

# WORKING PAPERS ON CAPITAL AS POWER

No. 2018/02

## **A Hierarchy Model of Income Distribution**

Blair Fix

April 2018

<http://www.capitalaspower.com/?p=2467>

# A Hierarchy Model of Income Distribution

Blair Fix\*

April 8, 2018

## 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.

---

\*Author contact: [blairfix@gmail.com](mailto:blairfix@gmail.com)

---

**Contents**

|          |                                                               |           |
|----------|---------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                           | <b>3</b>  |
| <b>2</b> | <b>A Hierarchy Model</b>                                      | <b>6</b>  |
| <b>3</b> | <b>A Capitalist Gradient Hypothesis</b>                       | <b>21</b> |
| <b>4</b> | <b>A Hierarchical Redistribution Hypothesis</b>               | <b>35</b> |
| <b>5</b> | <b>Conclusions: Modeling from the Top Down</b>                | <b>43</b> |
| <b>A</b> | <b>Sources and Methods</b>                                    | <b>48</b> |
| <b>B</b> | <b>Hierarchical Structure and Pay within Case-Study Firms</b> | <b>55</b> |
| <b>C</b> | <b>Compustat Data</b>                                         | <b>59</b> |
| <b>D</b> | <b>Hierarchy Model Equations</b>                              | <b>64</b> |
| <b>E</b> | <b>Restricting Parameters</b>                                 | <b>71</b> |
| <b>F</b> | <b>The Adjusted Hierarchy Model</b>                           | <b>83</b> |
| <b>G</b> | <b>A Null Effect Model for Top Incomes and Firm Size</b>      | <b>88</b> |
| <b>H</b> | <b>How Hierarchy Generates the Power-Law Tail</b>             | <b>90</b> |

## 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.<sup>1</sup>

---

<sup>1</sup> 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].

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.

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 hopeless 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 distribution. 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 to replicate (with 75% accuracy) the observed relation between US top income share and the dividend share of national income.

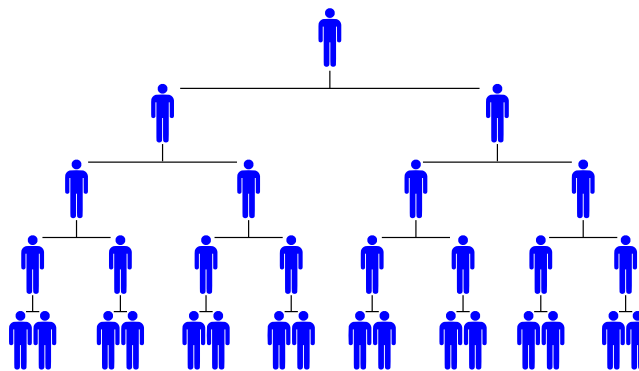
## Paper Layout

The remainder of the paper is divided into three sections. In section 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 3, I investigate if hierarchy can be used to unify the study of personal and functional income distribution. In section 4, I investigate if hierarchy can be used to unify our understanding of historical trends in income inequality.

## 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. 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.



**Figure 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 B).

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 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 B) 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.

## 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 B). 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. 2) :

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

(*inter-hierarchical* dispersion);

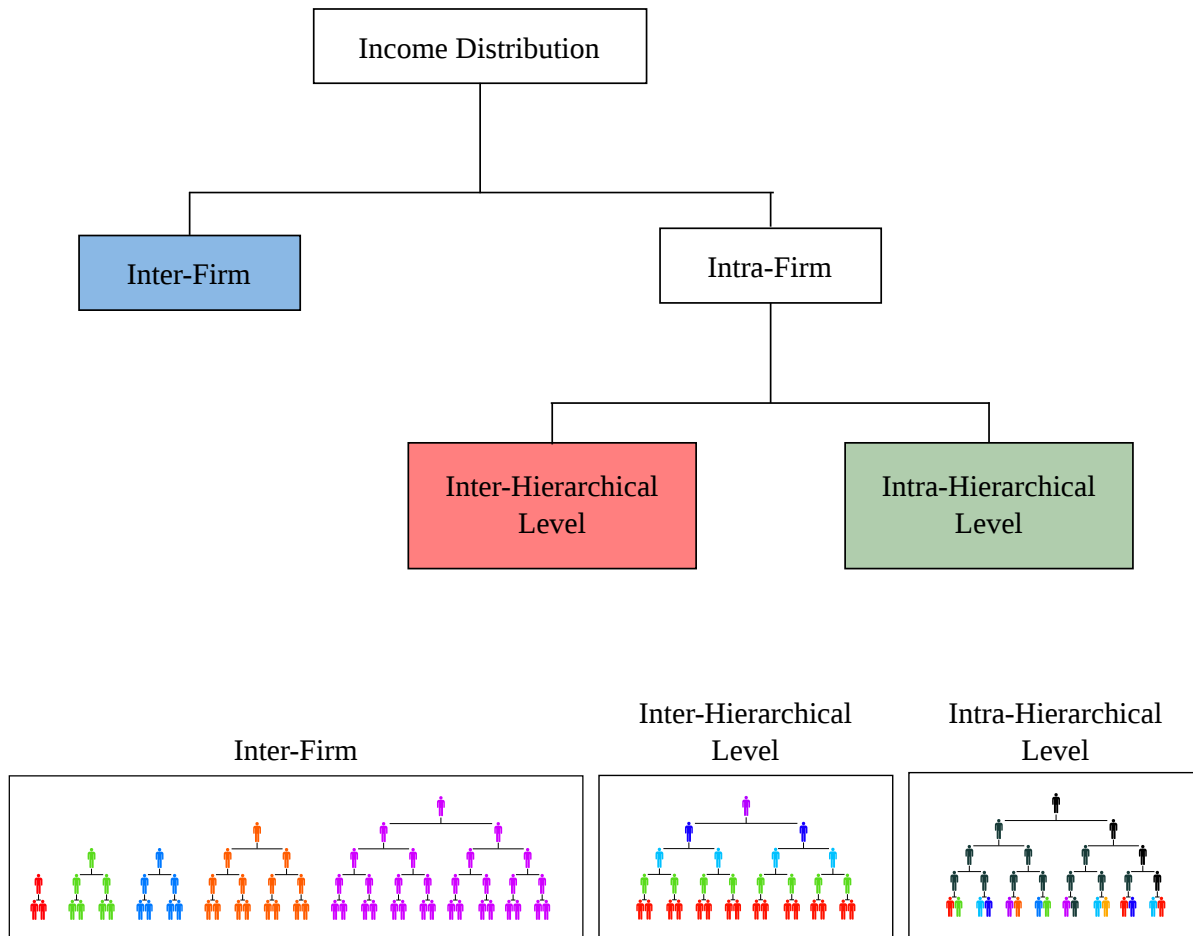
**Source 2:** Income dispersion *within* hierarchical levels of each firm

(*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



**Figure 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.

---

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 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  $\alpha$  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 E).

**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 B). 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 exponentially with hierarchical rank.

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 D and E.

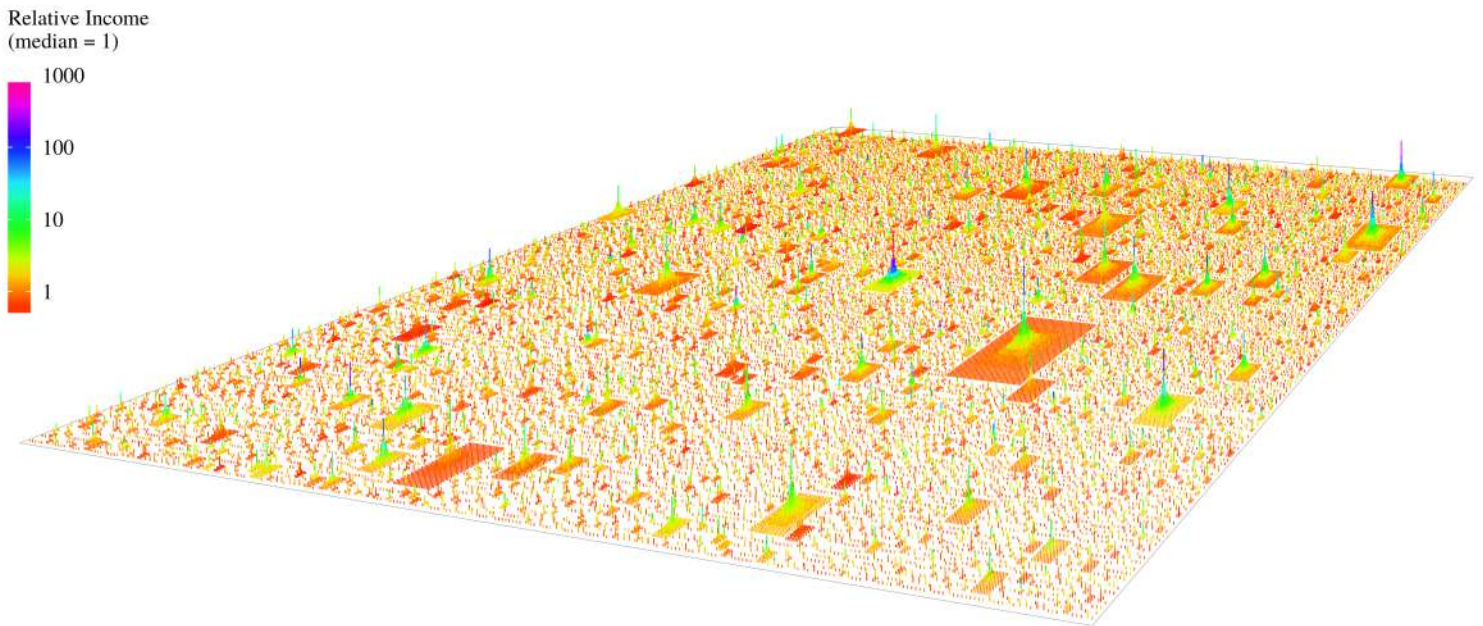
**Step 3: Endow Individuals with Income** After each firm has a hierarchical structure, we begin the most important part of the model, which is to assign every individual an income. Because the model has three dispersion mechanisms, this last step has three components, outlined below.

**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 B). 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 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 D and E.

**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 D and E.

**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 B). 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 parameter-



**Figure 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.

---

ized distribution to randomly generate income dispersion within all hierarchical levels of every firm. For more details, see Appendix [D](#) and [E](#).

### 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 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 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.

## 2.2 Testing the Hierarchy Model (Part 1)

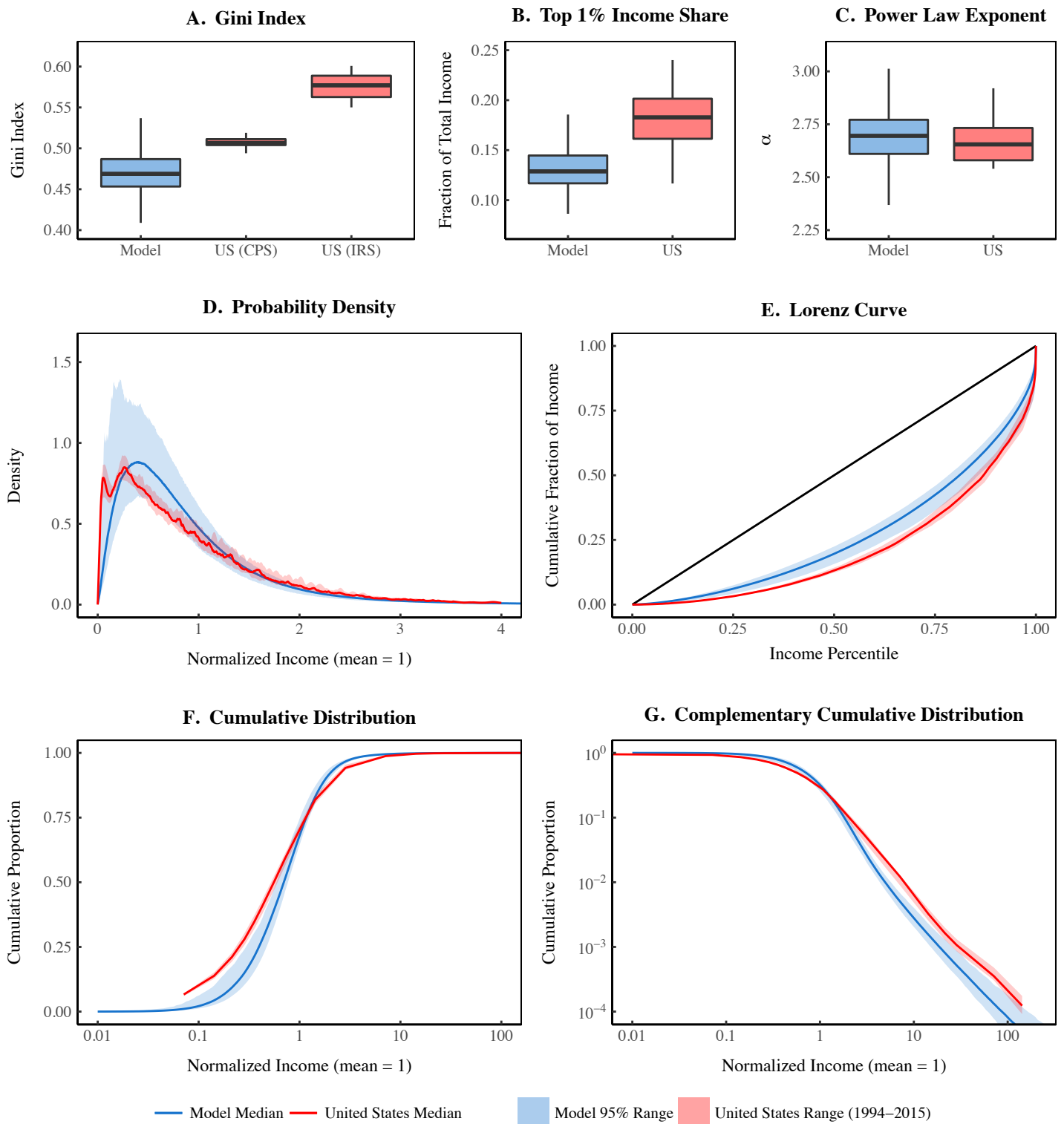
The purpose of the hierarchy model is to study the hierarchical structure of the United States economy. The first step, then, is to make sure that the model produces realistic results. To that end, Figure 4 compares the model's aggregate structure to US empirical data. Even though the model is an extrapolation 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. 4A) and the income share of the top 1% (Fig. 4B). What is the source of this discrepancy? Looking at the income probability density in Figure 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. 4F).

Why does this discrepancy occur? I demonstrate in Appendix F 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.

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: 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  $\alpha$  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 A.

tive distribution (Fig. 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  $\alpha$  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 4C shows the results of fitting a power law to the top 1% of incomes (for methods, see Appendix A). 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.

### 2.3 Testing the Hierarchy Model (Part 2)

When discussing the model visualization shown in Figure 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.

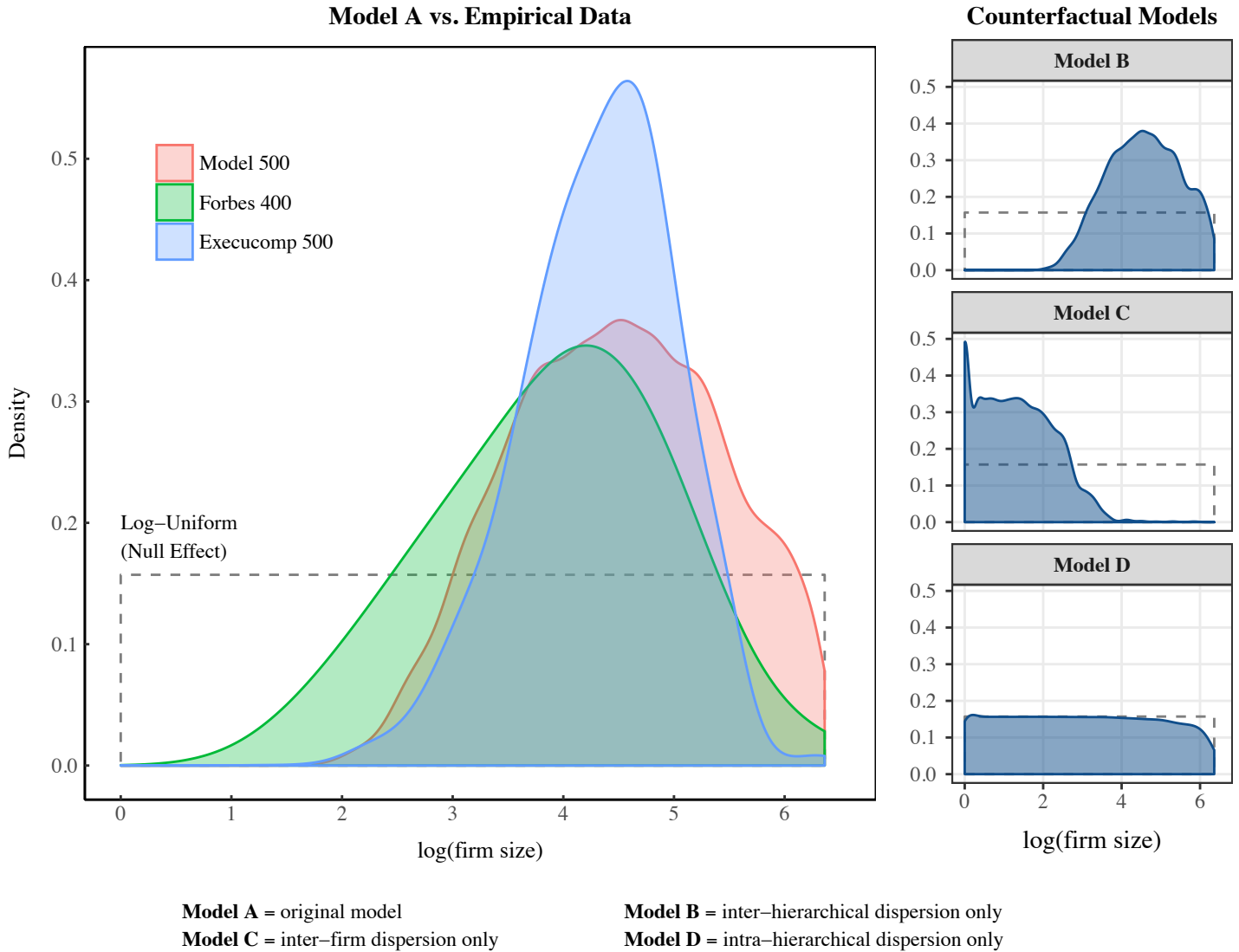
Before discussing the results of this test, it is instructive to know what a *null-effect* would look like. If there is absolutely no relation between income and firm membership, what sort of firm size distribution should be associated with top incomes? It turns out that for the United States, we should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

Results for the Fortune 400 and Execucomp 500 firm size distributions are shown in the main panel of Figure 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 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 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 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 5. This analysis indicates that it is *exclusively* inter-hierarchical income



**Figure 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 A.

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 5, it would seem that these models need rethinking.

