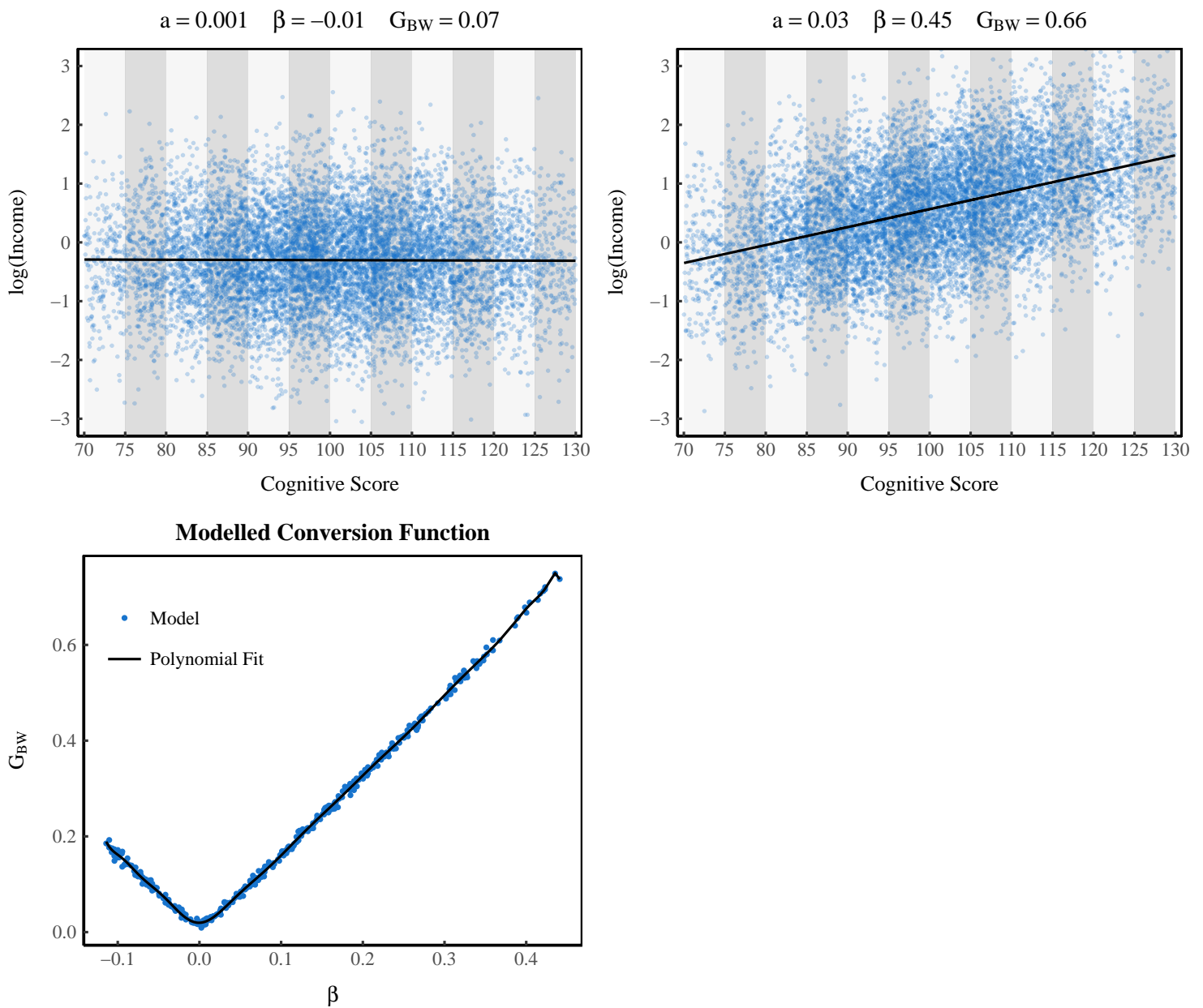
Figure B.1: Cognitive Score Method — Estimating the Between-Within Indicator (G_{BW}) from Normalized Regression Coefficients (β)

This figure shows an example of the model for converting cognitive score regression data from Bowles et al. [1] to the G_{BW} indicator. Using equations B.2-B.7, I create a stochastic scaling relation between the logarithm of individual income and cognitive score. The strength of this scaling relation is determined by the parameter a, and is quantified by the normalized regression coefficient β. The top left panel shows a weak scaling relation, while the top right shows a strong scaling relation. I then group individuals into cognitive score intervals of 5 (vertical grey bars) and calculate the G_{BW} metric. The bottom left panel shows the resulting relation between G_{BW} and β that is used to convert Bowles' data.

By taking the logarithm and solving for μ, Eq. B.5 can be transformed into the following:

$$\mu = \ln(\bar{I}) - \frac{1}{2}\sigma^2 \tag{B.6}$$

We then substitute Eq. B.3 into Eq. B.6 to define μ in terms of cognitive score:

$$\mu = a(S - 70) - \frac{1}{2}\sigma^2 \tag{B.7}$$

The algorithm for the model is as follows. We first generate a random cognitive score S, drawn from the normal distribution defined by Eq. B.2. We then take this score and use Eq. B.7 to define the parameter μ. Finally, we generate a random income for this cognitive score, drawn from the lognormal distribution defined by Eq. B.4. This process is then repeated as many times to generate a stochastic dataset relating income to cognitive score.

The model has 2 free parameters: a and σ. Parameter a affects the rate at which income scales with cognitive score, while σ determines the amount of dispersion around the mean income $\bar{I}$. The parameter σ strongly affects the level of 'global' inequality in the model, while a has only a slight effect. For this reason, it is important to chose σ such that the model has a realistic level of inequality. I chose $\sigma = 0.8$. Over the chosen range of $-0.007 < a < 0.03$, this produces global Gini indexes that range between 0.43 and 0.47, which is roughly consistent with US data for the second half of the 20th century.

For any given value of a, the model generates a stochastic relation between cognitive score and income I. Two examples are shown in Figure B.1. In Figure B.1A, the small value of a produces a very weak relation between income and cognitive score. In Figure B.1B, the larger value of a produces a stronger relation between income and cognitive score.

The strength of the relation is indicated by the beta coefficient β. The purpose of this model is to convert the values of β reported by Bowles et al. into the between-within Gini ratio that is used in this paper. To make this conversion, we must group individuals by their cognitive score. The bin-size of this grouping is arbitrary; I construct groupings of 5 point cognitive score intervals (indicated by the grey vertical bands in Fig. B.1A-B). For each group, we calculate the mean income and within-group Gini index. The between-within Gini metric G_{BW} is then calculated by the method outlined in the main paper.

I repeat this process for many different values of a, which produces the modeled relation between G_{BW} and β shown in Figure B.1C. I then fit this relation

with a high order polynomial that serves as the function for converting Bowles' β values into the G_{BW} values used in this paper.

Counties

US County data comes from the American Community survey for the years 2006-2015. County Gini indexes are from series B19083, while mean income is from series S1902.

Education

Mean income and within-group Gini indexes by educational level come from US Census tables PINC-03 over the years 1994-2014. Educational level is categorized into the following groups:

- Less Than 9th Grade
- 9th to 12th Nongrad
- High school Graduate (Incl GED)
- Some College
- Associate Degree

- Bachelor's Degree
- Master's Degree
- Professional Degree
- Doctorate Degree

Employees vs. Self-Employment

To calculate mean income and intra-group Gini indexes for employees and self-employed workers, I use US Census table PINC-07 between 1994 and 2015. This table contains three categories: *Government Wage And Salary Workers*, *Private Wage And Salary Workers*, and *Self-Employed Workers*. Table B.1 shows how I have mapped these categories onto the 'employees' and 'self-employed' sectors.

Table B.1: Grouping Categories of Census Table PINC-07

Employees	Self-Employed
• Government Wage And Salary Workers	• Self-Employed Workers
• Private Wage And Salary Workers	

Self-employed mean income and within-group Gini index come directly from PINC-07. To calculate the mean income of employees, I use the average of the means of government workers and private workers, weighted by the size of each group.

Since Gini indexes are not additive, I estimate the inequality among employees from binned data. I first add the binned income counts of both government and private wage/salary workers to get a binned income distribution for all 'employees'. From this binned data, I then use the the R 'binequality' package to estimate private sector Gini indexes.

I construct two different estimates: one using a parametric method and the other using the midpoint method. For the parametric method, I fit various theoretical distributions to the binned data. Gini indexes are then calculated from the best-fitting distribution. The midpoint method uses midpoints of the bins to estimate the Gini index. The midpoint of the upper bin (which has an open upper bound) is estimated from a best-fit power law (again, implemented in the R binequality package). Both Gini estimates are used in Figures 10 and 11 of the main paper. The R code implementing this method is included in the Supplementary Material.

Firms

Firm between-within inequality calculations use the Compustat database, and are a combination of empirical and modeled data. Firm mean income is calculated directly from Compustat data by dividing Total Staff Expenses (series XLR) by the number of employees (series EMP). Firm internal inequality is estimated using the Compustat Model. See Appendix B.2-B.7 for a detailed discussion.

Full and Part Time Workers

Full and part time worker mean income and within-group inequality data comes from US Census tables PINC-05 from 1994-2015.

Parent Income Percentile

'Parent income percentile' refers to grouping individuals by the income percentile of their parents. My calculations are done using Table 1 and 2 from the online data tables of Chetty et al. [2] — a seminal study of US intergenerational mobility. For every parent income percentile x, Table 1 gives the probability $p(x, y)$ that the corresponding child will have an income in percentile y. Table 2 gives the mean income ($\bar{I}_y$) of each child percentile y.

My method for estimating group mean incomes and within-group inequality is shown in equations B.8 and B.9. The first step is to convert the probability $p(x, y)$ into an integer $w(x, y)$ that can be used to weight incomes. Since the

probabilities in Table 1 contain 7 decimal places, I multiply $p(x, y)$ by 10^7 (Eq. B.8).

$$w(x, y) = p(x, y) \times 10^7 \tag{B.8}$$

For each each income percentile x, we then create a vector of child incomes ($\mathbf{I_x}$) by repeating each child percentile mean income $\bar{I}_y$ by the weighting factor $w(x, y)$. Here the notation $\overset{\times w(x,y)}{\cdots\cdots}$ indicates that the value $\bar{I}_y$ is repeated $w(x, y)$ times.

$$\mathbf{I_x} = \left(\bar{I}_1 \overset{\times w(x,1)}{\cdots\cdots}, \; \bar{I}_2 \overset{\times w(x,2)}{\cdots\cdots}, \dots, \; \bar{I}_{100} \overset{\times w(x,100)}{\cdots\cdots} \right) \tag{B.9}$$

We can think of $\mathbf{I_x}$ as an estimated income distribution for children of parents in income percentile x. Mean income and within-group inequality of parent group x are then estimated by calculating the mean and Gini index (respectively) of $\mathbf{I_x}$.

Note that this method neglects the income dispersion *within* each child income percentile (Chetty et al. do not provide this data). Thus, our estimated Gini index will have a slight *downward* bias. The R code implementing this method is included in the Supplementary Material.

Hierarchical Level — Heyman

This data comes from Fredrik Heyman's [3] study of 560 Swedish firms in the year 1995. His dataset includes only the top 4 levels of management. I include Heyman's results in the paper with the caveat that his data does not represent all hierarchical levels.

Heyman (Table A.1) provides the mean and standard deviation of the *logarithm* of incomes in each level. I estimate mean income ($\bar{I}$) and Gini index (G) by hierarchical level by assuming that intra-hierarchical level income is lognormally distributed. Under this assumption, the *mean* of log income is equal to the lognormal location parameter μ, while the *standard deviation* of log income is equal to the scale parameter σ. Equations B.10 and B.11 then define the mean income and Gini index (respectively) of each hierarchical level.

$$\bar{I} = e^{\mu + \frac{1}{2}\sigma^2} \tag{B.10}$$

$$G = \text{erf}\left(\frac{\sigma}{2}\right) \tag{B.11}$$
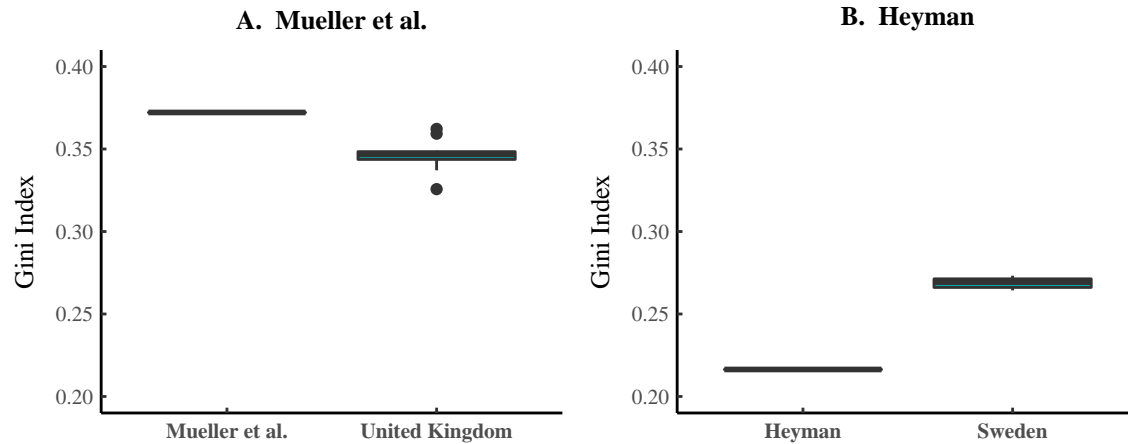
Figure B.2: Aggregate Inequality Implied by Hierarchy Data

This figure compares levels of inequality implied by the Mueller et al. and Heyman firm samples against the inequality in their respective countries. UK inequality data is over the period 2004-2013, the same as covered by Mueller's data. Heyman's study covers the year 1995, while Swedish data is from 2004-2013. UK and Sweden Gini data is from the World Bank, series SI.POV.GINI.

Figure B.2 shows how the implied aggregate inequality within the Heyman's sample compares to Swedish empirical data. Heymans's sample implies a bit less inequality than the empirical data. This is not surprising, however, as Heyman's data includes only the top 4 levels of management.

Hierarchical Level — Mueller et al.

This data comes from Mueller et al. [4], who study the hierarchical pay structure of 880 United Kingdom firms over the period 2004-2013. For each hierarchical level, Mueller et al. provide the mean income as well as the 25th, 50th, and 75th income percentiles. To estimate intra-level inequality, I adapt R code written by Andrie de Vries to find the best-fit theoretical distribution for each hierarchical level. Intra-hierarchical level inequality is then calculated from the best-fit distribution.

Figure B.2 shows how aggregate inequality within the Mueller et al. sample compares to UK data over the same period. Although the Mueller et al. data is slightly more unequal than the UK as a whole, it is a reasonably representative sample.

Hierarchical Level — Compustat Model

The Compustat model is discussed extensively in Appendix B.2-B.7.

Labor and Property Income

'Labor' income is defined as wages and salaries, while 'property' income is defined as the sum of interest, dividends, rents, royalties, and estates or trust income. Mean income and within-group inequality data comes from US Census tables PINC-08 from 2003-2015.

Occupation

Data for mean income and with-group inequality by occupation comes from US Census tables PINC-06 (income by occupation of longest job) between 2007 and 2015. This table classifies occupations by major type, minor type, and detailed type. I use *detailed* categories only, which amounts to between 53 to 55 different occupation groups (depending on the year).

The US Bureau of Labor Statistics also publishes occupational wage estimates (available at https://www.bls.gov/oes/tables.htm). For the sake of completeness, I analyze this data here, but do not use it for the results published in the paper. The BLS data differs from Census data in the ways shown in Table B.2.

Because the BLS does not report within-occupation Gini indexes directly, I estimate them via the reported values for 10th, 25th, 50th, 75th, and 90th income percentiles. Using an adaption of R code written by Andrie de Vries, I fit a variety of theoretical distributions to this percentile data. Within-occupation Gini indexes are calculated from the best-fit theoretical distribution.

The resulting between-within inequality indicator is shown in Figure B.3A, alongside the results from Census occupation data. The two calculations differ starkly. Census data indicates that between-occupation inequality is *less* than within-occupation inequality; however, the BLS data indicate the *reverse*.

Which result is correct? The answer to this question depends on the type of income inequality we are interested in explaining. The BLS data covers only full-time, non-self-employed workers earning labor income. Census data, on the other hand, includes *all* individuals. For the purposes of this paper, the Census data is a better choice.

To demonstrate the differences between BLS and Census data, we can calculate the aggregate inequality that is implied by the data. To do this, I make

Table B.2: Contrasting the US Census and BLS Occupational Income Data

Census Data	BLS Data
Includes self-employed workers	Does *not* include self-employed workers
Income for all full and part-time work	Income is for full-time equivalent workers only (hourly wage × 2080 hours).
Includes non-labor income	Does *not* include non-labor income
53-55 detailed occupation types	700-800 detailed occupational types
Reports Gini index directly	Reports 10th, 25th, 50th, 75th and 90th income percentiles

the simplifying assumption that all occupations have lognormal income distributions. Given the mean income ($\bar{I}$) and within-group Gini index (G) of a particular occupation, we can define the lognormal location (μ) and scale (σ) parameters:

$$\sigma = 2 \cdot \mathrm{erf}^{-1}(G) \tag{B.12}$$

$$\mu = \ln(\bar{I}) - \frac{1}{2}\sigma^2 \tag{B.13}$$

If the number of individuals engaged in this occupation is n, we can create a simulated occupational income distribution by generating n values from the lognormal distribution defined by μ and σ. We repeat this process for every occupation, and then aggregate all of the simulated occupational income distributions. The Gini index of this aggregated distribution is the level of inequality that is *implied* by the data.

The results of this analysis are shown in Figure B.3B. As expected, the inequality that is implied by Census data closely matches actual levels of inequality between all individuals. However, the inequality implied by BLS data is *much* lower — a clear result of the restrictions underlying the BLS methods. For the purposes of this paper, the Census data is the correct choice.

Owner vs. Renter

Mean income and intra-group Gini indexes by home-ownership status come from US Census table PINC-01 between 1994 and 2015. I use the following two categories: (1) Owner Occupied; and (2) Renter Occupied.

A. Conflicting Data

B. Empirical and Implied Inequality
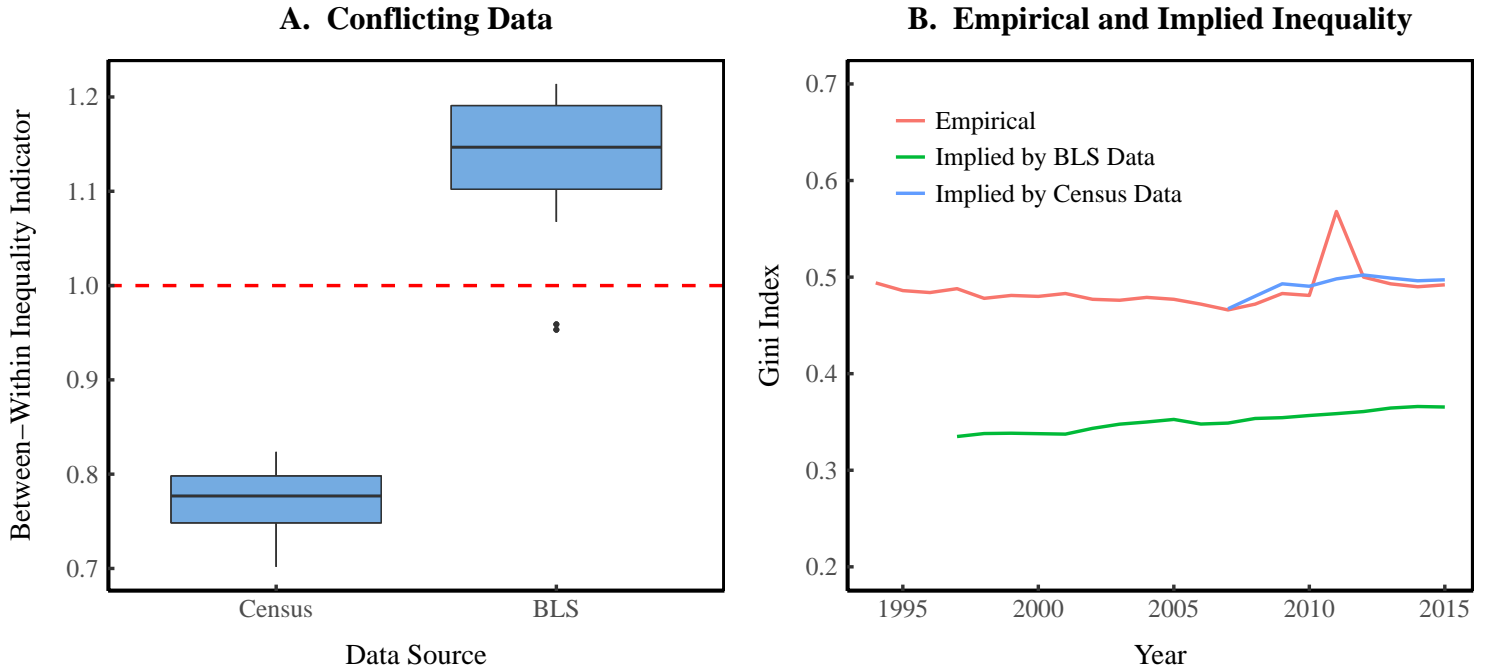
Figure B.3: Inequality by Occupation — Data Discrepancies

This figure shows differences in the occupation income data published by the US Census versus that published by the US Bureau of Labor Statistics (BLS). Panel A shows calculations of the between-within indicator (G_{BW}) for both BLS and Census data. The BLS data gives a much higher G_{BW} value, meaning between-occupation inequality is far greater (relative to within-occupation inequality) in BLS data than it is in the Census data. Why? The two datasets imply very different levels of aggregate (society-wide) inequality, as shown in panel B. This is because the BLS data includes only full-time wage/salary earners, while the Census data includes all individuals. The level of aggregate inequality implied by the Census data closely matches actual levels. I use Census data only in this paper.

Public vs. Private Sector

To calculate mean income and intra-group Gini indexes for public and private sector workers, I use US Census table PINC-07 between 1994 and 2015. This table contains three categories: *Government Wage And Salary Workers*, *Private Wage And Salary Workers*, and *Self-Employed Workers*. Table B.3 shows how I have mapped these categories onto the 'public' and 'private' sectors.

The mean income and Gini index of the public sector is thus equivalent to the values for government wage/salary workers. Private sector mean income is calculated as the average of the means of private wage/salary worker income

Table B.3: Grouping Categories of Census Table PINC-07

Public Sector	Private Sector
• Government Wage And Salary Workers	• Private Wage And Salary Workers
	• Self-Employed Workers

and self-employed worker income, weighted by the size of each group.

Since Gini indexes are not additive, I estimate the inequality of private sector income from binned data. I first add the binned income counts of both private wage/salary workers and self-employed workers to get a binned income distribution for the private sector. From this binned data, I then use the R 'binequality' package to estimate private sector Gini indexes.

I construct two different estimates: one using a parametric method and the other using the midpoint method. For the parametric method, I fit various theoretical distributions to the binned data. Gini indexes are then calculated from the best-fitting distribution. The midpoint method uses midpoints of the bins to estimate the Gini index. The midpoint of the upper bin (which has an open upper bound) is estimated from a best-fit power law (again, implemented in the R binequality package). Both Gini estimates are used in Figures 10 and 11 of the main paper. The R code implementing this method is included in the Supplementary Material.

Race

Data for mean income and within-group inequality by race comes from US Census tables PINC-01 between 1994 and 2015. Data for 2002-2015 contain the following four categories: Asian, Black, Hispanic, and White. Data for 1994-2001 contains only three categories: Black, Hispanic, and White.

Religion

Religion income data comes from the Pew Research Center 2007 U.S. Religious Landscape Survey (RLS). I use the following groups:

- Agnostic
- Atheist
- Baptist
- Buddhist

- Church of Christ, or Disciples of Christ
- Congregational or United Church of Christ

- Episcopalian or Anglican
- Hindu
- Holiness (Nazarenes, Wesleyan Church, Salvation Army)
- Jewish
- Lutheran
- Methodist
- Mormon
- Muslim

- Nondenominational or Independent Church
- Nothing in particular
- Orthodox
- Pentecostal
- Presbyterian
- Reformed (include Reformed Church in America; Christian Reformed; Calvinist)
- Roman Catholic

The RLS reports the binned income of each respondent. I use the R 'binequality' package to estimate group mean income and Gini indexes (using the midpoint method). Because some religions have a very small sample size, I use the bootstrap method [5] to estimate a plausible range of values for group mean incomes and intra-group income inequality.

Sex

Data for mean income and within-group inequality by sex (male/female only) comes from US Census tables PINC-01 between 1994 and 2015.

Urban vs. Rural

Data for urban/rural mean income and intra-group Gini index comes from US Census tables PINC-01 between 1994 and 2015. I define 'urban' as individuals inside metropolitan statistical areas, and 'rural' as individuals outside these areas.

B.2 Hierarchical Structure and Pay within Case-Study Firms

Purely based on worldly experience, most people would agree that firms are hierarchically organized, and that pay tends to increase as one moves up the hierarchy. But the exact structure of this hierarchy has not been widely studied.

Figure B.4 shows the hierarchical employment and pay structure of six different firms whose data has been made available to social scientists. The firms remain anonymous, and are named after the authors of the case-study papers (see Table B.6 for details). By and large, these studies confirm our basic intuition about firm structure. Although the exact shapes vary, all of the firms in Figure B.4 have a roughly pyramidal employment structure and inverse pyramid pay structure.

To analyze the structure of these firms in further detail, I define and calculate the three metrics shown in Table B.4. Results are shown in Figure B.5. Figure B.5A shows how the *span of control* changes as a function of hierarchical level. The data shows unambiguously that the span of control tends to *increase* as one moves up the hierarchy. Figure B.5B shows how the *inter-level pay ratio* changes as a function of hierarchical level. Again, this ratio tends to *increase* as one moves up the hierarchy. Figure B.5C shows the *intra-level Gini index* as a function of hierarchical level. Unlike the other two quantities, intra-level income inequality seems to be more-or-less constant across all hierarchical levels (a linear regression reveals no significant trend).

As well as single-firm case studies, a handful of studies exist that have analyzed the hierarchical structure of multiple firms. Although these aggregate studies offer more scope than case studies, they have one major shortcoming: they rarely study the structure of *entire* firms. Instead, these aggregate studies typically focus on the span of control and pay in the *top* hierarchical levels of a

Table B.4: Metrics of Firm Hierarchical Employment and Pay Structure

Name	Definition
Span of Control	Employment ratio between adjacent hierarchical levels.
Inter-Level Pay Ratio:	Ratio of mean pay between adjacent hierarchical levels.
Intra-Level Gini Index	The Gini index of income inequality within a specific hierarchical level of a firm.

Table B.5: Stylized Facts About Firm Employment and Pay

1. The span of control tends to *increase* with hierarchical level.

2. The the inter-level pay ratio tends to *increase* with hierarchical level.

3. Intra-level income inequality is approximately *constant* across all hierarchical levels.

---

firm (the CEO and adjacent upper management levels).

This is a problem. I have conceptualized firm hierarchy as a *bottom-up* ranking (see Fig. 3.8 in the main paper). Under this definition, a CEO in a small firm will be in a very different hierarchical level than a CEO in a large firm. Thus, when we compare the span of control between a CEO and his subordinates across firms of different size, we are likely comparing very different hierarchical levels.

As a result of this shortcoming, the aggregate studies summarized in Table B.7 are less useful than the case studies in Table B.6. However, keeping in mind their shortcomings, these aggregate studies still reveal the same trends as our case study data. Figure B.6 shows the analysis of these aggregate studies. Note that hierarchical level is counted from the *top down*, where level 0 is the CEO. Figure B.6A and B (respectively) indicate that there is still a tendency for the span of control and inter-level pay ratio to increase with hierarchical level.

From this evidence, I propose the 'stylized' facts shown in Table B.5 about firm employment and pay structure.

A. Firm Hierarchical Employment Structure

Percent of Employment

B. Firm Hierarchical Pay Structure

Average Pay (Base = 1)
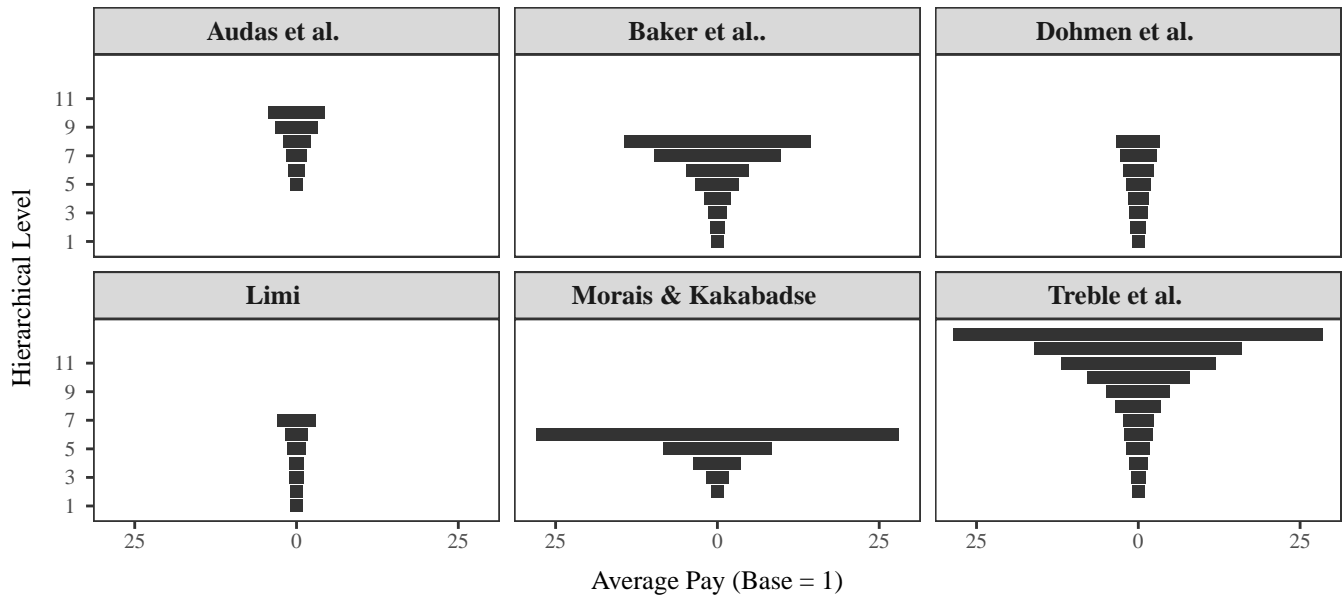
Figure B.4: The Hierarchical Employment and Pay Structure of Six Different Firms

This figure shows the pyramid structure of six different case study firms. Panel A shows the hierarchical structure of employment, while panel B shows the hierarchical pay structure.
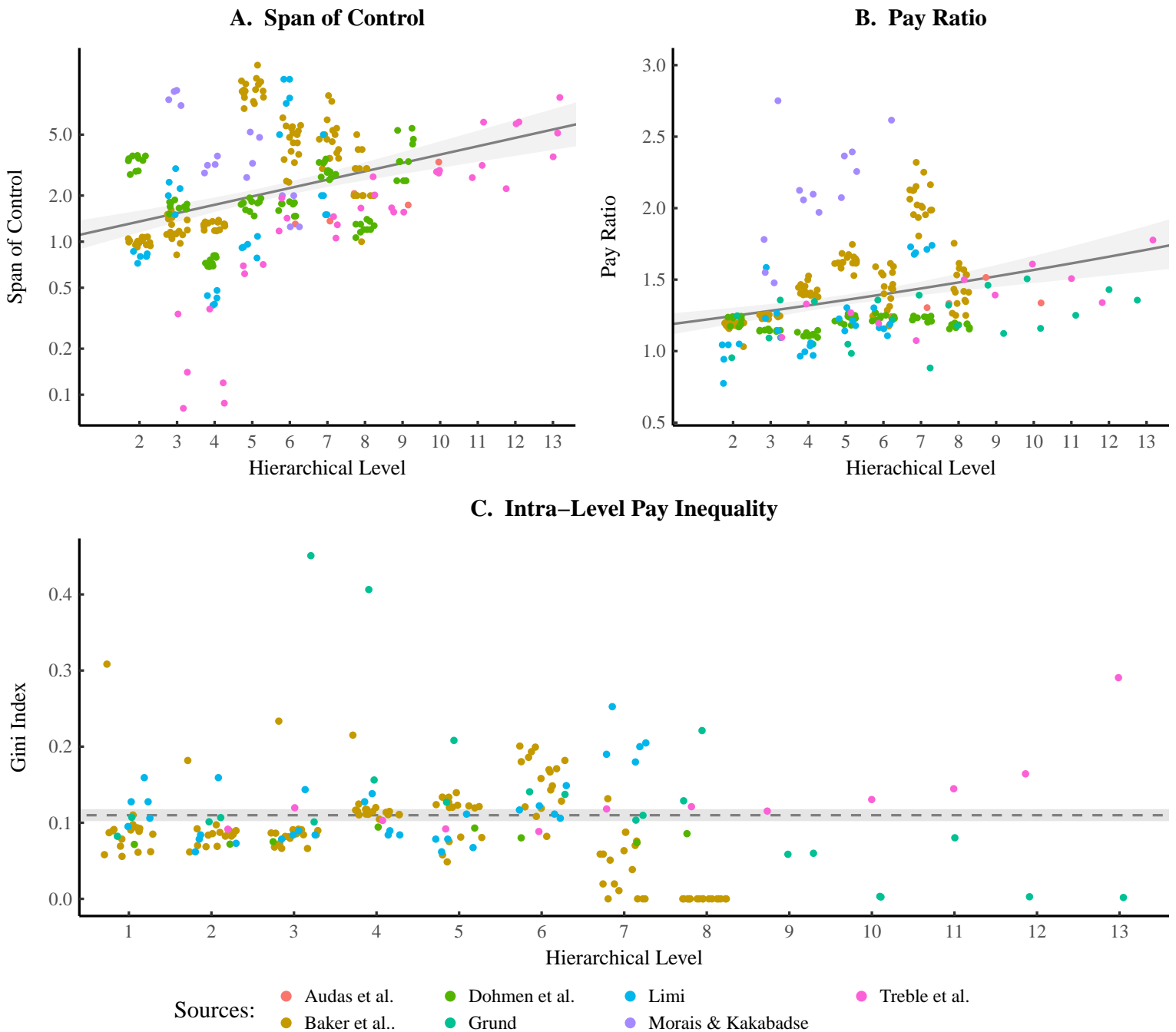
Figure B.5: Case Studies of Firm Hierarchical Structure

This figure shows data from 7 different single-firm case studies. Panel A shows how the span of control (the employment ratio between adjacent levels) relates to hierarchical level. Panel B shows how the pay ratio between adjacent levels varies with hierarchical level. In these two panels, span of control and pay ratios between two hierarchical levels, h and $h-1$, are plotted on the x-axis at level h. Panel C shows levels of income inequality *within* individual hierarchical levels of each firm. Note that horizontal 'jitter' has been introduced in all three plots in order to better visualize the data (hierarchical level is a discrete variable). Grey regions correspond to the 95% confidence interval for regressions (or in panel C, the mean).

