Fix, Blair

Working Paper

Evidence for a Power Theory of Personal Income Distribution

Working Papers on Capital as Power, No. 2017/03

Provided in Cooperation with:
The Bichler & Nitzan Archives

Suggested Citation: Fix, Blair (2017) : Evidence for a Power Theory of Personal Income Distribution, Working Papers on Capital as Power, No. 2017/03, Forum on Capital As Power - Toward a New Cosmology of Capitalism, s.l.,
http://www.capitalaspower.com/?p=2344

This Version is available at:
https://hdl.handle.net/10419/162938

WORKING PAPERS ON
CAPITAL AS POWER

No. 2017/03

Evidence for a Power Theory of Personal Income Distribution

Blair Fix

July 2017

http://www.capitalaspower.com/?p=2344

Evidence for a Power Theory of Personal Income Distribution

Blair Fix*

July 23, 2017

Abstract

This paper proposes a new 'power theory' of personal income distribution. Contrary to the standard assumption that income is proportional to productivity, I hypothesize that income is most strongly determined by *social power*, as indicated by one's position within an *institutional hierarchy*. While many theorists have proposed a connection between personal income and power, this paper is the first to *quantify* this relation. I propose that power can be quantified in terms of the number of subordinates below one's position in a hierarchy. Using this definition, I find that relative income within firms scales strongly with hierarchical power. I also find that hierarchical power has a stronger effect on income than any other factor for which data is available. I conclude that this is evidence for a power theory of personal income distribution.

*Author contact: blairfix@gmail.com. I would like to thank Jonathan Nitzan and Shai Gorsky for helpful comments on an earlier draft of this paper.

Contents

1 Introduction

Over the last decade, concerns about income inequality have risen to the forefront of public attention. As testament to this interest, Thomas Piketty's expansive treatise on inequality, *Capital in the Twenty-First Century* [1], became an unlikely best seller when it was published in 2014. Due in no small part to the work of Piketty and colleagues [2–5], *empirical* study of income inequality has flourished. But this plethora of new data has not led to a corresponding *theoretical* revolution.

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

I hypothesize that personal income is explained most strongly by *social power*, as manifested by one's rank in an *institutional hierarchy*. Using the common definition of power as the 'ability to influence or control others', I propose the novel approach of *quantifying* hierarchical power in terms of the number of subordinates under an individual's control. From this definition, it follows that power, unlike productivity, tends to be *very* unequally distributed within hierarchies — a natural consequence of the tree-like chain of command that concentrates control at the top. Thus, a power theory of income distribution naturally solves the under-explanation problem that is inherent in the 'productivist' approach.

This paper is the first (to my knowledge) to *quantitatively* connect power with personal income. I provide two key pieces of evidence in support of a power theory of personal income distribution. Firstly, using the available case study data, I demonstrate that relative income within firms scales tightly with my proposed metric for power — the number of subordinates under an individual's control. Secondly, I find that grouping individuals by hierarchical rank

(i.e. power) has the strongest effect on income, when compared to a wide range of other group types. I conclude that this is evidence for a power theory of personal income distribution.

The paper is organized into the following parts. In section 2, I review and critique existing theories of personal income distribution, and summarize the key failings of the dominant 'productivist' approach. In order to situate my proposed power theory of income distribution, in section 3 I look at the big picture of human inequality and how it relates to the formation of social hierarchies. In section 4, I outline the principles and motivations behind the power theory of value hypothesis. In section 5, I test the power-income hypothesis against empirical evidence. All methods and sources are documented in the Appendix.

2 Theories of Personal Income Distribution

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

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

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

Following the findings of Galton and Pareto, political economists have spent a century struggling to reconcile these two facts [8]. The process has been made difficult primarily because the two dominant theories of *functional* (class-based) income distribution assume a connection between individual productivity (hence ability) and income.

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

2.1 Stochastic Theories

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

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

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

Despite their successes, stochastic models have been mostly ignored by the economics profession. One reason is that the assumptions underlying this type

of theory (especially kinetic exchange models) are often unrealistc [19]. Kinetic exchange models imply a world in which money is conserved for all time, nothing is ever produced, there are no groups, institutions or classes of people, and the world exists in static equilibrium.

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

2.2 Productivist Theories

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

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

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

The logical implication of this theory is that income differences between *workers* — who all earn *labor* income — must be due to differences in *individual* productivity. Out of this line of reasoning came *human capital* theory, which attributes worker's productivity to some internal stock of 'human capital' [22–24].