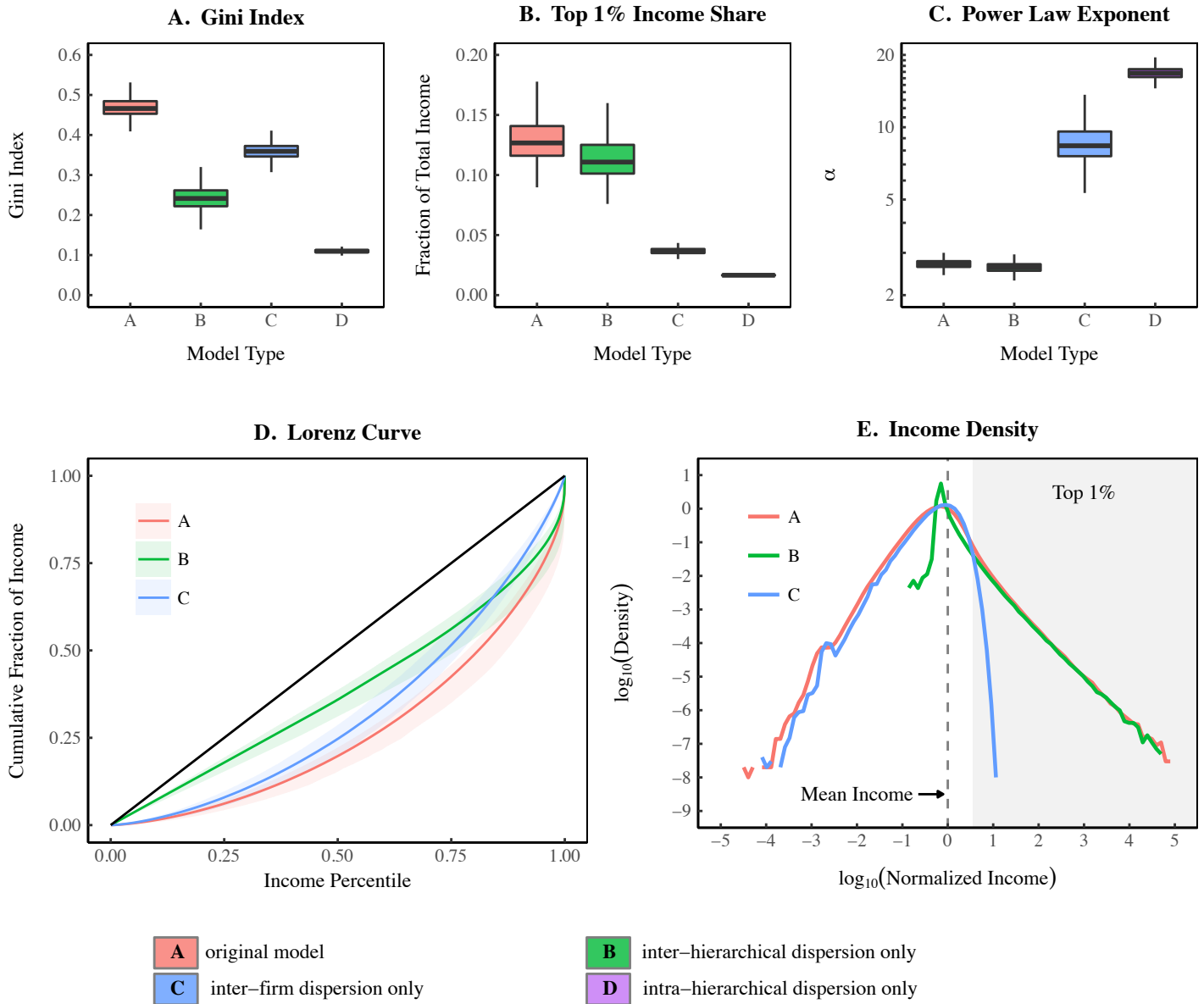
To conclude, the hierarchy model correctly predicts that top paid individuals should be associated with firms that are far larger than those of the general population. Moreover, the model indicates that this effect is purely a result of inter-hierarchical pay dispersion.

## 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 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 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. 6B). Now we find that the model with inter-hierarchical dispersion has inequality that is nearly identical



**Figure 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?<sup>2</sup>

To understand this apparent contradiction, let's look at the Lorenz curves for each model (Fig. 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. 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 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

---

<sup>2</sup> 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 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 H.)

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 E). 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.

### 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.

### 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 with 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.

### 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.<sup>3</sup> Capitalists earn income because their capital is *productive* — it contributes to economic output [75, 76]. This thinking is illustrated in Figure 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 8.

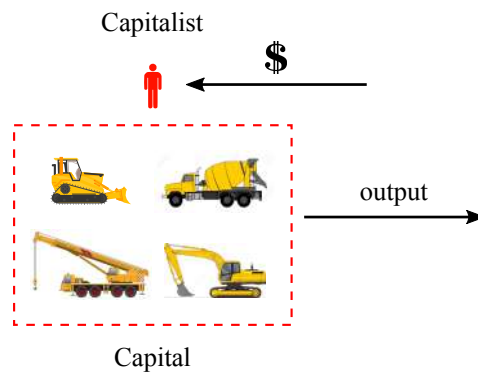
Both neoclassical and Marxist theories of capital keep their eyes firmly on the 'real' sphere of production — on the ownership of *things*. The 'capital as power' approach, proposed by Nitzan and Bichler, is quite different. This approach focuses on ownership as an *institutional* act. What is the difference? 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.

---

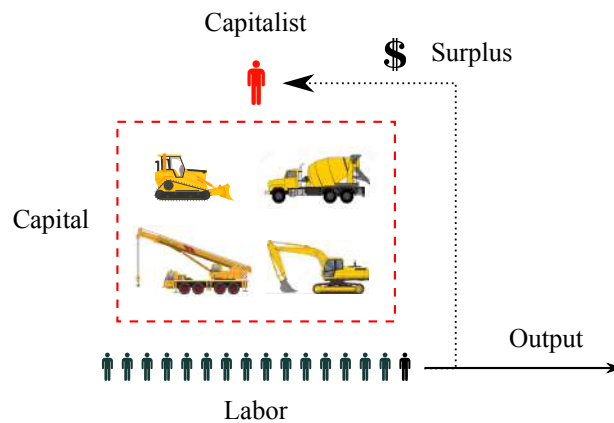
<sup>3</sup> Admittedly, neoclassical economists have significantly broadened their definition of 'capital' over the years. For instance, there is now 'human capital' [65–67], 'knowledge capital' [77, 78], and even 'cultural capital' [79]. However, what has *not* changed is the insistence that 'capital'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





**Figure 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 8: The Marxist Vision of Capitalist Income**

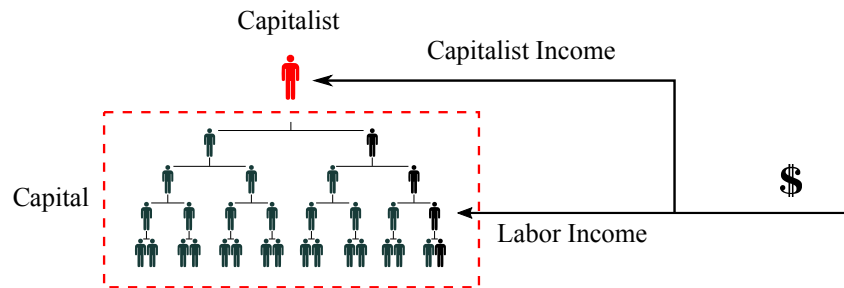
In Marxist theory, capitalists earn income because they own ‘means of production’. Unlike neoclassical theory, Marxists see labor as the source of all value, and capitalists as parasites. Because capitalists control the means of production (capital), they are able to extract a surplus from labor.

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 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 be 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*.

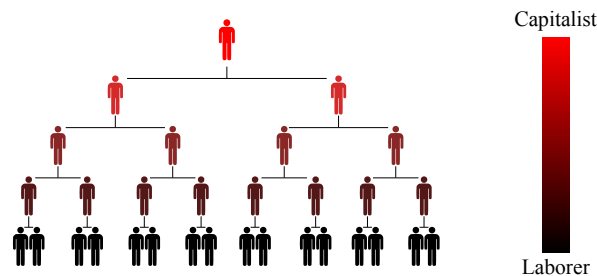
The problem with the separation thesis is that it acknowledges the rise of partial ownership, but insists on a traditional dichotomy between capitalists and laborers. The truth is that the line between being a capitalist and being a laborer has been blurred. Top managers often earn a large portion of 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 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.<sup>4</sup> This means



**Figure 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. 7 and 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 10: A Gradient Vision of Capitalist Income**

This figure shows a more nuanced (than Fig. 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 hierarchical class structure. This is the hypothesis that I test.

### 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:

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

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.

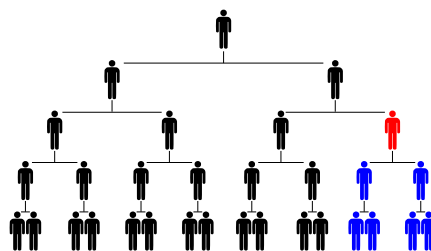


Figure 11: Measuring Hierarchical Power

<sup>4</sup> 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 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.

### 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*.

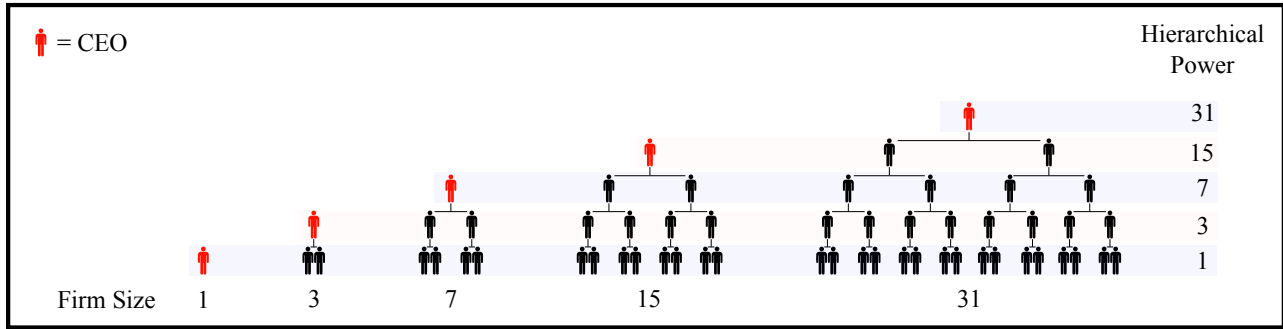
This thinking is visualized in Figure 12. If a firm has  $x$  employees,  $x - 1$  of them will be subordinate to the CEO. Since hierarchal power is defined as the number of subordinates plus one, the CEO's hierarchical power is simply firm size  $x$ . Thus, if we have data for firm size, we automatically 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:

$$\text{CEO Capitalist Income Fraction} = \frac{\text{Income from Stock Options}}{\text{Total Compensation}} \quad (2)$$

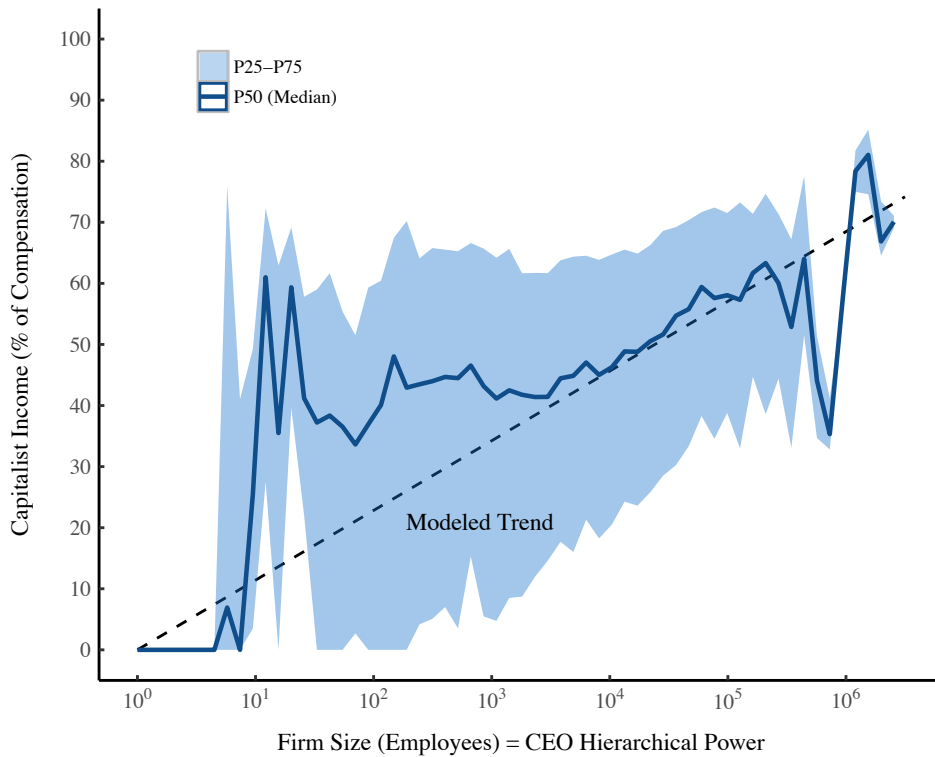
Unlike cash compensation, there are many different ways to value stock options [82–84]. This means that CEO capitalist income fraction has some inherent ambiguity. However, the *nuances* of stock option valuation do not concern me here. Instead, I am interested in *general trends* in CEO compensation. 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.

Figure 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 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. 1), CEOs have hierarchical power equal to firm size.



**Figure 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.

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.

### 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.

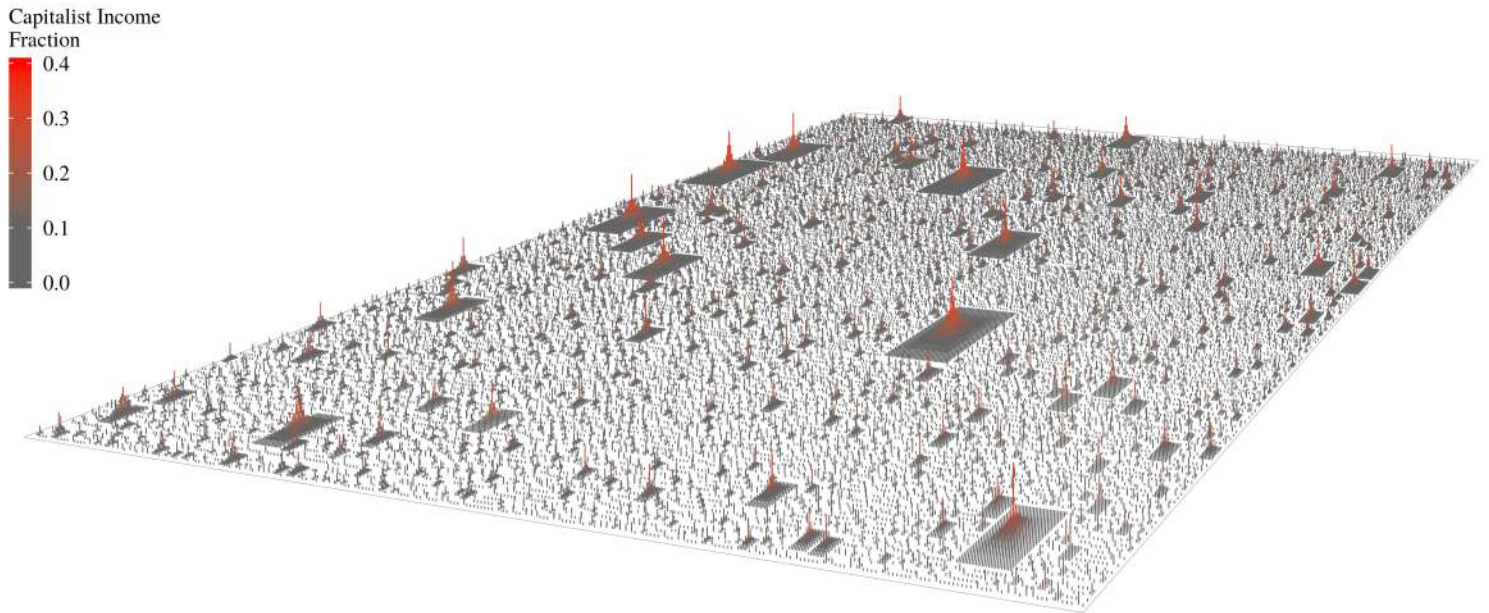
I do this by using the CEO data to create a simple function relating capitalist income fraction to hierarchical power. Once I have this function, I plug it into the hierarchy model and endow each individual with a capitalist income. I then check the model's results against US data. If the model produces 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 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 13. This gives the following function relating capitalist income fraction ( $K_{\text{frac}}$ ) to hierarchical power ( $P$ ):<sup>5</sup>

$$K_{\text{frac}} = 0.05 \ln(P) \quad (3)$$

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

<sup>5</sup> The discerning reader may note that, since a logarithmic function is uniformly increasing, Eq. 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. 3 yields a capitalist income fraction of about 0.7.