A. Span of Control

B. Pay Ratio

Ariga et al. Eriksson Leonard Mueller et al. Tao & Chen
Bell & Van Reenen Heyman Main et al. Rajan & Wulf
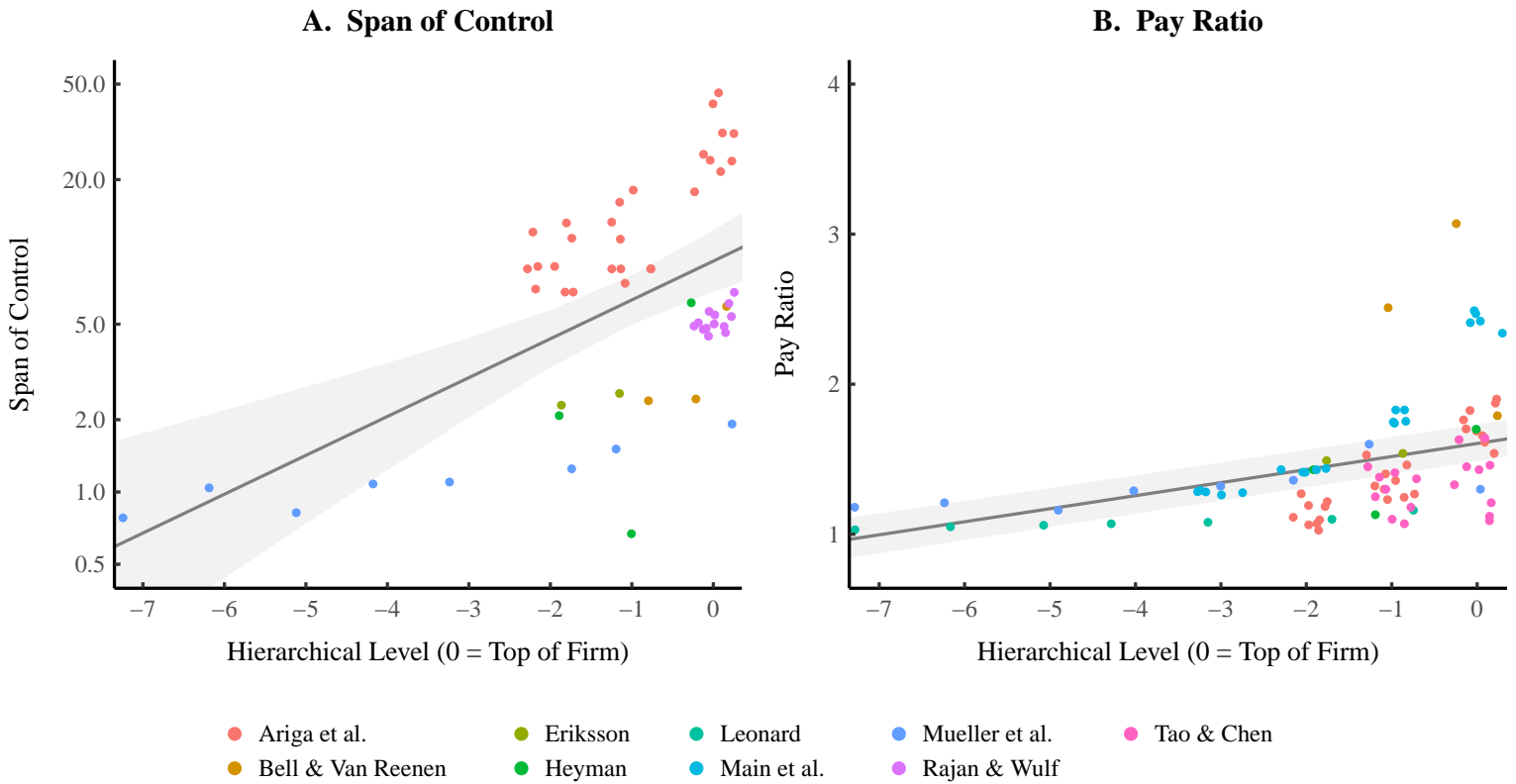
Figure B.6: Aggregate Studies of Firm Hierarchical Structure

This figure shows data from 9 different aggregate firm studies. Most of these studies only survey the top several hierarchical levels in each firm. Because of this, I order hierarchical levels from the top down, where the CEO is level 0, the level below is -1, etc. Panel A shows how the span of control (the employment ratio between adjacent levels) relates to hierarchical level. Panel B shows how the pay ratio between adjacent levels varies with hierarchical level. In both plots, horizontal 'jitter' has been introduced in order to better visualize the data (hierarchical level is a discrete variable). Grey regions correspond to the 95% confidence interval for regressions.

Table B.6: Firm Case Studies

Source	Years	Country	Firm Levels	Span of Control	Level Income	Level Income Dispersion
[6]	1992	Britain	All	✓	✓	
[7]	1969-1985	United States	Management	✓	✓	✓
[8]	1987-1996	Netherlands	All	✓	✓	✓
[9]	1995 & 1998	US and Germany	All		✓	✓
[10]	1991-1995	Portugal	All	✓	✓	✓
[11]*	2007-2010	Undisclosed	All	✓	✓	
[12]	1989-1994	Britain	All	✓	✓	✓

Notes: This table shows metadata for the firm case studies displayed in Fig. B.5. 'Firm Levels' refers to the portion of the firm that is included in the study. 'Management' indicates that only management levels were studied.
 *For the analysis conducted in this paper I discard (as an outlier) the bottom hierarchical level in Morais and Kakabadse's data.

Table B.7: Firm Aggregate Studies

Source	Years	Number of Firms	Country	Firm Levels	Span of Control	Level Income
[13]	1981-1989	unknown	Japan	All	✓	✓
[14]	2001-2010	552	United Kingdom	Top 3	✓	✓
[15]	1992-1995	210	Denmark	Management	✓	✓
[3]	1991,1995	560	Sweden	Management	✓	✓
[16]	1981-1985	439	United States	Top 9		✓
[17]	1980–1984	200	United States	Top 4		✓
[4]	2004-2013	880	United Kingdom	All	✓	✓
[18]	1986-1998	261	United States	Top 2	✓	
[19]	1986-1998	8101	Taiwan	Top 2		✓

Notes: This table shows metadata for the aggregate studies displayed in Fig. B.6. 'Firm Level's refers to the portion of the firm that is included in the study. 'Top 2', 'Top 3', etc. indicates that only the top *n* levels were included in the study (where the top level is the CEO).

Table B.8: Income Inequality Within Case Study Firms		
Source	Years	Mean Gini Index
[7]	1969-1985	0.32
[8]	1991	0.18
[10]	1991-1995	0.15
[11]	2007-2010	0.23
[12]	1989-1997	0.26

B.2.1   Inequality Within Case Study Firms

I report here my estimates for inequality within the case study firms. Of the seven case studies summarized in Table B.6, only one (Morais and Kakabadse) directly reports a firm Gini index. However, four other studies — Baker et al., Dohmen et al., Lima, and Treble et al. — provide enough data to allow estimates of firm internal inequality. I outline my calculation methods below. The resulting Gini estimates are shown in Table B.8.

Baker et al.

Baker et al. [7] have made their raw personnel data publicly available at the site below. I use this raw data to calculate the firm internal Gini index.

http://faculty.chicagobooth.edu/michael.gibbs/research/index.html

Dohmen et al.

Dohmen et al. [8] report the following data that I use to estimate the firm Gini index:

1. Fraction of employment by hierarchical level (Tbl. 1);
2. Density plots of income distribution by hierarchical level (Fig. 5).

I use the *Engauge Digitizer* program to digitize and pull data from the density plots. I then use the resulting numerical density functions to estimate the firm Gini index.

We define $f_h(x)$ as the income density function for hierarchical level h. The income density function of the entire firm $f_T(x)$ is then defined by Eq. B.14 – the

sum of the density functions for each hierarchical level, weighted by the fraction of total employment (E_h/E_T).

$$f_T(x) = \sum_{h=1}^{n} \frac{E_h}{E_T} \cdot f_h(x) \tag{B.14}$$

The firm Gini index is then defined by Eq. B.15-B.17. Equation B.15 defines the mean income of the firm ($\bar{I}$), while equation B.16 defines the cumulative income distribution function $F(x)$. Equation B.17 then defines the Gini index (G). I use numerical integration implemented in R to evaluate these integrals.

$$\bar{I} = \int_0^\infty x \cdot f_T(x) \, dx \tag{B.15}$$

$$F(x) = \int_0^\infty f_T \, dx \tag{B.16}$$

$$G = \frac{1}{\bar{I}} \int_0^\infty F(x)(1 - F(x)) dx \tag{B.17}$$

Grund

Grund [9] does not provide enough information to calculate firm-wide inequality. However, I am able to calculate intra-level income dispersion, (which appears in Fig. B.5C). I use data from Grund's Fig. 1, which shows mean income by level, as well as what I assume to be 5th and 95th percentiles. After digitizing this data, I use the best-fit theoretical distribution to estimate the Gini index.

Lima

Lima [10] provides the following summary statistics, which I use to estimate a firm Gini index:

1. Employment within each hierarchical level (Tbl. 1);
2. Mean pay within each hierarchical level (Fig. 2);
3. Wage coefficient of variation by hierarchical level (Tbl. 6).

I use the *Engauge Digitizer* program to digitize and pull data from Fig. 2. To calculate the firm Gini index, I assume income within each hierarchical level is lognormally distributed. For each hierarchical level h, I then use equation B.18

to define the lognormal scale parameter σ that produces a distribution with an equivalent coefficient of variation, c_v:

$$\sigma_h = \sqrt{\ln(c_v^2 + 1)} \tag{B.18}$$

Once we have σ_h, we use equation B.19 to calculate the lognormal location parameter μ for each hierarchical level . Here $\bar{I}_h$ is the mean pay in hierarchical level h (which Lima reports directly).

$$\mu_h = \ln(\bar{I}_h) - \frac{1}{2}\sigma_h^2 \tag{B.19}$$

Once we have the appropriate lognormal parameters for each hierarchical level, we use these distributions to create a simulated payroll. To do this, we draw E_h numbers (employment in level h) from each lognormal distribution $\ln \mathcal{N}(\mu_h, \sigma_h)$. I then calculate the Gini index from this simulated payroll.

Treble et al.

Treble et al. [12] report the following summary statistics, which I use to estimate a firm Gini index:

1. Employment within each hierarchical level (Fig. 2);
2. Mean pay within each hierarchical level (Fig. 3);
3. 5th and 95th wage percentile by hierarchical level (Fig. 4).

Again, I use *Engauge Digitizer* to pull data from all graphs. To estimate the intra-level Gini index, I adapt code writtent by Andrie de Vries to fit a parameterized distribution to the mean and 5th/95th percentiles.

Table B.9: Notation

Symbol	Definition
a	span of control parameter 1
b	span of control parameter 2
c_v	coefficient of variation
C	CEO to average employee pay ratio
E	employment
f	either (1) a generic function;
	or (2) a probability density function
F	cumulative distribution function
G	Gini index of inequality
h	hierarchical level
$\bar{I}$	average income
μ	lognormal location parameter
n	number of hierarchical levels in a firm
p	pay ratio between adjacent hierarchical levels
r	pay-scaling parameter
s	span of control
σ	lognormal scale parameter
T	total for firm
$\downarrow$	round down to nearest integer
$\prod$	product of a sequence of numbers
$\sum$	sum of a sequence of numbers

B.3   A Hierarchical Model of the Firm

In this section, I outline the mathematics underlying my hierarchical model of the firm. The model assumptions, outlined below, are based on the stylized facts gleaned from the real-world firm data in section B.2.

Model Assumptions

1. Firms are hierarchically structured, with a span of control that increases *exponentially* with hierarchical level.

2. The ratio of mean pay between adjacent hierarchical levels increases *exponentially* with hierarchical level.

3. Intra-hierarchical-level income is lognormally distribute and constant across all levels.

Using these assumptions, I first develop an algorithm that describes the hierarchical employment within a model firm, followed by an algorithm that describes the hierarchical pay structure.

B.3.1 Generating the Employment Hierarchy

To generate the hierarchical structure of a firm, we begin by defining the span of control (s) as the ratio of employment (E) between two consecutive hierarchical levels (h), where $h = 1$ is the *bottom* hierarchical level. It simplifies later calculations if we define the span of control in level 1 as $s = 1$. This leads to the following piecewise function:

$$s_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \dfrac{E_{h-1}}{E_h} & \text{if } h \geq 2 \end{cases} \tag{B.20}$$

Based on our empirical findings in Section B.2, we assume that the span of control is *not* constant; rather it increases *exponentially* with hierarchical level. I model the span of control as a function of hierarchical level (s_h) with a simple exponential function, where a and b are free parameters:

$$s_h = \begin{cases} 1 & \text{if } h = 1 \\ a \cdot e^{bh} & \text{if } h \geq 2 \end{cases} \tag{B.21}$$

As one moves up the hierarchy, employment in each consecutive level (E_h) *decreases* by $1/s_h$. This yields Eq. B.22, a recursive method for calculating E_h. In this model, we want employment to be *whole* numbers. To accomplish this I have included the $\downarrow$ symbol to indicate that the last step is to round *down* to the nearest whole number. By repeatedly substituting Eq. B.22 into itself, we can obtain a non-recursive formula (Eq. B.23). In product notation, Eq. B.23 can be written as Eq. B.24.

$$E_h = \downarrow \frac{E_{h-1}}{s_h} \qquad \text{for} \qquad h > 1 \tag{B.22}$$

$$E_h = \downarrow E_1 \cdot \frac{1}{s_2} \cdot \frac{1}{s_3} \cdot \ldots \cdot \frac{1}{s_h} \tag{B.23}$$

$$E_h = \downarrow E_1 \prod_{i=1}^{h} \frac{1}{s_i} \tag{B.24}$$

Total employment in the whole firm (E_T) is the sum of employment in all hierarchical levels. Defining n as the total number of hierarchical levels, we get Eq. B.25, which in summation notation, becomes Eq. B.26.

$$E_T = E_1 + E_2 + \ldots + E_n \tag{B.25}$$

$$E_T = \sum_{h=1}^{n} E_h \tag{B.26}$$

In practice, n is not known beforehand, so we define it using Eq. B.24. We progressively increase h until we reach a level of zero employment. The highest level n will be the hierarchical level directly *below* the first hierarchical level with zero employment:

$$n = \{h \mid E_h \geq 1 \quad \text{and} \quad E_{h+1} = 0\} \tag{B.27}$$

To summarize, the hierarchical employment structure of our model firm is determined by 3 free parameters: the span of control parameters a and b, and base-level employment E_1.

B.3.2 Generating Hierarchical Pay

To model the hierarchical pay structure of a firm, we begin by defining the inter-hierarchical pay-ratio (p_h) as the ratio of mean income ($\bar{I}$) between adjacent hierarchical levels. Again, it is helpful to use a piecewise function so that we can define a pay-ratio for hierarchical level 1:

$$p_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \dfrac{\bar{I}_h}{\bar{I}_{h-1}} & \text{if } h \geq 2 \end{cases} \tag{B.28}$$

Based on our empirical findings in Section B.2, we assume that the pay ratio

increases *exponentially* with hierarchical level. I model this relation with the following function, where r is a free parameter:

$$p_h = \begin{cases} 1 & \text{if } h = 1 \\ r^h & \text{if } h \geq 2 \end{cases} \tag{B.29}$$

Using the same logic as with employment (shown above), the mean income I_h in any hierarchical level is defined recursively by Eq. B.30 and non-recursively by Eq. B.31.

$$\bar{I}_h = \frac{\bar{I}_{h-1}}{p_h} \tag{B.30}$$

$$\bar{I}_h = \bar{I}_1 \prod_{i=1}^{h} p_i \tag{B.31}$$

Mean income for all employees ($\bar{I}_T$) is then the weighted average of hierarchical level mean income ($\bar{I}_h$) and hierarchical level employment (E_h):

$$\bar{I}_T = \sum_{h=1}^{n} \bar{I}_h \cdot \frac{E_h}{E_T} \tag{B.32}$$

We define the CEO as the person(s) in the top hierarchical level. Therefore, CEO pay is simply $\bar{I}_n$, average income in the top hierarchical level. The *CEO-to-average-employee pay ratio* is given by the equation below. For succinctness, I refer to this ratio as the 'CEO pay ratio' C:

$$C = \frac{\bar{I}_n}{\bar{I}_T} \tag{B.33}$$

To summarize, the hierarchical pay structure of our model firm is determined by 2 free parameters: the pay-scaling parameter r, and mean pay in the base level ($\bar{I}_1$)

B.3.3   Adding Intra-Level Pay Dispersion

Up to this point, we have modelled only the *mean* income within each hierarchical level of a firm. The last step in the modelling process is to make the firm more realistic by adding pay *dispersion* within each hierarchical level.
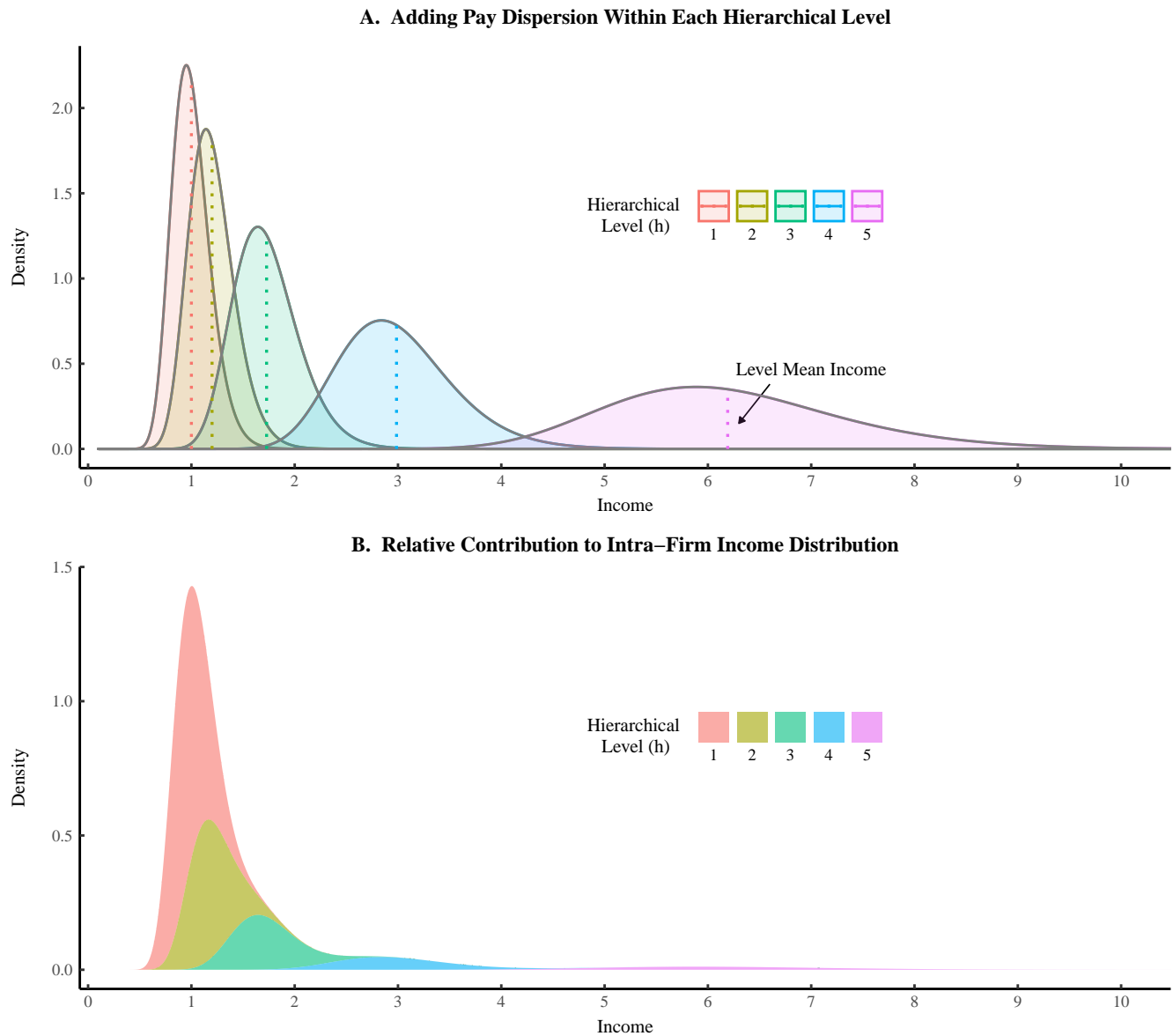
Figure B.7: Adding Intra-Level Pay Dispersion to a Model Firm

This illustrates a model firm with lognormal pay dispersion in each hierarchical level. The model firm has a pay-scaling parameter of $r = 1.2$ and an intra-level Gini index of 0.13. Panel A shows the separate distributions for each level, with mean income indicated by a dashed vertical line. Panel B shows contribution of each hierarchical level to the resulting income distribution for the whole firm (income density functions are summed while weighting for their respective employment. Span of control parameters are identical to those used Table B.10.

For this model, I assume that pay dispersion within hierarchical levels is *log-normally* distributed. This means that income (I) in hierarchical level h is described by the probability density function $\ln \mathcal{N}(I_h; \mu, \sigma)$, where μ and σ are the *location* and *scale* parameters, respectively:

$$\ln \mathcal{N}(I_h; \mu, \sigma) = \frac{1}{I_h \cdot \sigma \sqrt{2\pi}} \exp\left[-\frac{(\ln I_h - \mu)^2}{2\sigma^2}\right] \qquad \text{(B.34)}$$

Our empirical investigation of firm case studies indicated that pay dispersion with hierarchical levels is relatively constant (see Fig. B.5C). Given this finding, I assume *identical inequality* within all hierarchical levels. This means that the lognormal scale parameter σ is the same for all hierarchical levels.

To define μ, I use Eq. B.35, the formula for the mean income ($\bar{I}_h$) of our lognormal distribution. Solving for μ gives Eq. B.36.

$$\bar{I}_h = e^{\mu + \frac{1}{2}\sigma^2} \qquad \text{(B.35)}$$

$$\mu = \ln(\bar{I}_h) - \frac{1}{2}\sigma^2 \qquad \text{(B.36)}$$

Given a value for σ (which is a free parameter), we can define the pay distribution within any hierarchical level of a firm. This process is shown graphically in Figure B.7. Figure B.7A shows the lognormal income distributions for each hierarchical level of a 5-level firm with pay-scaling parameter $r = 1.2$. Figure B.7B shows the size-adjusted contribution of each hierarchical level to the overall intra-firm income distribution. Lower levels have more members, and thus dominate the overall distribution.

Once we have defined the probability distributions governing income in each hierarchical level, the last step is to simulate individual pay, and ultimately construct a firm payroll. We do this by defining income as a random lognormal variable:

$$I_h \sim \ln \mathcal{N}(\mu_h, \sigma) \qquad \text{(B.37)}$$

We construct a completed firm payroll by drawing E_h random numbers for each level h, and combining them all in the payroll vector I. Using subscripts to denote the hierarchical level and superscripts to denote the individual in that level (ranging from 1 to E_h) we get:

$$I = \{I_1^1, I_1^2, ..., I_1^{E_1}, I_2^1, I_2^2, ..., I_2^{E_2}, ..., \bar{I}_n\} \qquad \text{(B.38)}$$

Note that the last entry, the CEO pay $\bar{I}_n$, is *not* a random variable. In order to preserve the CEO pay ratio (dictated by the Compustat dataset) I do not allow this value to vary stochastically.

B.3.4   Example of the Model Algorithm

We begin by choosing the arbitrary values of a, b, E_1, r, and $\bar{I}_1$ shown in Table B.10. We then input these values into the hierarchy-building algorithm. Column A shows the hierarchical levels of the firm (h), where $h = 1$ is the base level. Using parameters a and b, we first calculate the span of control (column B), which defines the employment-ratio between adjacent hierarchical levels. In column C, we begin with the base level and use Eq. B.22 to calculate employment in each hierarchical level. In column D, we calculate the pay ratio p_h using the pay-scaling parameter r. Finally, in column E, we calculate mean income $\bar{I}_h$ in each hierarchical level

Once we have this table of values, we can calculate aggregate statics like total employment (the sum of column C) and mean pay (the mean of column E, weighted by column C). We can also calculate the CEO pay ratio. These results are shown at the bottom of Table B.10

The last step of the model is to generate a simulated payroll by adding lognormal dispersion to each hierarchical level. For large firms, this involves drawing many random numbers from a lognormal distribution. For example purposes, it is convenient to choose a small firm. Table B.11 shows a firm with the same span of control parameters as in Table B.10, but with a base size of 10. As before, we use the model algorithm to calculate mean pay in each level. We then use Eq. B.36 to calculate the lognormal location parameter in each level. The last step is to create the simulated payroll. For each hierarchical level h, we draw E_h random numbers from a lognormal distribution with parameters σ and μ_h. Note that we do not let income in the top hierarchical level vary stochastically — this preserves the CEO pay ratio on which the model is based.

Once we have the simulated payroll, we can calculate the firm's income inequality. The resulting Gini index will vary randomly, due to the stochastic nature of the model. For large firms (more than 1000 employees) this variation is negligible. For small firms, if we wish to know the 'true' Gini index that is predicted from the sum of the lognormal density functions, we can do two things:

1. Run the model many times and take the mean of resulting sample of Gini indexes;
2. Multiply all hierarchical employment E_h by a large, constant factor.

Table B.10: Example of the Model Algorithm

Parameters

a	b	E_1	r	$\bar{I}_1$
1	0.2	10 000	1.15	1

A	B	C	D	E
Hierarchical Level	Span of Control	Employment	Pay Ratio	Mean Income
h	$s_h = e^{0.2h}$	$E_h = \downarrow \frac{E_{h-1}}{s_h}$	$p_h = 1.15^h$	$\bar{I}_h = \bar{I}_{h-1} \cdot p_h$
10	7.39	0	–	–
9	6.05	1	3.52	468.5
8	4.95	8	3.06	133.2
7	4.06	44	2.66	43.5
6	3.32	182	2.31	16.4
5	2.72	607	2.01	7.1
4	2.23	1652	1.75	3.5
3	1.82	3678	1.52	2.0
2	1.49	6703	1.32	1.3
1	–	10000	–	1

Results

Total Employment	Mean Pay	CEO Pay	CEO Pay Ratio
$E_T = \sum_{h=1}^{n} E_h$	$\bar{I}_T = \sum_{h=1}^{n} \bar{I}_h \cdot \frac{E_h}{E_T}$	$\bar{I}_n$	$C = \bar{I}_n / \bar{I}_T$
22 875	1.87	468.5	250

Table B.11: Adding Intra-Level Pay Dispersion to a Firm

Parameters

a	b	E_1	r	$\bar{I}_1$	σ
1	0.2	10	1.2	1	0.24

Hierarchical Level	Pay Ratio	Mean Pay	Scale Parameter	Location Paramter
h	$p_h = 1.2^h$	$\bar{I}_h = \bar{I}_{h-1} \cdot p_h$	σ	$\mu_h = \ln(\bar{I}_h) - \frac{1}{2}\sigma^2$
4	1.73	2.99	0.24	1.07
3	1.44	1.73	0.24	0.52
2	1.20	1.20	0.24	0.15
1	–	1.00	0.24	-0.03

Generating a Simulated Payroll

Hierarchical Level	Employment	Simulated Payroll
h	E_h	$I_h \sim \ln\mathcal{N}(\mu_h, \sigma)$
4	1	$\{2.99\}$
3	3	$\{1.62, 1.88, 1.16\}$
2	6	$\{1.11, 0.94, 1.08, 1.15, 1.07, 2.13\}$
1	10	$\{0.75, 0.65, 1.04, 1.09, 0.96, 0.95, 0.97, 1.09, 0.75, 0.87\}$

Mathematically, these two approaches produce identical results. However, method 2 is computationally faster. In our example, the Gini of the simulated payroll is $G = 0.206$. The Gini predicted from the sum of lognormal density functions is $G = 0.217$.

B.4   The Compustat Data

To model US intra-firm income distribution, I use the Compustat data series shown in Table B.12. Selected statistics from this dataset are shown in Figure B.8.

Employment, staff expense, and executive compensation data available are for roughly 300 firms per year over the period 1992–2015 (Fig. B.8A). Although this is a small firm sample, the firms themselves are very large, with mean sizes of between 20,000 and 30,000 employees (Fig. B.8B). As a result, this firm sample accounts for roughly 5% of US employment (Fig. B.8C). Unlike the total US firm size distribution, which has a power-law shape [20], this Compustat firm sample is lognormally distributed (Fig. B.8D).