Unlike neoclassical theory, Marxist theory posits that labor is the *sole* producer of value. Therefore, both labor and capitalist income ultimately stem from worker's productivity. The Marxist twist is to treat capitalist income as *parasitic* – the result of the expropriation of surplus value created by workers. The relative balance between labor and capitalist income is then a function of the 'degree of exploitation' of workers. But when it comes to income distribution *among* workers, Marxists come to conclusions that are very similar to their neoclassical counterparts. Since labor is the sole source of value, skilled workers who earn more than unskilled workers must somehow be more *productive* [25].

This productivity-income hypothesis has made it difficult for neoclassical and Marxist theories to address the Galton-Pareto paradox. Since individual productivity is presumably related to *ability*, one cannot take the easy road and simply negate any relation between *ability* and *income*. Instead, one must explain why productivity is as unequally distributed as income, but ability is not. The most common resolution to the Galton-Pareto paradox has been to assume that different abilities, each normally distributed, somehow interact to have a *multiplicative* effect on productivity [26, 27].[1] This hypothesis is central to human capital theory, which proposes that investments in human capital yield multiplicative returns to productivity [22, 23].

But is this actually the case? Is productivity as unequally distributed as income? Unfortunately, this question is not as easily answered as it might seem. The problem is this: how do we compare the productivity of different workers who have *qualitatively different outputs*? For instance, how can we determine if a farmer, who produces potatoes, is more productive than a composer, who

[1] This multiplicative effect can be expressed as a production function in which a worker's output (Y) is an exponential function of the sum of different abilities (a_i): $Y = e^{a_1+a_2+\ldots+a_i}$