**Figure 14: A Landscape View of the Capitalist Gradient Model**

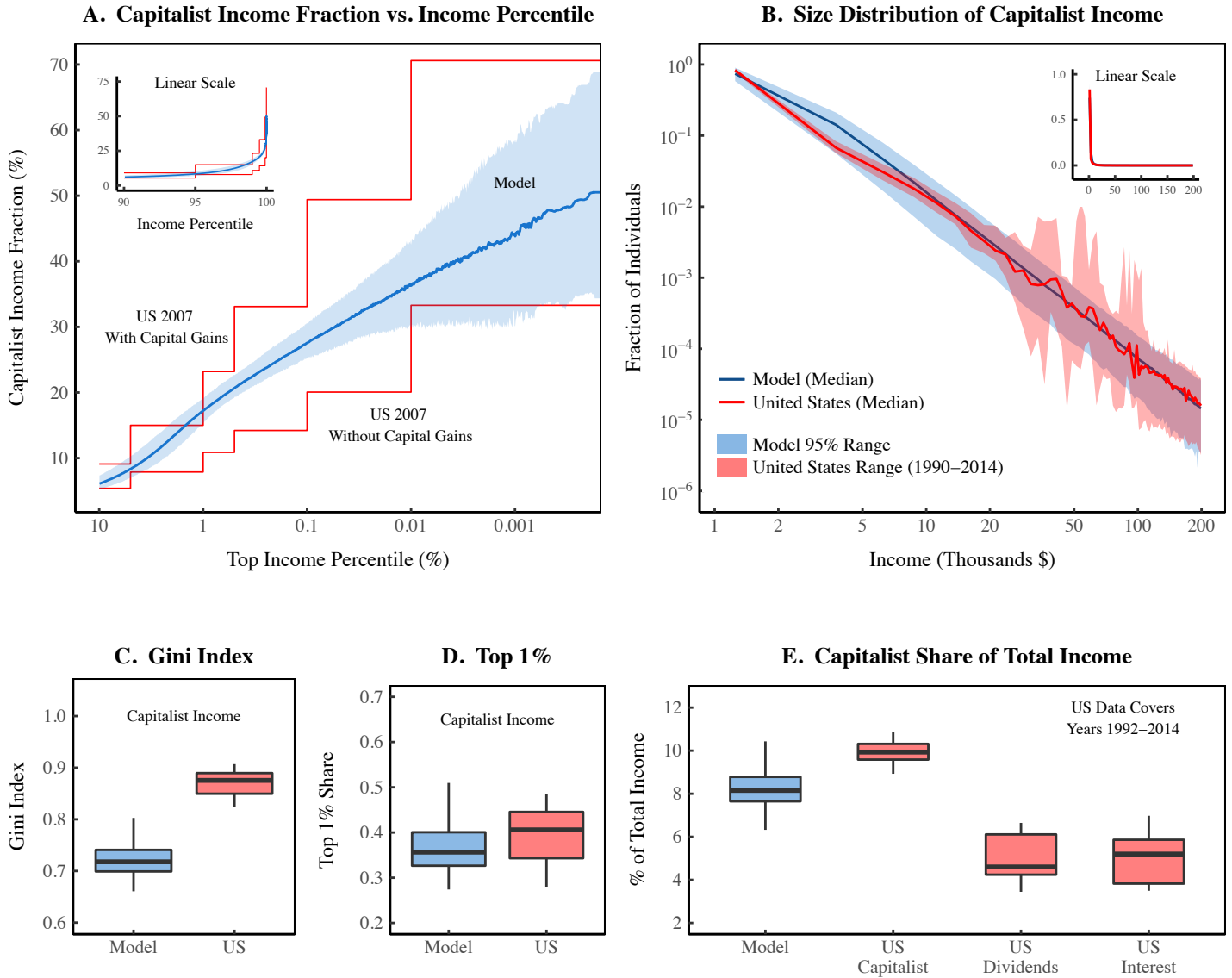
This figure visualizes the capitalist gradient model as a landscape of firms. Each pyramid represents a firm, with size indicating the number of employees. Hierarchical rank is indicated by height, and capitalist income fraction by color.

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 D) and then use Eq. 3 to calculate the capitalist fraction of their income. The resulting capitalist gradient model is visualized in landscape form in Figure 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 produce results that match US data. So does it? Figure 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 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*





**Figure 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 A.

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 15B shows the size distribution of US capitalist income. For this graph (as well as Fig. 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 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 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. 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 F for more details of this adjustment.

Lastly, Figure 15E shows the capitalist share of total (national) income. The model produces a capitalist income share that is slightly lower (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, 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.

### 3.6 Property, Power, and Income

The results shown in Figure 13 and 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 the 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 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.

### 4.1 The Evidence

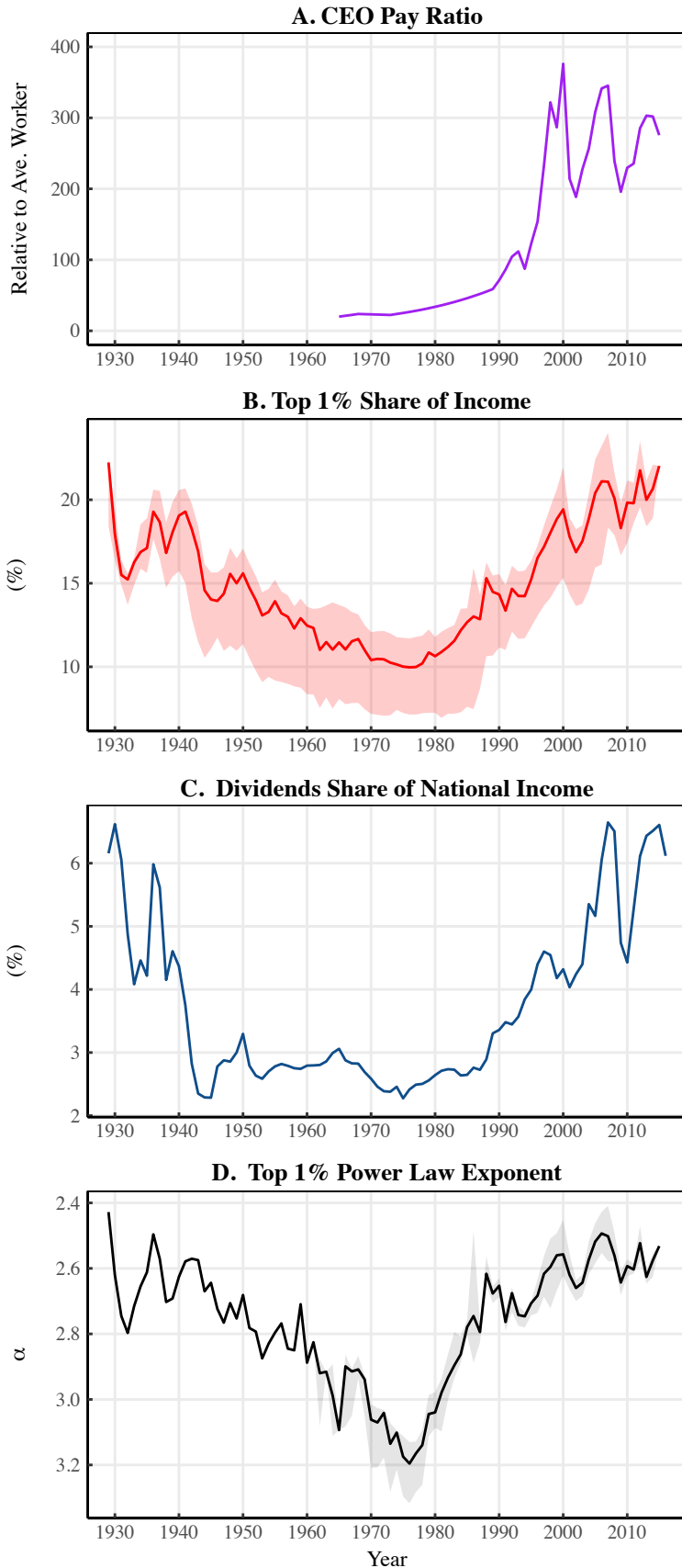
Let’s look at some evidence that hints at hierarchical redistribution. One trend that slaps us in the face is the post-1980 explosion in the CEO pay ratio. As shown in Figure 16A, this explosion corresponds closely with increases in the top 1% income share (Fig. 16B). Assuming that CEOs sit at the top of the corporate hierarchy, the increasing CEO pay ratio suggests that hierarchical redistribution has occurred.<sup>6</sup>

Figure 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 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

---

<sup>6</sup> 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 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 firm’s 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. 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 A.

redistribution of *functional* income. As shown in Figure 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 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.

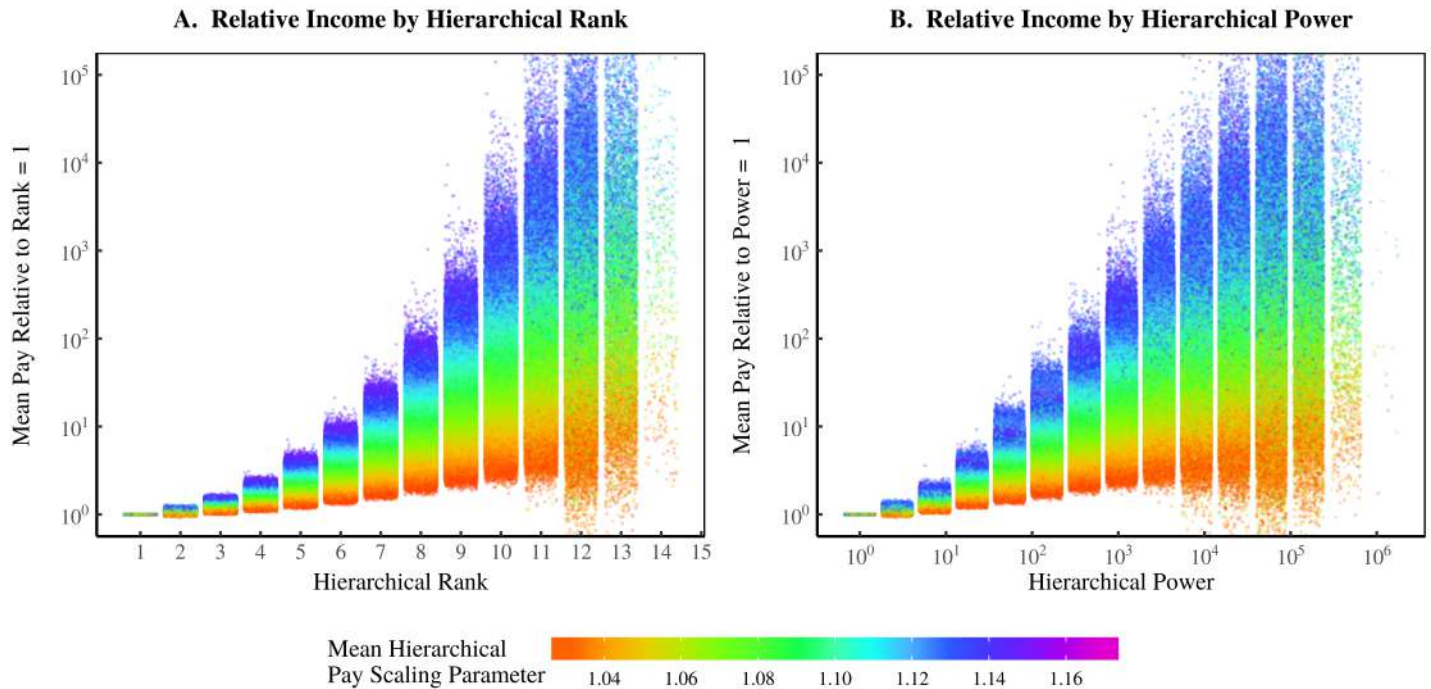
To summarize, the trends in Figure 16 suggest a hierarchical redistribution of income within firms. To test this hypothesis, I use the hierarchy model.

## 4.2 Methods

The idea behind my test of the hierarchical redistribution hypothesis is quite simple. If the trends in Figure 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 E 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 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.



**Figure 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. 3).

To model the data in Figure 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. 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 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). Because the model has no sectors, I use the average pay in the whole model to calculate average worker pay.

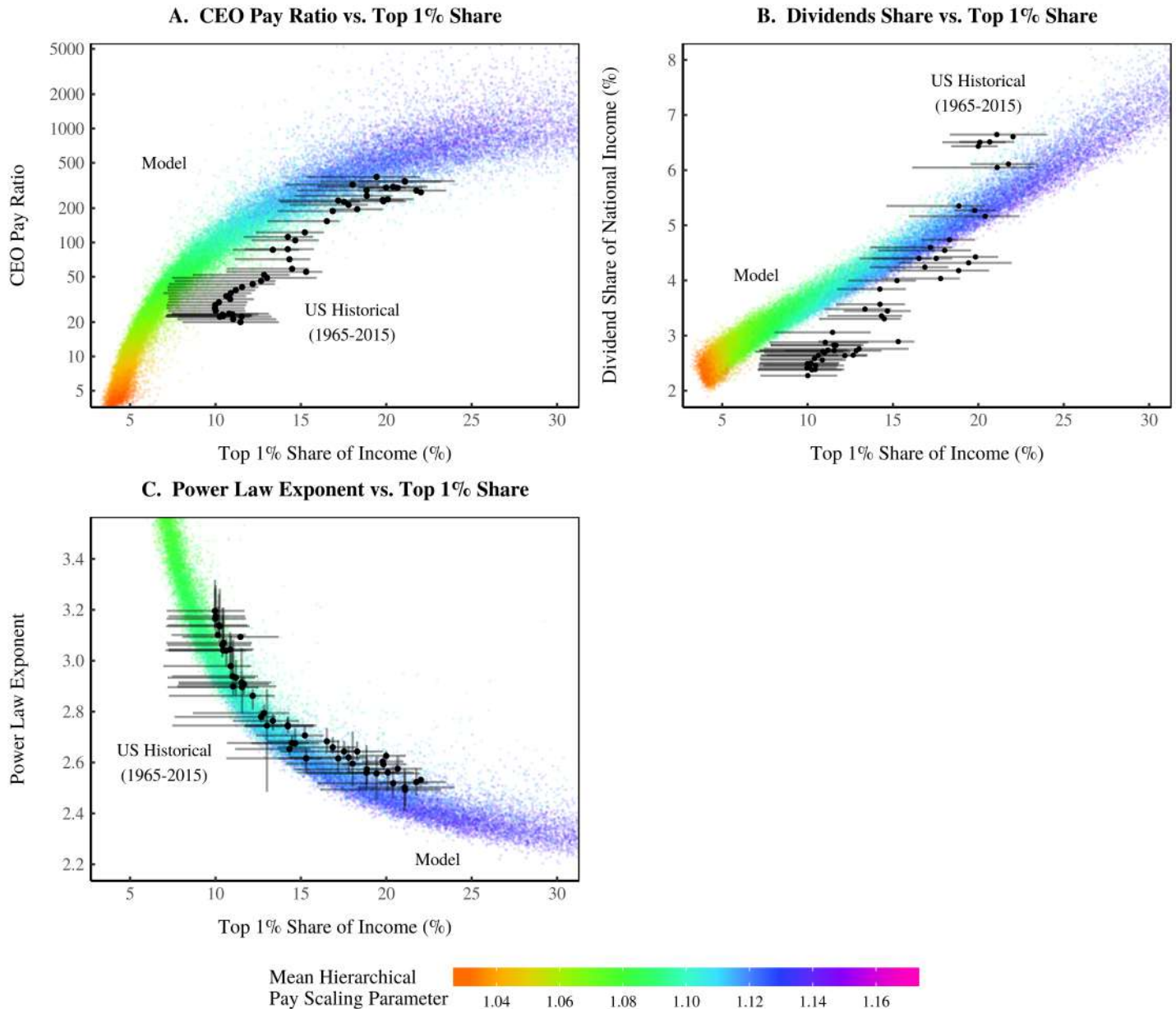
### 4.3 Results

Results of the hierarchical redistribution model are shown in Figure 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 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 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 F, 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 3), there is





**Figure 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 A.

tremendous uncertainty in this relation. More empirical research is needed to understand the source of this model discrepancy.