From the three data series shown in Table B.12, we can calculate the following:

$$\text{Employee Mean Income} = \frac{\text{Total Staff Expenses}}{\text{Employees}} \tag{B.39}$$

$$\text{CEO Pay Ratio} = \frac{\text{Top Exec Pay}}{\text{Employee Mean Income}} \tag{B.40}$$

Note that 'CEO pay' is defined as the income of the top-paid executive in a given firm. The CEO pay ratio of this sample is lognormally distributed (Fig. B.8E) with an average of between 50 and 150 (Fig. B.8F). The normalized firm mean pay distribution is shown in Figure B.8G (each firm's pay is divided by average of the annual sample). The resulting log distribution has bimodal structure. Figure B.8H shows how the mean pay of the Compustat sample compares to mean pay for all US workers. Employees in these Compustat firms earn slightly more than

---

Table B.12: Compustat Data Series

Database	Series ID	Description
ExecuComp	TDC1	Executive Total Compensation
Fundamentals Annual	XLR	Total Staff Expenses
Fundamentals Annual	EMP	Employees

Notes: Executive compensation series TDC1 = Salary + Bonus + Other Annual + Restricted StockGrants + LTIPPayouts + All Other + Value of Option Grants

the US population at large — a result that is consistent with the well-known firm size-wage gap [21].

Figure B.8I shows inter-firm income inequality — the Gini index of firm mean pay in our Compustat sample. Inter-firm income inequality in this sample tended to increase over the time period in question.
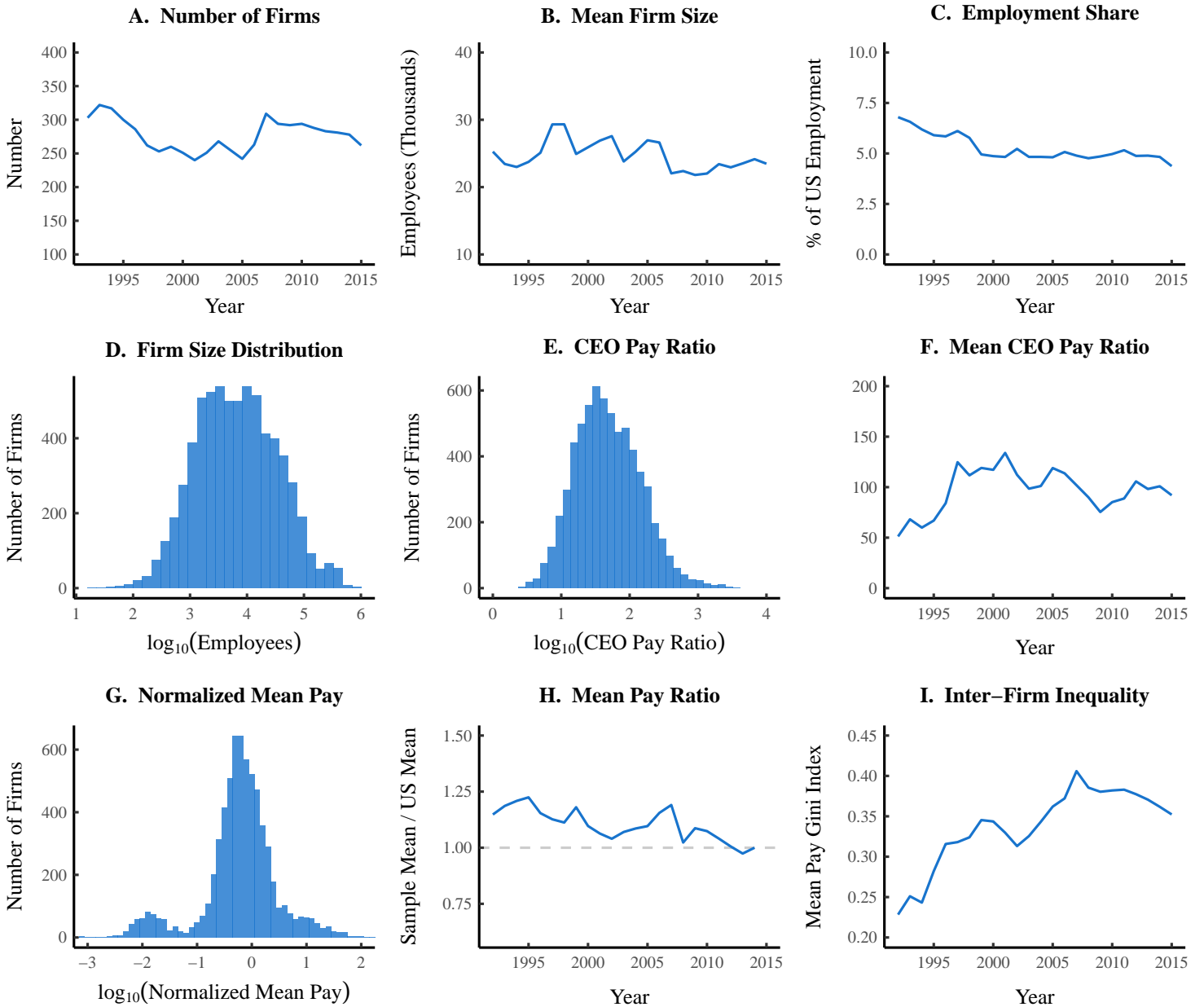
Figure B.8: Selected Statistics from the Compustat Firm Sample

This figure shows statistics for the Compustat firm sample, which consists of US firms for which the data series in Table B.12 are available. In panel H, US mean income per worker is calculated from national accounts (BEA Table 1.12, National Income by Type of Income) by dividing the sum of employee and proprietor income by the number of workers (BEA Table 6.8C-D, persons engaged in production).

Table B.13: Compustat Model Parameters

Parameter	Definition	Estimation Method
a, b	Span of control parameters	Estimated from case study data (Fig. B.5A). *Values are fixed for all modeled firms.*
σ	Intra-hierarchical level pay dispersion parameter	Estimated from case study data (Fig. B.5C). *Value is fixed for all modeled firms.*
E_1	Employment in base hierarchical level	Estimated numerically, given a, b, and Compustat firm employment E_T.
r	Pay-scaling parameter	Estimated from Compustat CEO pay ratio, given a, b, and E_1.
$\bar{I}_h$	Mean pay in base hierarchical level	Estimated from Compustat firm mean pay $\bar{I}_T$, given a, b, E_1, and r.

B.5 Estimating Compustat Model Parameters

We now apply the algorithm developed in Appendix B.3 to model intra-firm income distribution within the Compustat firm sample (Sec. B.4). The Compustat model is characterized by the 6 parameters shown in Table B.13. In order to model the internal income distribution of Compustat firms, we need to estimate these 6 parameters for each firm. My methods are summarized in Table B.13 and discussed in detail in the following sections.

B.5.1 Parameters Derived from Case-Study Data

The parameters a, b, and σ are estimated from the firm case-study data shown in Figure B.5 and assumed to be fixed for all Compustat firms. The parameters a and b, which together determine how the span of control changes with hierarchical level, come from an exponential regression on data in Fig. B.5A. The parameter σ determines the amount of pay dispersion within each hierarchical level of a firm (all levels are assumed to have the same amount of dispersion). This dispersion is modeled with a lognormal distribution, and σ is the 'scale' parameter (see Eq. B.34).

We estimate σ from the case-study data shown in B.5C. Note that this data uses the Gini index as the metric for dispersion. To estimate σ, we first calculate
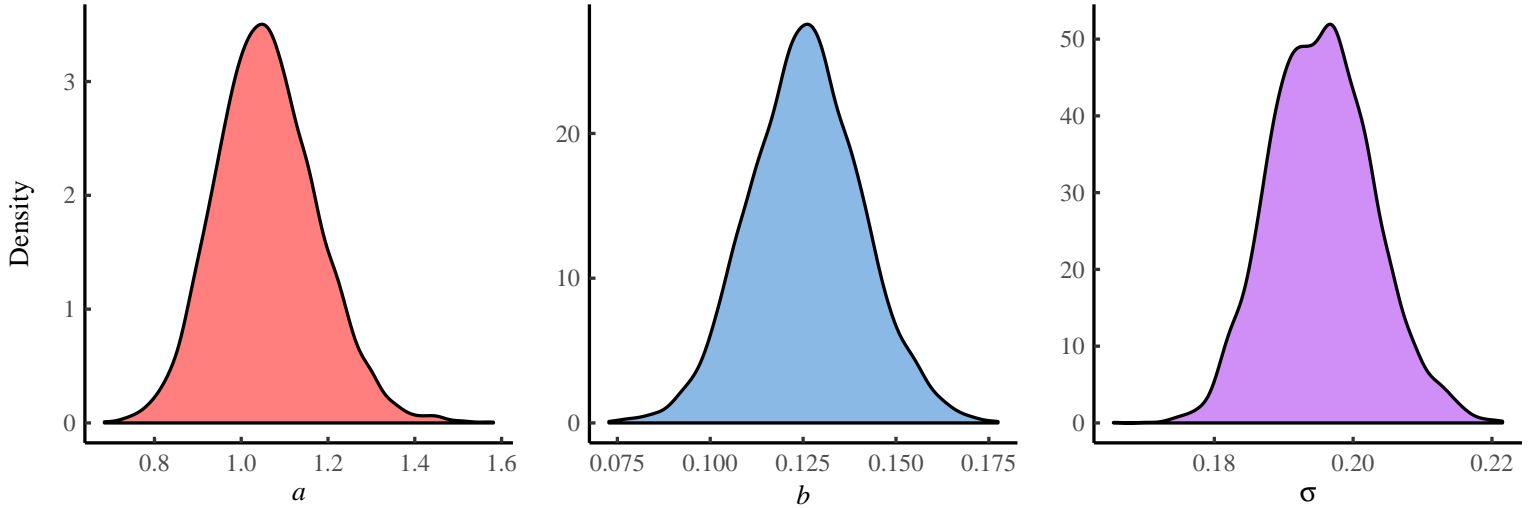
Figure B.9: Density Estimates for Case-Study Derived Parameters

This figure shows density estimates for the parameters a, b, and σ. Parameters a and b together determine the 'shape' of the firm hierarchy. The parameter σ determines the amount of income dispersion within each hierarchical level. Each parameter is determined from regressions on firm case-study data (Fig. B.5). The density functions are estimated using a bootstrap analysis, which involves resampling (with replacement) the case study data many times, and calculating the parameters a, b, and σ for each resample.

the mean Gini index of all data ($\bar{G}$). We then use Eq. B.41 to calculate the value σ, which corresponds to the lognormal scale parameter that would produce a lognormal distribution with an equivalent Gini index. This equation is derived from the definition of the Gini index of a lognormal distribution — $G = \mathrm{erf}(\sigma/2)$.

$$\sigma = 2 \cdot \mathrm{erf}^{-1}(\bar{G}) \tag{B.41}$$

Once a, b, and σ are estimated from case-study data, the model proceeds on the assumption that these parameters are *constant* across all Compustat firms. While real-world Compustat firms likely have parameters that vary widely, the hope is that our case-study regression provides a reasonable midpoint estimate.

Regressions on the case-study data provide a single best-fit estimate of the values of a, b, and σ. However, because the case-study sample size is small, there is considerable uncertainty in these values. This uncertainty can be quantified using the *bootstrap* method [5], which involves repeatedly resampling the data (with replacement) and then estimating the parameters a, b, and σ from
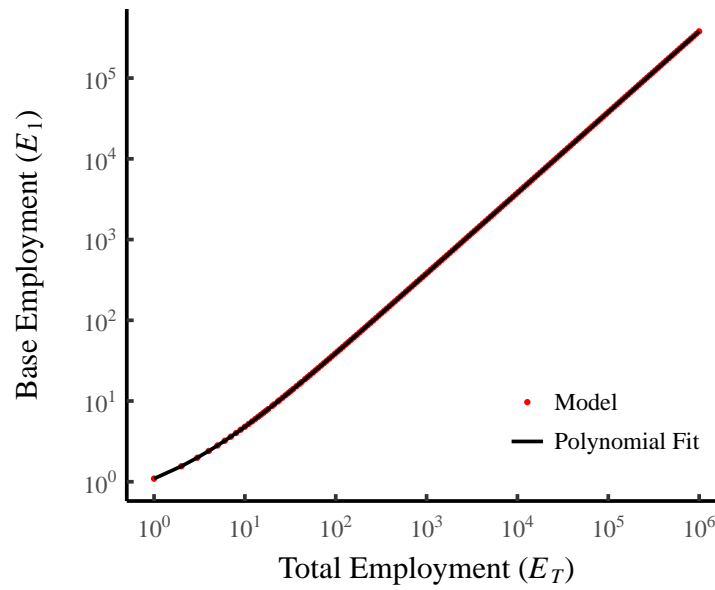
Figure B.10: Finding Base Level Employment From Total Employment

This figure shows the modeled relation between total employment in a firm (E_T) and base-level employment (E_1). This relation, defined by Eq. B.21, B.24, and B.26, depends on the span of control parameters a and b (here $a = 1.05$ and $b = 0.13$). I fit this numerical mapping with a high-order polynomial to allow fast (but accurate) estimation of E_1 from E_T. The R code for this procedure is available in the Supplementary Material.

---

this resampled data. Figure B.9 shows the probability density distributions resulting from this bootstrap analysis.

To incorporate this uncertainty into the model, I run the model many times — once for each bootstrapped estimate of a, b, and σ. In each iteration, we first resample the case-study data and calculate values of a, b, and σ. We then use these values (particularly a and b) to calculate all other model parameters. The results shown in this paper are based on 5000 bootstrap runs of the model.

B.5.2   Base Level Employment

Having estimated the span of control parameters a and b, the next step is to calculate base-level employment E_1 for each Compustat firm. We do this by using data for total employment E_T.

The modeled-relation between total employment E_T and base-level employment E_1 is determined by equations B.21, B.24, and B.26 (see Appendix B.3).

Given values for a and b, these equations produce a unique relation between E_1 and E_T.

$$E_T = f_{a,b}(E_1) \tag{B.42}$$

What we want is an inverse function that gives E_1 from E_T:

$$E_1 = f_{a,b}^{-1}(E_T) \tag{B.43}$$

Although there may be a way to define this inverse function analytically, it is beyond my mathematical abilities. Instead, I use the model to reverse engineer the problem. I define $f_{a,b}$ numerically by inputting a range of different values for E_1 into equations B.21, B.24, and B.26 and calculating E_T for each value. The result is a discrete mapping relating base-level employment to total employment (see Fig. B.10). I then fit this mapping with a high-order polynomial, which then serves as an approximation to the inverse function f^{-1}. This polynomial can then be used to quickly and accurately calculate E_1 from E_T for every Compustat firm.

B.5.3   Pay-Scaling Parameter

Once we have calculated base-level employment (E_1) for all Compustat firms, we can estimate their respective pay-scaling ratios (r) using the CEO-to-average-employee pay ratio (C). The pay-scaling ratio r determines the rate at which mean pay increases by hierarchical level.

Having estimated a, b, and E_1 for each Compustat firm, the model (specifically equations B.24, B.29, B.31, B.32, and B.33) produces a CEO pay ratio (C_{model}) that is a unique function of the pay-scaling parameter r:

$$C_{\text{model}} = f_{a,b,E_1}(r) \tag{B.44}$$

As with base-employment, I am not aware of an analytical method for defining the inverse function f_{a,b,E_1}^{-1}. Instead I use a numerical optimization method to solve for r. I define an error function $\epsilon(r)$ that quantifies the error between the actual value of a firm's CEO pay ratio ($C_{empirical}$) and the value predicted by the model (C_{model}) for a given value of r:

$$\epsilon(r) = \left| C_{\text{model}} - C_{\text{empirical}} \right| \tag{B.45}$$

For each firm, the correct value of r is that which *minimizes* this error function. I use the R non-linear optimization function 'nlminb' to solve this mini-

A. Fitted Pay–Scale Parameters   **B. Pay–Scaling Parameter Distribution**
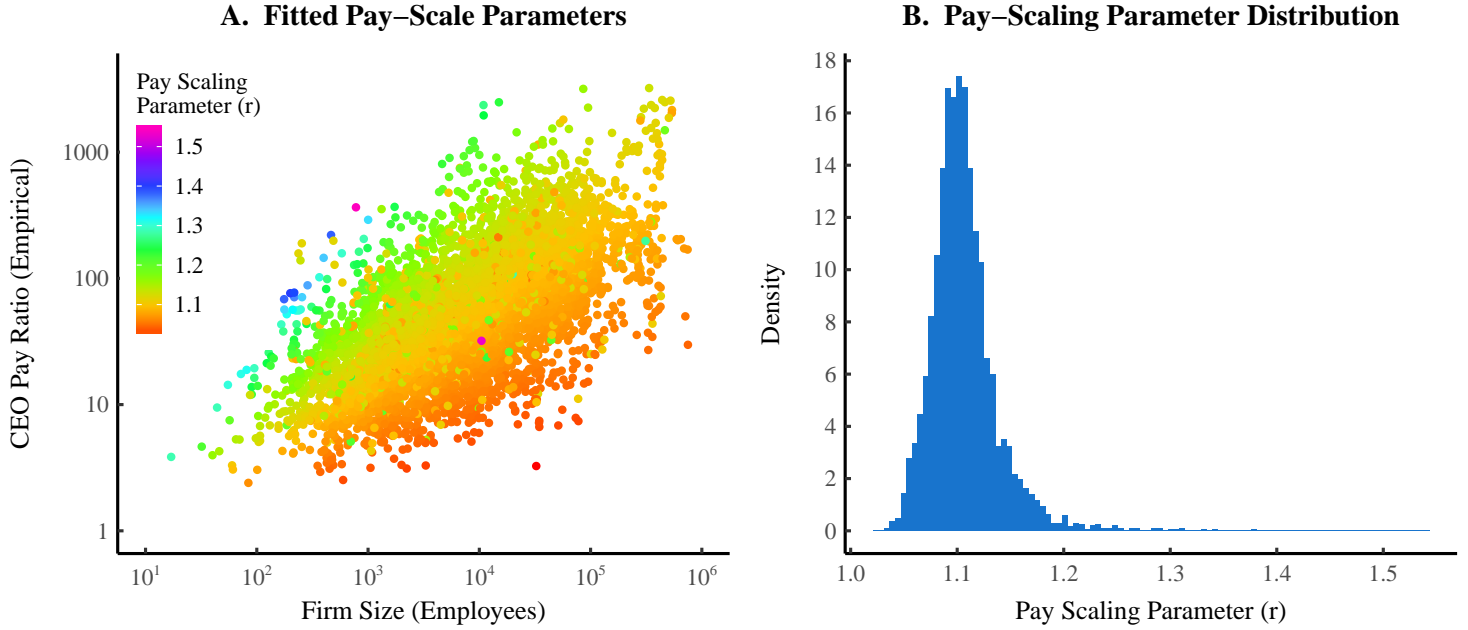
Figure B.11: Fitting Compustat Firms with a Pay-Scaling Parameter

This figure shows the fitted pay-scaling parameters (r) for all Compustat firms. Panel A shows the relation between the CEO pay ratio and firm size, with the fitted pay-scaling parameter indicated by color. The pay-scaling parameter distribution for all firms (and years) is shown in panel B. These results show the average of 5000 model runs, each with different bootstrapped parameters a, b, and σ.

---

mization problem. To ensure that there are no large errors, I discard Compustat firms for which the best-fit r parameter produces an error that is larger than 5% of $C_{\text{empirical}}$. Fitted results for r are shown in Figure B.11.

B.5.4 Base-Level Pay

Once we have the pay-scaling parameter r, we can estimate base-level pay for each Compustat firm. To do this, we set up a ratio between base level pay ($\bar{I}_1$) and firm mean pay ($\bar{I}_T$) for both the model and Compustat data:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{\bar{I}_1^{\text{model}}}{\bar{I}_T^{\text{model}}} \tag{B.46}$$

The modeled ratio between base pay and firm mean pay ($\bar{I}_1^{\text{model}}/\bar{I}_T^{\text{model}}$) is *independent* of the choice of base pay. This is because the modeled firm mean

pay is actually a *function* of base pay (see Eq. B.31 and B.32). If we run the model with $\bar{I}_1^{\text{model}} = 1$, then Eq. B.46 reduces to:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{1}{\bar{I}_T^{\text{model}}} \tag{B.47}$$

We can then rearrange Eq. B.47 to solve for an estimated base pay for each Compustat firm ($\bar{I}_1^{\text{Compustat}}$):

$$\bar{I}_1^{\text{Compustat}} = \frac{\bar{I}_T^{\text{Compustat}}}{\bar{I}_T^{\text{model}}} \tag{B.48}$$

B.6 Compustat Model Results

I review here the results of the Compustat model that are not discussed in the main paper. All results are generated using 5000 bootstrap model runs over different values for the parameters a, b, and σ. From the data generated by the model, many different calculations are possible. I review here the following: (1) estimates for income inequality within Compustat firms; (2) estimates for income by hierarchical level; and (3) aggregate inequality of all firms in the model.

B.6.1 Inequality Within Compustat Firms

Figure B.12 shows estimate of income inequality within Compustat firms. In Figure B.12A, I illustrate how firm Gini indexes are related to both the CEO Pay ratio and firm size. Note that the CEO pay ratio is a reliable indicator of firm inequality only for firms of the *same size*. A general feature of a hierarchical firm model is that when internal inequality is held constant, the CEO pay ratio nonetheless tends to increase with firm size (a feature first demonstrated by Herbert Simon [22]). In Figure B.12A, this feature is evident as color contours of constant firm inequality that scale with both firm size and the CEO pay ratio.

Figure B.12B shows the overall distribution of all firm Gini indexes. According to our model, 90% of Compustat firms have internal Gini indexes between 0.2 and 0.5. Note that the distribution is right-skewed — a small minority of firms have extremely unequal pay.

In Figure B.12C I compare firm inequality in the Compustat model to inequality within the case-study firms discussed in Appendix B.2. The results indicate that Compustat firms are slightly more unequal than the case study firms. However, because the case-study sample size is small, this difference is not statistically significant. A Kolmogorov-Smirnov test gives a p-value of 0.20, indicating that there is a reasonable (20%) probability that the two firm samples (model and case study) come from the *same* distribution. Thus, under the conventional 5% significance level, we cannot reject the null-hypothesis that these samples come from the same distribution.

But if the case study data and Compustat model produce firm internal Gini distributions that are statistically indistinguishable, why not simply use case study data for the test of hypothesis B (hierarchical power has the strongest effect on income)? There are several reasons the case study data cannot be used. Firstly, the case study sample size is extremely small. Secondly, the firms
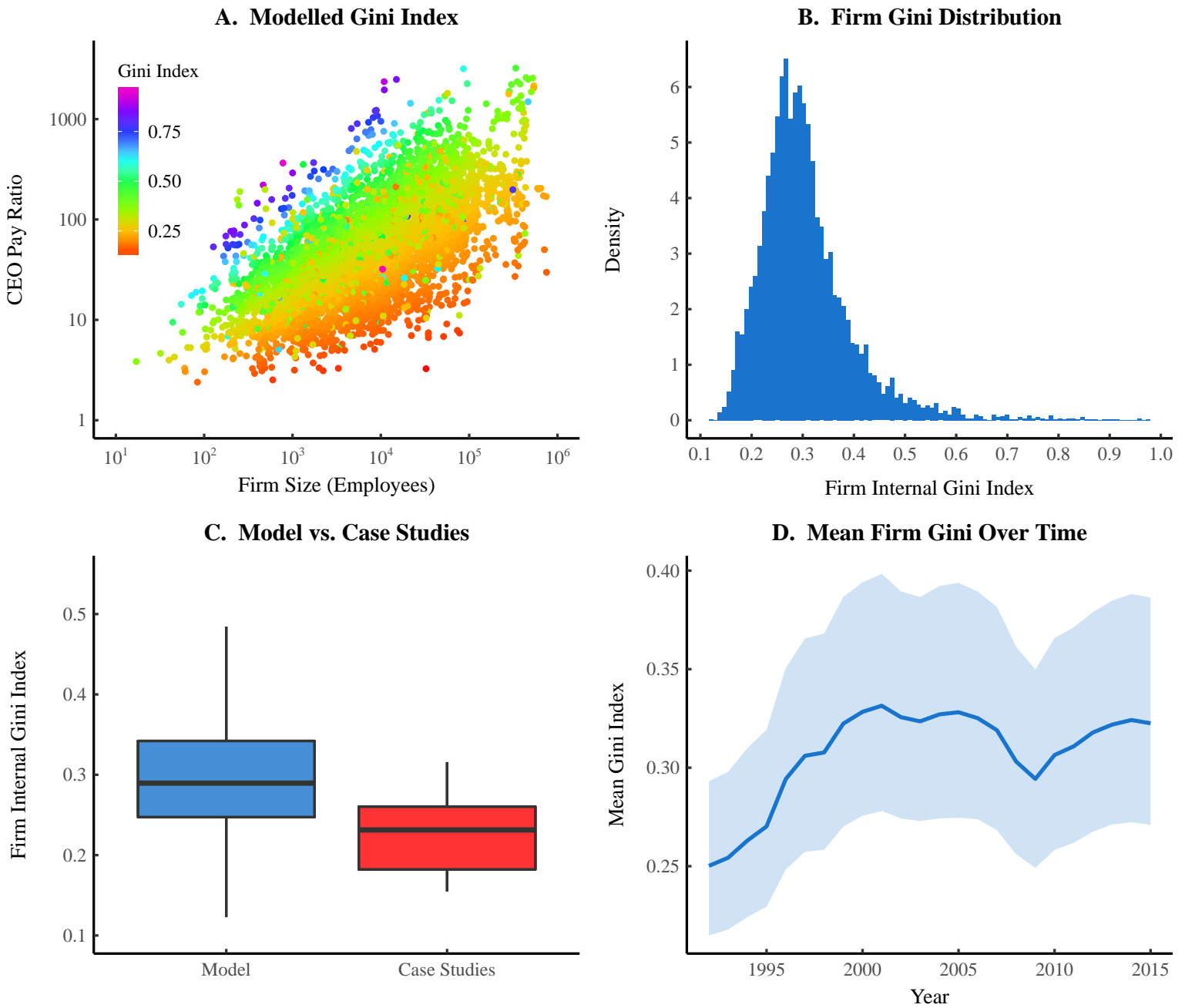
Figure B.12: Compustat Model Results for Intra-Firm Inequality

This figure shows the firm internal Gini index results of the Compustat model. Panel A shows how firm internal inequality (indicated by color) is related to the CEO pay ratio and firm size. Panel B shows the distribution of modeled Gini indexes for all firms. Panel C compares model results to the Gini index of case study firms (see section B.2.1 for case study methods). Panel D shows time evolution of the average Gini index of all modeled firms. The shaded region indicates the 95% confidence interval. All results are computed from 5000 model runs, each with different bootstrapped parameters a, b, and σ.

A. 50 Most Unequal Firms

Firm	
MICREL INC	
GRAND CASINOS INC	
CONSOL ENERGY INC	
SCORE BOARD INC	
APPROACH RESOURCES INC	
CUSTOMERS BANCORP INC	
POLYMEDICA CORP	
BLOCKBUSTER ENMNT CORP	
HCI GROUP INC	
FINANCIAL FEDERAL CORP	
GBC BANCORP/CA	
INTEGRATED HEALTH SVCS INC	
MONACO COACH CORP	
PROVIDIAN FINANCIAL CORP	
HUDSON CITY BANCORP INC	
TALMER BANCORP INC	
PHARMERICA CORP	
ORITANI FINANCIAL CORP	
BOFI HOLDING INC	
FOOTHILL GROUP INC –CL A	
APOLLO EDUCATION GROUP INC	
NORTH FORK BANCORPORATION	
TORCHMARK CORP	
ROSLYN BANCORP INC	
REWARDS NETWORKS INC	
TPI ENTERPRISES INC	
ONBANCORP INC	
COLONIAL PROPERTIES TRUST	
CHIPOTLE MEXICAN GRILL INC	
FIRST USA INC	
NYFIX INC	
WATERHOUSE INVESTORS SVCS	
LEUCADIA NATIONAL CORP	
EAST WEST BANCORP INC	
OAK INDUSTRIES INC	
DELPHI FINANCIAL GROUP INC	
BANK OF THE OZARKS INC	
CEC ENTERTAINMENT INC	
TRUSTCO BANK CORP/NY	
NATIONAL DISC BROKERS INC	
JACK IN THE BOX INC	
PHH CORP	
HCC INSURANCE HOLDINGS INC	
FUTUREFUEL CORP	
NUVEEN INVESTMENTS INC	
CATHAY GENERAL BANCORP	
NATIONAL COMMERCE FINANCIAL	
STONE ENERGY CORP	
FIRST NBC BANK HOLDING CO	
CNO FINANCIAL GROUP INC	

Gini Index Estimate

B. 50 Most Equal Firms

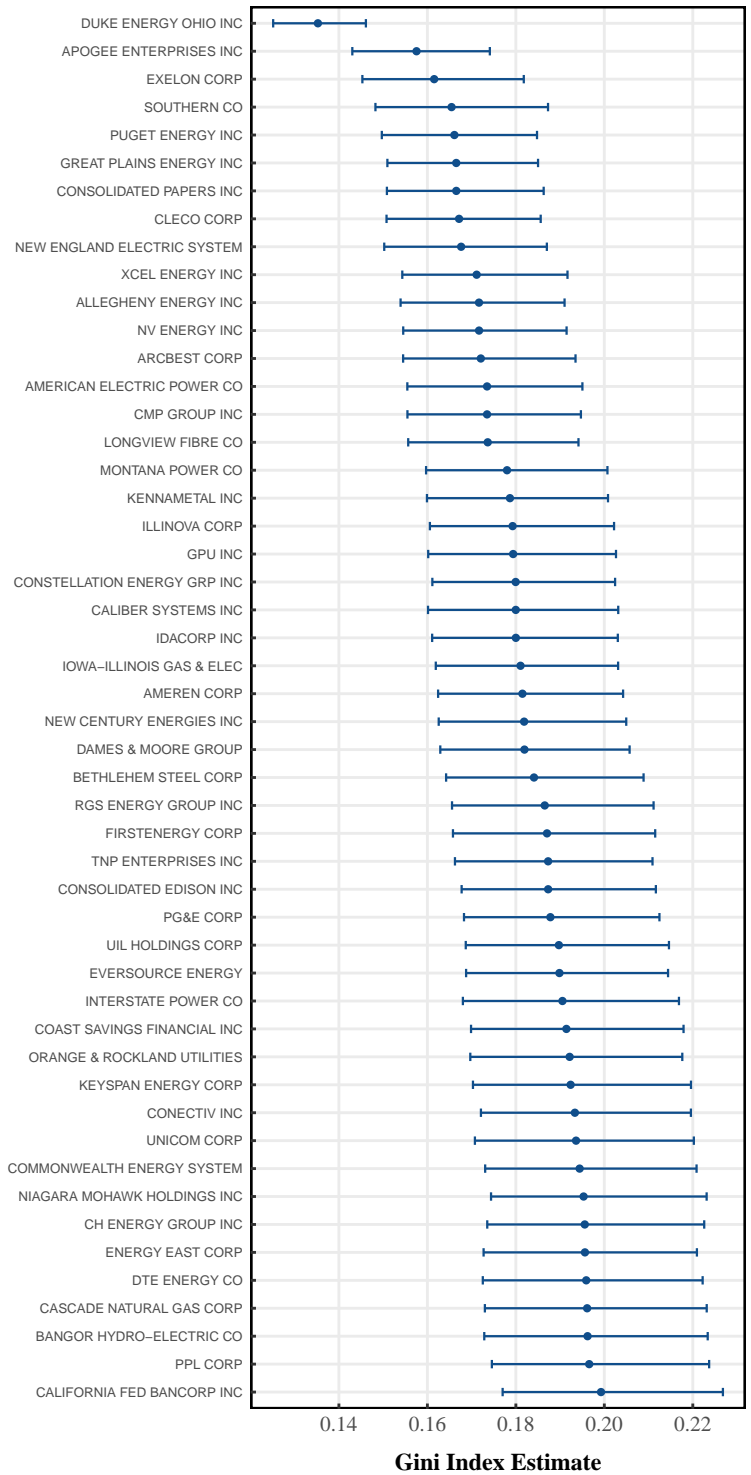| Firm	
DUKE ENERGY OHIO INC	
APOGEE ENTERPRISES INC	
EXELON CORP	
SOUTHERN CO	
PUGET ENERGY INC	
GREAT PLAINS ENERGY INC	
CONSOLIDATED PAPERS INC	
CLECO CORP	
NEW ENGLAND ELECTRIC SYSTEM	
XCEL ENERGY INC	
ALLEGHENY ENERGY INC	
NV ENERGY INC	
ARCBEST CORP	
AMERICAN ELECTRIC POWER CO	
CMP GROUP INC	
LONGVIEW FIBRE CO	
MONTANA POWER CO	
KENNAMETAL INC	
ILLINOVA CORP	
GPU INC	
CONSTELLATION ENERGY GRP INC	
CALIBER SYSTEMS INC	
IDACORP INC	
IOWA–ILLINOIS GAS & ELEC	
AMEREN CORP	
NEW CENTURY ENERGIES INC	
DAMES & MOORE GROUP	
BETHLEHEM STEEL CORP	
RGS ENERGY GROUP INC	
FIRSTENERGY CORP	
TNP ENTERPRISES INC	
CONSOLIDATED EDISON INC	
PG&E CORP	
UIL HOLDINGS CORP	
EVERSOURCE ENERGY	
INTERSTATE POWER CO	
COAST SAVINGS FINANCIAL INC	
ORANGE & ROCKLAND UTILITIES	
KEYSPAN ENERGY CORP	
CONECTIV INC	
UNICOM CORP	
COMMONWEALTH ENERGY SYSTEM	
NIAGARA MOHAWK HOLDINGS INC	
CH ENERGY GROUP INC	
ENERGY EAST CORP	
DTE ENERGY CO	
CASCADE NATURAL GAS CORP	
BANGOR HYDRO–ELECTRIC CO	
PPL CORP	
CALIFORNIA FED BANCORP INC	

Gini Index Estimate

Figure B.13: The Most Equal and Unequal Compustat Firms

This figure shows the 50 most unequal (panel A) and 50 most equal firms (panel B). Points indicate the mean Gini index for each firm, while the error bars show the 95% confidence interval calculated from 5000 bootstrap model runs.

cover many different countries (not just the US, the desired country). Thirdly, the observation years often do not overlap. To test hypothesis B, we need a *large* firm sample from a *single* country in a *single* year. While model dependent, the results inferred from Compustat data satisfy these conditions, while case study data does not.