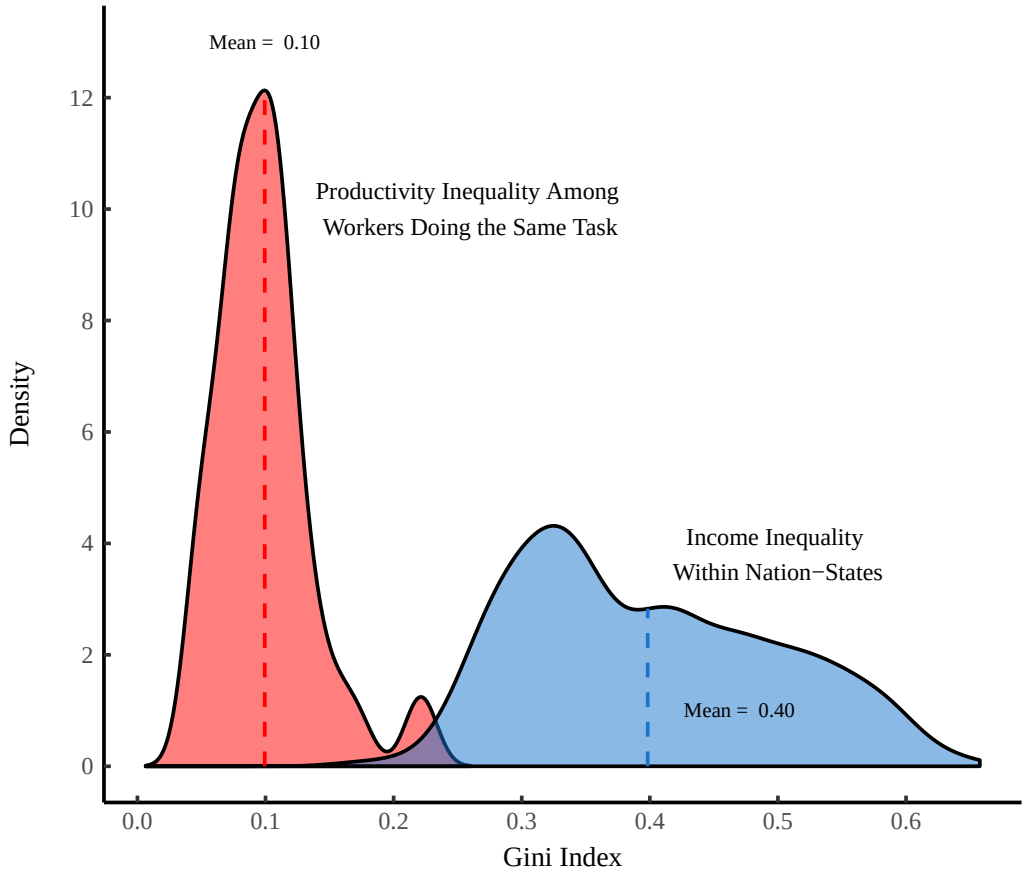


Figure 1: Labor Productivity Inequality vs. Income Inequality

Using a Gini index, this figure compares the inequality of worker productivity to income inequality within nation-states. Data for the former comes from Hunter et al. [28], who report the coefficient of variation of productivity among workers conducting the *same task*. Data plotted here shows the distribution of productivity inequality for 55 different tasks. I convert Hunter's data to a Gini index by assuming that worker productivity is lognormally distributed. The Gini index (G) of a lognormal distribution with a coefficient of variation c_v is $G = \text{erf}(\frac{1}{2}\sqrt{\log(c_v^2 + 1)})$. I plot the resulting distribution against the distribution of Gini indexes of income inequality for all country-year observations in World Bank database (series SI.POV.GINI).

produces music? Any such comparison of qualitatively different outputs inevitably requires choosing a common unit of analysis. But the choice of this unit is *subjective*, and different units will lead to different results. The logical implication is that there are no *objective* grounds for comparing the productivity of workers with qualitatively different types of output.[2] Unfortunately, admitting this fact means that one can compare productivity only between workers who have *exactly the same output*.

Hunter et al. [28] have compiled data that does exactly that — they report differences in productivity among workers doing the *same task*. In Figure 1, I take this data and convert it into a Gini index of 'productivity inequality' so that it is directly comparable to income inequality within nation states. I find that differences in productivity are *systematically* too small to account for observed levels of inequality.

But the logic that qualitatively different outputs cannot be objectively compared is not widely accepted among economists. Instead, they sidestep the problem by adopting *monetary value* as a common unit of comparison for measuring different outputs. Thus, labor productivity is generally measured in terms of sales or value-added per worker [32–38]. The problem with this approach is that it relies on circular logic. According to theory, income is *explained* by productivity. But when the theory is tested, productivity is *measured* in terms of income. And based purely on accounting principles, we expect wages to be correlated with sales/value-added per worker.[3]

[2] The same problem occurs when attempting to measure the productivity of *capital*: one can only compare capitalists with exactly the same output. There are other measurement problems inherent in marginal productivity theory. These include the inability to objectively measure capital [21, 29, 30], as well as the inability to isolate the effect on output caused by changes in capital vs. changes in labor. See Pullen [31] for a good review.

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

$$\frac{Y}{L} = \frac{W+K}{L} = w + \frac{K}{L} \tag{1}$$

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

$$\frac{S}{L} = \frac{W+K+C}{L} = w + \frac{K+C}{L} \tag{2}$$

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

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

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

3 A Big Picture View of Human Inequality

As Gerhard Lenski put it, the study of income distribution is about understanding "*who gets what and why*?" [39]. Like Lenski — and later Nitzan and Bichler [21] — I hypothesize that income distribution is primarily about *power*. And like many critical political economists, I believe that income distribution theory should be historically informed. Thus, in this section I trace the historical link between social hierarchy, power, and human inequality.

In short, I argue that the human urge to seek power (and to use this power to gain preferential access to resources) likely has deep evolutionary origins as a means for *increasing reproductive success*. But while the urge to seek power has likely *always* been part of the human psyche, inequality has *not*. Using a wide variety of different evidence, I argue that the rise of hierarchy and inequality during the neolithic era was the result of a social transformation that radically changed the balance of power between the sexes. The result was that males were suddenly free to use power and status to achieve enormous increases in differential reproductive success.

3.1 The Origins of Inequality

From the scattered evidence that is available, there is general agreement that humanity spent the majority of its existence (perhaps as long as 300,000 years [40, 41]) in small-scale hunter-gather societies that were relatively egalitarian [42–45]. Then around 10,000 years ago, signs of inequality begin to appear in the archaeological record [46]. By 5000 years ago the first states began to form, and inequality became pervasive [47]. Archaeologist Gary Feinman puts it bluntly: "In the history of the human species, there is no more significant transition than the emergence and institutionalization of inequality" [48].

Even if one is primarily concerned with modern industrial societies, as I am in this paper, the fact that inequality has a specific *origin* ought to be taken seriously. Why? If one proposes that income/resource distribution is primarily determined by some factor x, it follows that prior to the onset of inequality, this factor was missing (or somehow different). Based on the available evidence, I suggest that the origin of significant resource/income inequality corresponds to the introduction of *institutional hierarchy*.