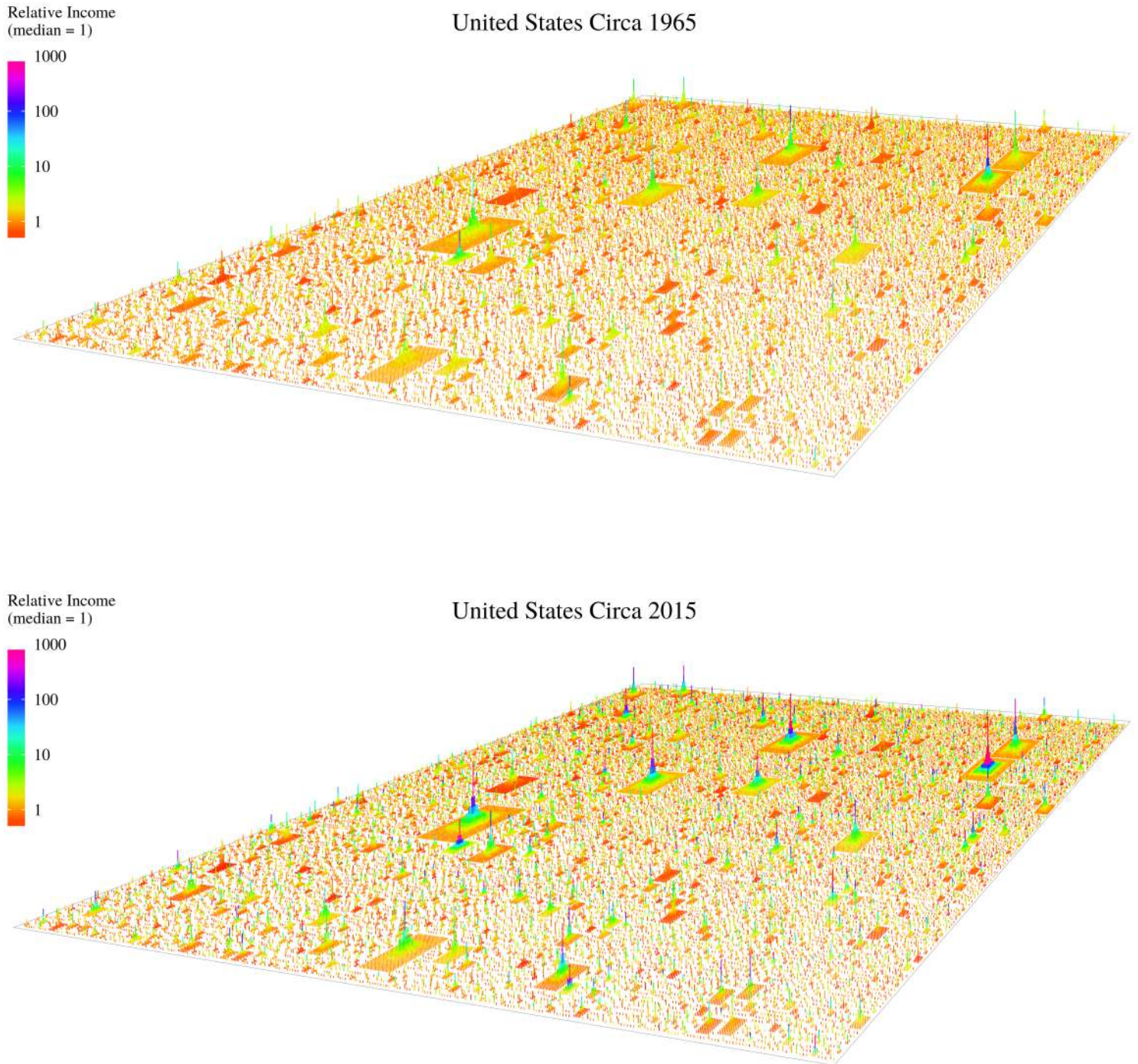
#### 4.4 Discussion

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 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



**Figure 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.

top-ranked individuals. If the model is correct, we can conclude that the US has undergone a massive hierarchical redistribution of income in the last 30 years.

## 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.

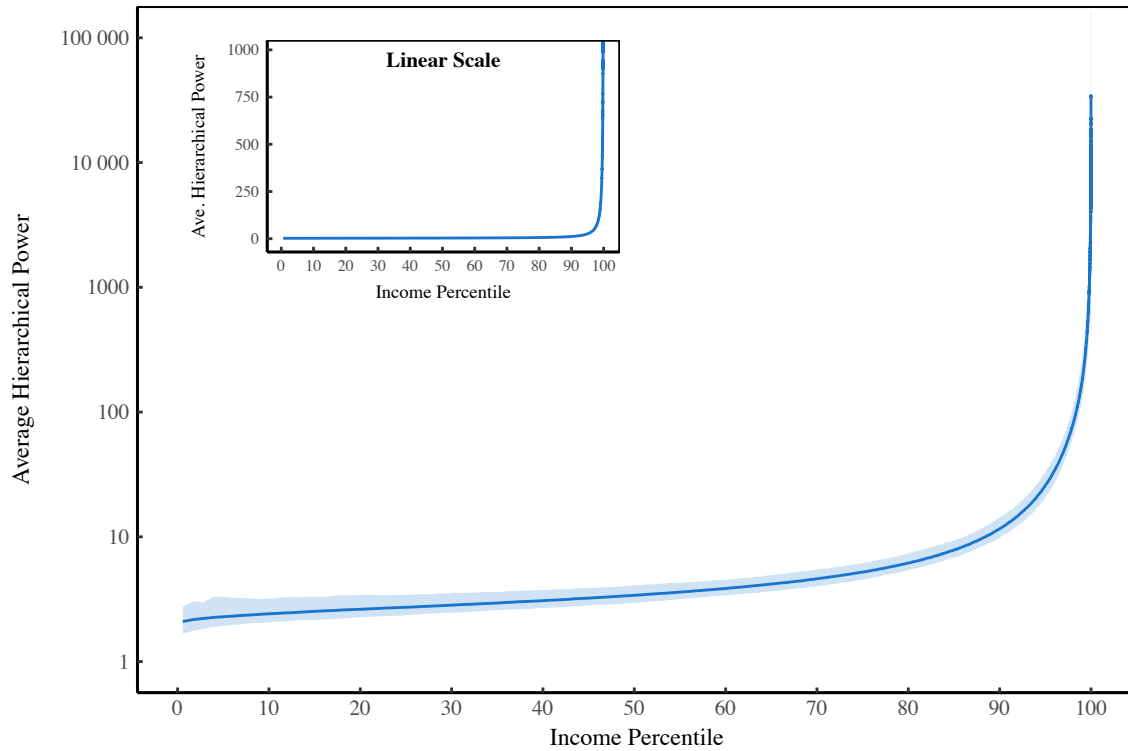
The alternative to the bottom-up approach is to model from the *top down*. What does this mean? Instead of having social structure emerge from the bottom-up actions of individuals, we (the modelers) *impose* structure from the top down. In essence, we impose structure on society and then explore the 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



**Figure 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 indicate 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.

problem” [57]. If the hierarchy model is correct, concentrated power is not an 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 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.

# Appendices

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.



## A Sources and Methods

### Fig. 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” [96].

#### 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 1: Power Law Cutoff Boundaries in US Data**

| Year | Percentile | $\alpha$ |
|------|------------|----------|
| 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 [97]:

$$\mathcal{L} = n(\alpha - 1) \cdot \ln b_{\min} + \sum_{i=\min}^k h_i \ln \left[ b_i^{(1-\alpha)} - b_{i+1}^{(1-\alpha)} \right] \quad (4)$$

Here  $\alpha$  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  $\alpha$  is the value that maximizes the log-likelihood function ( $\mathcal{L}$ ). Since there is no closed-form solution to this maximization problem, I solve for  $\alpha$  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 1.

To find the power law exponent in modeled data, I use the following maximum likelihood estimator:

$$\hat{\alpha} = 1 + n \left[ \sum_i^n \ln \frac{x_i}{x_{\min}} \right]^{-1} \quad (5)$$

Here  $\hat{\alpha}$  is the best-fit power law exponent,  $x_i$  is the  $i$ th 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 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 ( $\mathbf{I}$ ) using bin midpoints. Each midpoint income  $M_i$  is repeated  $F_i$  times, where  $F_i$  is the frequency count for the  $i$ th bin. I then normalize  $\mathbf{I}$  by dividing all elements by the mean income  $\bar{I}$ .

$$\mathbf{I} = \frac{\left( M_1 \cdot \overset{\times F_1}{\dots}, M_2 \cdot \overset{\times F_2}{\dots}, \dots, M_i \cdot \overset{\times F_i}{\dots} \right)}{\bar{I}} \quad (6)$$

Lastly, I fit the simulated income distribution ( $\mathbf{I}$ ) 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 2.

**Table 2: US Top 1% Income Share Sources**

| Series                   | Info                                                                                   | Source |
|--------------------------|----------------------------------------------------------------------------------------|--------|
| sfainc992j               | Pre-tax factor income   equal-split adults   Share   Adults   share of total (ratio)   | [98]   |
| sfainc996i               | Pre-tax factor income   individuals   Share   20 to 64   share of total (ratio)        | [98]   |
| sfainc999i               | Pre-tax factor income   individuals   Share   All Ages   share of total (ratio)        | [98]   |
| sfainc999t               | Pre-tax factor income   tax unit   Share   All Ages   share of total (ratio)           | [98]   |
| sfiinc992j               | Fiscal income   equal-split adults   Share   Adults   share of total (ratio)           | [98]   |
| sfiinc992t               | Fiscal income   tax unit   Share   Adults   share of total (ratio)                     | [98]   |
| sfiinc996i               | Fiscal income   individuals   Share   20 to 64   share of total (ratio)                | [98]   |
| sfiinc999i               | Fiscal income   individuals   Share   All Ages   share of total (ratio)                | [98]   |
| sfiinc999t               | Fiscal income   tax unit   Share   All Ages   share of total (ratio)                   | [98]   |
| sptinc992j               | Pre-tax national income   equal-split adults   Share   Adults   share of total (ratio) | [98]   |
| sptinc996i               | Pre-tax national income   individuals   Share   20 to 64   share of total (ratio)      | [98]   |
| sptinc999i               | Pre-tax national income   individuals   Share   All Ages   share of total (ratio)      | [98]   |
| sptinc999t               | Pre-tax national income   tax unit   Share   All Ages   share of total (ratio)         | [98]   |
| sfiinc_z_US              | World Top Incomes Legacy Series                                                        | [99]   |
| lakner                   | Calculated from micro data                                                             | [100]  |
| piketty_book_no_kgains   | Legacy data from Capital in the 21st Century                                           | [71]   |
| piketty_book_with_kgains | Legacy data from Capital in the 21st Century                                           | [71]   |

### **Fig. 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. 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.

### **Fig. 15: Comparing the Capitalist Gradient Model to US Data**

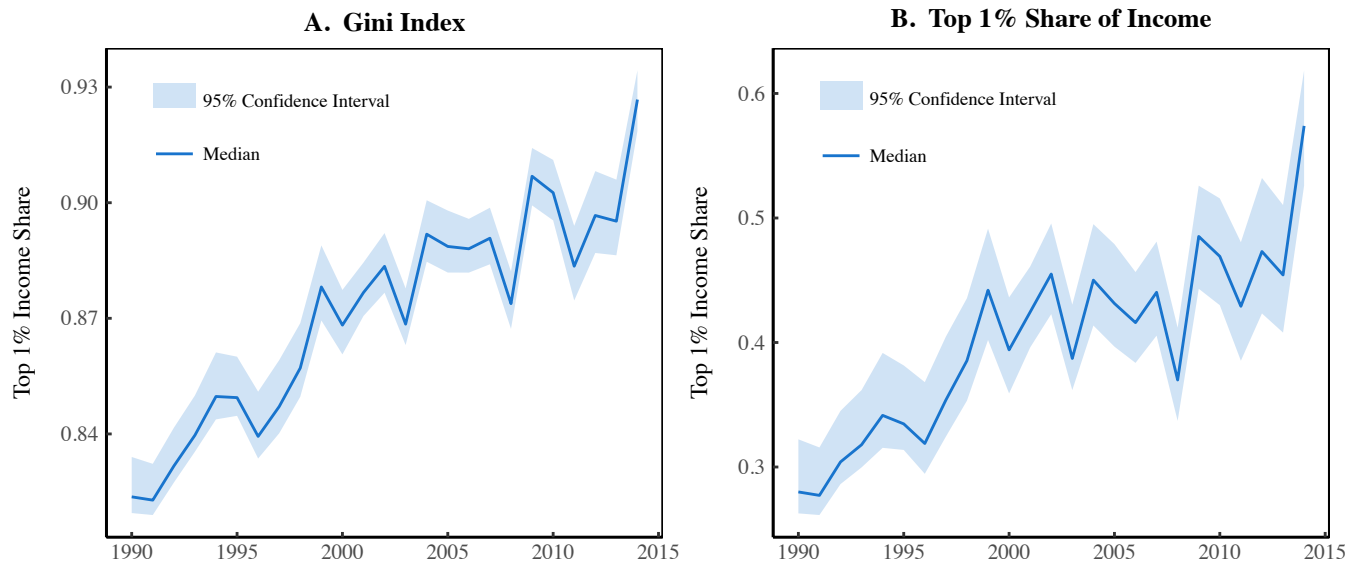
#### **Capitalist Income Fraction vs. Income Percentile**

US data is for the year 2007 and comes from Piketty, Fig. 8.10 [71]. Data is available at [piketty.pse.ens.fr/en/capital21c2](http://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. [101]. 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.



**Figure 21: 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 [102] 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 21 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. 16: Historical Income Distribution Trends in the United States**

CEO pay ratio data comes from Mishel and Schieder [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 firm's respective industry. Top 1% income share data sources are shown in Table 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 [97]. I use income threshold data from the World Wealth and Income Database (see Table 3)

**Table 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   | [98]   |
| tfiinc992j | Fiscal income   equal-split adults   Threshold   Adults   constant 2015 local currency           | [98]   |
| tfiinc992t | Fiscal income   tax unit   Threshold   Adults   constant 2015 local currency                     | [98]   |
| tptinc992j | Pre-tax national income   equal-split adults   Threshold   Adults   constant 2015 local currency | [98]   |

## B 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 neo-classical 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 4 summarizes the case studies used in this paper, while Figure 22 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 22 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 4: Summary of Firm Case Studies**

| Source | Years             | Country        | Firm Levels | Span of Control | Level Income | Level Income Dispersion |
|--------|-------------------|----------------|-------------|-----------------|--------------|-------------------------|
| Audas  | [103] 1992        | Britain        | All         | ✓               | ✓            |                         |
| Baker  | [104] 1969-1985   | United States  | Management  | ✓               | ✓            | ✓                       |
| Dohmen | [105] 1987-1996   | Netherlands    | All         | ✓               | ✓            | ✓                       |
| Grund  | [106] 1995 & 1998 | US and Germany | All         |                 | ✓            | ✓                       |
| Lima   | [107] 1991-1995   | Portugal       | All         | ✓               | ✓            | ✓                       |
| Morais | [108] 2007-2010   | Undisclosed    | All         | ✓               | ✓            |                         |
| Treble | [109] 1989-1994   | Britain        | All         | ✓               | ✓            | ✓                       |

**Notes:** This table shows metadata for the firm case studies displayed in Fig. 23. ‘Firm Levels’ refers to the portion of the firm that is included in the study. ‘Management’ indicates that only management levels were studied.



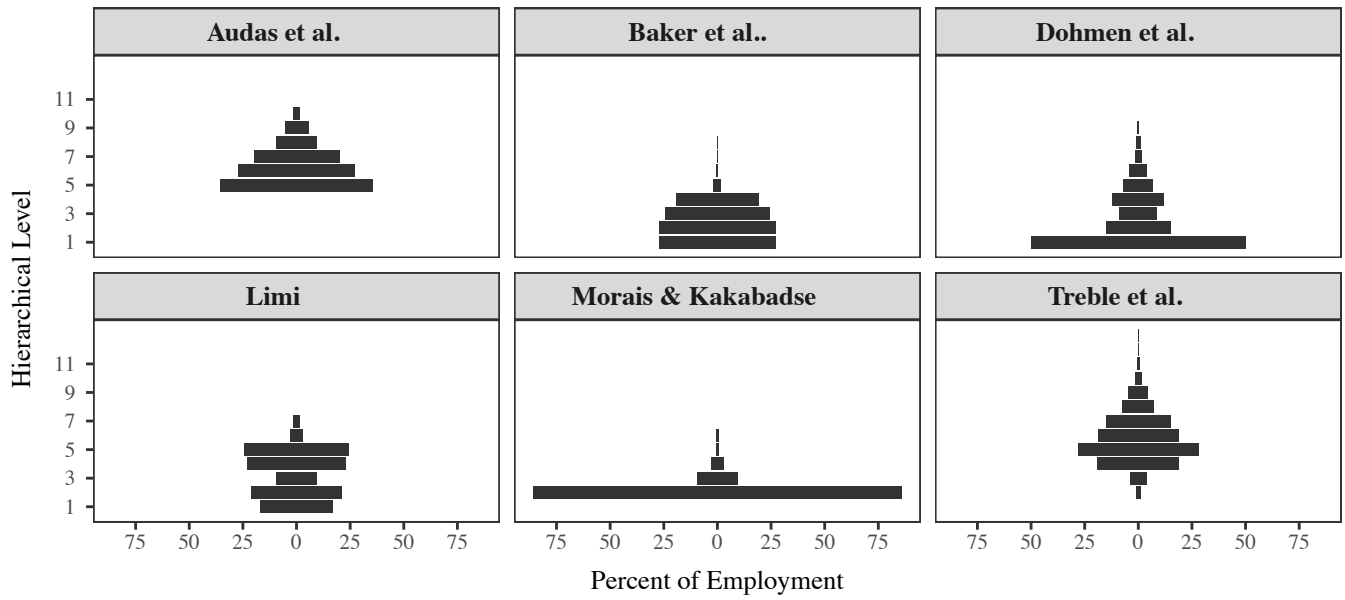
Figure 23 shows data for these metrics for the 6 case study firms. Figure 23A shows how the *span of control* changes as a function of hierarchical level. The data shows unambiguously that the span of control tends to *increase* as one moves up the hierarchy. Figure 23B shows how the *inter-level pay ratio* changes as a function of hierarchical level. Again, this ratio tends to *increase* as one moves up the hierarchy. Figure 23C 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:

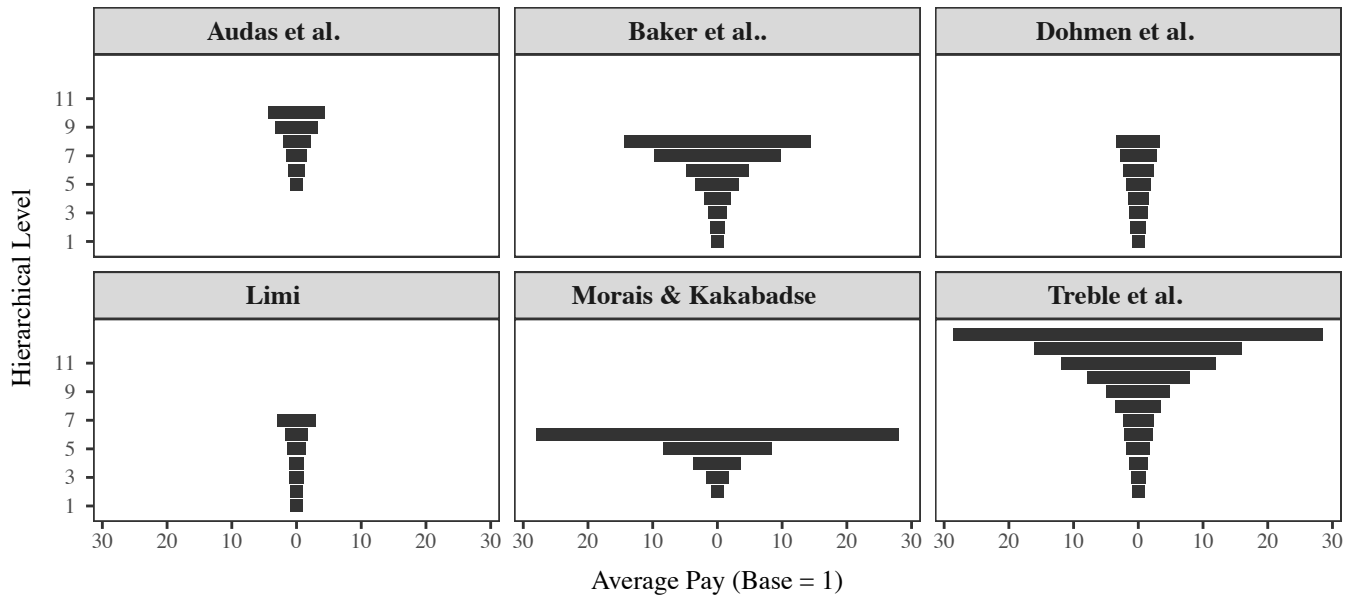
1. The span of control tends to *increase* with hierarchical level.
2. The inter-level pay ratio tends to *increase* with hierarchical level.
3. Intra-level income inequality is approximately *constant* across all hierarchical levels.