Figure B.12D shows the time-evolution of average inequality within Compustat firms. During the late 1990s inequality rapidly increased, followed by relative stability from 2000 onward. While the *trend* is clear, there is significant uncertainty in the *absolute* level of inequality (as indicated by the shaded region). This uncertainty is due to the small case-study sample size on which key model parameters are based (see Appendix B.5).

Finally, Figure B.13 shows Gini index estimates for the 50 most equal and 50 most unequal firms. What is most interesting about these results is the *sectoral composition* of the 50 most equal firms. The vast majority (80%) are energy/utility companies. In the United States, firms in the utility sector are highly regulated, which leads to far more scrutiny over executive pay. Previous studies have found similar results — executives in regulated firms earn far less than those in unregulated firms [23]. This finding has important implications for a power theory of income distribution. It suggests that government regulation serves as a *check* on power, limiting the degree to which elites are able to use their status to amass wealth.

B.6.2   Income By Hierarchical Level

Besides estimating firm internal inequality, I use the Compustat model to estimate income and inequality by hierarchical level. To do this, we group all individuals by their hierarchical level, regardless of firm membership (see Fig. 8 of main paper).

Model results are shown in Figure B.12, and are compared to the UK data documented by Mueller et al. [4]. Figure B.12A shows how mean income changes by hierarchical level. In both the Compustat model and Mueller's data, mean income increases super-exponentially with hierarchical level — that is, it increases *faster* than an exponential function, which would appear as a straight line on the log-linear scale. Figure B.12B shows how intra-level income inequality changes by hierarchical level. For hierarchical levels 1-10, both the Compustat model and Mueller's data show similar trends.

The similarities between the model and Mueller's data lend credence to the model. However, what explains the differences? One key factor is that the
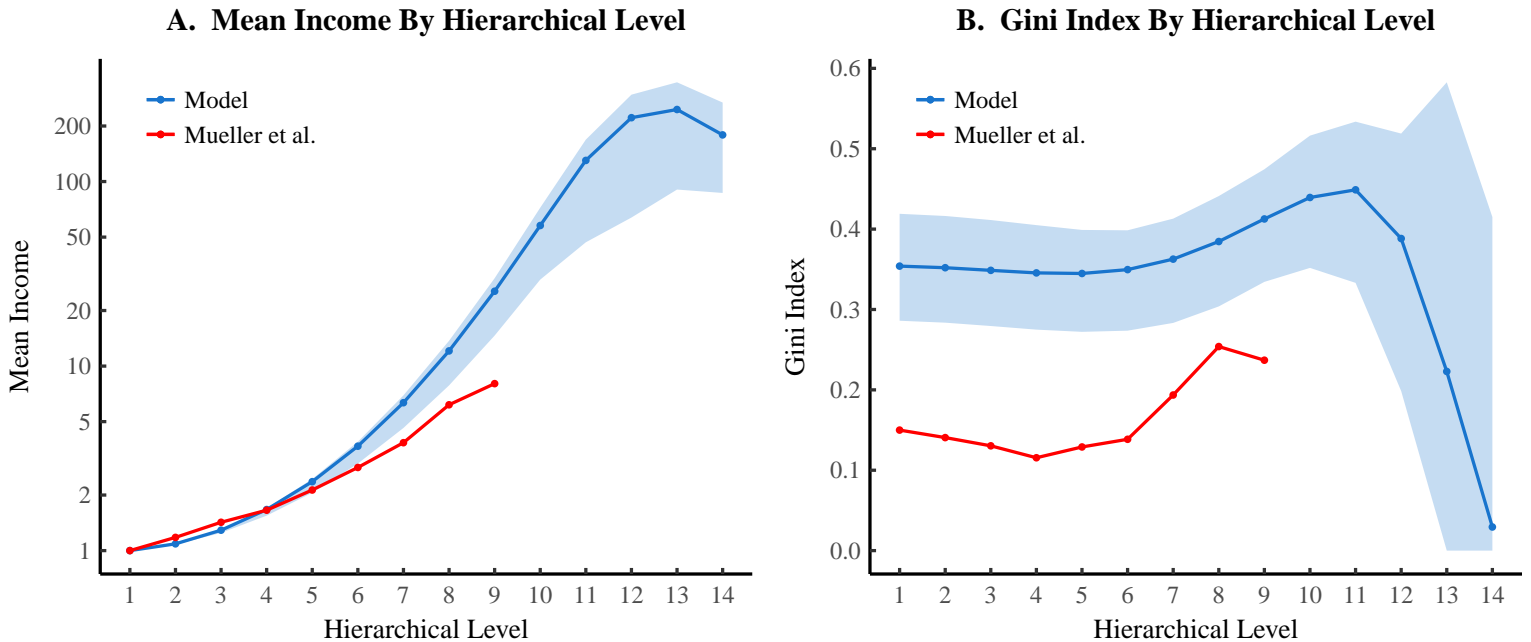
A. Mean Income By Hierarchical Level

B. Gini Index By Hierarchical Level

Figure B.14: Compustat Model Results for Income by Hierarchical Level

This figure compares the results of the Compustat model to the UK data from Mueller et al. [4]. Panel A shows average income by hierarchical level (across all firms) indexed to pay in level 1. Panel B shows how intra-level inequality changes by hierarchical level. Shaded regions indicate the 95% confidence region of the model, estimated from 5000 bootstrap runs (see Appendix B.5).

United States has much greater income inequality than the United Kingdom, and the Compustat firm sample comes from the former and Mueller's sample the latter. As it turns out, both the Compustat model and Mueller's data imply aggregate levels of inequality that are consistent with their respective national Gini indexes (see Fig. B.2 and B.15).

In this light, the results in Figure B.12A make sense — in the more unequal United States, income scales more rapidly with hierarchical level than in the United Kingdom. The results in Figure B.12B can be similarly explained — in the more unequal United States, intra-hierarchical level income dispersion is greater than in the UK.

Another interesting result in Figure B.12 is the conspicuous change in model trends for hierarchical levels above 11. Above this level, mean income no longer increases with hierarchical level, and intra-level inequality declines precipitously. The former result may simply be an artifact of the particular firm sample. Going back to Figure B.12A, note that the four largest firms have particularly low

CEO pay ratios. Given the model's assumptions, only the very largest firms will have more than 11 hierarchical levels. Since the 4 largest firms have particularly low CEO pay ratios, resulting mean income in hierarchical levels 12-14 will be relatively low.

The precipitous drop in intra-level inequality for hierarchical levels 12-14 is likely due to the convergence to a size of one. This is because there is often only one firm with 12 or more hierarchical levels, and the top level of this firm will contain only one individual. By definition, there is zero inequality in a sample size of one.

B.6.3   Aggregate Inequality

An important test of the Compustat model is to see if it produces aggregate levels of inequality that are comparable to US empirical data. Figure B.15 shows the results of such a test. Here I plot the time-series trends in both US historical inequality and aggregate inequality in the Compustat model. This latter metric is calculated by aggregating (by year) all individuals in the model into a single sample, and then calculating the inequality of the resulting income distribution.

Figure B.15A compares the model's aggregate Gini index against three different types of data published by the US Census: Gini by *individual*, *family*, and *household*. Two findings are evident. Firstly, the model is roughly consistent with the US empirical data over the period 2000-2015. However, the model produces too little inequality during the 1990s. Secondly, the US empirical data shows *contradictory* trends — roughly *constant* inequality among individuals, but secularly *increasing* inequality among families and households. The model reproduces the secular trend. But which empirical data should we believe? My vote is that the secular increase is the correct trend.

Largely in response to his dissatisfaction with official inequality statistics, Thomas Piketty [24] has focused on measuring inequality in the tail of the income distribution. Figure B.15B and C show Piketty's series for the top 10% and 1% income share in the United States. Both series show secularly increasing inequality over the period in question. The model reproduces these *trends* quite accurately, but at a lower absolute level of inequality.

How can it be that the model more or less matches US Gini index data, but gives much less inequality than Piketty's metrics? A plausible explanation is that official data simply *underestimates* inequality. However, the validity (or lack their of) of official inequality statistics is not something that this paper is concerned with. Rather, I simply take official data as a given, and use it to test my power-

income hypothesis. As such, the important take-home finding here is that the Compustat model produces a level of inequality that is consistent with official US data. This means that it is fair to compare the model's results to other results derived from official data.
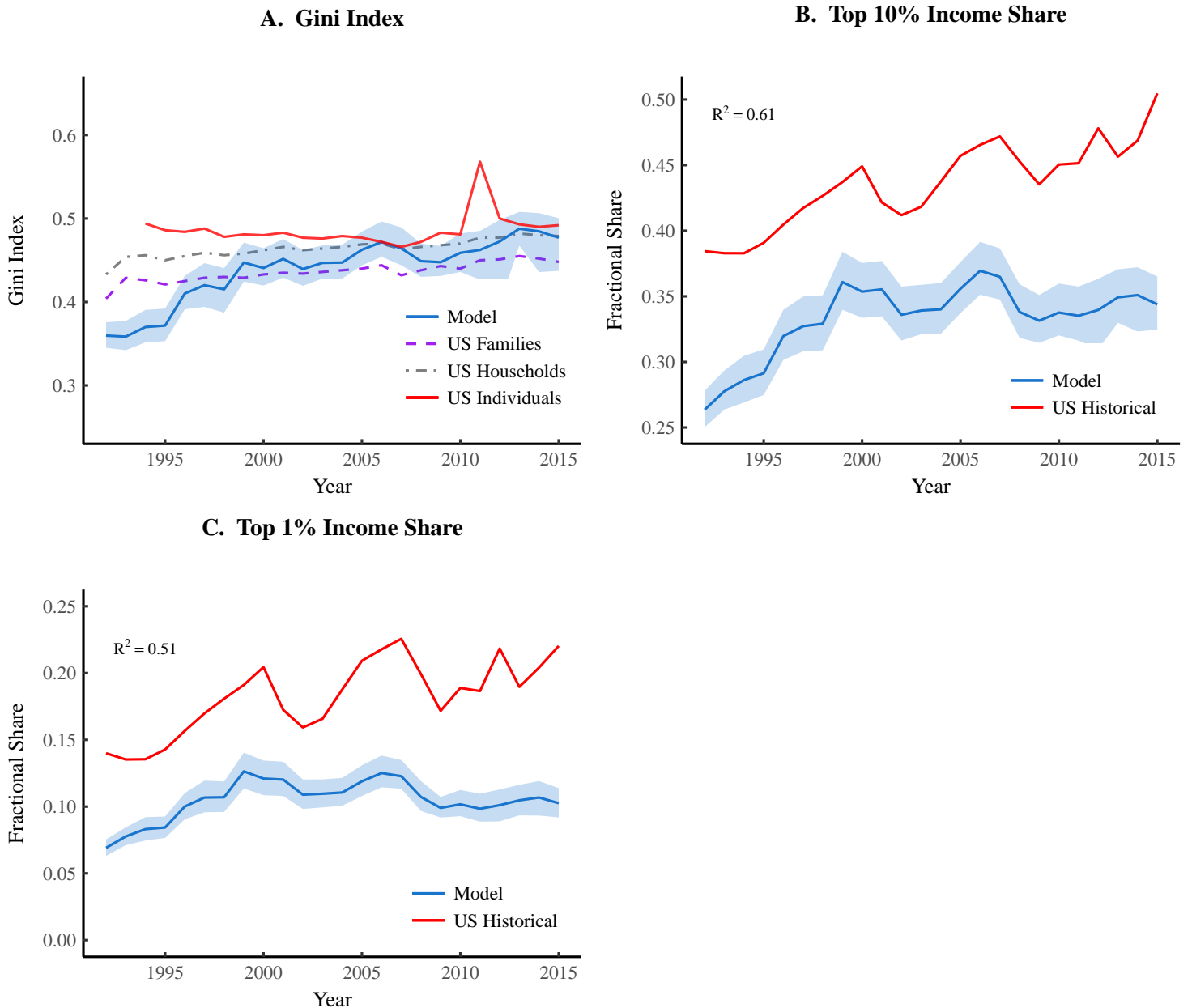
A. Gini Index

B. Top 10% Income Share

C. Top 1% Income Share

Figure B.15: Compustat Model Aggregate Inequality vs. US Historical Data

This figure compares estimates of aggregate income inequality in the Compustat model to US historical data. Panel A compares the model GIni index to three different US measures (the Gini of individuals, families and households). Panel B shows the income share of the top 10%, while panel C shows the top 1%. The shaded regions indicate the 95% confidence interval of the model, estimated over 5000 bootstrap runs. US Gini index data is for individuals, and comes from US Census table PINC-05. The 2011 outlier in US data is likely a statistical error. Families and Household Gini indexes are from the Federal Reserve Bank, series GINIALLRF and GINIALLRH, respectively. US top 10% and top 1% share data is from the World Wealth and Income Database, series *sptinc992j*.

B.7   A Sensitivity Analysis of the Compustat Model

The Compustat model relies on three parameters — a, b, and σ — that are determined from regressions on firm case study data (see Appendix B.5). Parameters a and b define the span of control, and ultimately determine the 'shape' of a firm's hierarchy. The parameter σ determines the level of income inequality within each hierarchical level of a firm.

Unfortunately, our case study analysis contains only seven firms, and we have no way of knowing if it is a representative sample on which to base our model. Given this ambiguity, it is important to understand the 'sensitivity' of our model results to changes in the parameters a, b, and σ.

Recall that the model results presented in the paper are based on 5000 bootstrap runs of the model — each run uses a different value of a, b, and σ generated by running regressions on resampled (with replacement) case-study data. I use this bootstrapped data to analyze how each parameter affects the following metrics.

1. The G_{BW} metric for hierarchical levels;
2. The G_{BW} metric for firms;
3. Aggregate levels of inequality within the entire model

Figure B.16 shows the results of this analysis. We can immediately conclude that the model is not sensitive to the value of σ, which has virtually no effect on any of the above metrics. However, the model appears to be highly sensitive to the parameters a and b. This sensitivity is least pronounced for hierarchical level results. But for firm G_{BW} and aggregate inequality, changes in a and b have a strong effect on model results.

This is an important finding. It suggests that our hierarchical level results (used to test the power-income hypothesis) are relatively robust. A different firm case-study sample would likely not lead to significant changes in our findings. Our firm results, however, are less robust. A different firm case-study sample could lead to very different results.
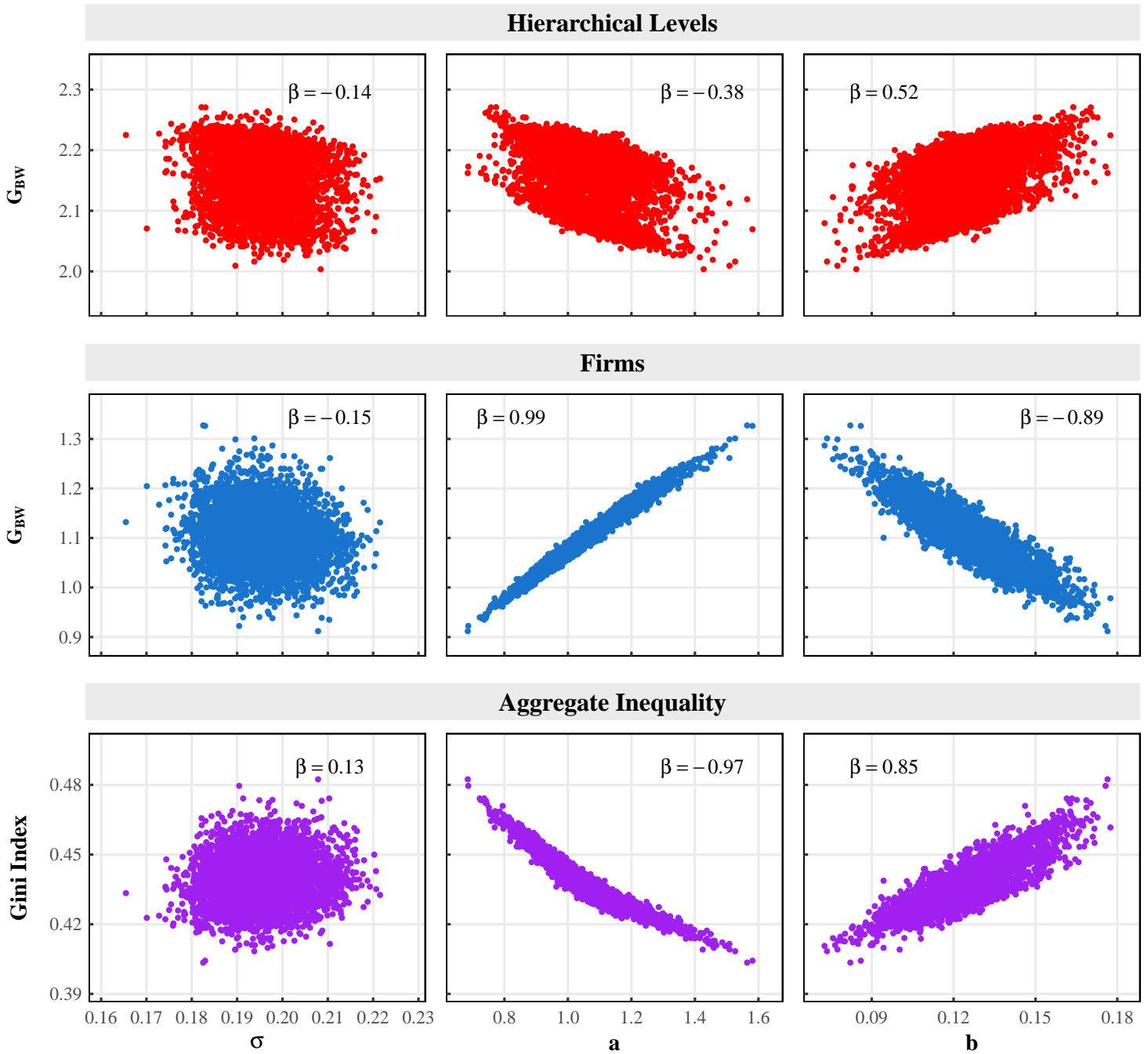
Figure B.16: A Sensitivity Analysis of Model Parameters

This figure shows the results of a sensitivity analysis of Compustat model parameters. The top row shows the effect that each parameter has on the G_{BW} metric (the between-group vs. within-group Gini index) for hierarchical levels. The middle row shows the same for firms, and the bottom row for aggregate inequality in the model. Each plotted point indicates a different parameter combination. The normalized regression coefficient β (which can range from -1 to 1) quantifies the model sensitivity.

B.8   The Between-Within Gini Metric and Effect Size

In this section, I discuss what is meant by 'effect size' (in the context of this paper) and how my between-within Gini metric relates to more standard measures of effect size.

B.8.1   What is Meant By 'Effect Size'

In the context of income distribution, there are two possible ways that we might define effect size:

Definition A: How much a factor effects *total inequality*.
Definition B: How much a factor effects *individual income*.

Effect size definition A refers to what we might call 'inequality accounting'. For instance, we might ask: how much do differences in pay between two groups contribute to total inequality? The point is that this definition attempts to measure how a given factor effects *total inequality*. In general, inequality accounting depends crucially on the *size* of the various groups.

Figure B.17 shows this phenomena. In both panels, groups A and B have equal differences in mean income and equal within-group income dispersion. However the total inequality obtained by merging the two groups varies dramatically depending on the relative size of A to B. In Fig. B.17A, the two groups are of equal size, while in Fig. B.17B, group B is 50 times smaller than group A. The resulting merger of A and B produces much more inequality when the two groups are equal size than when they are not.

Effect size definition B is concerned only with the effect on *individual income*, not on accounting for total inequality. The key difference is that for definition A, we care about group size, while for definition B we do not. In more technical terms, effect size definition B should be calculated by drawing *equal sized samples* from each group. Perhaps the simplest and most intuitive metric of effect size definition B is *Cohen's d* (Eq. B.49), defined as the difference in means ($\bar{x}$) between two group samples (A and B), divided by the within-group standard deviation (s_W).

$$d = \frac{\bar{x}_B - \bar{x}_A}{s_W} \tag{B.49}$$

Cohen's d can be interpreted as a *signal-to-noise ratio*. The 'signal' is the difference in means (the effect we want to measure), while the 'noise' is the
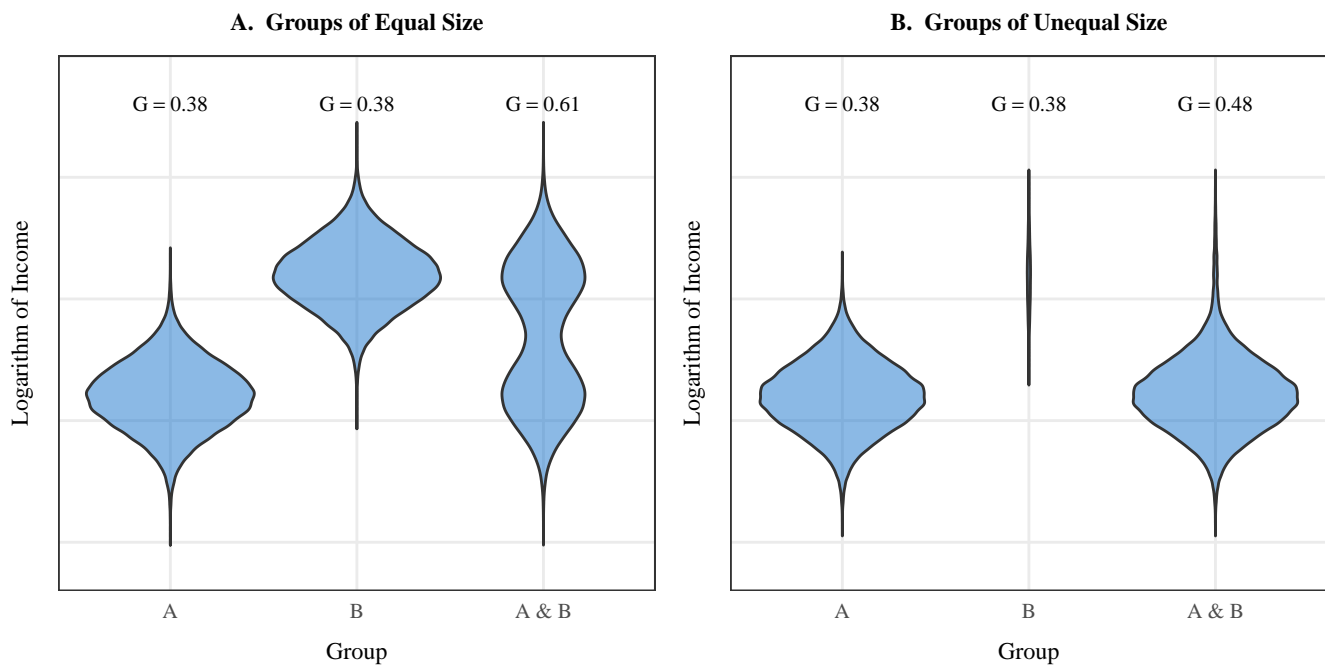
Figure B.17: How Group Size Affects Inequality Accounting

This figure shows how differences in group size effect total inequality. In both panels, the income distribution of two different groups (A and B) are shown. The distributions are displayed as 'violin' plots, where the thickness of the violin indicates the number of individuals with that income. In both panels, groups A and B have identical differences in mean income, and identical within-group income dispersion (Gini indexes are shown above each violin). In the left panel, both groups have the same size. In the right panel, group B is 50 times smaller than group A. The rightmost violin plot in each panel shows the income distribution produced by merging groups A and B. Far more inequality is produced when the two groups are of equal size than when there are large differences in size.

dispersion within groups, as measured by the standard deviation. In the case of income, the size of the signal-to-noise ratio indicates how accurately we can predict someone's income based only on knowledge of their group membership (either A or B). The larger the signal-to-noise ratio, the more accurate the prediction.

In the example shown in Figure B.17, group A and B have the same difference in means and the same within-group standard deviation in both the left and right panel. Therefore Cohen's d would measure an identical effect size. To be clear, this is the effect on *individual income* (definition B), not the effect on inequality (definition A).

B.8.2   Measuring Effect Size

In this paper, I am concerned only with effect size definition B — the effect on individual income. I have proposed the between-within Gini metric (G_{BW}) as a measure of this type of effect size. This metric is defined by equation B.50, where G_B is the Gini index of group means and $\bar{G}_W$ is the mean of all within-group Gini indexes:

$$G_{BW} = \frac{G_B}{\bar{G}_W} \tag{B.50}$$

How does this metric relate to more standard measures of effect size? It amounts to a signal-to-noise ratio that is similar to Cohen's f^2 measure, the latter of which is a generalization of Cohen's d to many different groups. Cohen's d uses the *difference* between means in the numerator. In order to generalize to many groups, f^2 uses the *sum of squared differences* (SS). To obtain f^2 (Eq. B.51), we divide the sum of squares between-groups (SS_B) by the sum of squares within groups (SS_W). See Fleishman [25] and Steiger [26] for a more detailed discussion of the f^2 metric.[1]

$$f^2 = \frac{SS_B}{SS_W} \tag{B.51}$$

To be clear, SS_B is the sum of squared differences between each group mean ($\bar{x}_i$) and the grand mean ($\bar{x}_{GM}$), multiplied by group size n. Similarly, SS_W is the sum of squared differences between each observation (x_{ij}) and its group mean ($\bar{x}_i$). This double sum operates over each of the k groups and n observations

within each group. Lastly, i indexes groups, and j indexes observations within each group.

$$SS_B = n \sum_{i=1}^{k} (\bar{x}_i - \bar{x}_{GM})^2 \tag{B.52}$$

$$SS_W = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \tag{B.53}$$

Like Cohen's d, the f^2 metric is a signal-to-noise ratio. The 'signal' is the sum of squares between groups, while the 'noise' is the sum of squares within groups. When applied to income, the size of f^2 indicates the accuracy with which we can predict individual income from group membership.

Comparing the form of f^2 and G_{BW}, we see that the two measures of effect size are very similar. Both are signal-to-noise ratios, consisting of a ratio of between-group dispersion to within-group dispersion. The difference is that f^2 uses the sum of squares to measure dispersion, while G_{BW} uses the Gini index. Given the similarity between G_{BW} and f^2, there should be some relation between the two measures.

Rather than attempt to show this similarity analytically, I use simulated data. I build a model based on the following assumptions:

1. Income within groups is lognormally distributed.
2. Within-group income dispersion is the same for all groups (but can vary over different model iterations).
3. Mean income between groups is lognormally distributed (and can vary between iterations).
4. Total inequality is (roughly) constant for all iterations.
5. The size of each group is constant.
6. The number of groups varies (between iterations) from 2 to 100.

For each iteration of the model, we define the mean income of each group by drawing randomly from a lognormal distribution. We then simulate individuals within each group by drawing randomly from (a different) lognormal distribution. The model has 2 key parameters: the lognormal scale parameter that defines the dispersion *between* groups, and the lognormal scale parameter that

---

[1]A more common formula for this metric is $f^2 = \eta^2/(1 - \eta^2)$, where $\eta = SS_B/SS_T$, the sum of squares between groups divided by the total sum of squares. Since $SS_T = SS_B + SS_W$, simple algebraic substitution can prove that the two definitions of f^2 are equivalent.

determines dispersion *within* groups. Varying these parameters changes the size of the group-income effect. For consistency, I use only parameter combinations that produce roughly the same level of total inequality (a Gini index of 0.5).

For each set of simulated data, I calculate both G_{BW} and f^2. Because analysis of variance typically assumes that within-group data is normally distributed, I calculate f^2 using the *logarithm* of income. The results are shown in Figure B.18. As expected, there is an extremely strong relation between the two effect-size measures. This indicates that an f^2 test of the power-income effect would likely give very similar results to the G_{BW} findings shown in Figure 10 of the main paper. To reiterate, I do not conduct such an f^2 test in this paper because the relevant data is not available.

B.8.3   A Note on Group Size

The simulation shown in Figure B.18 shows a special case where all groups have the same size. However, for the vast majority of the income-affecting factors studied in this paper, the groups do *not* have the same size. For instance, there are *vastly* more people in lower hierarchical levels than in upper hierarchical levels. How do we deal with this situation?

The key ingredient of effect-size definition B is that it weights each different group *equally* (rather than weighting a group by its size). One way to achieve this equal weighting is to draw *equal-sized samples* from each group and calculate Cohen's f^2 using equations B.51-B.53. But what if we do not have raw data? What if we have only summary statistics such as the mean and standard deviation of each group?