This is the road taken by anthropologists/archaeologists such as Price and Feinman who see inequality as "the organizing principle of hierarchical structure in human society" [47]. Institutional hierarchy has also been the focus of many sociological theories of inequality [39, 49–51]. However, the starting point for my approach is the work of economists Herbert Simon and Harold Lydall who both focus on the *branching* nature of modern institutional hierarchies, in which each superior has control over *multiple* subordinates [52, 53].

3.2 Hierarchy in Human Evolution

The focus on branching hierarchy is important, because non-branching (linear) hierarchies have probably always played a role in the distribution of resources in human societies (and hence cannot be used to explain the emergence of inequality). In evolutionary terms, humans are but one of a vast number of social mammal species, virtually all of which form linear dominance hierarchies, or 'pecking orders' [54–59]. A key characteristic of these dominance hierar-

chies is that social status is associated with preferential access to resources, particularly sexual mates [60–64].

Similar tendencies appear to be present among humans. Several studies have shown that children and adolescents spontaneously form linear dominance hierarchies when placed into small groups [65–67]. There is also ample evidence (from many different societies and time periods) indicating that human male reproductive success increases with social status [68]. This suggests that, at least to some degree, humans have inherited an instinct for linear hierarchy formation.

Modern institutional branching hierarchies, however, are starkly different from linear dominance hierarchies. The key difference between the two is that the number of subordinates grows *linearly* with rank in a linear hierarchy, and *exponentially* with rank in a branching hierarchy (see Fig. 2). The effect of a branching hierarchy is thus a profound concentration of power in the hands of the few. If this power is then used to gain preferential access to resources, it will result in levels of inequality that would be otherwise impossible within a linear hierarchy.

3.3 The Origin of Branching Institutional Hierarchies

How and why did humans develop this branching hierarchical structure that is so different from the dominance hierarchies exhibited by other animals? Addressing this question is important because it will inform our view of the present.

If social hierarchies provide net benefits to the whole population, as the *functionalist* school proposes [69, 70], then a neoclassical approach to income distribution might make sense. We might suppose that the high income of elites is due to the large services they provide to the rest of society. However, if the *conflict* school is correct and social hierarchies mostly benefit elites [39, 49, 50, 71, 72], then a power-based approach to income distribution makes more sense. Under such a theory, we suppose that the high income of elites is due to their status and power.

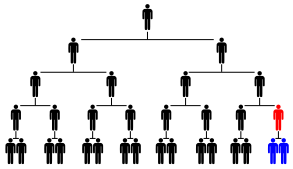
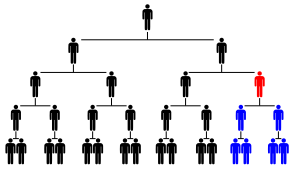
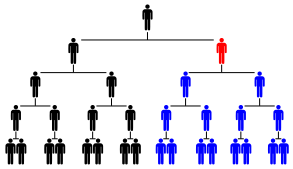
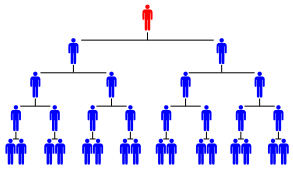
Hierarchical Rank	Linear Hierarchy	Number of Subordinates	Branching Hierarchy	Number of Subordinates
2		1		2
3		2		6
4		3		14
5		4		30

Figure 2: How the Number of Subordinates Changes with Hierarchical Rank in Linear vs. Branching Hierarchies

This figure visualizes the difference between the linear hierarchies observed in animals, and the branching hierarchies that exist in modern human societies. Hierarchical levels are numbered from the bottom up. For illustration purposes, I use a branching hierarchy with a span of control of 2. Subordinates are indicated by blue, while the individual in question is shown in red.

Anthropologist Antonio Gilman argues that the most compelling case against the functionalist theory of social stratification is the ubiquity of *inherited status* in human history:

> A shared feature of the few archaic states for which adequate documentary sources exist is a hereditary nobility: *alii* (Hawaii), *pilli* (Aztec), *orejones* (Inca), etc. Membership in these groups is by ascription and grants a small minority wealth disproportionate to their numbers (i.e., preferential access to resources). ... The functionalist account of the development of elites may be criticized at once for its *failure to explain the hereditary character of the [upper class]*. [73] (emphasis added)