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 23A, while parameters for intra-level income dispersion are determined from the mean of data in Figure 23C. For a detailed discussion of the model algorithm and parameter fitting procedure, see Sections D and E.

**A. Firm Hierarchical Employment Structure**

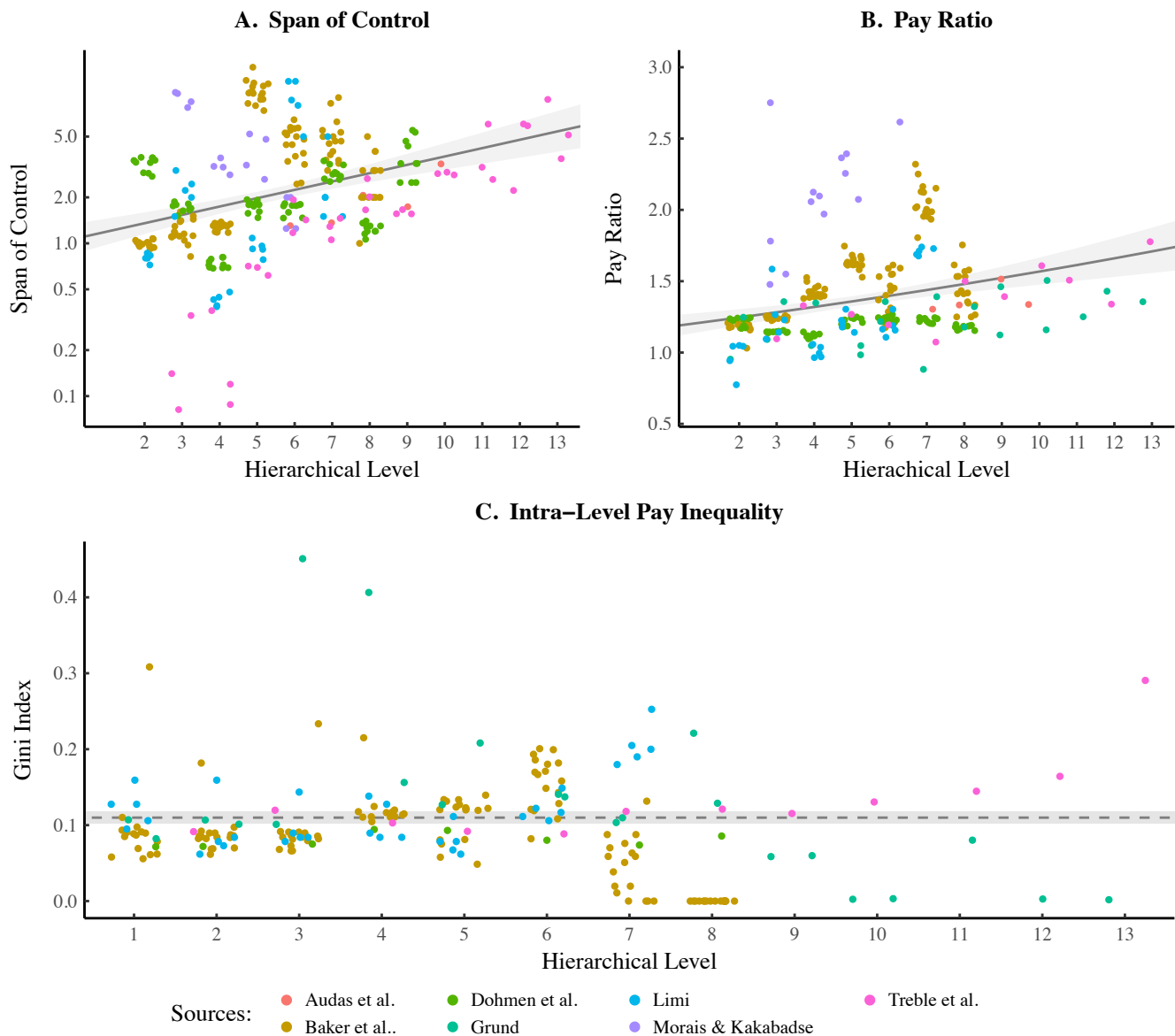


**B. Firm Hierarchical Pay Structure**



**Figure 22: 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 23: 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 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.1 Firm Mean Income

Firm mean income is calculated by dividing total staff expenses (Compustat Series XLR) by total employment (Compustat Series EMP):

$$\text{Firm Mean Income} = \frac{\text{Total Staff Expenses}}{\text{Total Employment}} \quad (7)$$

### C.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 5).
2. Of these executives, take the subset whose title does *not* contain any of the words in the ‘Subordinate Titles’ list (Table 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:

$$\text{CEO Pay Ratio} = \frac{\text{CEO Pay}}{\text{Firm Mean Income}} \quad (8)$$

**Table 5: Titles Used to Identify the ‘CEO’**

| <b>CEO Titles:</b>      | <b>Subordinate Titles</b> |
|-------------------------|---------------------------|
| 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   |

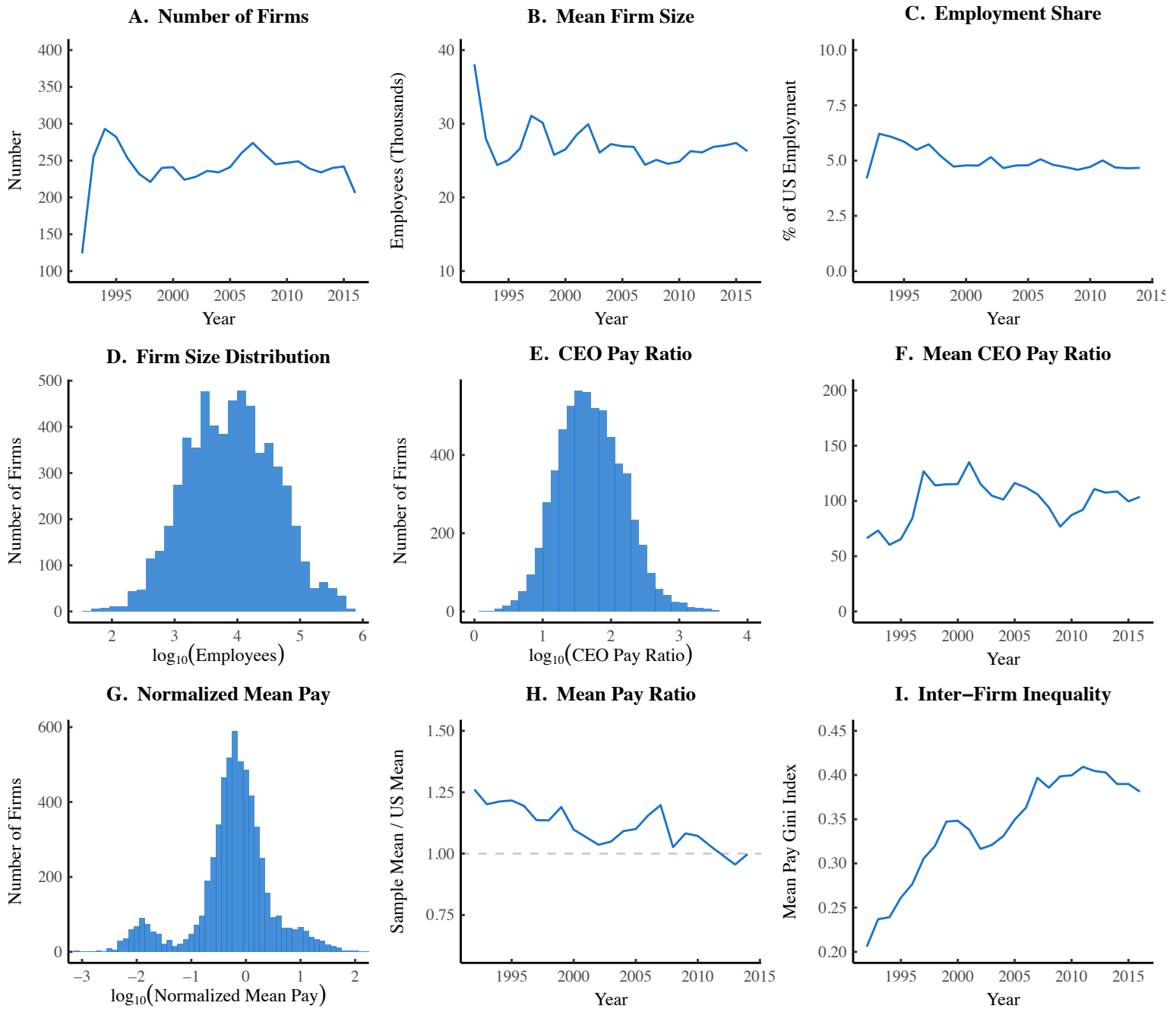
Notes: This table shows the Execucomp titles used to identify the CEO of each company. CEOs are deemed to be those whose title contains words in the left column, but not those in the right column. Titles such as ‘president-’ and ‘president of’ are included in the subordinate list because they typically 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 E) . Figure 24 shows selected summary statistics of this dataset.

### C.3 Capitalist Income Share of Executives

I define the capitalist income share of executives ( $K_{\text{frac}}$ ) as the ratio of stock-options income to total income:

$$K_{\text{frac}} = \frac{\text{Stock Options}}{\text{Total Income}} \quad (9)$$



**Figure 24: 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.

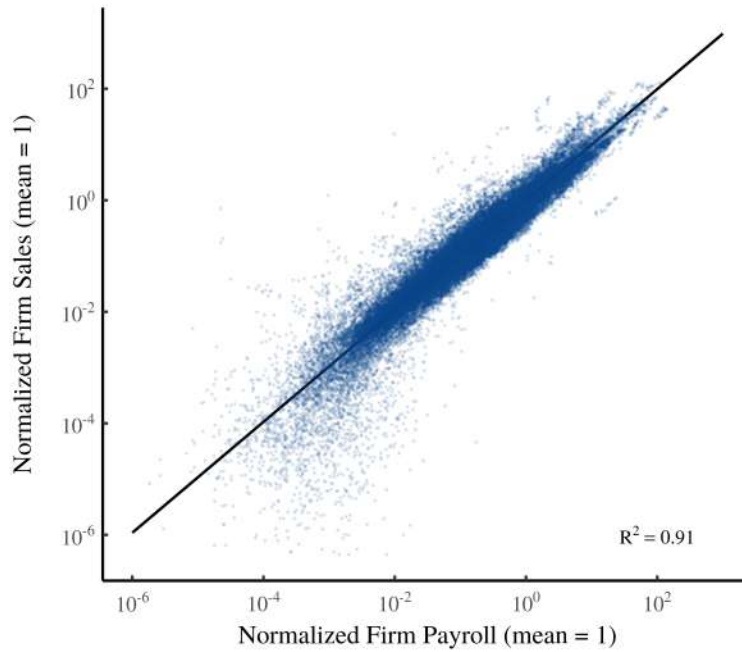
**Table 6: Data Used to Calculate Executive Capitalist Income Fraction**

| 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 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:

$$K_{\text{frac}_{1992}} = \frac{\text{RSTKGRNT} + \text{OPTION\_AWARDS\_BLK\_VALUE}}{\text{TDC1}} \quad (10)$$

$$K_{\text{frac}_{2006}} = \frac{\text{STOCK\_AWARDS\_FV} + \text{OPTION\_AWARDS\_FV}}{\text{TDC1}} \quad (11)$$



**Figure 25: 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.4 Firm Sales vs. Firm Payroll

In section 4, I use the hierarchy model to reproduce historical trends in the CEO pay ratio. The empirical data from Mishel and Schieder [89] 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. 25), the former is a good proxy for the latter.



## D 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 B.

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 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                                 |
| $\mu$        | 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                                |
| $\sigma$     | 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                   |

## D.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:

$$s_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \frac{E_h}{E_{h-1}} & \text{if } h \geq 2 \end{cases} \quad (12)$$

Based on our empirical findings in Section B, 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:

$$s_h = \begin{cases} 1 & \text{if } h = 1 \\ a \cdot e^{bh} & \text{if } h \geq 2 \end{cases} \quad (13)$$

As one moves up the hierarchy, employment in each consecutive level ( $E_h$ ) *decreases* by  $1/s_h$ . This yields Eq. 14, 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. 14 into itself, we can obtain a non-recursive formula (Eq. 15). In product notation, Eq. 15 can be written as Eq. 16.

$$E_h = \downarrow \frac{E_{h-1}}{s_h} \quad \text{for } h > 1 \quad (14)$$

$$E_h = \downarrow E_1 \cdot \frac{1}{s_2} \cdot \frac{1}{s_3} \cdot \dots \cdot \frac{1}{s_h} \quad (15)$$

$$E_h = \downarrow E_1 \prod_{i=1}^h \frac{1}{s_i} \quad (16)$$

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. 17, which in summation notation, becomes Eq. 18.

$$E_T = E_1 + E_2 + \dots + E_n \quad (17)$$

$$E_T = \sum_{h=1}^n E_h \quad (18)$$

In practice,  $n$  is not known beforehand, so we define it using Eq. 16. 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:

$$n = \{h \mid E_h \geq 1 \text{ and } E_{h+1} = 0\} \quad (19)$$

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.

## D.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:

$$p_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \frac{\bar{I}_h}{\bar{I}_{h-1}} & \text{if } h \geq 2 \end{cases} \quad (20)$$

Based on our empirical findings in Section B, 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:

$$p_h = \begin{cases} 1 & \text{if } h = 1 \\ r^h & \text{if } h \geq 2 \end{cases} \quad (21)$$

Using the same logic as with employment (shown above), the mean income  $I_h$  in any hierarchical level is defined recursively by Eq. 22 and non-recursively by Eq. 23.

$$\bar{I}_h = \frac{\bar{I}_{h-1}}{p_h} \quad (22)$$

$$\bar{I}_h = \bar{I}_1 \prod_{i=1}^h p_i \quad (23)$$

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.