We can proceed by noting that an alternative way to define Cohen's f^2 is as the ratio of between-group variance σ_B^2 and mean within-group variance $\bar{\sigma}_W^2$:

$$f^2 = \frac{\sigma_B^2}{\bar{\sigma}_W^2} \tag{B.54}$$

$$\sigma_B^2 = \sum_{i=1}^{k} \frac{(\bar{x}_i - \bar{x}_{GM})^2}{k} \tag{B.55}$$

$$\bar{\sigma}_W^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} \frac{(x_{ij} - \bar{x}_i)^2}{nk} = \sum_{i=1}^{k} \frac{\sigma_i^2}{k} \tag{B.56}$$

Here, σ_B^2 and $\bar{\sigma}_W^2$ are derived by dividing SS_B and SS_W (respectively) by nk.
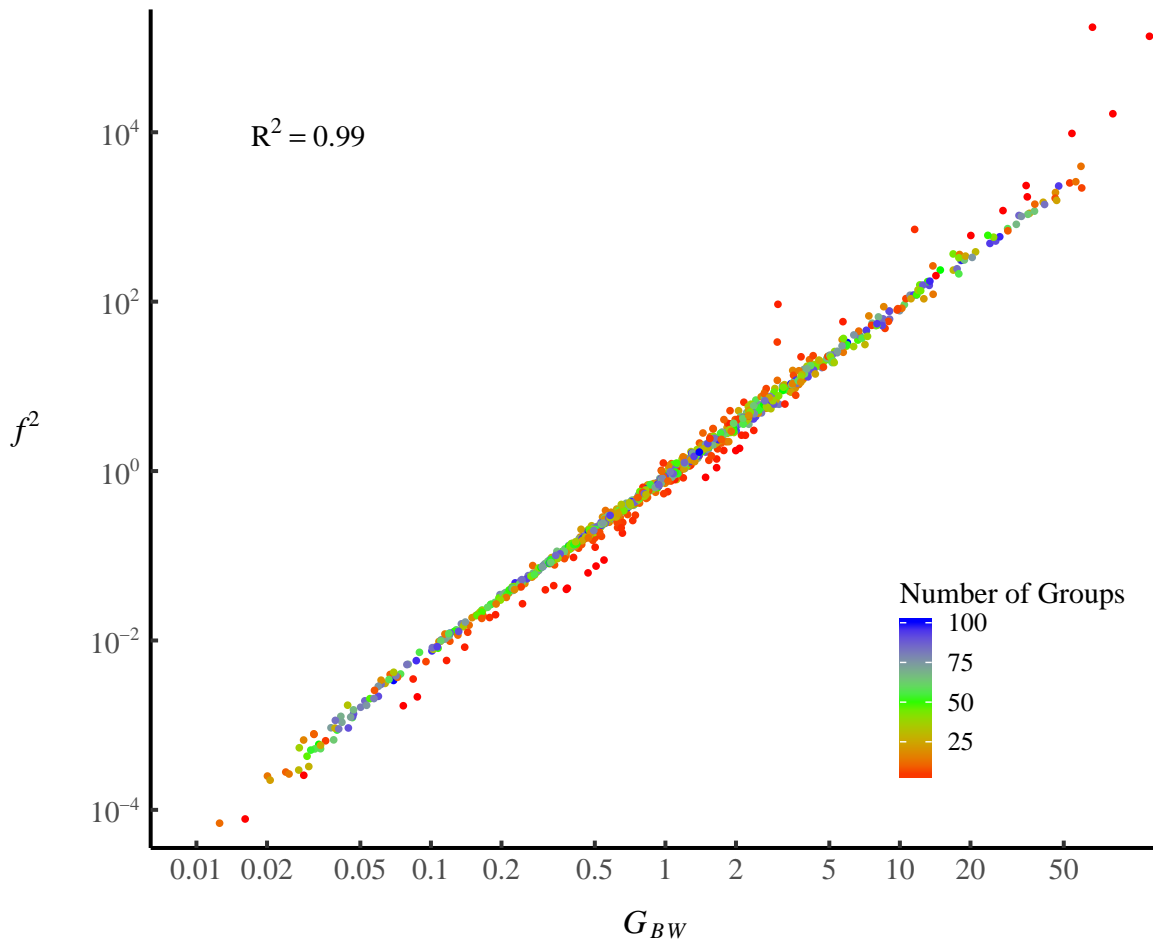
Figure B.18: Standard Effect Size Measure f^2 vs. the Gini Metric G_{BW}

This figure compares Cohen's f^2 metric of effect size (Eq. B.51) to my between-within Gini metric, G_{BW} (Eq. B.50). The comparison uses simulated data, and each data point represents different parameter combinations (see model assumptions above). Color indicates the number of groups used in each iteration. R code for the model is available in the Supplementary Material.

Given group means ($\bar{x}_i$) and within-group standard deviations(σ_i), we can use this alternative formula to calculate f^2.

What equation B.54 does is give *identical weight* to each group's summary statistics. This accomplishes the same thing as if we took equal sized samples from raw data and used Eq. B.51-B.53 to calculate f^2. This same logic applies to my construction of the G_{BW} metric: it calculates a signal-to-noise ratio from summary statistics by giving equal weight to each group.

References

1. Bowles S, Gintis H, Osborne M. The determinants of earnings: A behavioral approach. Journal of economic literature. 2001;39(4):1137–1176.

2. Chetty R, Hendren N, Kline P, Saez E. Where is the land of opportunity? The geography of intergenerational mobility in the United States. The Quarterly Journal of Economics. 2014;129(4):1553–1623.

3. Heyman F. Pay inequality and firm performance: evidence from matched employer–employee data. Applied Economics. 2005;37(11):1313–1327.

4. Mueller HM, Ouimet PP, Simintzi E. Within-Firm Pay Inequality. SSRN Working Paper. 2016;.

5. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: CRC press; 1994.

6. Audas R, Barmby T, Treble J. Luck, effort, and reward in an organizational hierarchy. Journal of Labor Economics. 2004;22(2):379–395.

7. Baker G, Gibbs M, Holmstrom B. Hierarchies and compensation: A case study. European Economic Review. 1993;37(2-3):366–378.

8. Dohmen TJ, Kriechel B, Pfann GA. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. Journal of Population Economics. 2004;17(2):193–228.

9. Grund C. The wage policy of firms: comparative evidence for the US and Germany from personnel data. The International Journal of Human Resource Management. 2005;16(1):104–119.

10. Lima F. Internal Labor Markets: A Case Study. FEUNL Working Paper. 2000;378.

11. Morais F, Kakabadse NK. The Corporate Gini Index (CGI) determinants and advantages: Lessons from a multinational retail company case study. International Journal of Disclosure and Governance. 2014;11(4):380–397.

12. Treble J, Van Gameren E, Bridges S, Barmby T. The internal economics of the firm: further evidence from personnel data. Labour Economics. 2001;8(5):531–552.

13. Ariga K, Brunello G, Ohkusa Y, Nishiyama Y. Corporate hierarchy, promotion, and firm growth: Japanese internal labor market in transition. Journal of the Japanese and International Economies. 1992;6(4):440–471.

14. Bell B, Van Reenen J. Firm performance and wages: evidence from across the corporate hierarchy. CEP Discussion paper. 2012;1088.

15. Eriksson T. Executive compensation and tournament theory: Empirical tests on Danish data. Journal of Labor Economics. 1999;17(2):262–280.

16. Leonard JS. Executive pay and firm performance. Industrial and Labor Relations Review. 1990;43(3):13–28.

17. Main BG, O'Reilly III CA, Wade J. Top executive pay: Tournament or teamwork? Journal of Labor Economics. 1993;11(4):606–628.

18. Rajan RG, Wulf J. The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. The Review of Economics and Statistics. 2006;88(4):759–773.

19. Tao HL, Chen IT. The level of technology employed and the internal hierarchical wage structure. Applied Economics Letters. 2009;16(7):739–744.

20. Axtell RL. Zipf distribution of US firm sizes. Science. 2001;293:1818–1820.

21. Brown C, Medoff J. The employer size-wage effect. Journal of political Economy. 1989;97(5):1027–1059.

22. Simon HA. The compensation of executives. Sociometry. 1957;20(1):32–35.

23. Joskow P, Rose N, Shepard A, Meyer JR, Peltzman S. Regulatory constraints on CEO compensation. Brookings Papers on Economic Activity Microeconomics. 1993;1993(1):1–72.

24. Piketty T. Capital in the Twenty-first Century. Cambridge: Harvard University Press; 2014.

25. Fleishman AI. Confidence intervals for correlation ratios. Educational and Psychological Measurement. 1980;40(3):659–670.

26. Steiger JH. Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. Psychological methods. 2004;9(2):164–182.

Appendix C

Appendices For A Hierarchy Model of Income Distribution

Supplementary materials for this paper are available at the Open Science Framework repository:

https://osf.io/3bsvt/

The supplementary materials include:

1. Data for all figures appearing in the paper;
2. Raw source data;
3. R code for all analysis;
4. Hierarchy model code.

C.1 Sources and Methods

Fig. 4.4: Modeled Income Distribution vs. US Data

Complementary Cumulative Distribution

The US complementary cumulative distribution is calculated from data in the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015.

Cumulative Distribution

The US cumulative distribution is calculated from data in the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015.

Gini Index

I use two sources for the US Gini index. The first source is the US Current Population Survey, Table PINC-08 (available from the US Census) over the years 1994 to 2015. The second source is the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015. I estimate the Gini index by constructing a Lorenz curve from the reported cumulative frequency data. R code implementing this method is available in the Supplementary Material.

The Census and IRS data are *not* mutually consistent. IRS data is based on tax units, not individuals. The advantage of the IRS data is that it is an administrative record. Current Population Survey (CPS) data, on the other hand, is obtained by *interview*. The advantage of the CPS data is that it explicitly counts individuals. The disadvantage is that "there is a tendency in household surveys for respondents to *under report* their income" [1].

Lorenz Curve

The US Lorenz curve is calculated from data in the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015.

Power Law Exponents

I estimate the power law exponent of the income distribution tail using the maximum likelihood method. US empirical data comes from the IRS Individual Complete Report (Publication 1304), Table 1.1. Since this data is reported in

Table C.1: Power Law Cutoff Boundaries in US Data

Year	Percentile	α
1996	0.987	2.92
1997	0.985	2.89
1998	0.996	2.58
1999	0.996	2.58
2000	0.995	2.54
2001	0.996	2.63
2002	0.996	2.67
2003	0.996	2.65
2004	0.995	2.59
2005	0.994	2.54
2006	0.993	2.54
2007	0.993	2.54
2008	0.994	2.66
2009	0.995	2.78
2010	0.994	2.73
2011	0.994	2.74
2012	0.992	2.64
2013	0.993	2.74
2014	0.992	2.70
2015	0.991	2.72

binned form, I use the binned log-likelihood equation developed by Virkar and Clauset [2]:

$$\mathscr{L} = n(\alpha - 1) \cdot \ln b_{\min} + \sum_{i=\min}^{k} h_i \ln \left[b_i^{(1-\alpha)} - b_{i+1}^{(1-\alpha)} \right] \qquad \text{(C.1)}$$

Here α is the power law exponent, b_i and b_{i+1} are consecutive bin boundaries, h_i and h_{i+1} are consecutive bin counts, k is the number of bins, and n is the sum of bin counts above $b_{\min}$ (the cutoff point for the power law). The best-fit exponent α is the value that maximizes the log-likelihood function ($\mathscr{L}$). Since there is no closed-form solution to this maximization problem, I solve for α numerically. To determine the power law exponent for the top 1% of incomes in each year, I set the power law cutoff boundary ($b_{\min}$) to the empirical bin that is closest to the 99th percentile. Results are shown in Table C.1.

To find the power law exponent in modeled data, I use the following maximum likelihood estimator:

$$\hat{\alpha} = 1 + n \left[\sum_{i}^{n} \ln \frac{x_i}{x_{\min}} \right]^{-1} \tag{C.2}$$

Here $\hat{\alpha}$ is the best-fit power law exponent, x_i is the ith data point, $x_{\min}$ is the lower bound of the power law, and n is the number of data points above $x_{\min}$. To ensure compatibility with empirical power law estimates, I estimate the model's power law exponent using the *empirical* cutoff values. For each model run, I set $x_{\min}$ by randomly selecting a percentile value from Table C.1.

All data and code are available in the Supplementary Material.

Probability Density Function

I estimate the normalized probability density function for US income using data from Current Population Survey Table PINC-08 (available from the US Census) over the years 1994 to 2015. This table reports binned data.

To estimate the normalized probability density function in each year, I first create a simulated income distribution (**I**) using bin midpoints. Each midpoint income M_i is repeated F_i times, where F_i is the frequency count for the ith bin. I then normalize **I** by dividing all elements by the mean income $\bar{I}$.

$$\mathbf{I} = \frac{\left(M_1 \overset{\times F_1}{\cdots\cdots}, \ M_2 \overset{\times F_2}{\cdots\cdots}, \dots, \ M_i \overset{\times F_i}{\cdots\cdots} \right)}{\bar{I}} \tag{C.3}$$

Lastly, I fit the simulated income distribution (**I**) with a numerical density function. R code implementing this method is available in the Supplementary Material.

Top 1% Income Share

Sources for top 1% income share data are shown in Table C.2.

Table C.2: US Top 1% Income Share Sources

Series	Info	Source
sfainc992j	Pre-tax factor income | equal-split adults | Share | Adults | share of total (ratio)	[3]
sfainc996i	Pre-tax factor income | individuals | Share | 20 to 64 | share of total (ratio)	[3]
sfainc999i	Pre-tax factor income | individuals | Share | All Ages | share of total (ratio)	[3]
sfainc999t	Pre-tax factor income | tax unit | Share | All Ages | share of total (ratio)	[3]
sfiinc992j	Fiscal income | equal-split adults | Share | Adults | share of total (ratio)	[3]
sfiinc992t	Fiscal income | tax unit | Share | Adults | share of total (ratio)	[3]
sfiinc996i	Fiscal income | individuals | Share | 20 to 64 | share of total (ratio)	[3]
sfiinc999i	Fiscal income | individuals | Share | All Ages | share of total (ratio)	[3]
sfiinc999t	Fiscal income | tax unit | Share | All Ages | share of total (ratio)	[3]
sptinc992j	Pre-tax national income | equal-split adults | Share | Adults | share of total (ratio)	[3]
sptinc996i	Pre-tax national income | individuals | Share | 20 to 64 | share of total (ratio)	[3]
sptinc999i	Pre-tax national income | individuals | Share | All Ages | share of total (ratio)	[3]
sptinc999t	Pre-tax national income | tax unit | Share | All Ages | share of total (ratio)	[3]
sfiinc_z_US	World Top Incomes Legacy Series	[4]
lakner	Calculated from micro data	[5]
piketty_book_no_kgains	Legacy data from Capital in the 21st Century	[6]
piketty_book_with_kgains	Legacy data from Capital in the 21st Century	[6]

Fig. 4.5: Firm Size Distributions Associated With Top Incomes and Wealth

Forbes 400 data is from the year 2014. Firm size data was collected by the author. For public companies, firm size data comes from Compustat. For private companies, data comes from firm websites and annual reports. The Execucomp 500 consists of the 500 top paid US executives in the Execucomp database in each year from 1992 to 2015.

Fig. 4.13 Capitalist Income Fraction of US CEOs

CEO pay data comes from Execucomp, while firm size data comes from Compustat. For the methods used to identify firm CEOs and the methods used to calculate capitalist income fraction, see Appendix C.3.

Fig. 4.15: Comparing the Capitalist Gradient Model to US Data

Capitalist Income Fraction vs. Income Percentile

US data is for the year 2007 and comes from Piketty, Fig. 8.10 [6]. Data is available at piketty.pse.ens.fr/en/capital21c2.

Capitalist Income Gini Index, Top 1% Share, and Size Distribution

Data for US capitalist income Gini index, top 1% share, and size distribution all come from the IPUMS database. I define capitalist income as the sum of income from dividends and interest. (Dividends = series *INCDIVID*, Interest = series *INCINT*).

The main challenge with this dataset is that it *censors* income above $100,000. All incomes above this threshold are replaced with a 'topcode' value. To deal with this censoring, I use the method proposed by Jenkins et al. [7]. The gist of this method is that you fit the uncensored data with a parametric distribution. You then replace the censored (topcoded) data with stochastic values drawn from the fitted parametric distribution (above the censor threshold). This gives a partially synthetic dataset on which you compute whatever statistic you desire. Because the process is stochastic, you repeat it many times, giving a range of values for the given statistic.
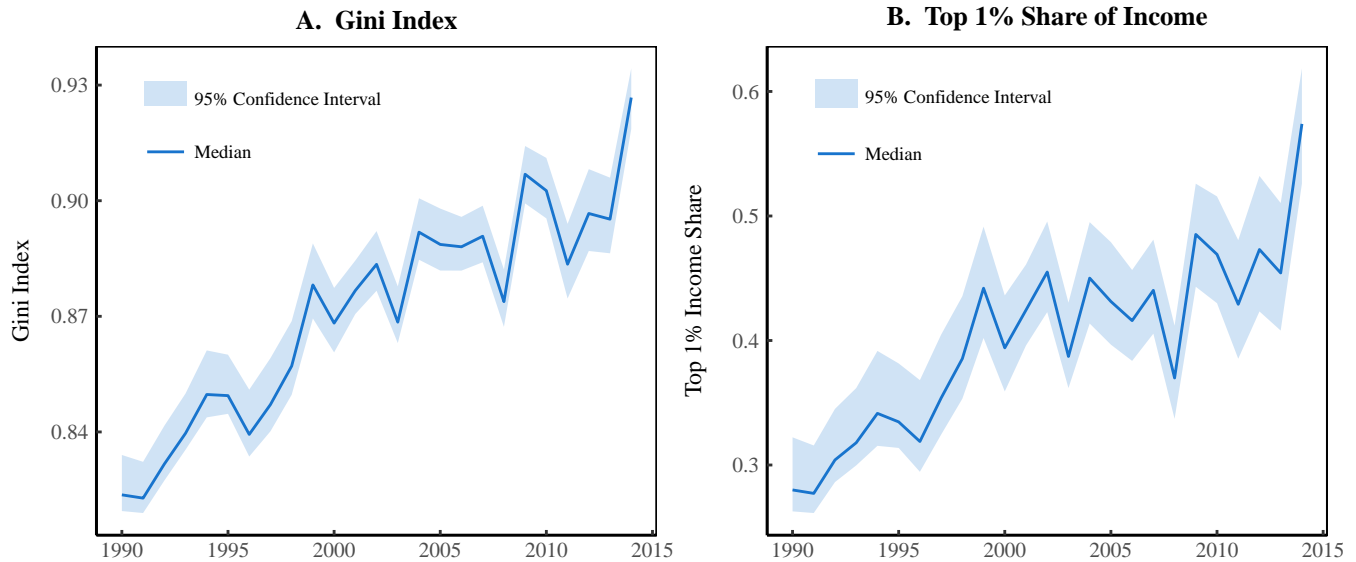
A. Gini Index

B. Top 1% Share of Income

Figure C.1: US Capitalist Income Inequality Estimates

This figure shows estimates of inequality in US capitalist income distribution. Data comes from the IPUMS database CPS public micro data. Capitalist income is the sum of dividends (series INCDIVID) and interest (series INCINT). Confidence intervals indicate the uncertainty in the estimate that arises from the stochastic method used to replace topcoded values.

I follow Jenkins et al. by using the GB2 distribution (generalized beta distribution of the second kind) to fit uncensored data. I use the R GB2 package [8] to fit both the dividends and interest data with a GB2 distribution. After replacing topcoded values with synthetic data, I sum dividends and interest income to estimate capitalist income. Figure C.1 show the resulting estimates for the Gini index and top 1% share of capitalist income. Although there is uncertainty in each annual estimate, the actual *range* of inequality values is *dominated* by the secular trend.

All code and data used for this analysis are provided in the Supplementary Material.

Capitalist Share of Total (National) Income

US Capitalist income share data comes from the Bureau of Economic Analysis, Table 1.12. (National Income by Type of Income). Capitalist income is defined as the sum of net dividends and net interest.

Fig. 4.16: Historical Income Distribution Trends in the United States

CEO pay ratio data comes from Mishel and Schieder [9]. This ratio is calculated using CEO income in the 350 largest US firms (ranked by sales), compared to the average income of workers in the firm's respective industry. Top 1% income share data sources are shown in Table C.2. The dividend share of national income is calculated using data from the Bureau of Economic Analysis, Table 1.12. (National Income by Type of Income).

Power law exponents for the top 1% of incomes are estimated on binned data using the method outlined by Virkar and Clauset [2]. I use income threshold data from the World Wealth and Income Database (see Table C.3)

Table C.3: US Top 1% Power Law Exponent Data Sources

Series	Info	Source
tfainc992j	Pre-tax factor income | equal-split adults | Threshold | Adults | constant 2015 local currency	[3]
tfiinc992j	Fiscal income | equal-split adults | Threshold | Adults | constant 2015 local currency	[3]
tfiinc992t	Fiscal income | tax unit | Threshold | Adults | constant 2015 local currency	[3]
tptinc992j	Pre-tax national income | equal-split adults | Threshold | Adults | constant 2015 local currency	[3]

C.2   Hierarchical Structure and Pay within Case-Study Firms

Based on worldly experience, most people would agree that firms are hierarchically organized, and that pay tends to increase as one moves up the hierarchy. However, the exact structure of this hierarchy has not been widely studied. This is likely due in part to the lack of scholarly interest (hierarchy is not part of neoclassical economic theory), but also the difficulty of obtaining firm payroll data, which is usually proprietary. Nonetheless, a handful of case-studies exist that have documented the hierarchical employment and pay structure of firms.

Table C.4 summarizes the case studies used in this paper, while Figure C.2 shows the hierarchical employment and pay structure of these firms. The firms remain anonymous, and are named after the authors of the case-study papers. By and large, these studies confirm our basic intuition about firm structure. Although the exact shapes vary, all of the firms in Figure C.2 have a roughly pyramidal employment structure and inverse pyramid pay structure.

To analyze the structure of these firms in further detail, I define and calculate the following three metrics: the *span of control*, the *inter-level mean pay ratio*, and the *intra-level Gini index*. The span of control is defined as the employment ratio between adjacent hierarchical levels. The inter-level mean pay ratio is the ratio of mean pay between adjacent hierarchical levels. Lastly, the intra-level Gini index is the Gini index of income inequality within a specific hierarchical level of a firm.

Table C.4: Summary of Firm Case Studies

Source		Years	Country	Firm Levels	Span of Control	Level Income	Level Income Dispersion
Audas	[10]	1992	Britain	All	✓	✓	
Baker	[11]	1969-1985	United States	Management	✓	✓	✓
Dohmen	[12]	1987-1996	Netherlands	All	✓	✓	✓
Grund	[13]	1995 & 1998	US and Germany	All		✓	✓
Lima	[14]	1991-1995	Portugal	All	✓	✓	✓
Morais	[15]	2007-2010	Undisclosed	All	✓	✓	
Treble	[16]	1989-1994	Britain	All	✓	✓	✓

Notes: This table shows metadata for the firm case studies displayed in Fig. C.3. 'Firm Levels' refers to the portion of the firm that is included in the study. 'Management' indicates that only management levels were studied.

Figure C.3 shows data for these metrics for the 6 case study firms. Figure C.3A shows how the *span of control* changes as a function of hierarchical level. The data shows unambiguously that the span of control tends to *increase* as one moves up the hierarchy. Figure C.3B shows how the *inter-level pay ratio* changes as a function of hierarchical level. Again, this ratio tends to *increase* as one moves up the hierarchy. Figure C.3C shows the *intra-level Gini index* as a function of hierarchical level. Unlike the other two quantities, intra-level income inequality seems to be more-or-less constant across all hierarchical levels (a linear regression reveals no significant trend).

This case study data plays a central role in the hierarchical model developed in this paper. From the case study evidence, I propose the following 'stylized' facts about firm employment and pay structure:

1. The span of control tends to *increase* with hierarchical level.
2. The inter-level pay ratio tends to *increase* with hierarchical level.
3. Intra-level income inequality is approximately *constant* across all hierarchical levels.

The case-study evidence informs the basic structure of the model, and also some of its key parameters. Parameters for span of control are determined from regressions on data in Figure C.3A, while parameters for intra-level income dispersion are determined from the mean of data in Figure C.3C. For a detailed discussion of the model algorithm and parameter fitting procedure, see Sections C.4 and C.5.

A. Firm Hierarchical Employment Structure
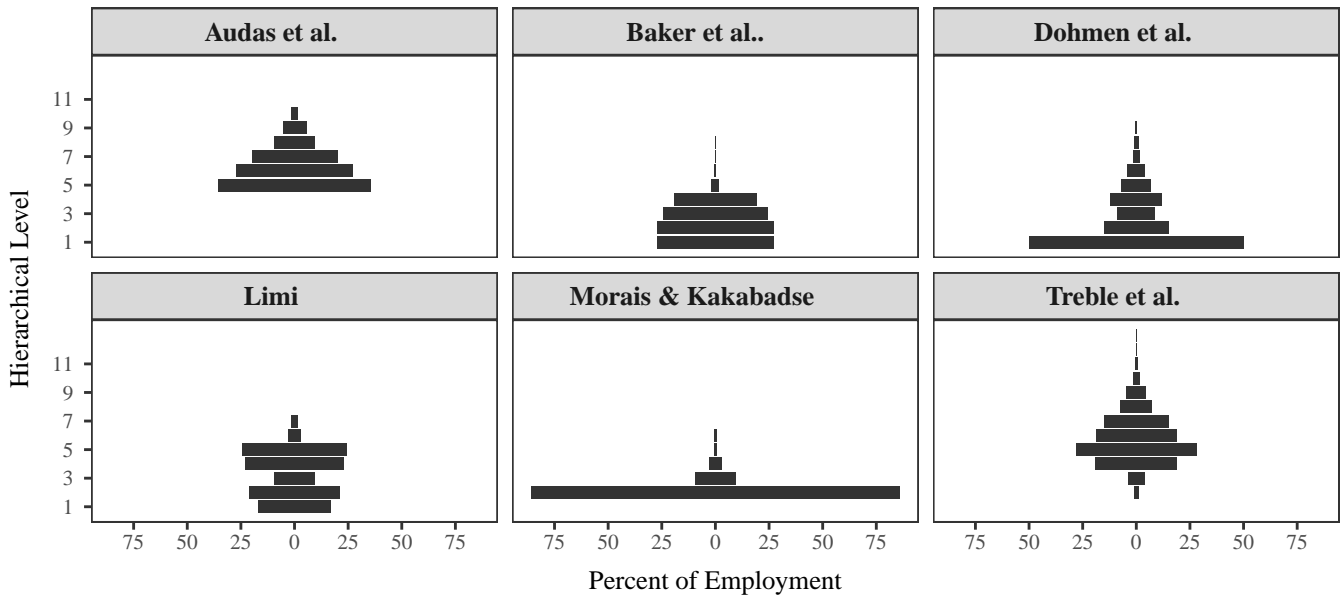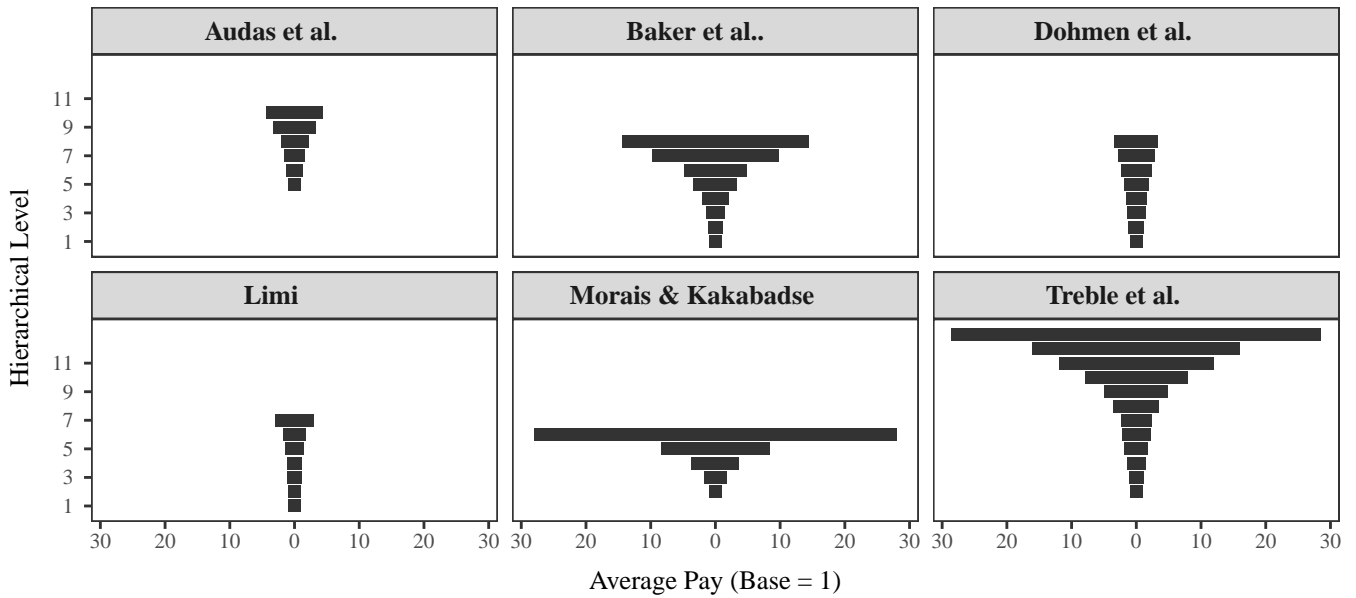

B. Firm Hierarchical Pay Structure

Figure C.2: The Hierarchical Employment and Pay Structure of Six Different Firms

This figure shows the pyramid structure of six different case study firms. Panel A shows the hierarchical structure of employment, while panel B shows the hierarchical pay structure.
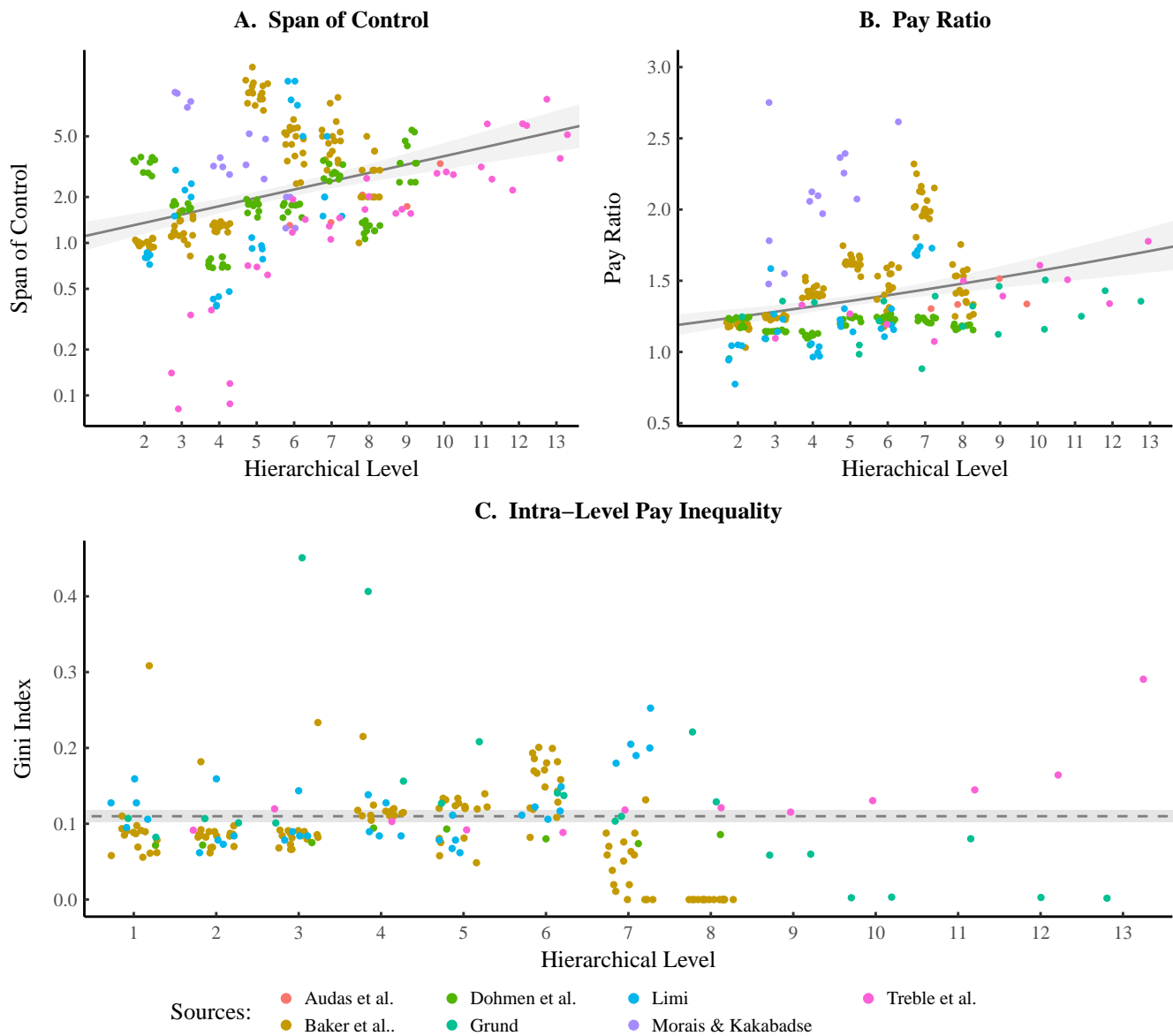
Figure C.3: Case Studies of Firm Hierarchical Structure

This figure shows data from 7 different single-firm case studies. Panel A shows how the span of control (the employment ratio between adjacent levels) relates to hierarchical level. Panel B shows how the ratio of mean pay between adjacent levels varies with hierarchical level. In these two panels, the *x*-axis corresponds to the *upper* hierarchical level in the ratio. Panel C shows levels of income inequality *within* individual hierarchical levels of each firm. Note that horizontal 'jitter' has been introduced in all three plots in order to better visualize the data (hierarchical level is a discrete variable). Grey regions correspond to the 95% confidence interval for regressions (or in panel C, the mean).