The historical and archaeological record clearly indicates that inherited status has deep roots. There is tentative archaeological evidence beginning in the neolithic era [74–76], and widespread evidence for inherited status beginning in the bronze age around 5000 years ago [77–81]. It is around this time that the first Egyptian dynasty formed [82], followed later by dynasties in Mesopotamia [83] and China [84]. Since then, as Gaetano Mosca observes, the existence of a hereditary ruling class has been the norm:

> There is practically no country of longstanding civilization that has not had a hereditary aristocracy at one period or another in its history. We find hereditary nobilities during certain periods in China and ancient Egypt, in India, in Greece before the wars with the Medes, in ancient Rome, among the Slays, among the Latins and Germans of the Middle Ages, in Mexico at the time of the Discovery and in Japan down to a few years ago. [85]

The ubiquity of inherited status certainly favors a conflict approach to the formation of social hierarchy. But if hierarchy primarily benefits elites, then why did it arise?

3.4 A Darwinian Approach

Based on Darwinian reasoning, I propose the following hypothesis:

> **Hypthesis**: Institutional hierarchies arose as a means for individuals to differentially increase their *reproductive success*.

From the purview of evolutionary biology, an organism exists solely to propagate its genes into future generations [86]. If a certain type of behavior reliably leads to greater reproductive success (over many generations), we expect that organisms will develop an instinct for this behavior. This reasoning can explain why virtually all social animals form dominance hierarchies: individuals have developed an instinct to seek status because status reliably leads to increased reproductive success.

Darwinian theory, as adapted by W.D. Hamilton, can also explain why humans would instinctively seek *inherited* status, rather than a meritocracy. In addition to propagating their genes *directly* by reproducing, Hamilton observed that individuals could *indirectly* propagate the same genes by aiding the reproduction of close relatives [87]. Given this theory of 'inclusive fitness', we expect that humans would instinctively seek high status for themselves *and* for their relatives, because both will aid in the propagation of their genes.

To summarize, the tenets of evolutionary biology allow us to explain both hierarchy and inherited status as social structure that develops out of individuals' attempts to increase their reproductive success.

3.5 Kinship Ranking as a Possible Origin of Institutional Hierarchy

Although the social practices of prehistoric humans are difficult to determine, I think it is plausible that the origin of institutional hierarchies can be traced to the tradition of *kinship ranking*, a practice common to many non-state societies [88–95].

In this type of system, status is determined by proximity in descent to a common ancestor. The resulting social structure is exactly what we would expect from Darwinian theory if dominant individuals created a system to preferentially benefit the reproductive success of themselves and their close relatives. In Figure 3, I explore how the practice of kinship ranking can produce the bottom heavy social structure typical of modern institutions.

We begin with a 5 generation ranked line of descent from a founding ancestor (Fig. 3A) . For simplicity, we assume that only the *last* generation is alive,