### D.2.1 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 ( $E_h$ ):

$$\bar{I}_T = \sum_{h=1}^n \bar{I}_h \cdot \frac{E_h}{E_T} \quad (24)$$

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:

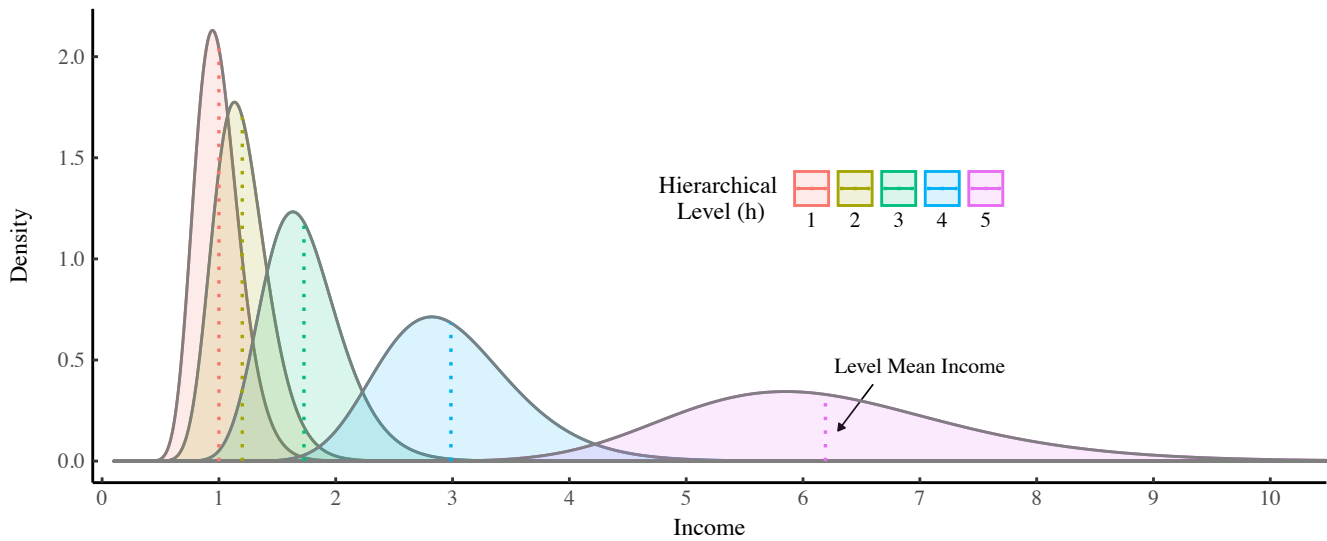
$$C = \frac{\bar{I}_n}{\bar{I}_T} \quad (25)$$

### D.3 Adding Intra-Level Pay Dispersion

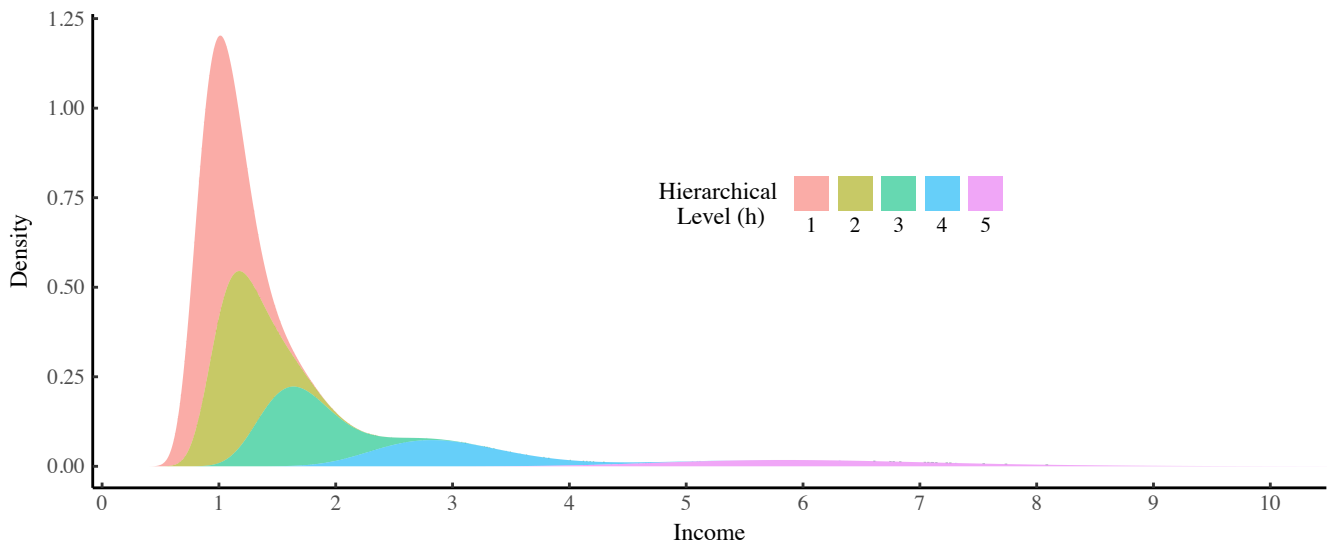
Up to this point, we have modeled only the *mean* income within each hierarchical level of a firm. The last step in the modeling process is to add pay *dispersion* within each hierarchical level.

I assume that pay dispersion within hierarchical levels is *lognormally* distributed. The lognormal distribution is defined by location parameter  $\mu$  and scale parameter  $\sigma$ . Our empirical investigation of firm case studies indicated that pay dispersion with hierarchical levels is relatively constant (see Fig. 23C). Given this finding, I assume *identical inequality* within all hierarchical levels. This means that the lognormal scale parameter  $\sigma$  is the same for all hierarchical levels.

**A. Adding Pay Dispersion Within Each Hierarchical Level**



**B. Relative Contribution to Intra-Firm Income Distribution**



**Figure 26: 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. 26. 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  $\mu$  be defined by Eq. 27.

$$I_h = \bar{I}_h \cdot \ln \mathcal{N}(\mu, \sigma) \quad (26)$$

$$\mu = -\frac{1}{2}\sigma^2 \quad (27)$$

Given a value for  $\sigma$  (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 26. Figure 26A shows the lognormal income distributions for each hierarchical level of a 5-level firm. Figure 26B 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.

#### D.4 Calculating Hierarchical Power

I define an individual's hierarchical power as the number of subordinates ( $S$ ) under their control, plus 1:

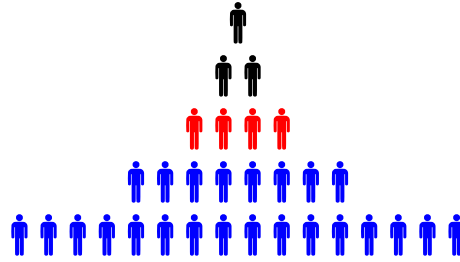
$$P = S + 1 \quad (28)$$

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:

$$\bar{P}_h = \bar{S}_h + 1 \quad (29)$$

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:

$$\bar{S}_h = \sum_{i=1}^{h-1} \frac{E_i}{E_h} \quad (30)$$



**Figure 27: Calculating the Average Number of Subordinates**

As an example, consider the hierarchy in Figure 27. The average number of subordinates below each individual in hierarchal level 3 (red) would be:

$$\bar{S}_3 = \frac{E_1 + E_2}{E_3} = \frac{16 + 8}{4} = 6 \quad (31)$$

Therefore, these individuals would all be assigned a hierarchical power of 7.

**Table 8: Model Parameters**

| Parameter   | Definition                                        | Action                                                                                    | Scope                    |
|-------------|---------------------------------------------------|-------------------------------------------------------------------------------------------|--------------------------|
| $\alpha$    | 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.   |
| $\sigma$    | Intra-hierarchical level pay dispersion parameter | Determines the level of inequality within hierarchical levels of a firm.                  | Identical for all firms. |

## E Restricting Parameters

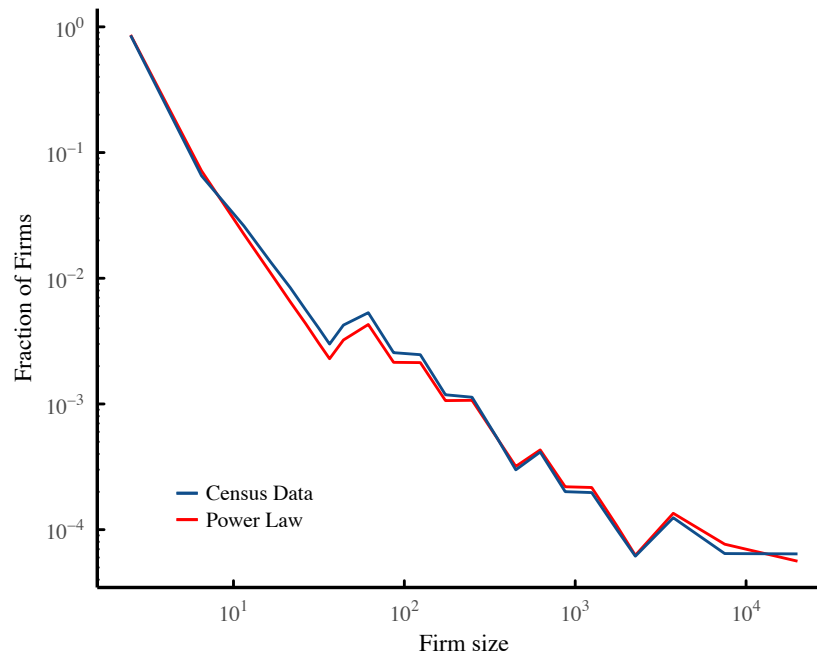
As discussed in section D, the hierarchy model has many ‘free’ parameters. Table 8 summarizes all of the parameters used in this model. While free to take on any value, I restrict these parameters exclusively using empirical data. In the following sections, I outline the methods used for this restriction.

### E.1 Firm Size Distribution

Recent studies have found that firm size distributions in the United States [68] and other G7 countries [70] can be modeled accurately with a power law. A power law has the simple form shown in Eq. 32, where the probability of observation  $x$  is inversely proportional to  $x$  raised to some exponent  $\alpha$ :

$$p(x) \propto \frac{1}{x^\alpha} \quad (32)$$





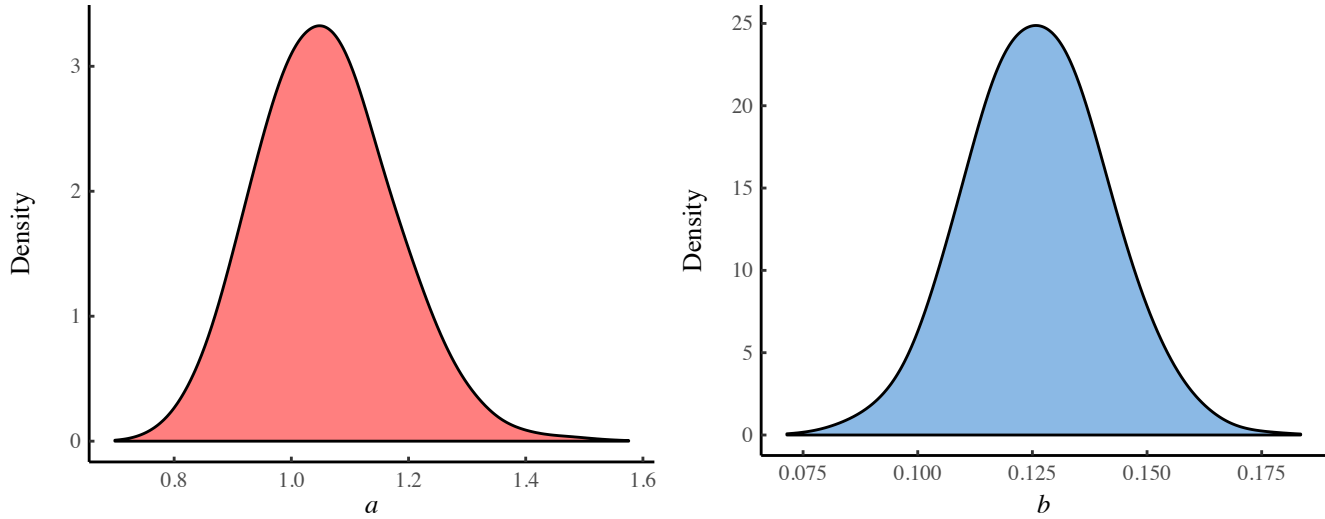
**Figure 28: The United States Firm Size Distribution**

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.

Figure 28 compares the US firm size distribution with a power law of exponent  $\alpha = 2.01$ . 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  $\alpha$  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



**Figure 29: 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. 23). 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.

the C++ header file `rp1d.h`, located in the Supplementary Material. This code is an adaption of Collin Gillespie’s discrete power law generator found in the R `powerLaw` package [110] (which is, in turn, an adaption of the algorithm outline by Clauset [111]).

## E.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. 23A). 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 [112], which involves repeatedly resampling the case-study data (with replacement) and then estimating the parameters  $a$  and  $b$  from this resample. Figure 29 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`.

### E.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 D to define an analytic function relating total employment to base level employment, this is beyond my mathematical abilities.

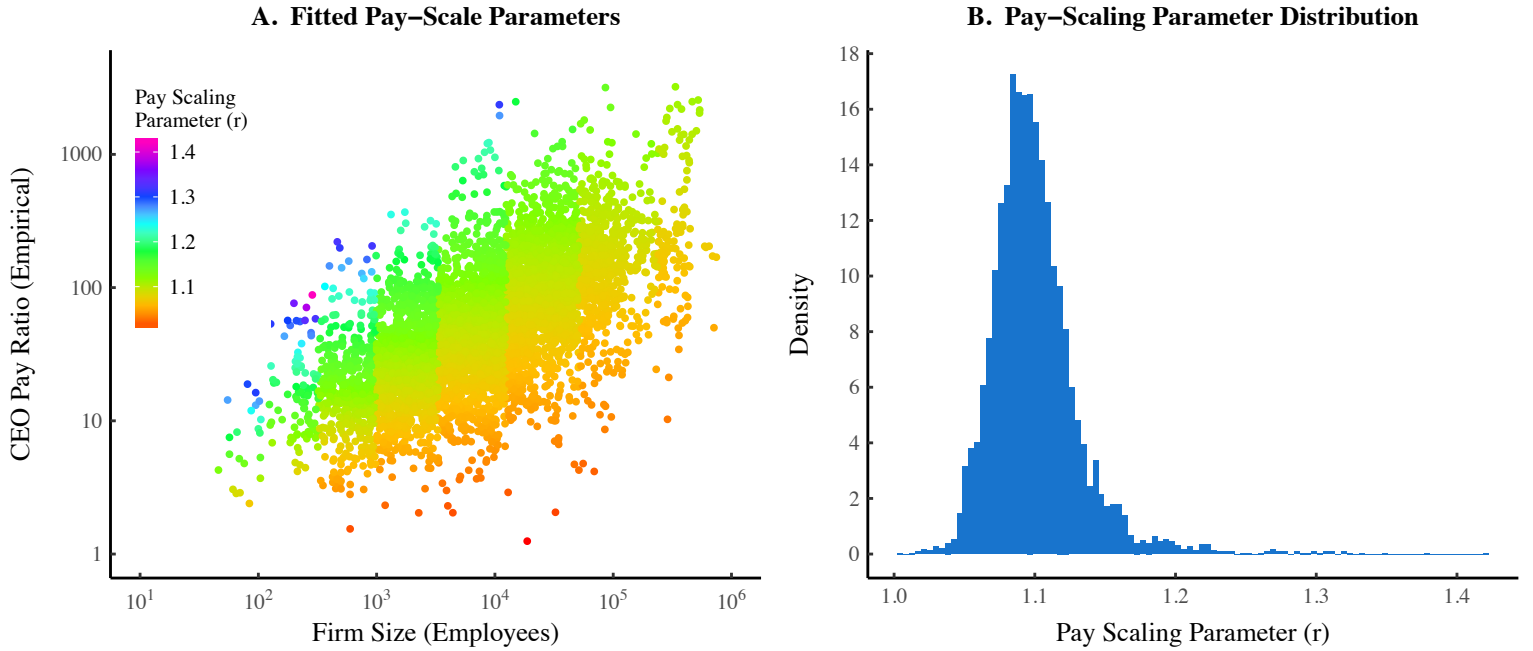
Instead, I use the model to reverse engineer the problem. I input a range of different base employment values into equations 13, 16, and 18 and calculate total employment for each value. The result is a discrete mapping relating base-level employment to total employment. I then use the C++ Armadillo 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.

### E.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.

#### E.4.1 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.



**Figure 30: 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.

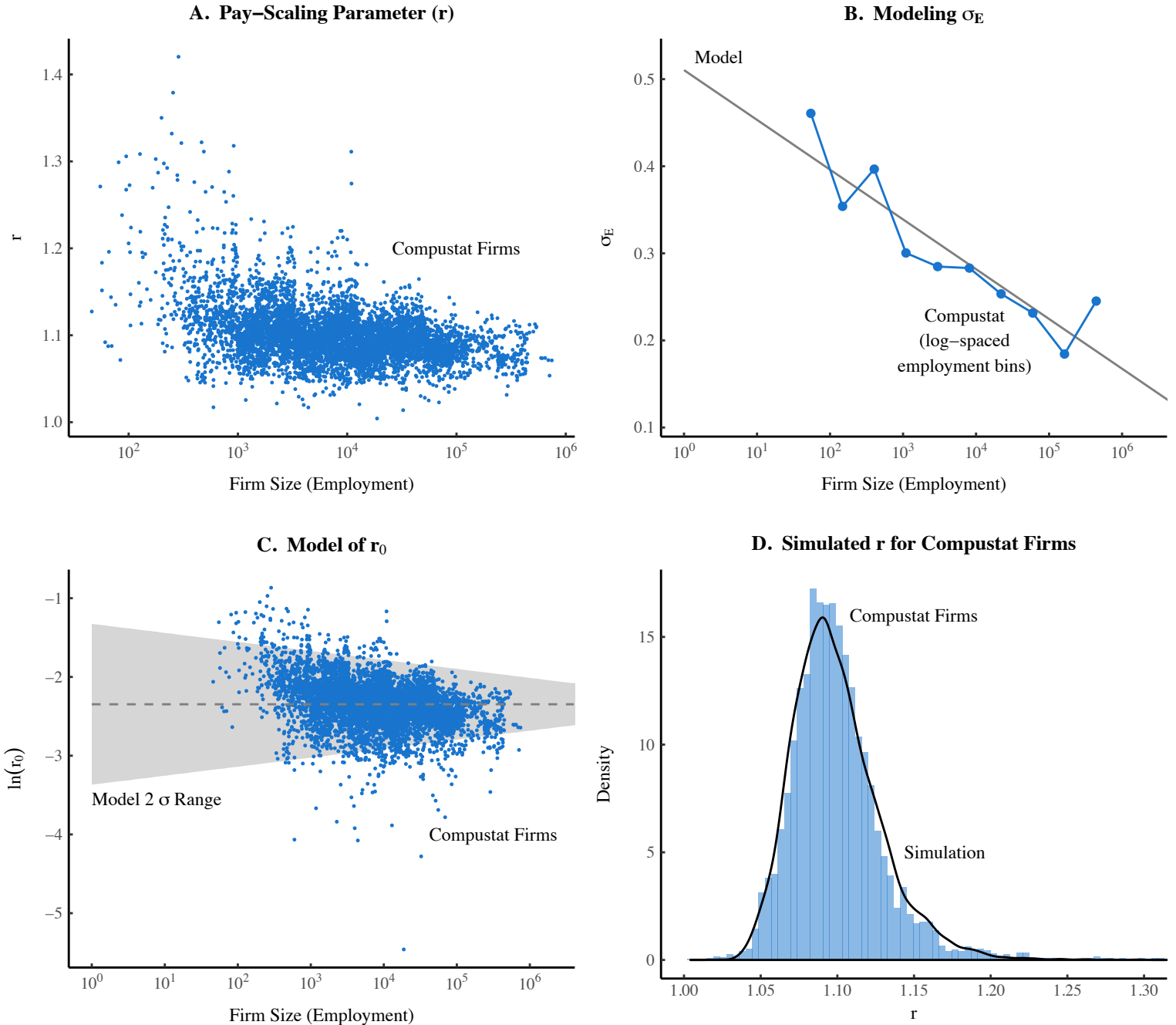
To solve for this  $r$  value, I use numerical optimization (the bisection method) to minimize the error function shown in Eq. 33. Here  $C_{\text{Compustat}}$  and  $C_{\text{model}}$  are Compustat and modeled CEO pay ratios, respectively.

$$\epsilon(r) = |C_{\text{model}} - C_{\text{Compustat}}| \quad (33)$$

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 30. Code implementing this method can be found in the C++ header file `fit_model.h`, located in the Supplementary Material.