C.3   Compustat Data

This paper makes extensive use of the *Compustat* and *Execucomp* databases. Compustat contains data for most publicly traded US companies, while Execucomp contains data for executive compensation. Three key statistics used throughout this paper are calculated from this data: *firm mean income*, the *CEO-to-average-employee pay ratio*, and the *capitalist income fraction of executives*. I discuss the data and methods used for these calculations in the following sections.

C.3.1   Firm Mean Income

Firm mean income is calculated by dividing total staff expenses (Compustat Series XLR) by total employment (Compustat Series EMP):

$$\text{Firm Mean Income} = \frac{\text{Total Staff Expenses}}{\text{Total Employment}} \tag{C.4}$$

C.3.2   CEO Pay Ratio

Throughout this paper, I use the term 'CEO' to refer to the executive at the top of the corporate hierarchy. I identify CEOs using the titles contained in the Execucomp series TITLEANN. Because titles vary greatly by company, identifying the top executive is not always a simple task. While a manual search would be most accurate, this is unrealistic given that the Execucomp database contains over 275 000 entries. Instead, I use the following three-step algorithm to identify the 'CEO':

1. Find all executives whose title contains one or more of the words in the 'CEO Titles' list (Table C.5).
2. Of these executives, take the subset whose title does *not* contain any of the words in the 'Subordinate Titles' list (Table C.5).
3. If this search returns more than one executive per firm per year, chose the executive with the highest pay.

After identifying the CEO (and matching CEO pay data with firm data contained in the Compustat database), I calculate the CEO pay ratio using the following equation:

$$\text{CEO Pay Ratio} = \frac{\text{CEO Pay}}{\text{Firm Mean Income}} \tag{C.5}$$

Table C.5: Titles Used to Identify the 'CEO'

CEO Titles:	Subordinate Titles
president	vp
chairman	v-p
CEO	cfo
Chief Executive Officer	vice
chmn	chief finance officer
	president of
	coo
	division
	div
	president-
	group president
	chairmain-
	co-president
	deputy chairman
	pres.-
	Chief Financial Officer

Notes: This table shows the Execucomp titles used to identify the CEO of each company. CEOs are deemed to be those whose title contains words in the left column, but not those in the right column. Titles such as 'president-' and 'president of' are included in the subordinate list because they typically refer to a president of a division with the company: i.e. 'president of western division' or 'president-western hemisphere'.

CEO pay ratio and firm mean income data are collectively available for roughly 6000 firm-year observations over the period 1992-2016. I use this data to 'tune' my hierarchical model of the firm (see Section C.5) . Figure C.4 shows selected summary statistics of this dataset.

C.3.3 Capitalist Income Share of Executives

I define the capitalist income share of executives (K_{frac}) as the ratio of stock-options income to total income:

$$K_{\text{frac}} = \frac{\text{Stock Options}}{\text{Total Income}} \tag{C.6}$$
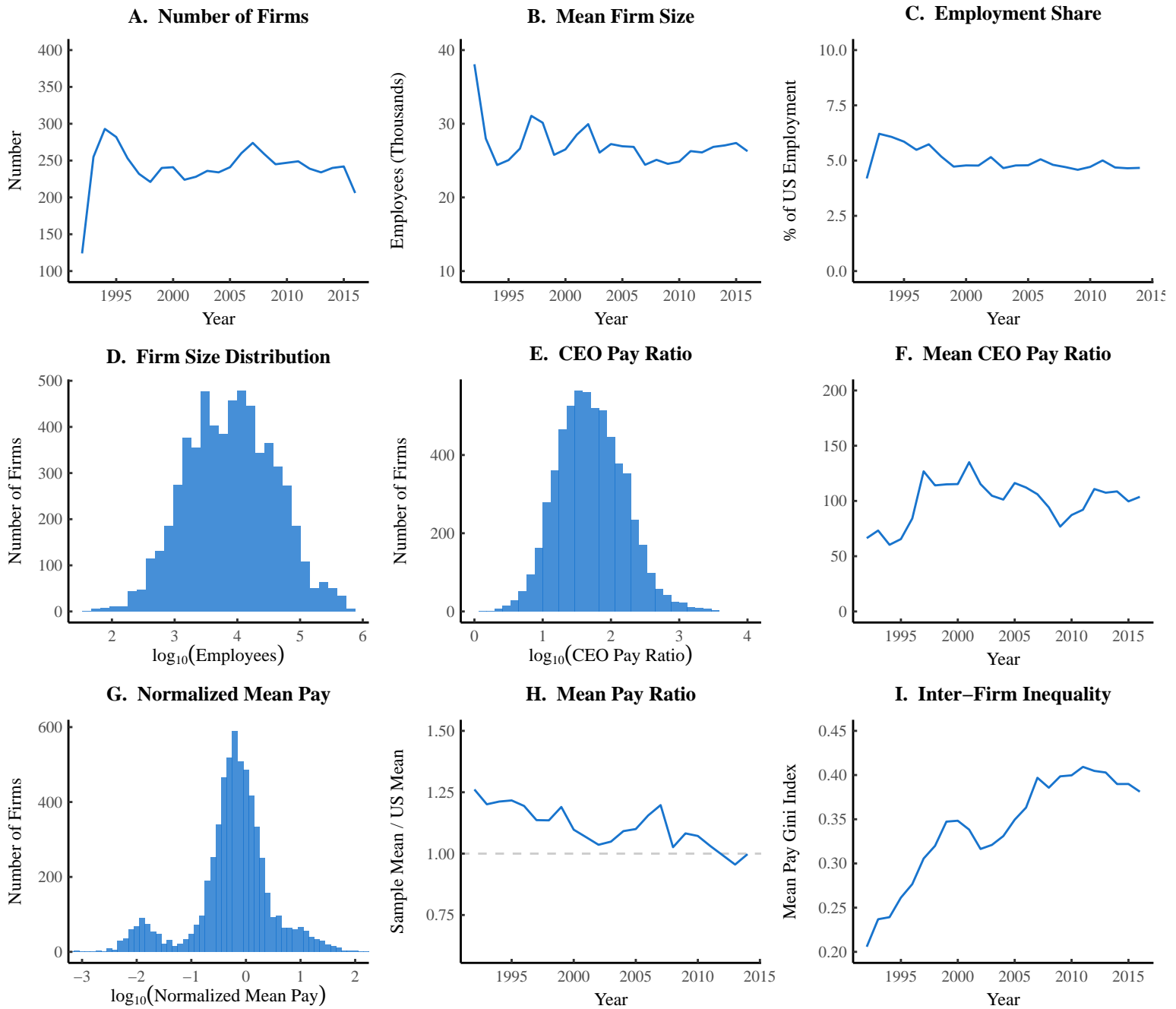
Figure C.4: Selected Statistics from the Firm Sample Used for Model Tuning

This figure shows statistics for the Compustat firm sample used to tune my hierarchical model. Panel A shows the number of firms in the sample over time, Panel B the average firm size, and Panel C the share of US employment held by these firms. Panel D shows the logarithmic distribution of firm size, and Panel E shows the logarithmic distribution of the CEO pay ratio. Panel F shows the mean CEO pay ratio of all firms over time. Panel G shows the logarithmic distribution of normalized mean pay (mean pay divided by the average pay of the firm sample in each year). Panel H shows the ratio of mean pay in the Compustat sample relative to the US average (calculated from BEA Table 1.12 by dividing the sum of employee and proprietor income by the number of workers in BEA Table 6.8C-D. Panel I shows the Gini index of firm mean pay over time.

Table C.6: Data Used to Calculate Executive Capitalist Income Fraction

Series	Description	Reporting Format
RSTKGRNT	The value of restricted stock granted during the year (determined as of the date of the grant).	1992
OPTION_AWARDS_BLK_VALUE	The aggregate value of stock options granted to the executive during the year as valued using Standard & Poor's Black-Scholes methodology.	1992
TDC1	Salary, Bonus, Other Annual, Total Value of Restricted Stock Granted, Total Value of Stock Options Granted (using BlackScholes), Long-Term Incentive Payouts, and All Other	1992
STOCK_AWARDS_FV	Fair value of all stock awards during the year as detailed in the Plan Based Awards table. Valuation is based upon the grant-date fair value as detailed in FAS 123R.	2006
OPTION_AWARDS_FV	Fair value of all options awarded during the year as detailed in the Plan Based Awards table. Valuation is based upon the grant-date fair value as detailed in FAS 123R.	2006
TDC1	Salary, Bonus, Non-Equity Incentive Plan Compensation, Grant-Date Fair Value of Option Awards, Grant-Date Fair Value of Stock Awards, Deferred Compensation Earnings Reported as Compensation, and Other Compensation.	2006

The Execucomp database contains two main accounting methods for valuing stock options: a '1992' reporting format that applies from 1992 to 2005', and a '2006' reporting format that applies from 2006 onward. These series are summarized in Table C.6. For both reporting formats, the relevant total income series (TDC1) remains the same. I calculate the capitalist income fraction of executives using the following two formulas for 1992 format and 2006 format, respectively:

$$K_{\text{frac\_1992}} = \frac{\text{RSTKGRNT} + \text{OPTION\_AWARDS\_BLK\_VALUE}}{\text{TDC1}} \tag{C.7}$$

$$K_{\text{frac\_2006}} = \frac{\text{STOCK\_AWARDS\_FV} + \text{OPTION\_AWARDS\_FV}}{\text{TDC1}} \tag{C.8}$$
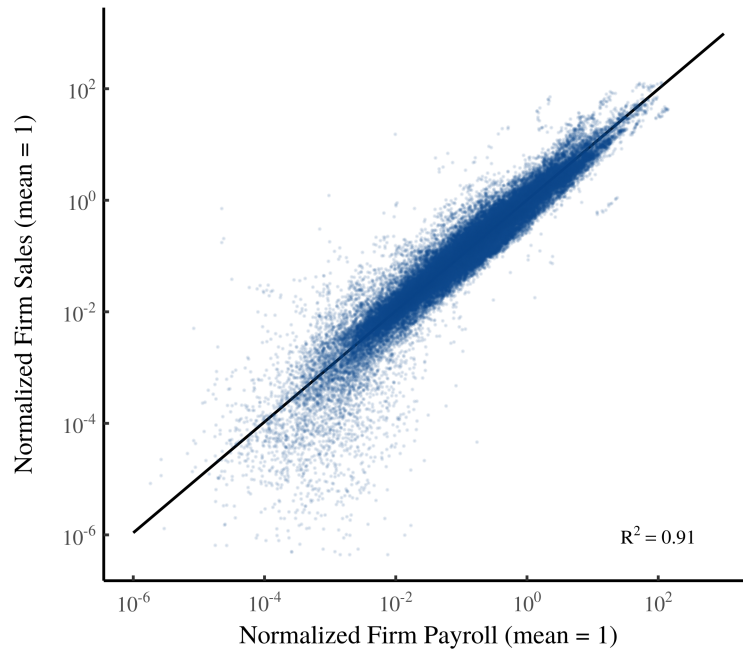
Figure C.5: Firm Sales vs. Payroll in the Compustat US Database

This figure plots normalized firm sales against normalized firm payroll for every firm-year observation in the Compustat US database from 1950 to 2015. Each dot is a specific firm in a specific year. To adjust for inflation, I divide sales and payroll by the database averages in the respective year.

C.3.4   Firm Sales vs. Firm Payroll

In section 4.4, I use the hierarchy model to reproduce historical trends in the CEO pay ratio. The empirical data from Mishel and Schieder [9] uses the CEOs in the top 350 US firms ranked by sales. Since the hierarchy model does not have sales, I calculate the CEO pay ratio by ranking firms by *total payroll*. Since payroll is highly correlated with firm sales (Fig. C.5), the former is a good proxy for the latter.

C.4   Hierarchy Model Equations

In this section, I outline the mathematics underlying my hierarchical model of the firm. The model assumptions, outlined below, are based on the stylized facts gleaned from the real-world firm data in section C.2.

1. Firms are hierarchically structured, with a span of control that increases *exponentially* with hierarchical level.
2. The ratio of mean pay between adjacent hierarchical levels increases *exponentially* with hierarchical level.
3. Intra-hierarchical-level income is lognormally distributed and constant across all levels.

Using these assumptions, I first develop an algorithm that describes the hierarchical employment within a model firm, followed by an algorithm that describes the hierarchical pay structure.

Table C.7: Notation

Symbol	Definition
a	span of control parameter 1
b	span of control parameter 2
C	CEO to average employee pay ratio
E	employment
F	cumulative distribution function
G	Gini index of inequality
h	hierarchical level
$\bar{I}$	average income
μ	lognormal location parameter
n	number of hierarchical levels in a firm
p	pay ratio between adjacent hierarchical levels
r	pay-scaling parameter
s	span of control
σ	lognormal scale parameter
T	total for firm
$\downarrow$	round down to nearest integer
$\prod$	product of a sequence of numbers
$\sum$	sum of a sequence of numbers

C.4.1   Generating the Employment Hierarchy

To generate the hierarchical structure of a firm, we begin by defining the span of control (s) as the ratio of employment (E) between two consecutive hierarchical levels (h), where $h = 1$ is the *bottom* hierarchical level. It simplifies later calculations if we define the span of control in level 1 as $s = 1$. This leads to the following piecewise function:

$$s_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \dfrac{E_{h-1}}{E_h} & \text{if } h \geq 2 \end{cases} \tag{C.9}$$

Based on our empirical findings in Section C.2, we assume that the span of control is *not* constant; rather it increases *exponentially* with hierarchical level. I model the span of control as a function of hierarchical level (s_h) with a simple exponential function, where a and b are free parameters:

$$s_h = \begin{cases} 1 & \text{if } h = 1 \\ a \cdot e^{bh} & \text{if } h \geq 2 \end{cases} \tag{C.10}$$

As one moves up the hierarchy, employment in each consecutive level (E_h) *decreases* by $1/s_h$. This yields Eq. C.11, a recursive method for calculating E_h. Since we want employment to be *whole* numbers, we round down to the nearest integer (notated by $\downarrow$). By repeatedly substituting Eq. C.11 into itself, we can obtain a non-recursive formula (Eq. C.12). In product notation, Eq. C.12 can be written as Eq. C.13.

$$E_h = \downarrow \frac{E_{h-1}}{s_h} \qquad \text{for} \qquad h > 1 \tag{C.11}$$

$$E_h = \downarrow E_1 \cdot \frac{1}{s_2} \cdot \frac{1}{s_3} \cdot \ldots \cdot \frac{1}{s_h} \tag{C.12}$$

$$E_h = \downarrow E_1 \prod_{i=1}^{h} \frac{1}{s_i} \tag{C.13}$$

Total employment in the whole firm (E_T) is the sum of employment in all hierarchical levels. Defining n as the total number of hierarchical levels, we get Eq. C.14, which in summation notation, becomes Eq. C.15.

$$E_T = E_1 + E_2 + \ldots + E_n \tag{C.14}$$

$$E_T = \sum_{h=1}^{n} E_h \tag{C.15}$$

In practice, n is not known beforehand, so we define it using Eq. C.13. We progressively increase h until we reach a level of zero employment. The highest level n will be the hierarchical level directly *below* the first hierarchical level with zero employment:

$$n = \{h \mid E_h \geq 1 \text{ and } E_{h+1} = 0\} \tag{C.16}$$

To summarize, the hierarchical employment structure of our model firm is determined by 3 free parameters: the span of control parameters a and b, and base-level employment E_1. Code for this hierarchy generation algorithm can be found in the C++ header files `hierarchy.h` and `exponents.h`, located in the Supplementary Material.

C.4.2   Generating Hierarchical Pay

To model the hierarchical pay structure of a firm, we begin by defining the inter-hierarchical pay-ratio (p_h) as the ratio of mean income ($\bar{I}$) between adjacent hierarchical levels. Again, it is helpful to use a piecewise function so that we can define a pay-ratio for hierarchical level 1:

$$p_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \dfrac{\bar{I}_h}{\bar{I}_{h-1}} & \text{if } h \geq 2 \end{cases} \tag{C.17}$$

Based on our empirical findings in Section C.2, we assume that the pay ratio increases *exponentially* with hierarchical level. I model this relation with the following function, where r is a free parameter:

$$p_h = \begin{cases} 1 & \text{if } h = 1 \\ r^h & \text{if } h \geq 2 \end{cases} \tag{C.18}$$

Using the same logic as with employment (shown above), the mean income I_h in any hierarchical level is defined recursively by Eq. C.19 and non-recursively by Eq. C.20.

$$\bar{I}_h = \frac{\bar{I}_{h-1}}{p_h} \tag{C.19}$$

$$\bar{I}_h = \bar{I}_1 \prod_{i=1}^{h} p_i \tag{C.20}$$

To summarize, the hierarchical pay structure of our model firm is determined by 2 free parameters: the pay-scaling parameter r, and mean pay in the base level ($\bar{I}_1$). Code for generating hierarchical pay can be found in the C++ header files `model.h`, located in the Supplementary Material.

Useful Statistics

Two statistics are used repeatedly within the model: mean firm pay, and the CEO-to-average-employee pay ratio.

Mean income for all employees ($\bar{I}_T$) is equal to the average of hierarchical level mean incomes ($\bar{I}_h$) weighted by the respective hierarchical level employment (E_h):

$$\bar{I}_T = \sum_{h=1}^{n} \bar{I}_h \cdot \frac{E_h}{E_T} \tag{C.21}$$

To calculate the CEO pay ratio, we define the CEO as the person(s) in the top hierarchical level. Therefore, CEO pay is simply $\bar{I}_n$, average income in the top hierarchical level. The CEO pay ratio (C) is then equal to CEO pay divided by average pay:

$$C = \frac{\bar{I}_n}{\bar{I}_T} \tag{C.22}$$

C.4.3 Adding Intra-Level Pay Dispersion

Up to this point, we have modeled only the *mean* income within each hierarchical level of a firm. The last step in the modeling process is to add pay *dispersion* within each hierarchical level.

I assume that pay dispersion within hierarchical levels is *lognormally* distributed. The lognormal distribution is defined by location parameter μ and scale parameter σ. Our empirical investigation of firm case studies indicated that pay dispersion with hierarchical levels is relatively constant (see Fig. C.3C). Given this finding, I assume *identical inequality* within all hierarchical levels. This means that the lognormal scale parameter σ is the same for all hierarchical levels.

A. Adding Pay Dispersion Within Each Hierarchical Level
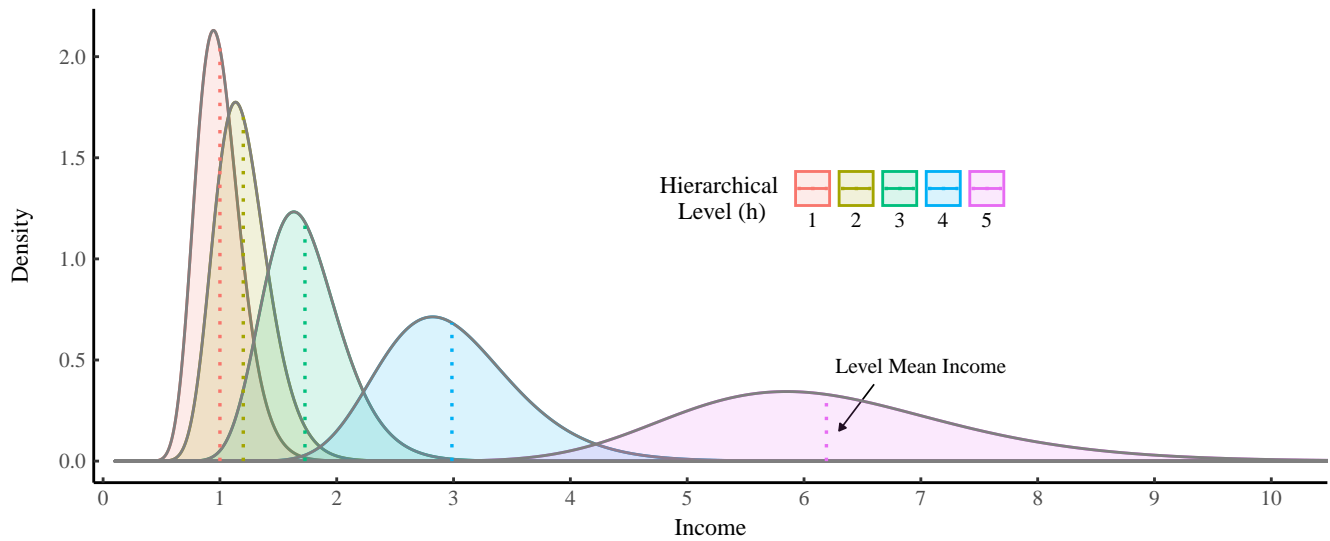
B. Relative Contribution to Intra–Firm Income Distribution
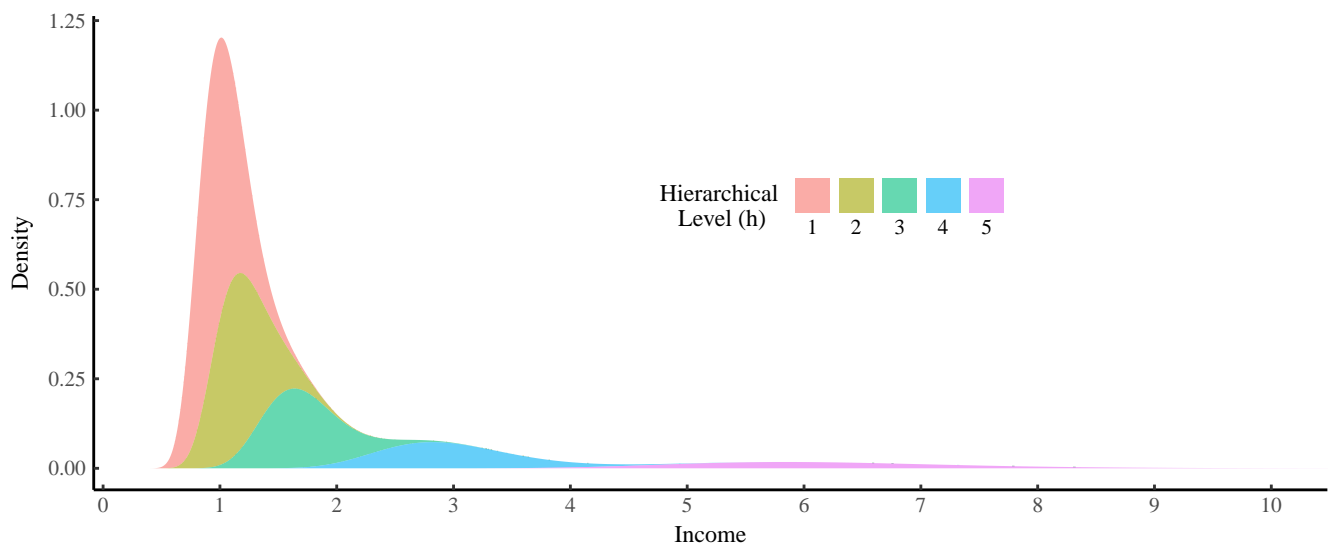

Figure C.6: Adding Intra-Level Pay Dispersion to a Model Firm

This illustrates a model firm with lognormal pay dispersion in each hierarchical level. The model firm has a pay-scaling parameter of $r = 1.2$ and an intra-level Gini index of 0.13. Panel A shows the separate distributions for each level, with mean income indicated by a dashed vertical line. Panel B shows contribution of each hierarchical level to the resulting income distribution for the whole firm (income density functions are summed while weighting for their respective employment).

In order to add dispersion within each hierarchical level, I multiply mean pay $\bar{I}_h$ by a lognormal random variate with an expected mean of one. Formally, this is represented by Eq. C.23. Since the mean of a lognormal distribution is equal to $e^{\mu + \frac{1}{2}\sigma^2}$, I leave it to the reader to show that a mean of one requires that μ be defined by Eq. C.24.

$$I_h = \bar{I}_h \cdot \ln \mathcal{N}(\mu, \sigma) \tag{C.23}$$

$$\mu = -\frac{1}{2}\sigma^2 \tag{C.24}$$

Given a value for σ (which is a free parameter), we can define the pay distribution within any hierarchical level of a firm. This process is shown graphically in Figure C.6. Figure C.6A shows the lognormal income distributions for each hierarchical level of a 5-level firm. Figure C.6B shows the size-adjusted contribution of each hierarchical level to the overall intra-firm income distribution. Lower levels have more members, and thus dominate the overall distribution. The code implementing this method can be found in the C++ header file `model.h`, located in the Supplementary Material.

C.4.4   Calculating Hierarchical Power

I define an individual's hierarchical power as the number of subordinates (S) under their control, plus 1:

$$P = S + 1 \tag{C.25}$$

Because the hierarchy model simulates only the *aggregate* structure of firms (employment by hierarchical level), hierarchical power is calculated as an *average* per rank. For hierarchical rank h, the average hierarchical power ($\bar{P}_h$) is defined as the average number of subordinates ($\bar{S}_h$) plus 1:

$$\bar{P}_h = \bar{S}_h + 1 \tag{C.26}$$

Each individual with rank h is assigned the average power $\bar{P}_h$. The average number of subordinates $\bar{S}_h$ is equal to the sum of employment (E) in all subordinate levels, divided by employment in the level in question:

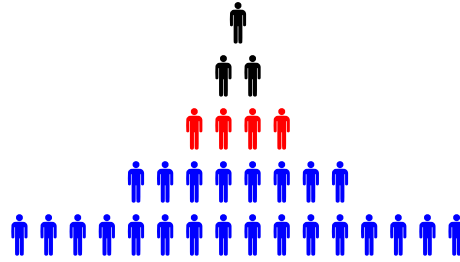$$\bar{S}_h = \sum_{i=1}^{h-1} \frac{E_i}{E_h} \tag{C.27}$$

Figure C.7: Calculating the Average Number of Subordinates

As an example, consider the hierarchy in Figure C.7. The average number of subordinates below each individual in hierarchal level 3 (red) would be:

$$\bar{S}_3 \;=\; \frac{E_1 + E_2}{E_3} \;=\; \frac{16 + 8}{4} = 6 \tag{C.28}$$

Therefore, these individuals would all be assigned a hierarchical power of 7.

Table C.8: Model Parameters

Parameter	Definition	Action	Scope
α	Firm size distribution exponent	Determines the skewness of the firm size distribution	—
a, b	Span of control parameters	Determines the shape of the firm hierarchy.	Identical for all firms.
E_1	Employment in base hierarchical level	Used to build the employment hierarchy from the bottom up. Determines total employment.	Specific to each firm.
r	Pay-scaling parameter	Determines the rate at which mean income (within a firm) increases by hierarchical level.	Specific to each firm.
$\bar{I}_h$	Mean pay in base hierarchical level	Sets the base level income of the firm, which determines firm average pay.	Specific to each firm.
σ	Intra-hierarchical level pay dispersion parameter	Determines the level of inequality within hierarchical levels of a firm.	Identical for all firms.

C.5   Restricting Parameters

As discussed in section C.4, the hierarchy model has many 'free' parameters. Table C.8 summarizes all of the parameters used in this model. While free to take on any value, I restrict these parameters exclusively using empirical data. In the following sections, I outline the methods used for this restriction.

C.5.1   Firm Size Distribution

Recent studies have found that firm size distributions in the United States [17] and other G7 countries [18] can be modeled accurately with a power law. A power law has the simple form shown in Eq. C.29, where the probability of observation x is inversely proportional to x raised to some exponent α:

$$p(x) \propto \frac{1}{x^\alpha} \tag{C.29}$$

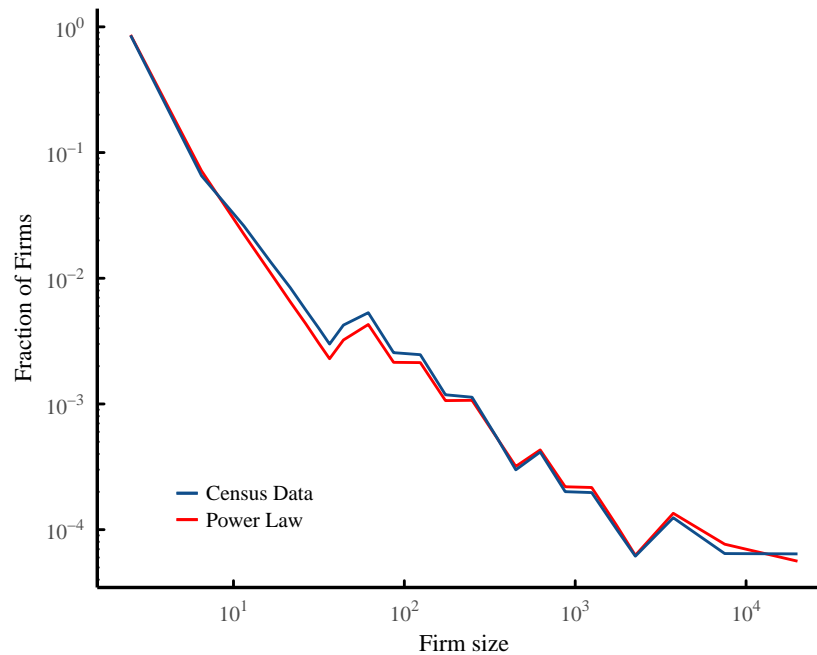Figure C.8 compares the US firm size distribution with a power law of ex-

Figure C.8: The United States Firm Size Distribution

This figure shows the US firm size distribution compared to a power law distribution with exponent $\alpha = 2.01$ (a simulation with 15 million firms) . The US histogram combines data for 'employer' firms with data for unincorporated self-employed workers. Data for 'employer' firms is from the US Census Bureau, Statistics of U.S. Businesses (using data for 2013). This data is augmented with Bureau of Labor Statistics data for unincorporated self-employed workers (series LNU02032185 and LNU02032192). The histogram preserves Census firm-size bins, with self-employed data added to the first bin. The last point on the histogram consists of all firms with more than 10,000 employees.

---

ponent $\alpha = 2.01$. Although not perfect, the fit is good enough for modeling purposes. I assume that the firm sizes can be modeled with a discrete power law random variate. I model the US firm size distribution with $\alpha = 2.01$.