#### E.4.2 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 31: 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  $\sigma_E$  (defined by Eq. 38) 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  $2\sigma$  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 31A). I model  $r$  as a shifted function of the lognormal variate  $r_0$ :

$$r = 1 + \ln \mathcal{N}(r_0) \quad (34)$$

The lognormal variate  $r_0$  is defined by location parameter  $\mu$  and scale parameter  $\sigma$ . While  $\mu$  is assumed to be constant for all firms,  $\sigma$  is a function of firm size  $E$ :

$$r_0(E) = \ln \mathcal{N}(r_0; \mu, \sigma_E) \quad (35)$$

I use the tuned Compustat data to solve for the parameters  $\mu$  and  $\sigma$ . We first transform Compustat  $r$  values using Eq. 36 to get the Compustat distribution of  $r_0$ :

$$r_0 = r - 1 \quad (36)$$

The best-fit value for  $\mu$  is defined by taking the mean of  $\ln(r_0)$ :

$$\mu = \overline{\ln(r_0)} \quad (37)$$

Similarly, we can solve for the best-fit value for  $\sigma$  by taking the standard deviation of  $\ln(r_0)$ . However, unlike  $\mu$ , the value  $\sigma$  will depend on the size range of firms ( $E$ ):

$$\sigma_E = SD[ \ln(r_0) ]_E \quad (38)$$

Figure 31B plots  $\sigma_E$  vs.  $E$  for logarithmically spaced size groupings of Compustat firms. I model this relation using a log-linear regression. Figure 31C 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  $\mu$  and  $\sigma$  to the tuned Compustat data, we can generate  $r$  values for simulated firms using equations 34 and 35. 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 31D 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. 18), I vary the mean of the pay-scaling distribution by multiplying the fitted lognormal component by a random constant  $c$ :

$$r = 1 + c \cdot \ln \mathcal{N}(r_0) \quad (39)$$

## E.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.

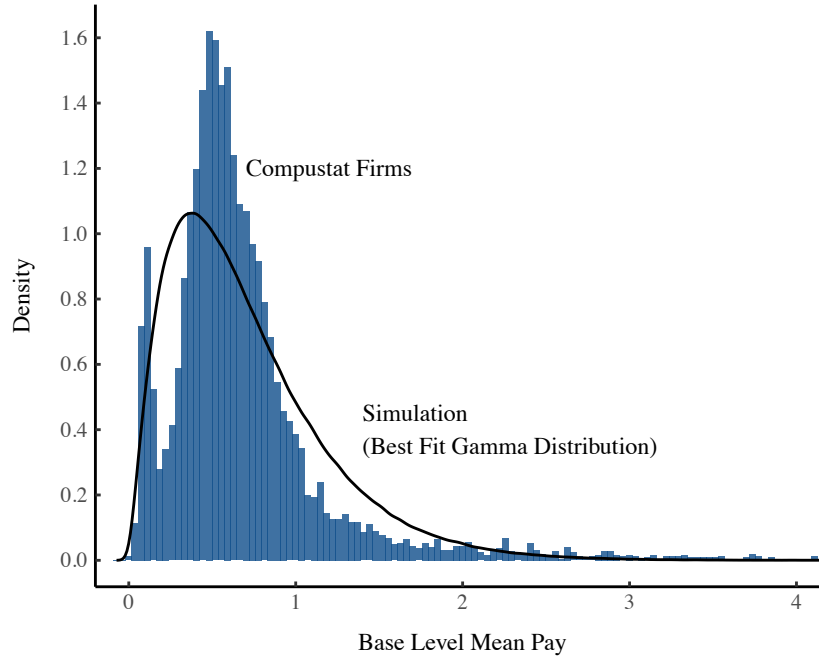
### E.5.1 Fitting Compustat Base Level Pay

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:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{\bar{I}_1^{\text{model}}}{\bar{I}_T^{\text{model}}} \quad (40)$$

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. 23 and 24). If we run the model with  $\bar{I}_1^{\text{model}} = 1$ , then Eq. 40 reduces to:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{1}{\bar{I}_T^{\text{model}}} \quad (41)$$



**Figure 32: 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. 41 to solve for an estimated base pay for each Compustat firm ( $\bar{I}_1^{\text{Compustat}}$ ):

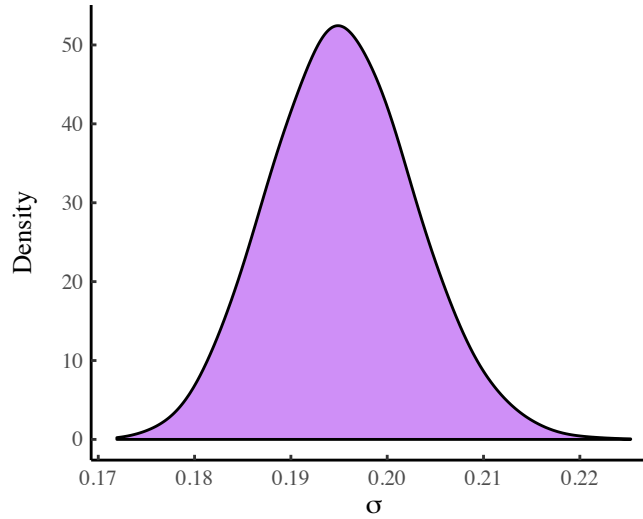
$$\bar{I}_1^{\text{Compustat}} = \frac{\bar{I}_T^{\text{Compustat}}}{\bar{I}_T^{\text{model}}} \quad (42)$$

Code implementing this method is found in the C++ header file `fit_model.h`, located in the Supplementary Material.

### E.5.2 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 33: Density estimates for Intra-Hierarchical Level Pay Dispersion Parameter  $\sigma$**

This figure shows the distribution of the lognormal scale parameter  $\sigma$ , 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. 32). 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).

## E.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  $\sigma$ . I estimate  $\sigma$  from the case-study data shown in Figure 23C. This data uses the Gini index as the metric for dispersion.

To estimate  $\sigma$ , we first calculate the mean Gini index of all data ( $\bar{G}$ ). We then use Eq. 43 to calculate the value  $\sigma$ , 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:  $G = \text{erf}(\sigma/2)$ .

$$\sigma = 2 \cdot \text{erf}^{-1}(\bar{G}) \quad (43)$$

The model proceeds on the assumption that  $\sigma$  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 [112], which involves repeatedly resampling the case-study data (with replacement) and then estimating the parameter  $\sigma$  from this resampled data.

Figure 33 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  $\sigma$ . Code implementing this method can be found in the C++ header file `boot_sigma.h`, located in the Supplementary Material.

## E.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 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.

Table 9: Model High-Level Structure

| Step | Action                                                                | Reference Section        | Parameter(s)   | Header File(s)              |
|------|-----------------------------------------------------------------------|--------------------------|----------------|-----------------------------|
| 1    | Bootstrap case-study data                                             | <a href="#">E.2, E.6</a> | $a, b, \sigma$ | boot_span.h<br>boot_sigma.h |
| 2    | Get Compustat base-level employment                                   | <a href="#">E.3</a>      | $E_1$          | base_fit.h                  |
| 3    | Fit Compustat pay-scaling parameters                                  | <a href="#">E.4.1</a>    | $r$            | fit_model.h                 |
| 4    | Get Compustat base-level mean pay                                     | <a href="#">E.5.1</a>    | $\bar{I}_1$    | fit_model.h                 |
| 5    | Generate power law firm size distribution                             | <a href="#">E.1</a>      | $\alpha$       | rp1d.h                      |
| 6    | Get simulation base-level employment                                  | <a href="#">E.3</a>      | $E_1$          | base_fit.h                  |
| 7    | Simulate pay-scaling parameter distribution by fitting Compustat data | <a href="#">E.4.2</a>    | $r$            | r_sim.h                     |
| 8    | Simulate base mean pay distribution by fitting Compustat data         | <a href="#">E.5.2</a>    | $\bar{I}_1$    | base_pay_sim.h              |
| 9    | Run hierarchy model                                                   | <a href="#">D</a>        | all            | model.h                     |

Notes: Model code makes extensive use of Armadillo, an open-source C++ linear algebra library [[113](#)].

## F 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 E), 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 E, inter-firm income dispersion is modeled by fitting a gamma distribution to Compustat data. The gamma distribution has the following probability density function:

$$p(x) = \frac{1}{\Gamma(k)\theta^k} \cdot x^{k-1} \cdot e^{-x/\theta} \quad (44)$$

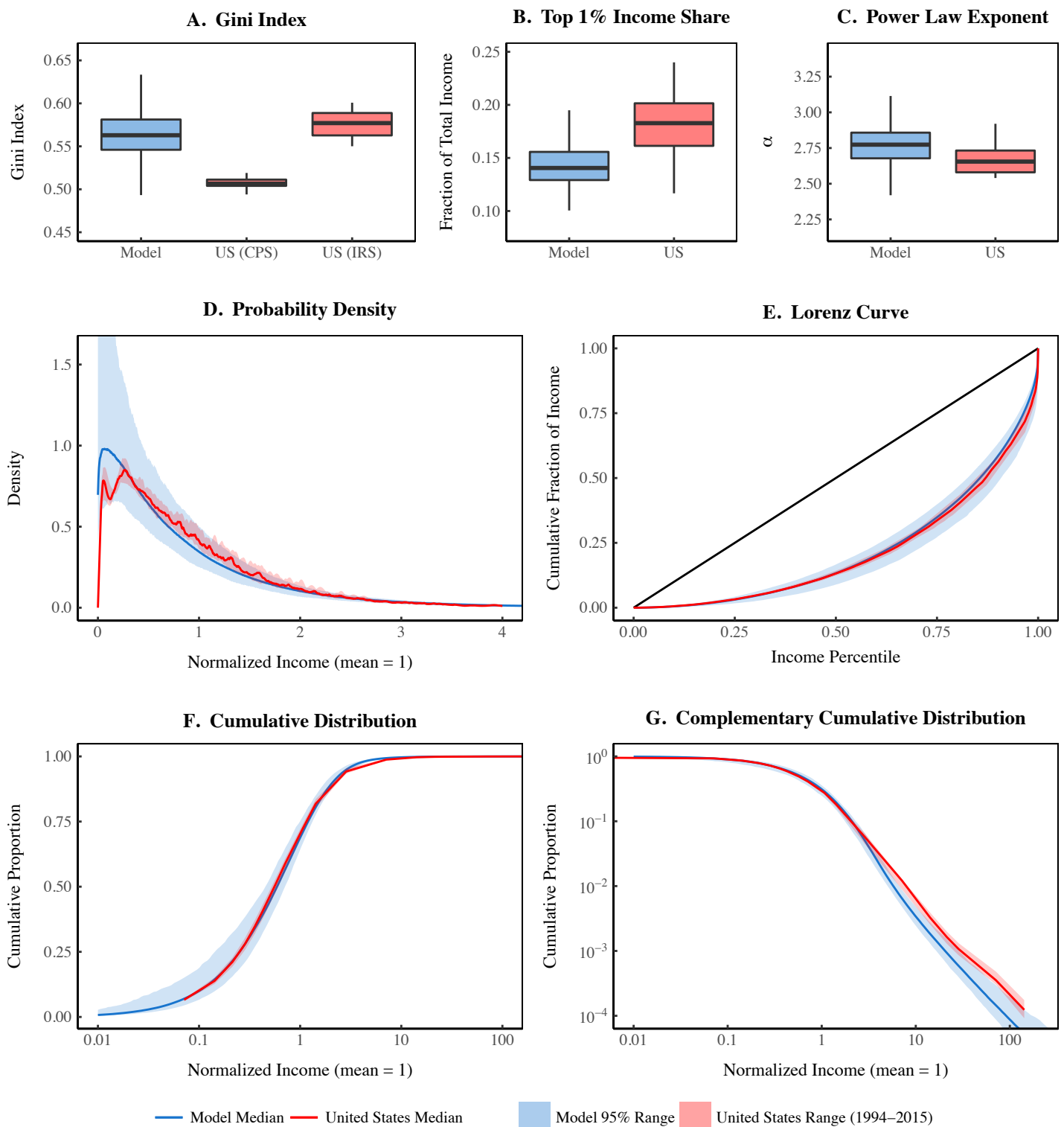
In the original model, the parameters  $k$  and  $\theta$  are both determined by empirical data. In the adjusted model, I introduce a fudge-factor  $c$  that allows me to adjust the fitted  $k$  parameter by a constant amount:

$$k_{\text{adjust}} = c \cdot k_{\text{fit}} \quad (45)$$

The adjusted model then uses the parameter  $k_{\text{adjust}}$  instead of  $k_{\text{fit}}$ . 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 34 in the same format as the original model was presented in Figure 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).

### F.1 Hierarchical Redistribution with the Adjusted Model

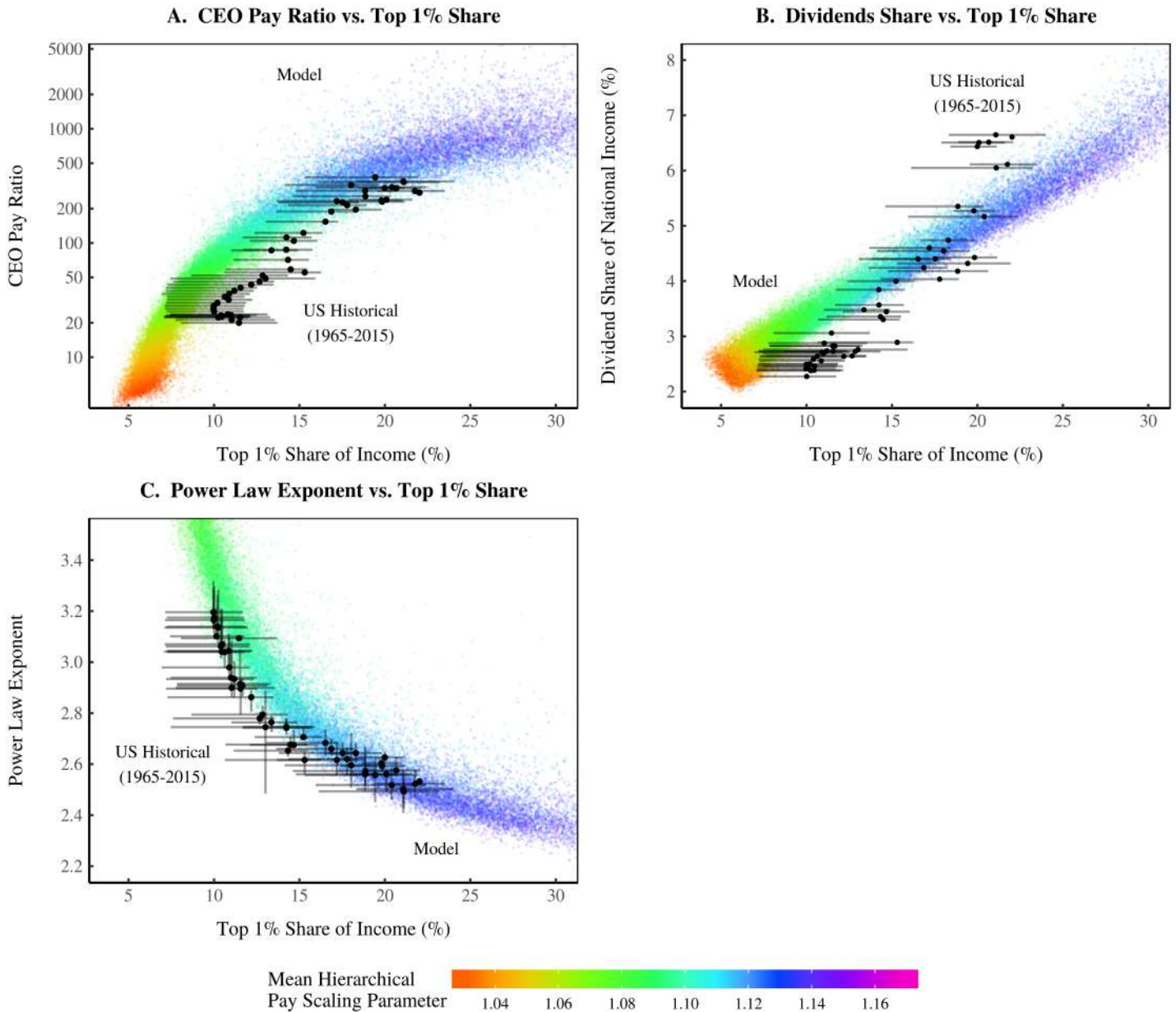
Because the hierarchy model tends to underestimate US income inequality, the hierarchical redistribution model tends to be shifted to the left relative to US empirical data (see Fig. 18). As shown in Figure 35, by using the adjusted model to calculate hierarchical redistribution, this problem disappears. Note,



**Figure 34: 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  $\alpha$  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 A.



**Figure 35: 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 A.

however, that increasing inter-firm dispersion does not fix the model/empirical discrepancy in the slope of the dividends vs. top 1% relation (Fig. 35B).

## E.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 ( $K_{\text{frac}}$ ):

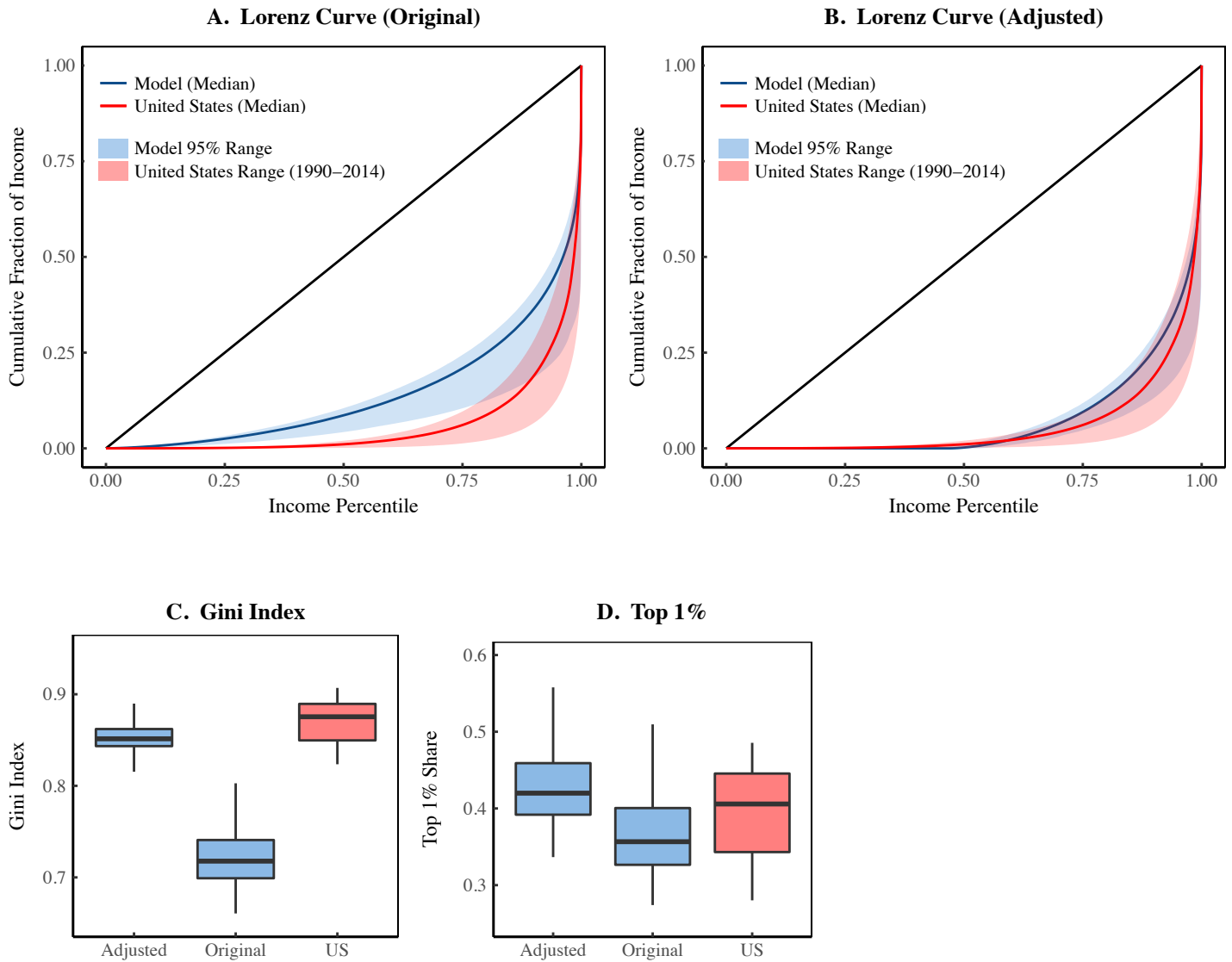
$$K_{\text{frac}} = 0.05 \ln(P) \quad (46)$$

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:

$$K_{\text{frac}} = 0.05 \ln(P) + \epsilon \quad (47)$$

Here  $\epsilon$  is a constant very close to zero. Its effect is only felt when  $P = 1$ . Instead of getting  $K_{\text{frac}} = 0$ , we get  $K_{\text{frac}} = \epsilon$ . 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 36, the adjusted model better matches the US data.



**Figure 36: 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 A.



## G 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 37, 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.

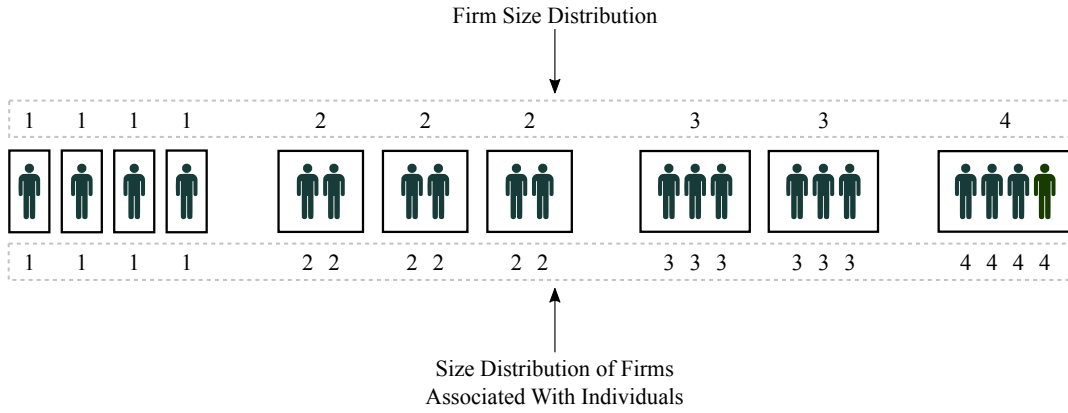
If we randomly select an individual from the private sector population, let  $p(i_x)$  be the probability that this individual is associated with a firm of size  $x$ . This probability will determine the size distribution of firms associated with a random sample of individuals. Let  $p(x)$  be the probability of randomly selecting a firm of size  $x$  from the firm population. Using Figure 37 for guidance, we can see that  $p(i_x)$  is given by:

$$p(i_x) \sim x \cdot p(x) \tag{48}$$

If we know  $p(x)$  — the probability distribution of firms — we can use Eq. 48 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 E). Substituting this into Eq. 48 gives:

$$p(i_x) \sim x^{-1} \tag{49}$$

Because firm sizes generally span many orders of magnitude, it is more convenient to look at the log transformation of Eq. 49. Therefore, we want to know



**Figure 37: 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 the individual population). Noting that each firm size  $x$  appears  $x$  times in the individual-to-firm mapping, we can state that  $p(i_x) \propto x \cdot p(x)$ .

the probability density for  $p(\ln i_x)$ . To find this, we use the standard change-of-variable function for a probability density:

$$f_y = f_x(x(y)) \cdot |x'(y)| \tag{50}$$

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. 50 gives:

$$f_y = c \cdot (e^y)^{-1} \cdot e^y = c \tag{51}$$

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.

## H 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 [63] 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 [114]. The generative mechanism works as follows. Suppose we have some quantity  $y$  that is exponentially distributed (here  $a$  is a negative constant):

$$p(y) \sim e^{ay} \quad (52)$$

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 1 for an idealized picture.

Suppose that we have another variable,  $x$ , that is *also* exponentially related to  $y$ :

$$x = e^{by} \quad (53)$$

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 [62]. My own view is that it is caused by the power asymmetries that are innate to hierarchical organization [115]. 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  $x$ :

$$f_x = f_y(y(x)) \cdot |y'(x)| \quad (54)$$

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:

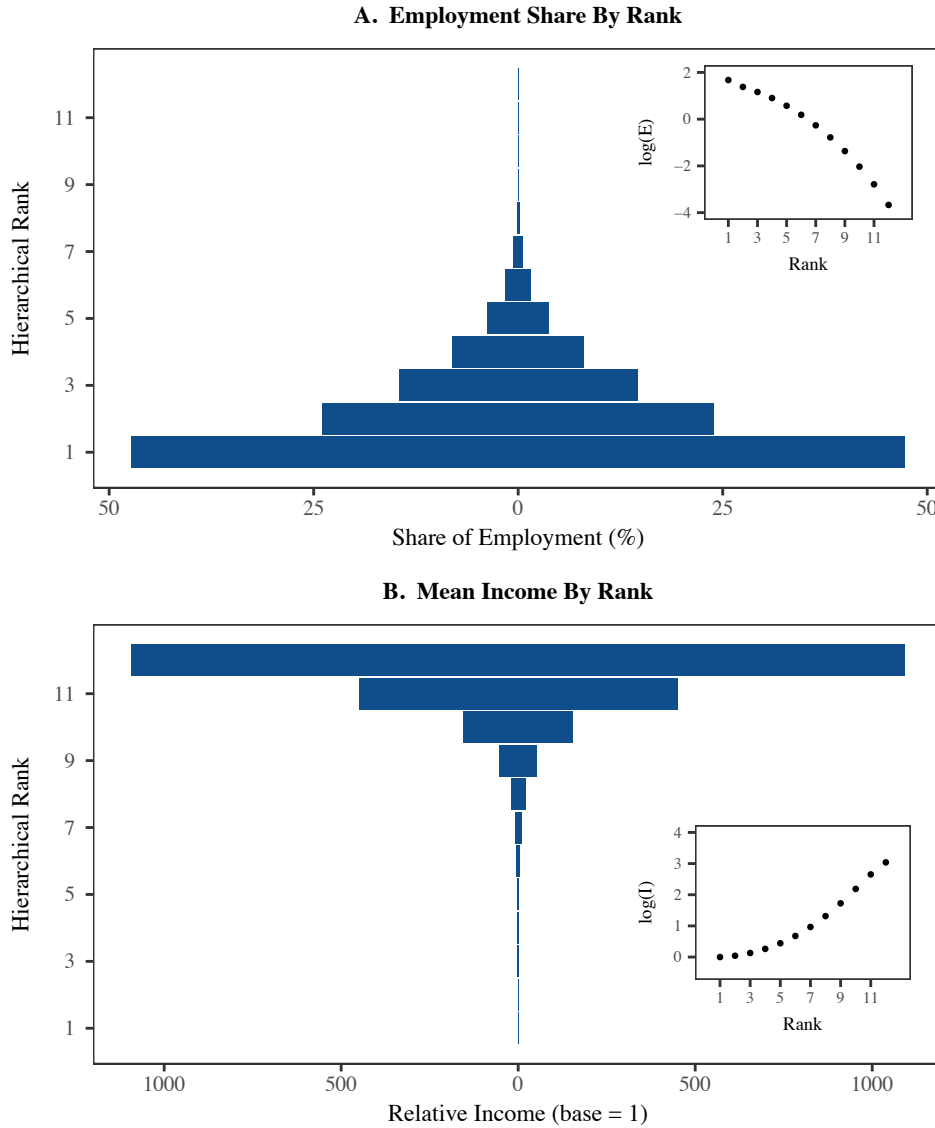
$$f_x = e^{\frac{a}{b} \ln x} \cdot \frac{1}{bx} = \frac{1}{b} x^{a/b-1} \quad (55)$$

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 38 highlights this contrapuntal behavior in the hierarchy model. As expected, the hierarchical employment distribution has a bottom-heavy pyramid shape (Fig. 38A). 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.

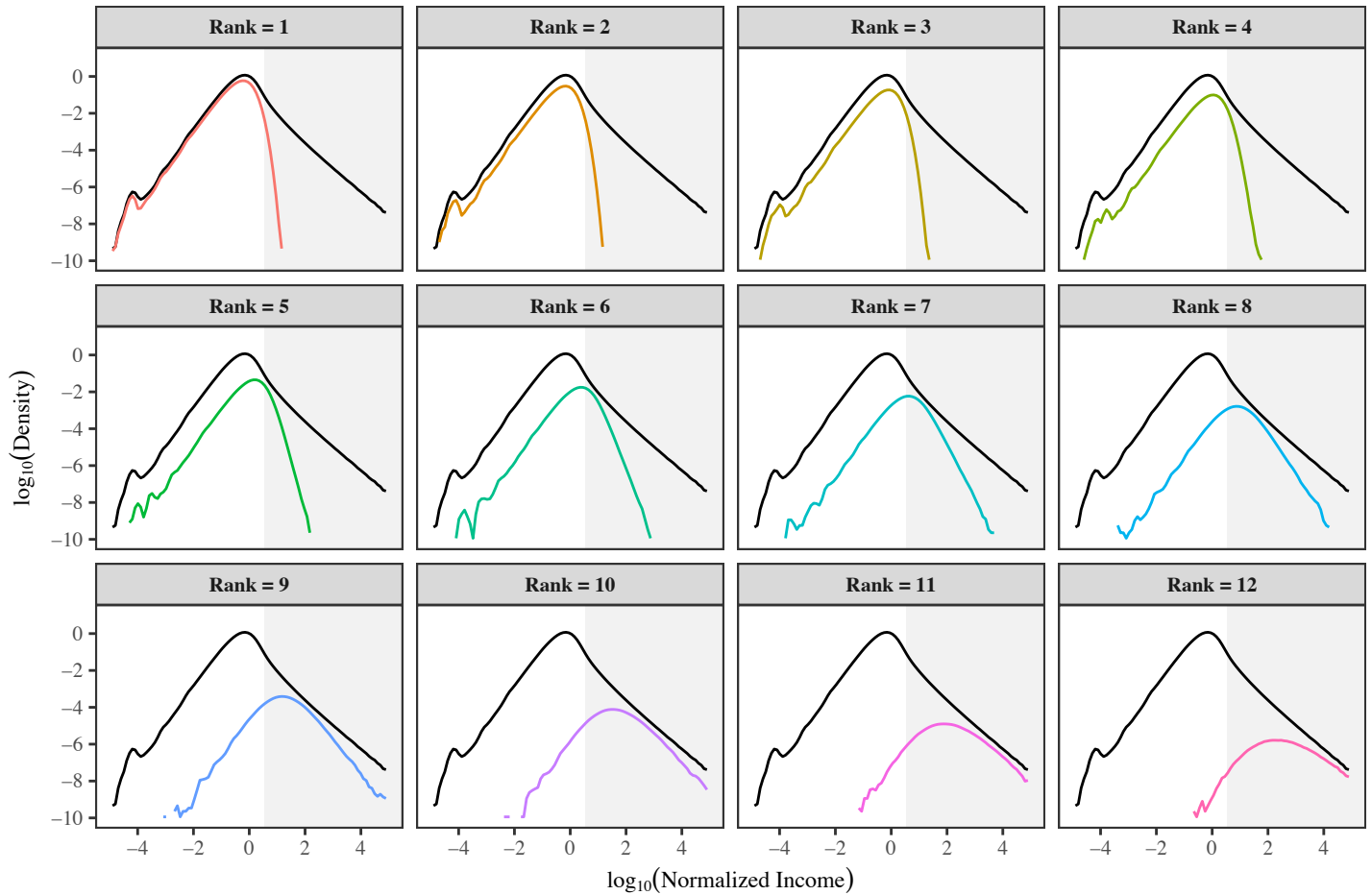
Figure 38B shows the model's hierarchical *pay* structure. To make comparison easy, I have normalized all income so that the base-level income is equal to one. As expected, hierarchical pay has an inverted pyramid shape. Average income at the top of the hierarchy dwarfs (by several orders of magnitude) 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 B and D for details). In any case, when we combine these two supra-exponential tendencies, the result still seems to be (roughly) a power law distribution of income in the model's tail. (I have not, as yet, worked out how this happens.)



**Figure 38: 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 39: 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 get a better picture of how this process works, we turn to Figure 39. 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. Pareto V. Cours d'économie 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. Urich 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.

96. Current Population Survey. Annual Social and Economic Supplement; 2017.
97. Virkar Y, Clauset A. Power-law distributions in binned empirical data. *The Annals of Applied Statistics*. 2014;8(1):89–119.
98. Alvaredo F, Atkinson AB, Piketty T, Saez E, Zucman G. The world wealth and income database. Website: <http://www.wid.world>. 2016;.
99. Alvaredo F, Atkinson T, Piketty T, Saez E. The World Top Incomes Database; 2013. Available from: <http://topincomes.parisschoolofeconomics.eu/#>.
100. 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.
101. 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.
102. Graf M, Nedyalkova D. GB2: Generalized Beta Distribution of the Second Kind: properties, likelihood, estimation. R package version. 2012;1.
103. Audas R, Barmby T, Treble J. Luck, effort, and reward in an organizational hierarchy. *Journal of Labor Economics*. 2004;22(2):379–395.
104. Baker G, Gibbs M, Holmstrom B. Hierarchies and compensation: A case study. *European Economic Review*. 1993;37(2-3):366–378.
105. 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.
106. 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.
107. Lima F. Internal Labor Markets: A Case Study. FEUNL Working Paper. 2000;378.
108. 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.

109. Treble J, Van Gasteren E, Bridges S, Barmby T. The internal economics of the firm: further evidence from personnel data. *Labour Economics*. 2001;8(5):531–552.
110. Gillespie CS. Fitting heavy tailed distributions: the powerLaw package. arXiv preprint arXiv:14073492. 2014;.
111. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM review*. 2009;51(4):661–703.
112. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. London: CRC press; 1994.
113. Sanderson C, Curtin R. Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software*. 2016;.
114. Newman ME. Power laws, Pareto distributions and Zipf's law. *Contemporary physics*. 2005;46(5):323–351.
115. Fix B. Evidence for a Power Theory of Personal Income Distribution. *Working Papers on Capital as Power*. 2017;No. 2017/03:1–125.