A characteristic property of power law distributions is that as α approaches 2, the mean becomes *undefined*. In the present context, this means that the model can produce firm sizes that are extremely large — far beyond anything that exists in the real world. To deal with this difficulty, I *truncate* the power law distribution at a maximum firm size of 2.3 million. This happens to be the present size of Walmart, the largest US firm in existence.

Code for the discrete power law random number generator can be found in the C++ header file `rpld.h`, located in the Supplementary Material. This code
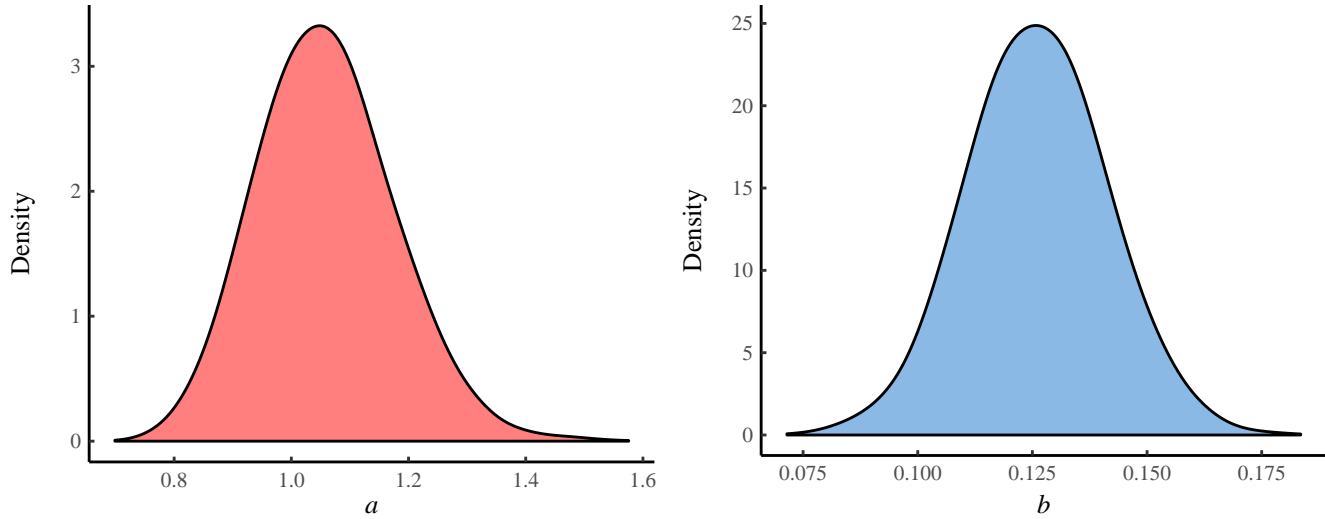
Figure C.9: Density Estimates for Span of Control Parameters

This figure shows density estimates for the parameters a and b, which together determine the 'shape' of the firm hierarchy. These parameters are determined from regressions on firm case-study data (Fig. C.3). The density functions are estimated using a bootstrap analysis, which involves resampling (with replacement) the case study data many times, and calculating the parameters a and b for each resample.

---

is an adaption of Collin Gillespie's discrete power law generator found in the R poweRlaw package [19] (which is, in turn, an adaption of the algorithm outline by Clauset [20]).

C.5.2   Span of Control Parameters

The parameters a and b together determine the shape of firm employment hierarchy. These parameters are estimated from an exponential regression on case study data (Fig. C.3A). The model proceeds on the assumption that these parameters are *constant* across all firms.

Because the case-study sample size is small, there is considerable uncertainty in these values. I incorporate this uncertainty into the model using the *bootstrap* method [21], which involves repeatedly resampling the case-study data (with replacement) and then estimating the parameters a and b from this resample. Figure C.9 shows the probability density distribution resulting from this bootstrap analysis. I run the model many times, each time with a and b determined by a bootstrap resample of case-study data.

Code implementing this bootstrap can be found in the C++ header file `boot_span.h`.

C.5.3 Base Level Employment

Given span of control parameters a and b, each firm hierarchy is constructed from the bottom hierarchical level up. Thus, we must know base level employment. In practice, however, we don't know this value — instead we are given *total* employment for a particular firm. While it may be possible to use the equations in section C.4 to define an analytic function relating total employment to base level employment, this is beyond my mathematical abilities.

Instead, I use the model to reverse engineer the problem. I input a range of different base employment values into equations C.10, C.13, and C.15 and calculate total employment for each value. The result is a discrete mapping relating base-level employment to total employment. I then use the C++ Armadillo interpolation function to linearly interpolate between these discrete values. This allows us to predict base level E_1, given total employment E_T. Code implementing this method can be found in the C++ header file `base_fit.h`, located in the Supplementary Material.

C.5.4 Pay-Scaling Parameter

The pay-scaling ratio r determines the rate at which mean pay increases by hierarchical level. Unlike the span of control parameters, the pay-scaling parameter is allowed to vary between firms. But how should it vary? I restrict the variation of this parameter in a two-step process. I first 'tune' the model to Compustat data. This results in a distribution of pay-scaling parameters specific to Compustat firms. I then fit this data with a parameterized distribution, from which simulation parameters are randomly chosen.

Fitting Compustat Pay-Scaling Parameters

I fit the pay-scaling parameter r to Compustat firms using the CEO-to-average-employee pay ratio (C). The first step of this process is to build the employment hierarchy for each Compustat firm using parameters a, b, and E_1 (the latter is determined from total employment). Given this hierarchical employment structure, the CEO pay ratio in the modeled firm is uniquely determined by the parameter r. Thus, we simply choose r such that the model produces a CEO pay ratio that is equivalent to the empirical ratio.

A. Fitted Pay–Scale Parameters

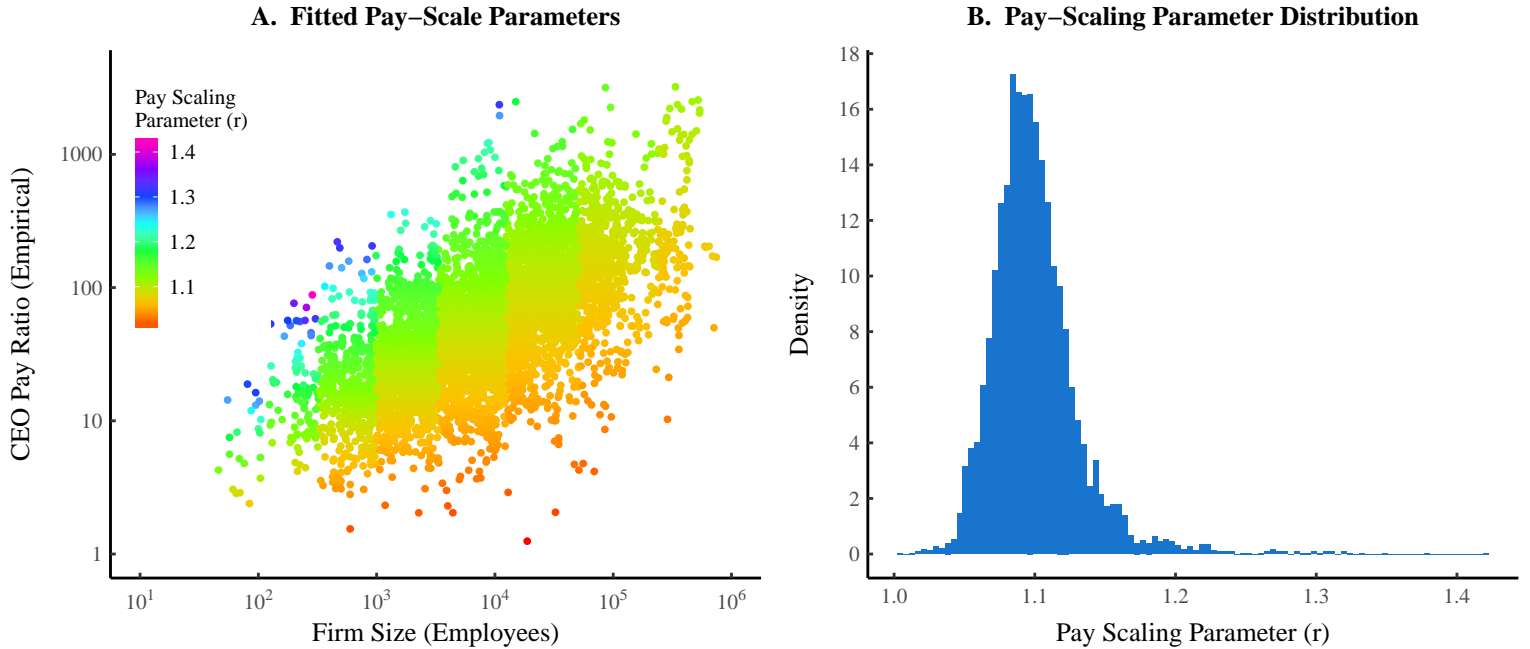## B. Pay–Scaling Parameter Distribution

Figure C.10: Fitting Compustat Firms with a Pay-Scaling Parameter

This figure shows the fitted pay-scaling parameters (r) for all Compustat firms. Panel A shows the relation between the CEO pay ratio and firm size, with the fitted pay-scaling parameter indicated by color. The discrete changes in color (evident as vertical lines) correspond to changes in the number of hierarchical levels within firms. The pay-scaling parameter distribution for all firms (and years) is shown in panel B.

To solve for this r value, I use numerical optimization (the bisection method) to minimize the error function shown in Eq. C.30. Here $C_{\text{Compustat}}$ and C_{model} are Compustat and modeled CEO pay ratios, respectively.

$$\epsilon(r) = \Big| C_{\text{model}} - C_{\text{Compustat}} \Big| \tag{C.30}$$

For each firm, the fitted value of r minimizes this error function. To ensure that there are no large errors, I discard Compustat firms for which the best-fit r parameter produces an error that is larger than $\epsilon = 0.01$). Fitted results for r are shown in Figure C.10. Code implementing this method can be found in the C++ header file `fit_model.h`, located in the Supplementary Material.

Generating a Pay Scaling Distribution

Once we have generated r parameters for every Compustat firm, the next step is to fit a parameterized distribution to this data. For Compustat firms, the dis-
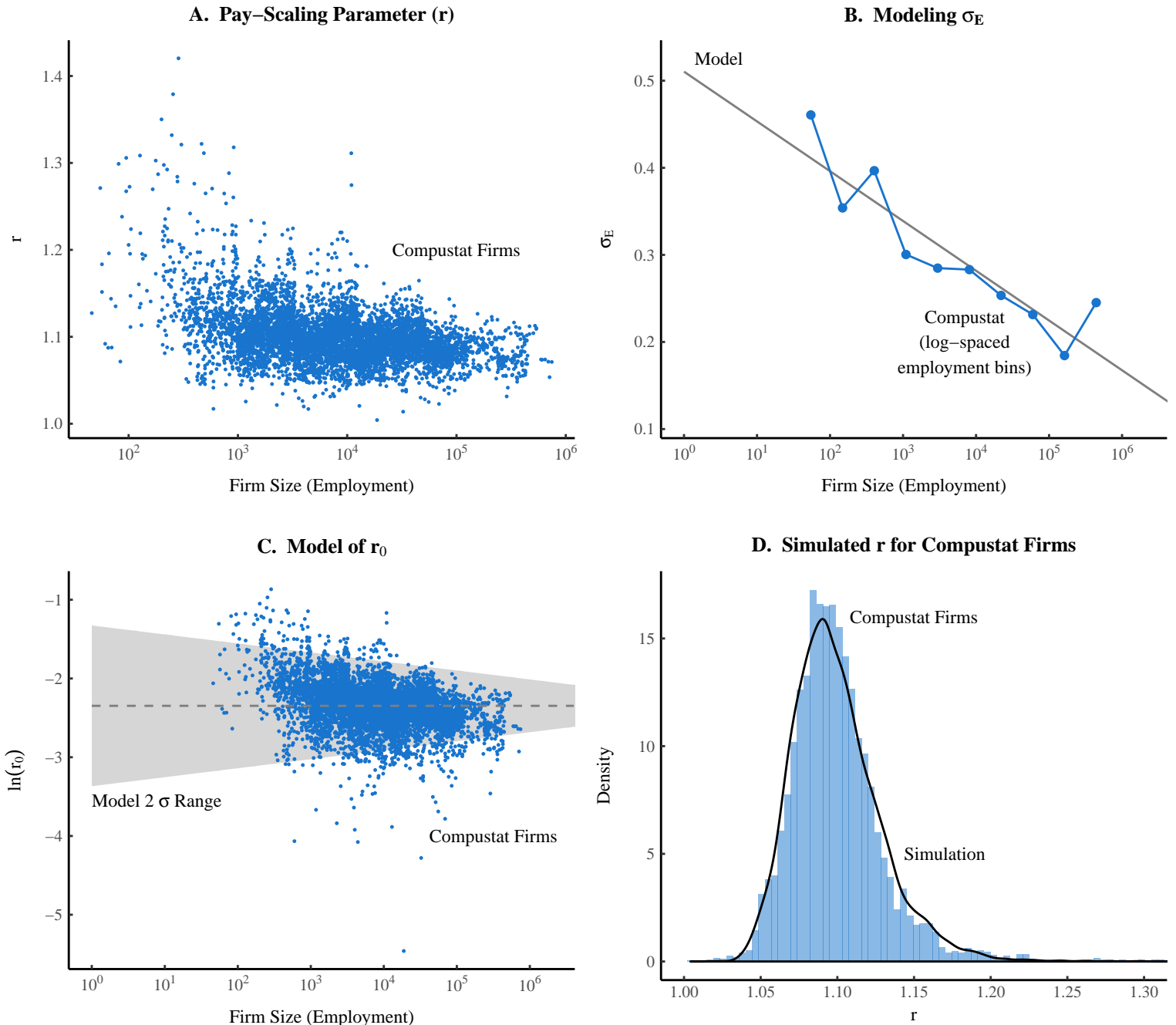
Figure C.11: Modeling the Firm Pay Scaling Distribution

This figure visualizes the model used to simulate firm pay-scaling parameters (r). Panel A shows the relation between r and firm employment for Compustat firms. For the simulation, the distribution of r is modeled with the lognormal variate r_0. Panel B shows how the lognormal scale parameter σ_E (defined by Eq. C.35) changes with firm size. The straight line indicates the modeled relation. Panel C shows how the modeled dispersion of $\ln(r_0)$ declines with firm size, and how this relates to Compustat r_0 data. The 2σ range indicates 2 standard deviations from the mean (on log-transformed data). Panel D shows how the distribution of r for Compustat firms compares to the *simulated* distribution achieved by applying the model to the same Compustat firms.

persion of r is approximately lognormal, and tends to decline with firm size (see Figure C.11A). I model r as a shifted function of the lognormal variate r_0:

$$r = 1 + \ln \mathcal{N}(r_0) \tag{C.31}$$

The lognormal variate r_0 is defined by location parameter μ and scale parameter σ. While μ is assumed to be constant for all firms, σ is a function of firm size E:

$$r_0(E) = \ln \mathcal{N}(r_0; \mu, \sigma_E) \tag{C.32}$$

I use the tuned Compustat data to solve for the parameters μ and σ. We first transform Compustat r values using Eq. C.33 to get the Compustat distribution of r_0:

$$r_0 = r - 1 \tag{C.33}$$

The best-fit value for μ is defined by taking the mean of $\ln(r_0)$:

$$\mu = \overline{\ln(r_0)} \tag{C.34}$$

Similarly, we can solve for the best-fit value for σ by taking the standard deviation of $\ln(r_0)$. However, unlike μ, the value σ will depend on the size range of firms (E):

$$\sigma_E = \text{SD}\left[\ \ln(r_0)\ \right]_E \tag{C.35}$$

Figure C.11B plots σ_E vs. E for logarithmically spaced size groupings of Compustat firms. I model this relation using a log-linear regression. Figure C.11C shows how the modeled dispersion in r_0 varies with firm size, and how this compares to Compustat data.

Once we have fitted the parameters μ and σ to the tuned Compustat data, we can generate r values for simulated firms using equations C.31 and C.32. Although the model is simple, it produces reasonably accurate results. To test this accuracy, we can apply the model to the same Compustat firms for which it is 'tuned'. For each Compustat firm, we use the method outlined above to stochastically generate a pay-scaling value r. As Figure C.11D shows, the resulting simulated distribution of r fairly accurately reproduces the original data.

When we move from simulating Compustat firms to a real-world distribution of firms, this model involves significant extrapolations for small firms. Why?

The Compustat firm sample has very few observations for firms smaller than 100. And those small firms that are included in the sample are likely *not* representative of the wider population, since they are small *public* firms. In the real world, virtually all small firms are *private*. As with all extrapolations, we simply do the best with the data that is available, while noting that better data might render the extrapolation moot. The code implementing this model can be found in the C++ header file `r_sim.h`, located in the Supplementary Material.

Note: When attempting to reproduce historical trends in US income inequality (Fig. 4.18), I vary the mean of the pay-scaling distribution by multiplying the fitted lognormal component by a random constant c:

$$r = 1 + c \cdot \ln \mathcal{N}(r_0) \tag{C.36}$$

C.5.5 Base-Level Mean Pay

As with the pay-scaling parameter, base level mean pay varies across firms. How should it vary? Again, I restrict the variation of this parameter in a two-step process. I first 'tune' the model to Compustat data. This results in a distribution of base pay specific to Compustat firms. I then fit this data with a parameterized distribution, from which simulation parameters are randomly chosen.

Fitting Compustat Base Level Pay

Having already fitted a hierarchical pay structure to each Compustat firm (in the process of finding r), we can use this data to estimate base pay for each firm. To do this, we set up a ratio between base level pay ($\bar{I}_1$) and firm mean pay ($\bar{I}_T$) for both the model and Compustat data:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{\bar{I}_1^{\text{model}}}{\bar{I}_T^{\text{model}}} \tag{C.37}$$

The modeled ratio between base pay and firm mean pay ($\bar{I}_1^{\text{model}}/\bar{I}_T^{\text{model}}$) is *independent* of the choice of base pay. This is because the modeled firm mean pay is actually a *function* of base pay (see Eq. C.20 and C.21). If we run the model with $\bar{I}_1^{\text{model}} = 1$, then Eq. C.37 reduces to:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{1}{\bar{I}_T^{\text{model}}} \tag{C.38}$$
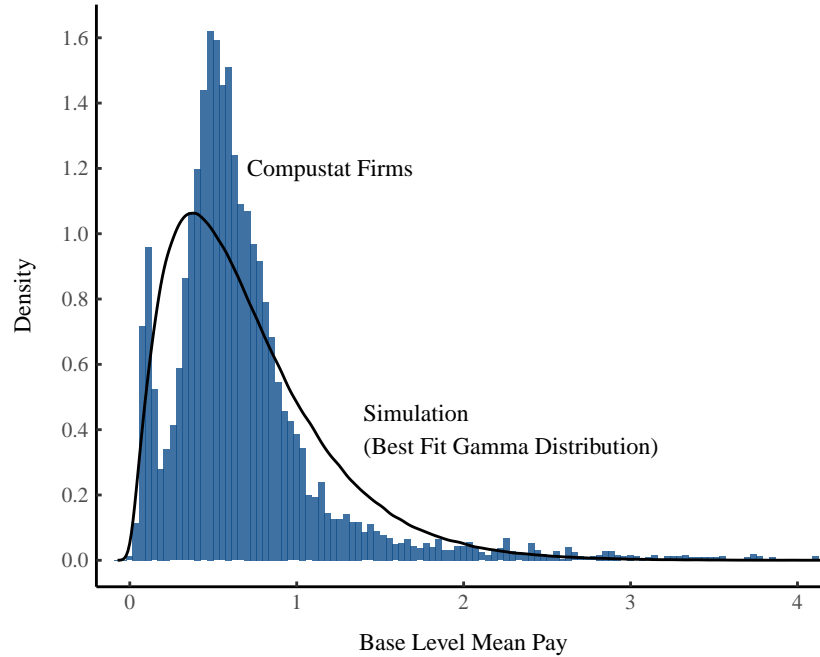
Figure C.12: Modeling Firm Base Level Mean Pay

This figure shows the distribution of fitted base level mean pay for Compustat firms (histogram). I model this data with a gamma distribution, from which simulated firm base mean pay is randomly drawn.

We can then rearrange Eq. C.38 to solve for an estimated base pay for each Compustat firm ($\bar{I}_1^{\text{Compustat}}$):

$$\bar{I}_1^{\text{Compustat}} = \frac{\bar{I}_T^{\text{Compustat}}}{\bar{I}_T^{\text{model}}} \tag{C.39}$$

Code implementing this method is found in the C++ header file `fit_model.h`, located in the Supplementary Material.

Generating a Base Pay Distribution

Once each Compustat firm has a fitted value for base-level mean pay, we fit this data with a parametric distribution which is then used to stochastically generate base-level mean pay for the simulation. Since Compustat data is comprised of observations over *multiple* years, in order to aggregate this data into a single distribution, we must account for inflation. Rather than use a price index like the GDP deflator, I divide all firm mean pay data by the average Compustat mean pay in the appropriate year. Since our simulation is concerned only with relative
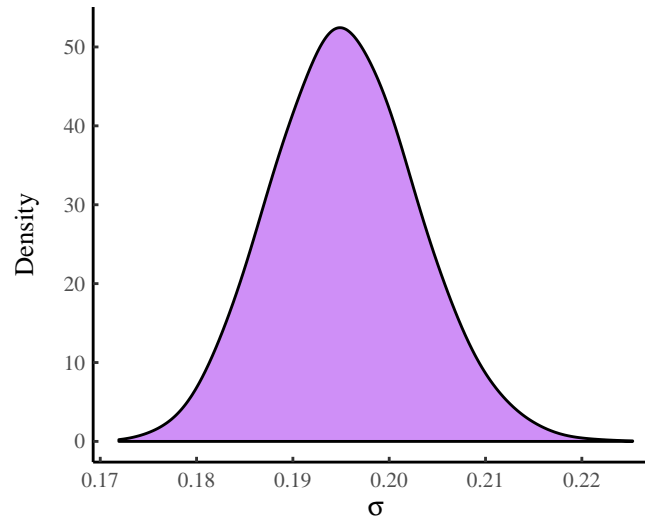
Figure C.13: Density estimates for Intra-Hierarchical Level Pay Dispersion Parameter σ

This figure shows the distribution of the lognormal scale parameter σ, which determines pay dispersion within all hierarchical levels of all firms. The distribution is calculated using the bootstrap method.

---

incomes (rather than absolute incomes) no pertinent information is lost in this process.

I model the Compustat firm base pay distribution with a gamma distribution (Fig. C.12). Note that because the Compustat data has a bimodal structure (that I do not aim to replicate), the gamma distribution is not a particularly strong fit. Nonetheless the gamma model closely replicates the inequality of firm base pay (which has a Gini index of roughly 0.35). Code implementing this model can be found in the C++ header file `base_pay_sim.h` (in the Supplementary Material).

C.5.6 Intra-Hierarchical Level Income Dispersion

Intra-hierarchical level income dispersion is modeled with a lognormal distribution, with the amount of inequality determined by the scale parameter σ. I estimate σ from the case-study data shown in Figure C.3C. This data uses the Gini index as the metric for dispersion.

To estimate σ, we first calculate the mean Gini index of all data ($\bar{G}$). We then use Eq. C.40 to calculate the value σ, which corresponds to the lognormal scale parameter that would produce a lognormal distribution with an equivalent

Gini index. This equation is derived from the definition of the Gini index of a lognormal distribution: $G = \text{erf}(\sigma/2)$.

$$\sigma = 2 \cdot \text{erf}^{-1}(\bar{G}) \tag{C.40}$$

The model proceeds on the assumption that σ is constant for all hierarchical levels within all firms. Because the case-study sample size is small, there is considerable uncertainty in these values. I quantify this uncertainty using the *bootstrap* method [21], which involves repeatedly resampling the case-study data (with replacement) and then estimating the parameter σ from this resampled data.

Figure C.13 shows the probability density distribution resulting from this bootstrap analysis. In order to incorporate this uncertainty, I run the model many times, with each run using a different bootstrapped value for σ. Code implementing this method can be found in the C++ header file `boot_sigma.h`, located in the Supplementary Material.

C.5.7   Summary of Model Structure

The model is implemented in C++ using a modular design. Each major task is carried out by a separate function that is defined in a corresponding header file. Table C.9 summarizes this structure sequentially in the order that functions are called. In each step, I briefly summarize the action that is performed, giving reference to the section where this action is described in detail.

Step	Action	Reference Section	Parameter(s)	Header File(s)
	Table C.9: Model High-Level Structure			
1	Bootstrap case-study data	C.5.2, C.5.6	a, b, σ	`boot_span.h` `boot_sigma.h`
2	Get Compustat base-level employment	C.5.3	E_1	`base_fit.h`
3	Fit Compustat pay-scaling parameters	C.5.4	r	`fit_model.h`
4	Get Compustat base-level mean pay	C.5.5	$\bar{I}_1$	`fit_model.h`
5	Generate power law firm size distribution	C.5.1	α	`rpld.h`
6	Get simulation base-level employment	C.5.3	E_1	`base_fit.h`
7	Simulate pay-scaling parameter distribution by fitting Compustat data	C.5.4	r	`r_sim.h`
8	Simulate base mean pay distribution by fitting Compustat data	C.5.5	$\bar{I}_1$	`base_pay_sim.h`
9	Run hierarchy model	C.4	all	`model.h`

Notes: Model code makes extensive use of Armadillo, an open-source C++ linear algebra library [22].

C.6   The Adjusted Hierarchy Model

The hierarchy model tends to underestimate US income inequality. I think that this is caused by the model's reliance on Compustat Firm data (see Appendix C.5), which is biased towards large firms. The result is that the model likely has too little inter-firm income dispersion. Here I present the results of an *adjusted* model in which inter-firm income dispersion is increased so that the model closely reproduces US macro-level data.

As outlined in Appendix C.5, inter-firm income dispersion is modeled by fitting a gamma distribution to Compustat data. The gamma distribution has the following probability density function:

$$p(x) = \frac{1}{\Gamma(k)\theta^k} \cdot x^{k-1} \cdot e^{-k/\theta} \tag{C.41}$$

In the original model, the parameters k and θ are both determined by empirical data. In the adjusted model, I introduce a fudge-factor c that allows me to adjust the fitted k parameter by a constant amount:

$$k_{\text{adjust}} = c \cdot k_{\text{fit}} \tag{C.42}$$

The adjusted model then uses the parameter k_{adjust} instead of k_{fit}. All of the model's other parameters remain constant. Note that for $c > 1$, inter-firm dispersion is *decreased* (relative to the original model). For $c < 1$, inter-firm dispersion is *increased*. I choose the value c so that the adjusted model produces the best match to US data. Model results for $c = 0.5$ are shown in Figure C.14 in the same format as the original model was presented in Figure 4.4. By increasing inter-firm dispersion, we significantly improve the fit of the model to the body of the US distribution of income. Note that the adjusted model's Gini index is significantly higher than in the original model, and now better matches US data. Results in the tail remain virtually unchanged. (This is expected, since hierarchy shapes the tail).

C.6.1   Hierarchical Redistribution with the Adjusted Model

Because the hierarchy model tends to underestimate US income inequality, the hierarchical redistribution model tends to be shifted to the left relative to US empirical data (see Fig. 4.18). As shown in Figure C.15, by using the adjusted model to calculate hierarchical redistribution, this problem disappears. Note,
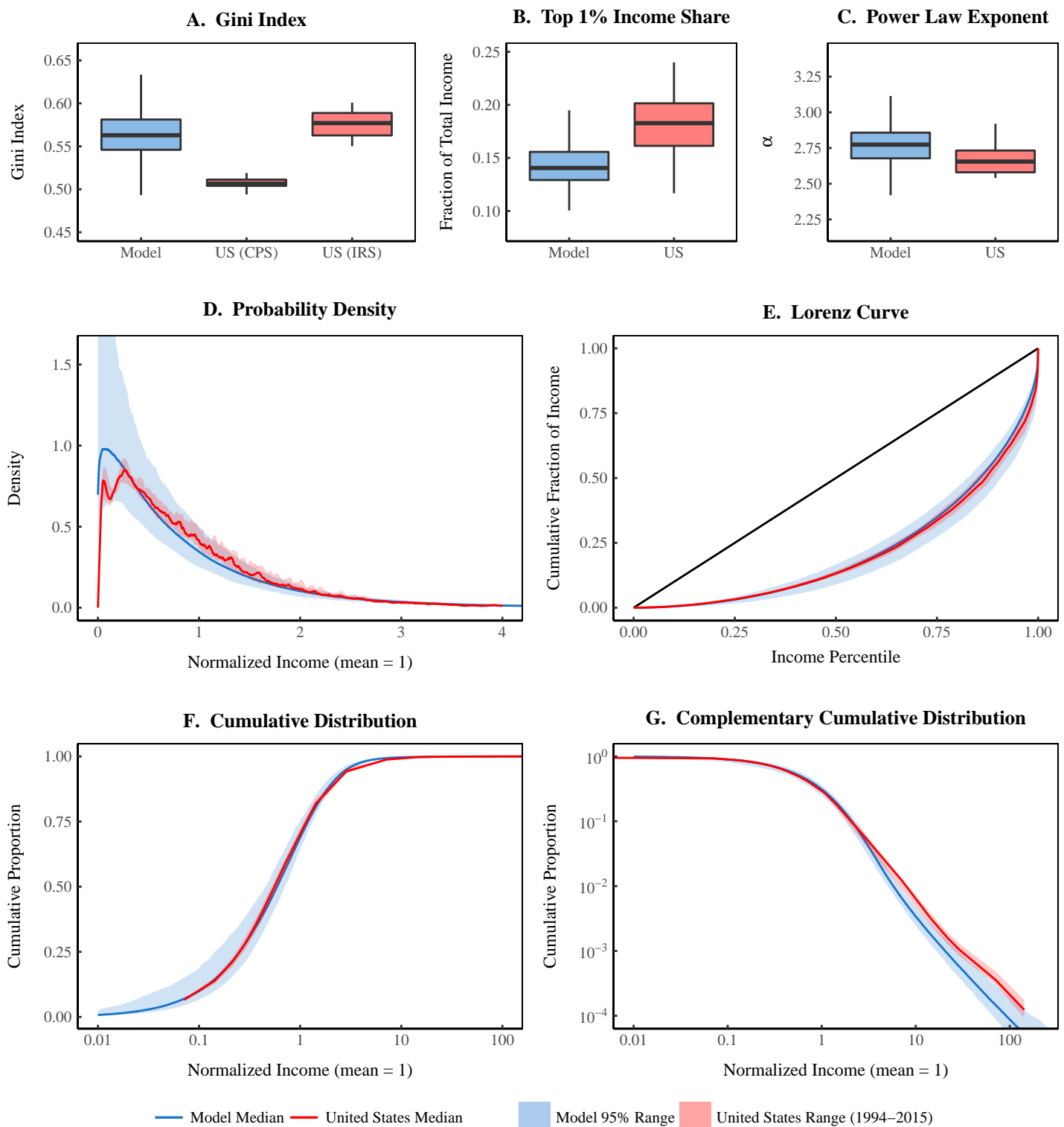
Figure C.14: Adjusted Model Income Distribution vs. US Data

This figure compares various aspects of the *adjusted* model's income distribution to US data over the years 1992-2015. The adjusted model has increased inter-firm income dispersion relative to the original model. Panel A shows the Gini index, with two different US sources — the Current Population Survey (CPS) and the Internal Revenue Service (IRS). Panel B shows the top 1% income share, using data from 17 different time series. Panel C shows the results of fitting a power law distribution to the top 1% of incomes (where α is the scaling exponent). Panel D plots the income density curve with mean income normalized to 1 (using data from the CPS). Panels E, F, and G use IRS data to construct the Lorenz curve, cumulative distribution, and complementary cumulative distribution (respectively). The cumulative distribution shows the proportion of individuals with income *less* than the given x value. The complementary cumulative distribution shows the proportion of individuals with income *greater* than the given x value. Note the log scale on the x-axis for these last two plots. For sources and methods, see Appendix C.1.
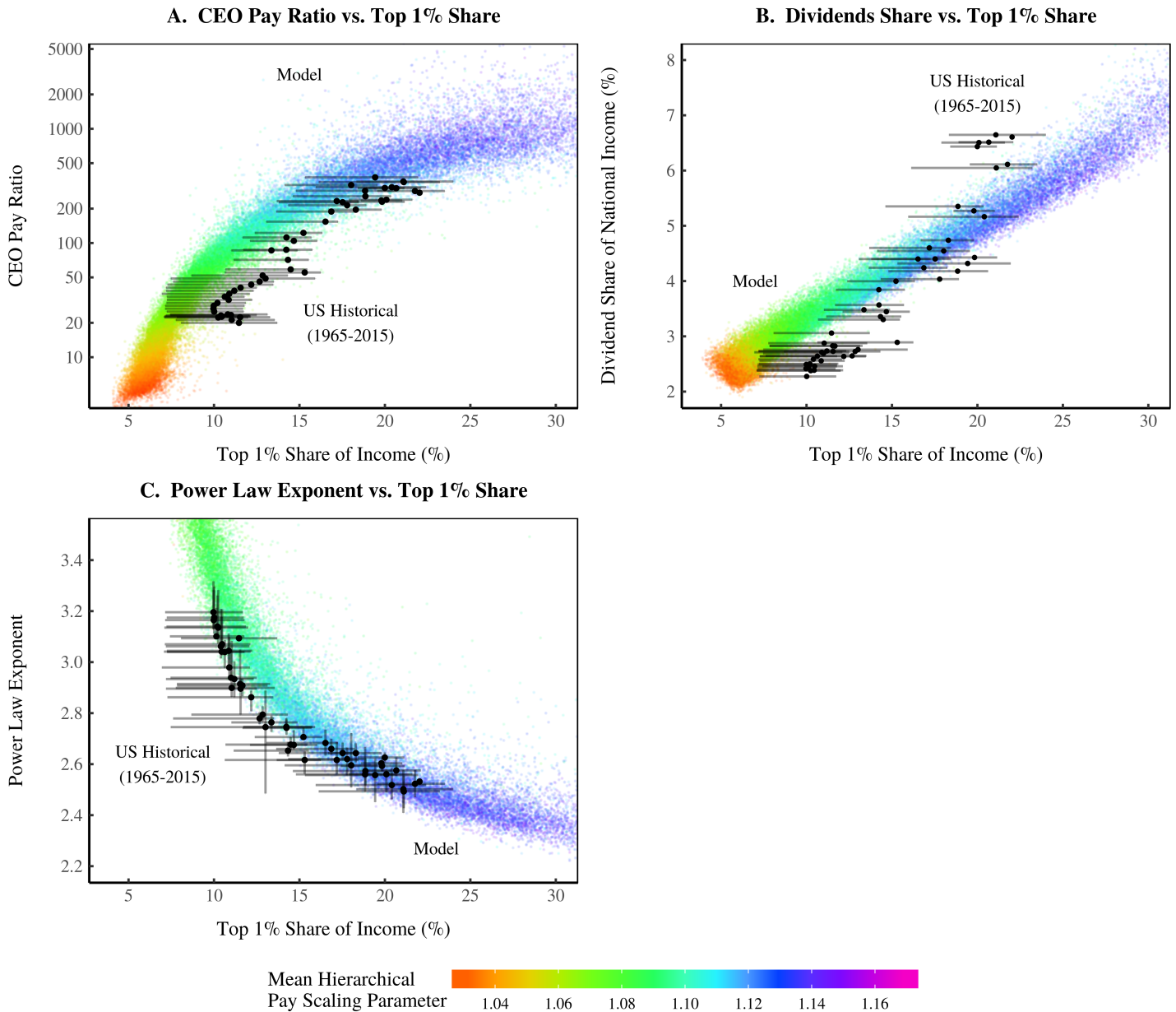
Figure C.15: The Adjusted Hierarchical Redistribution Model vs. US Data

This figure compares adjusted model results to historical trends in US income distribution. The adjusted model has increased inter-firm income dispersion relative to the original model. Model results are produced by varying the hierarchical pay-scaling parameter, indicated by color. Each colored point represents a single model iteration. US empirical data is shown in black, with horizontal error lines indicating the range of 17 different estimates for the top 1% income share. The point indicates the median of these estimates. Panel A plots the CEO pay ratio against the top 1% share, while panel B plots the dividend share of national income against the top 1% share. Panel C plots the fitted power law exponent of the top 1% of incomes against the top 1% income share. For sources and methods, see Appendix C.1.

however, that increasing inter-firm dispersion does not fix the model/empirical discrepancy in the slope of the dividends vs. top 1% relation (Fig. C.15B).

C.6.2   Adjusting the Capitalist Gradient Model

The capitalist gradient model is built on the following functional relation between hierarchical power (P) and capitalist income fraction (K_{frac}):

$$K_{frac} = 0.05 \ln(P) \tag{C.43}$$

Recall that hierarchical power is defined as the number of subordinates $+ 1$. All individuals with no subordinates therefore have hierarchical power $P = 0$. Since $\ln(1) = 0$, all these individuals will have *exactly* zero capitalist income. By convention, income distribution is usually only tabulated for non-zero incomes. Thus these individuals are excluded.

In the adjusted capitalist gradient model, I introduce an adjustment to the capitalist income fraction equation:

$$K_{frac} = 0.05 \ln(P) + \epsilon \tag{C.44}$$

Here ϵ is a constant very close to zero. Its effect is only felt when $P = 1$. Instead of getting $K_{frac} = 0$, we get $K_{frac} = \epsilon$. What does this do? It effectively endows individuals who previously had zero capitalist income with a tiny amount of capitalist income (a few dollars). The effect may seem insignificant, but it has an important impact on the capitalist income distribution. As shown in Figure C.16, the adjusted model better matches the US data.
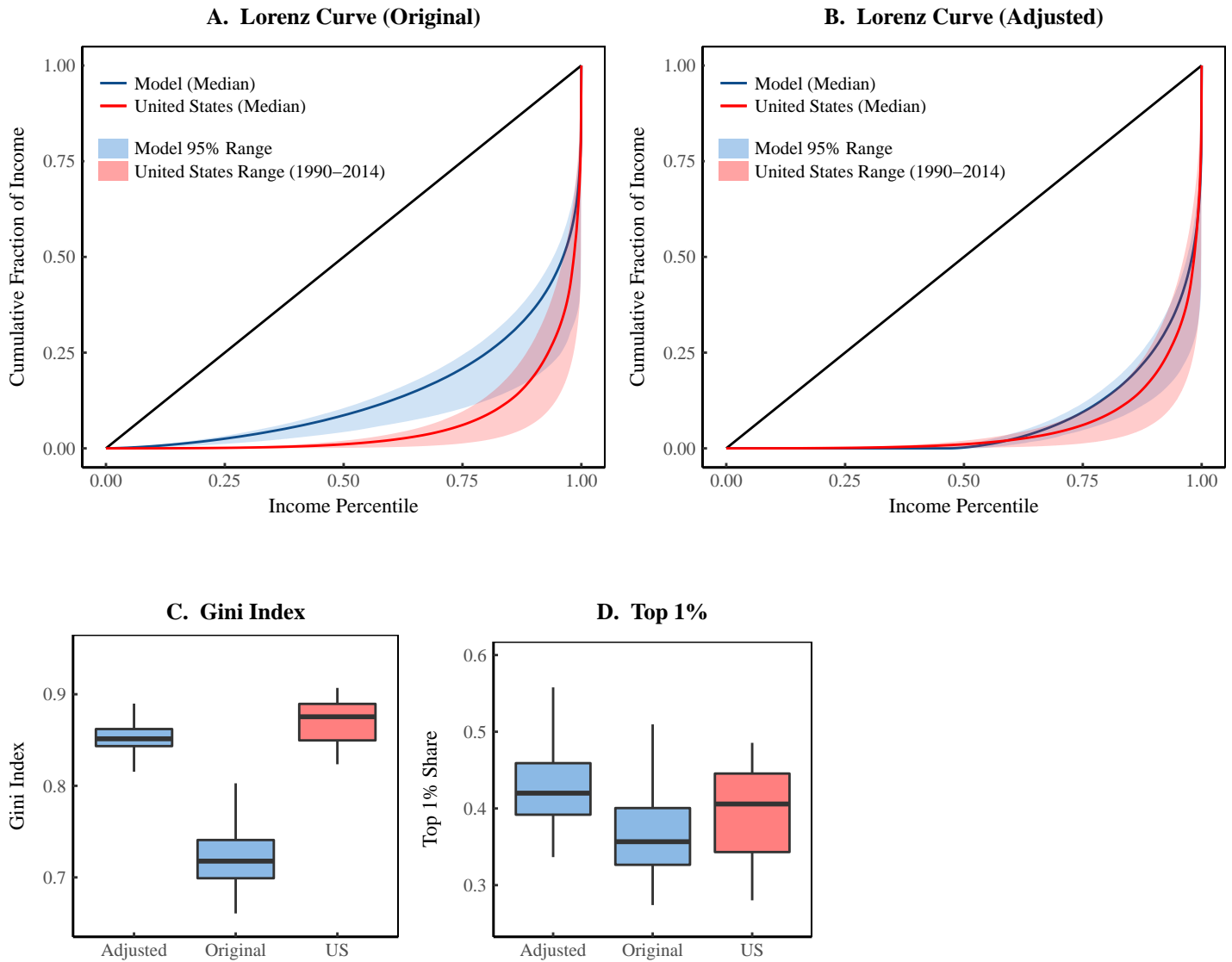
Figure C.16: The Adjusted Capitalist Gradient Model vs. US Data

This figure shows the results of an adjusted capitalist gradient model. The adjusted model allows individuals with a hierarchical power of 1 to have a small capitalist income. This significantly changes the model's Lorenz curve and Gini index. Panel A shows the original capitalist gradient model's Lorenz curve plotted against US data. Panel B shows the adjusted capitalist gradient model's Lorenz curve. Panel C compares the Gini indexes of the original and adjusted models to US data. Panel D shows the top 1% income share. US data is from the IPUMS database. See Appendix C.1.

C.7   A Null Effect Model for Top Incomes and Firm Size

One of the predictions of the hierarchy model is that top incomes should be concentrated at the top of large institutions. To test this prediction, I look at the size distribution of firms associated with top incomes. Here I develop a null-effect model. This model is what we would expect to find if there is absolutely no relation between firm membership and income. In the null-effect case, we should find that the size distribution of firms associated with *top earners* is exactly the same as the size distribution of firms associated with the general population.

To determine the null-effect we must find the size distribution of firms associated with the general population. Before doing so, some clarification is in order. What we are talking about is the size distribution of firms associated with *individuals*. As shown in Figure C.17, this is quite different from the firm size distribution. To determine the firm size distribution, each firm is counted *once*. However, when we map firm size to individuals, each firm is weighted by the number of individuals within it. When we do this, we are really looking at the distribution of *employment* by firm size. So what is this distribution? Let's find out.

If we randomly select an individual from the private sector population, let $p(i_x)$ be the probability that this individual is associated with a firm of size x. This probability will determine the size distribution of firms associated with a random sample of individuals. Let $p(x)$ be the probability of randomly selecting a firm of size x from the firm population. Using Figure C.17 for guidance, we can see that $p(i_x)$ is given by:

$$p(i_x) \sim x \cdot p(x) \tag{C.45}$$

If we know $p(x)$ — the probability distribution of firms — we can use Eq. C.45 to predict the firm size distribution associated with a random sample of individuals. Let's do so for the United States. The US firm size distribution can be approximated by the power law distribution $p(x) \sim x^{-2}$ (see Appendix C.5). Substituting this into Eq. C.45 gives:

$$p(i_x) \sim x^{-1} \tag{C.46}$$

Because firm sizes generally span many orders of magnitude, it is more convenient to look at the log transformation of Eq. C.46. Therefore, we want to
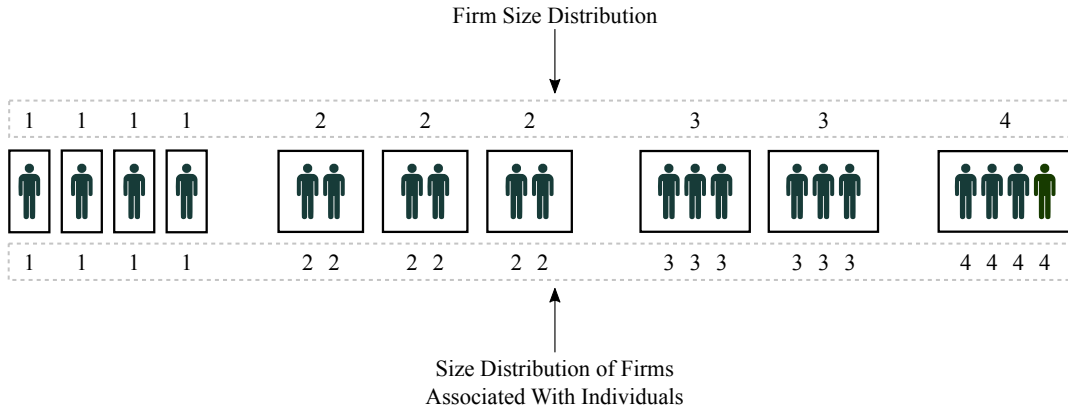
Figure C.17: Mapping Firm Sizes to Individuals

This figure illustrates the mapping of firm size to individuals. Each box represents a firm, with size indicated above. The mapping of firm size to individuals appears below each firm. Let $p(x)$ be the probability of randomly selecting a *firm* of size x from the firm population. Let $p(i_x)$ be the probability of randomly selecting an *individual* associated with a firm of size x (from th individual population). Noting that each firm size x appears x times in the individual-to-firm mapping, we can state that $p(i_x) \propto x \cdot p(x)$.

know the probability density for $p(\ln i_x)$. To find this, we use the standard change-of-variable function for a probability density:

$$f_y = f_x\big(x(y) \big) \cdot \big| x'(y) \big| \tag{C.47}$$

We let $f_y = p(\ln i_x)$ and $f_x = c \cdot x^{-1}$ (where c is constant). The transformation function is $y = \ln x$. We then note that $x(y) = e^y$ and $x'(y) = e^y$. Substituting into Eq. C.47 gives:

$$f_y = c \cdot (e^y)^{-1} \cdot e^y = c \tag{C.48}$$

Since $f_y = p(\ln i_x)$, we can state that $p(\ln i_x) = c$, the uniform distribution. If we randomly draw a sample of individuals from the US private sector, we predict that their associated firm size distribution will be *log-uniform*. This is the null-effect. If there is absolutely no relation between income and firm membership, we should find that the size distribution of firms associated with top incomes (in the US) is log-uniformly distributed.

C.8 How Hierarchy Generates the Power-Law Tail

Although the hierarchy model is not tuned to do so, it reproduces (with good accuracy) the power-law scaling of top US incomes. What is the mechanism at work here? It turns out that the basic mechanism was theorized by Lydall [23] in the late 1950s (and then largely ignored thereafter). It relies on the two contrapuntal exponential tendencies of hierarchical organization: (1) the share of employment tends to *decrease* exponentially with hierarchical rank; (2) income tends to *increase* exponentially with rank. These two opposing tendencies interact to produce a power law distribution of income (in the tail).

This mechanism is a specific case of a more general method. A power law will be created any time we *exponentially* transform an *exponential* distribution [24]. The generative mechanism works as follows. Suppose we have some quantity y that is exponentially distributed (here a is a negative constant):

$$p(y) \sim e^{ay} \tag{C.49}$$

In the case of hierarchical class structure, this would be the probability of finding someone with a hierarchical rank y. What causes employment to be distributed (approximately) exponentially by rank? It is a generic result of branching hierarchical structure, in which each superior has control over multiple subordinates. If the span of control is constant, employment will *decrease* exponentially with rank as one moves up the hierarchy. See Figure 4.1 for an idealized picture.

Suppose that we have another variable, x, that is *also* exponentially related to y:

$$x = e^{by} \tag{C.50}$$

In the context of hierarchical organization, x would be income, which *increases* exponentially with rank. Why does income have this scaling behavior? Herbert Simon suggests that it results from social norms [25]. My own view is that it is caused by the power asymmetries that are innate to hierarchical organization [26]. Hierarchical power (measured by the number of subordinates) tends to increase exponentially with rank. If income is a function of hierarchical power, then it too should increase exponentially with rank.

Moving on with our derivation, the question we want to know is this: how

is income x distributed? To find out, we use the change of variable formula to get f_x, the density function of x:

$$f_x = f_y\big(\, y(x)\,\big) \cdot \big|\, y'(x)\,\big| \tag{C.51}$$

We let $f_y = e^{ay}$. Since $x = e^{by}$, we note that $y(x) = \frac{1}{b}\ln x$ and $y'(x) = 1/bx$. Substituting into the change of variable formula gives:

$$f_x \;=\; e^{\frac{a}{b}\ln x} \cdot \frac{1}{bx} \;=\; \frac{1}{b}x^{a/b-1} \tag{C.52}$$

Thus the variate x (income) has a power law distribution with exponent $\alpha = a/b - 1$. A caveat here is that the derivation assumes that both x and y are continuous. If y represents rank, then it will be a discrete variable. This will result in a non-continuous distribution of x.

To reiterate, hierarchical organization creates a power law distribution because of two contrapuntal, exponential tendencies: (1) employment tends to *decrease* exponentially with rank; and (2) income tends to *increase* exponentially with rank. Figure C.18 highlights this contrapuntal behavior in the hierarchy model. As expected, the hierarchical employment distribution has a bottom-heavy pyramid shape (Fig. C.18A). The vast majority of people work in low ranks, and only a tiny elite occupy top positions. The inset panel highlights the *exponential* nature of the employment distribution. Here, the logarithm of hierarchical employment share is plotted on the y-axis, against rank on the x-axis. With this log transformation, a pure exponential function will appear as a straight line.

Figure C.18B shows the model's hierarchical *pay* structure. To make comparison easy, I have normalized all income so that the base-level income is equal to one. As expected, hierarchical pay has an inverted pyramid shape. Average income at the top of the hierarchy dwarfs (by several orders of magnitude) that at the bottom. To highlight the exponential nature of this relation, the inset plot shows the logarithm of income plotted against rank. Again, a pure exponential function will appear as a straight line.

Note that neither relative employment nor pay has a *purely* exponential relation with rank. This is a design feature of the model, stemming from case study evidence. In this data, income tends to *increase* supra-exponentially (faster than an exponential) with rank. Conversely, employment tends to *decrease* supra-exponentially with rank (see Appendix C.2 and C.4 for details). In any case, when we combine these two supra-exponential tendencies, the result still seems
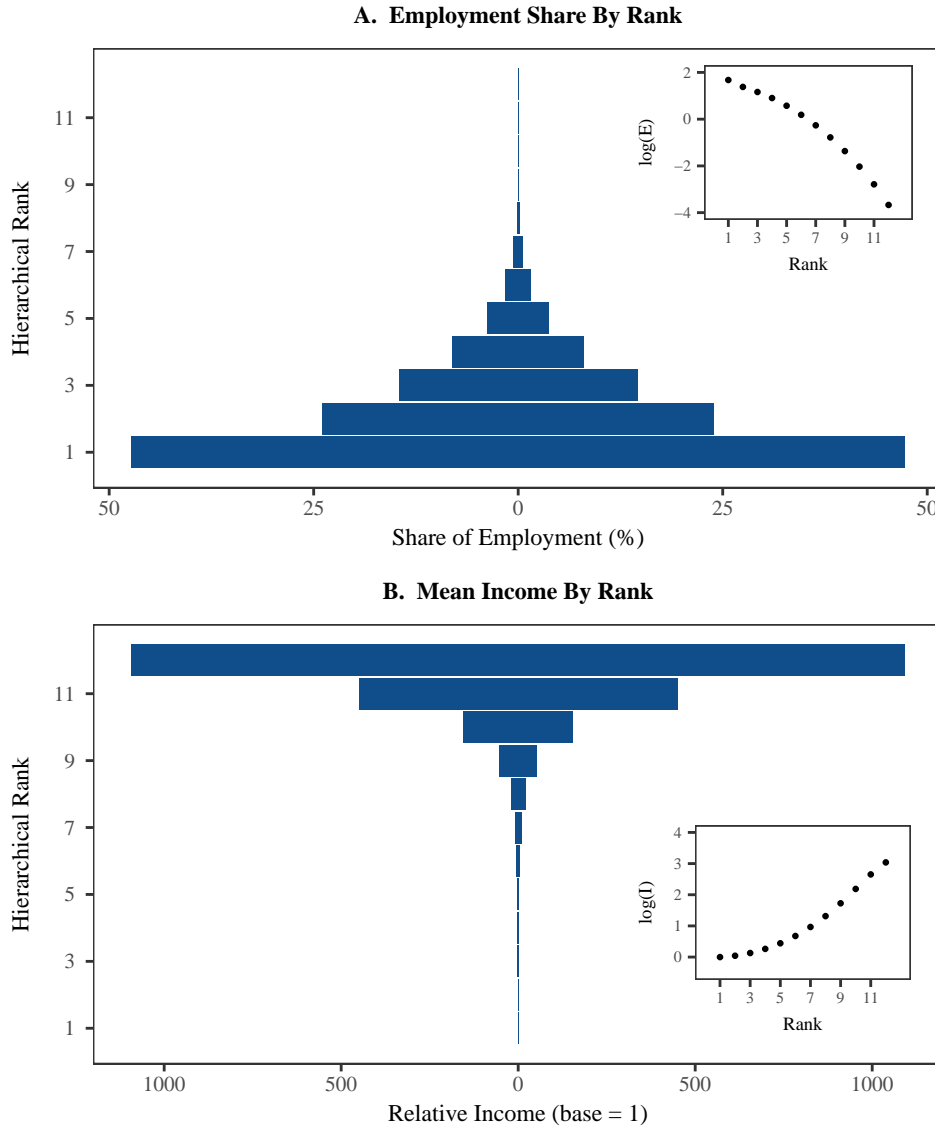
A. Employment Share By Rank

B. Mean Income By Rank

Figure C.18: The Hierarchy Model's Contrapuntal Exponential Tendencies

This figure shows the two contrapuntal exponential tendencies associated with the hierarchy model's class structure. Panel A shows the model's aggregate distribution of employment by hierarchical rank. The bottom-heavy shape results from firm's hierarchical structure (in conjunction with the firm size distribution). The inset graph shows the logarithm of employment share $\log(E)$, plotted against rank. The curved relation indicates that employment declines with rank slightly faster than an exponential function. Panel B shows the model's mean pay by hierarchical rank (normalized so that the base level $=1$). The inset graph shows the logarithm of income $\log I$ against rank. The curved relation indicates that income increases with rank slightly faster than an exponential function.
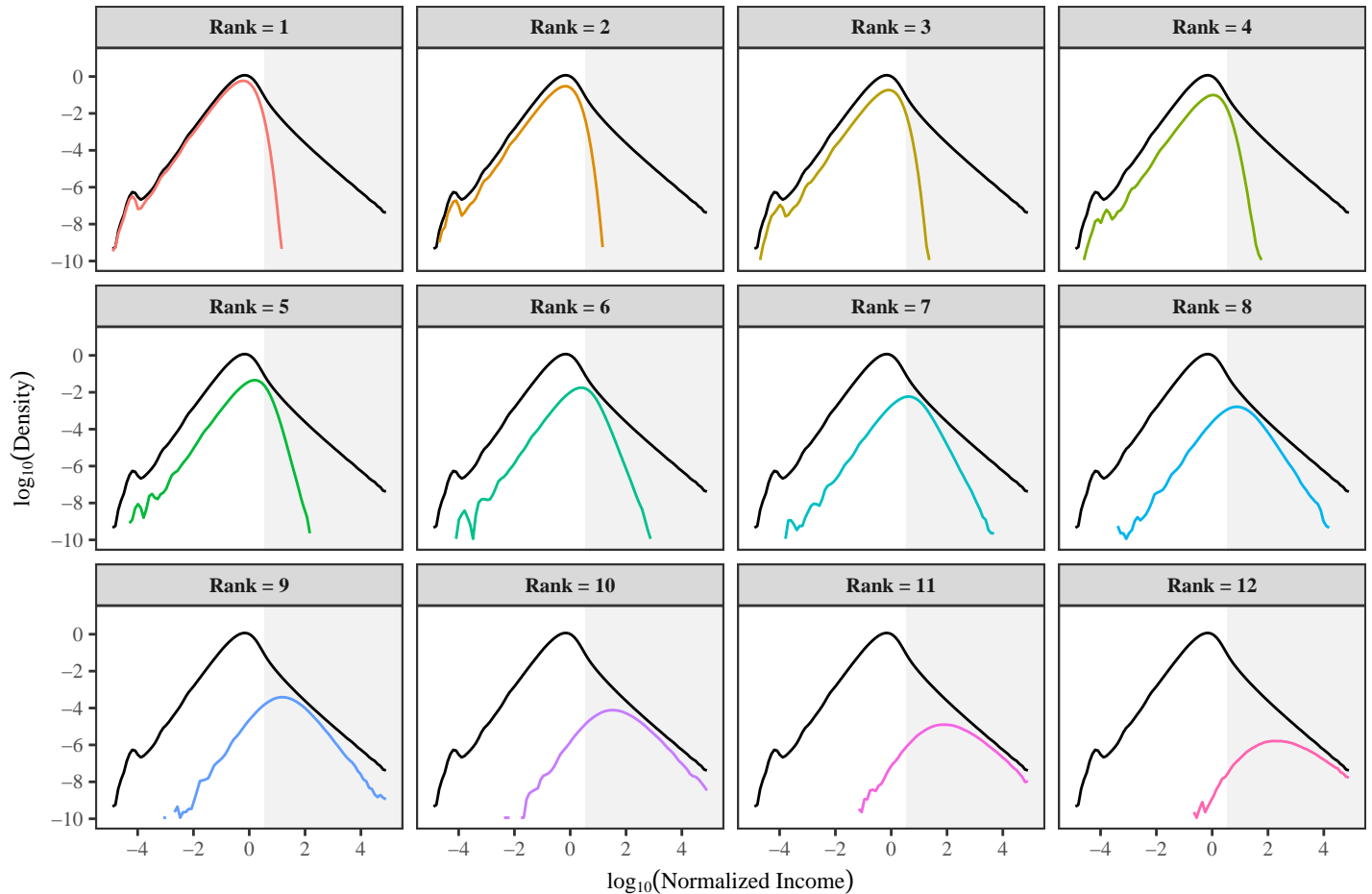
Figure C.19: Hierarchical Class Structure and the Distribution of Income

This figure shows the distribution of income for each hierarchical rank in the hierarchy model. To clearly show the distribution tail, I have used a log-log transformation. In each panel, a rank-specific income distribution is shown in color. For comparison, I also show the model's aggregate income distribution (black). The shaded region indicates the top 1% of incomes (in the aggregate model distribution). To interpret this plot, look at how closely each rank-specific distribution comes to the aggregate distribution. The closer the two are, the greater the rank's contribution to income distribution at that point. The power law right tail (evident as the straight line in the aggregate distribution) is jointly created by ranks five and up.

to be (roughly) a power law distribution of income in the model's tail. (I have not, as yet, worked out how this happens.)

To get a better picture of how this process works, we turn to Figure C.19. Here I show how the model's hierarchical class structure creates the tail of the income distribution. Each panel shows the distribution of income of a *specific* hierarchical rank. Note that I use a log-log transformation — this allows us to

better see the tail of the distribution. To allow comparison, every panel also shows the model's aggregate income distribution. How do we interpret this plot? Look at how closely each rank-based distribution comes to the main distribution. Where the two are close, it indicates that the particular hierarchical rank contributes a great deal to the distribution of income at that point. To get a sense for where the tail of the income distribution is located, I have shaded the top 1% of incomes (in the aggregate model distribution). We can see that the tail of the distribution is created by ranks 5 and above.

The take-home message here is that hierarchical class structure can serve as a generative mechanism for creating the well-recognized Pareto scaling of top incomes.

References

1. Current Population Survey. Annual Social and Economic Supplement; 2017.

2. Virkar Y, Clauset A. Power-law distributions in binned empirical data. The Annals of Applied Statistics. 2014;8(1):89–119.

3. Alvaredo F, Atkinson AB, Piketty T, Saez E, Zucman G. The world wealth and income database. Website: http://www wid world. 2016;.

4. Alvaredo F, Atkinson T, Piketty T, Saez E. The World Top Incomes Database; 2013. Available from: http://topincomes.parisschoolofeconomics.eu/#.

5. Atkinson A, Lakner C. Capital and labor: the factor income composition of top incomes in the United States, 1962-2006. World Bank Policy Research Working Paper. 2017;8268.

6. Piketty T. Capital in the Twenty-first Century. Cambridge: Harvard University Press; 2014.

7. Jenkins SP, Burkhauser RV, Feng S, Larrimore J. Measuring inequality using censored data: a multiple-imputation approach to estimation and inference. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2011;174(1):63–81.

8. Graf M, Nedyalkova D. GB2: Generalized Beta Distribution of the Second Kind: properties, likelihood, estimation. R package version. 2012;1.

9. Mishel L, Schieder J. Stock market headwinds meant less generous year for some CEOs: CEO pay remains up 46.5% since 2009. Washington, D.C.: Economic Policy Institute; 2016. 109799. Available from: http://www.epi.org/publication/ceo-and-worker-pay-in-2015/.

10. Audas R, Barmby T, Treble J. Luck, effort, and reward in an organizational hierarchy. Journal of Labor Economics. 2004;22(2):379–395.

11. Baker G, Gibbs M, Holmstrom B. Hierarchies and compensation: A case study. European Economic Review. 1993;37(2-3):366–378.

12. Dohmen TJ, Kriechel B, Pfann GA. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. Journal of Population Economics. 2004;17(2):193–228.

13. Grund C. The wage policy of firms: comparative evidence for the US and Germany from personnel data. The International Journal of Human Resource Management. 2005;16(1):104–119.

14. Lima F. Internal Labor Markets: A Case Study. FEUNL Working Paper. 2000;378.

15. Morais F, Kakabadse NK. The Corporate Gini Index (CGI) determinants and advantages: Lessons from a multinational retail company case study. International Journal of Disclosure and Governance. 2014;11(4):380–397.

16. Treble J, Van Gameren E, Bridges S, Barmby T. The internal economics of the firm: further evidence from personnel data. Labour Economics. 2001;8(5):531–552.

17. Axtell RL. Zipf distribution of US firm sizes. Science. 2001;293:1818–1820.

18. Gaffeo E, Gallegati M, Palestrini A. On the size distribution of firms: additional evidence from the G7 countries. Physica A: Statistical Mechanics and its Applications. 2003;324(1–2):117–123. doi:10.1016/S0378-4371(02)01890-3.

19. Gillespie CS. Fitting heavy tailed distributions: the poweRlaw package. arXiv preprint arXiv:14073492. 2014;.

20. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661–703.

21. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: CRC press; 1994.

22. Sanderson C, Curtin R. Armadillo: a template-based C++ library for linear algebra. Journal of Open Source Software. 2016;.

23. Lydall HF. The distribution of employment incomes. Econometrica: Journal of the Econometric Society. 1959;27(1):110–115.

24. Newman ME. Power laws, Pareto distributions and Zipf's law. Contemporary physics. 2005;46(5):323–351.

25. Simon HA. The compensation of executives. Sociometry. 1957;20(1):32–35.

26. Fix B. Evidence for a Power Theory of Personal Income Distribution. Working Papers on Capital as Power. 2017;No. 2017/03:1–125.