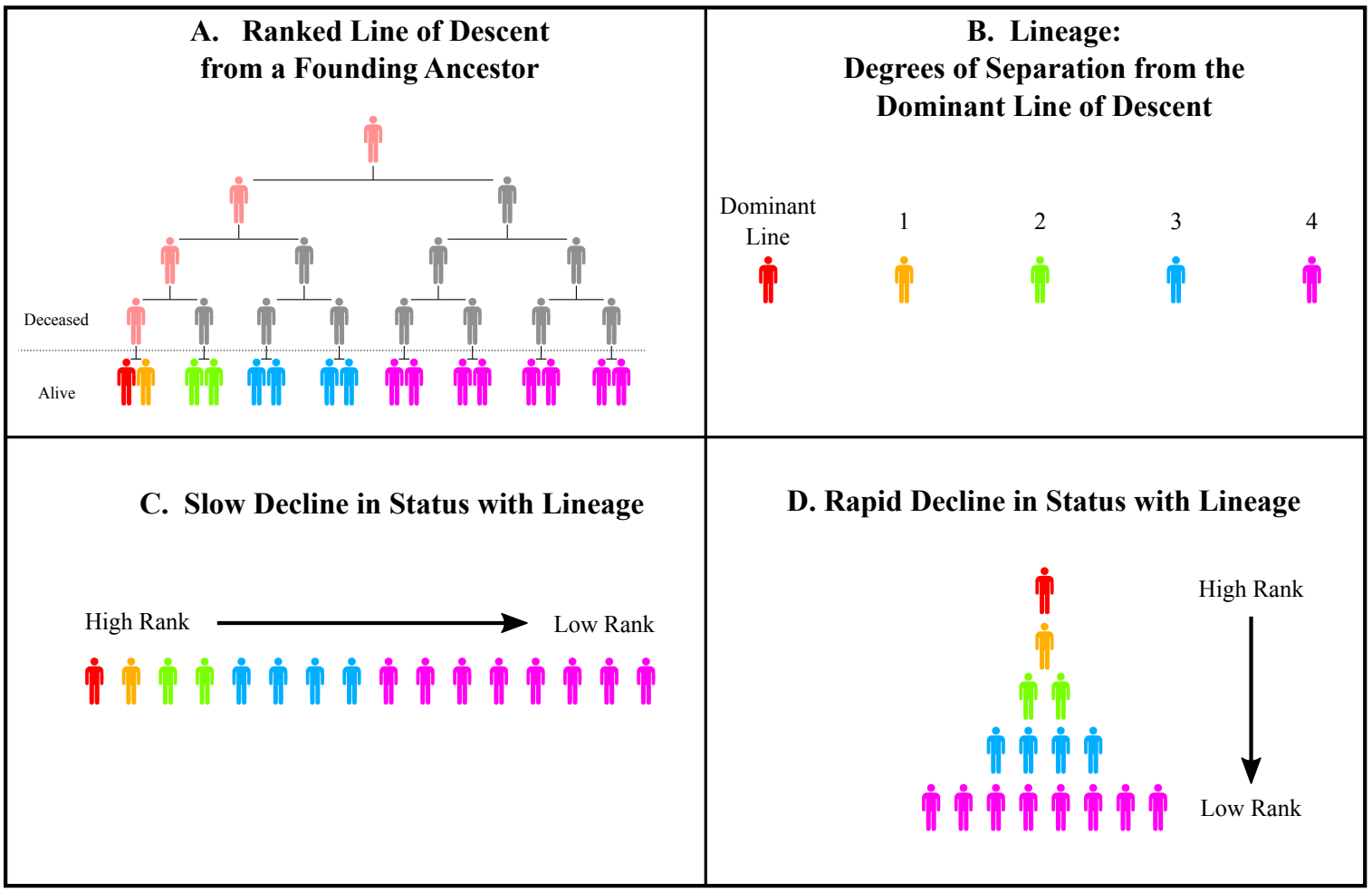


Figure 3: Kinship Ranking as a Plausible Origin of Branching Hierarchies

This figure shows how kinship ranking can generate a bottom heavy hierarchy that may explain the origin of human branching hierarchies. Panel A shows a unilineal family tree that begins with a founding ancestor. The dominant line of descent is shown in red, which in many societies would correspond to the first-born line. After 5 generations, we then rank individuals according to their separation from the dominant line. For simplicity, we assume that only the last generation is living. Panel B shows the color code corresponding to the degrees (generations) separating the living individual from the dominant line. The type of social structure produced by this kinship ranking system depends on the rate at which status declines with separation from the dominant line. If status declines *slowly*, as in panel C, then kinship ranking produces a relatively flat hierarchy, similar to the linear dominance hierarchies of non-human primates. However, if status declines rapidly, then kinship ranking produces a bottom-heavy structure that echoes the shape of modern institutional hierarchies.

and that individuals in all previous generations each had two children. I use color to indicate the degree of separation from the dominant line of descent among these living individuals (see code in Fig. 3B).

The resulting social structure depends on the *rate* at which status drops by descent. If status drops *slowly* with relatedness to the dominant line, then kinship ranking produces a linear hierarchy (Fig. 3C). However, if status drops *rapidly* with relatedness to the dominant line, kinship ranking generates a bottom-heavy hierarchy with the shape we expect of institutional branching hierarchies (Fig. 3D). What factors might affect the rate that status drops with lineage? As I explain in the following section, there is good evidence suggesting that this has to do with the particular sex (male/female) through which lineage is traced.

3.6 Matriliny, Patriliny, and Differential Reproductive Success

Kinship ranking seems to contains the seeds for both linear and branching hierarchies. Given this dualistic tendency, why then was the emergence of institutional hierarchy a *singular* event, and not something that occurred repeatedly over the last 100 000 years?

Again, I think this can be explained in Darwinian terms, but this time by looking at the differences between the *sexes*. Kinship ranking is almost always *unilineal*, meaning descent is traced through only one sex. Does it matter if descent lines are *patrilineal* (through the male line) versus *matrilineal* (through the female line)? From the standpoint of differential reproductive success, the answer is yes.

This is because there are profound differences in how status and wealth affect differential reproduction success in women versus men. For women, there is little reproductive reason to accumulate status and wealth beyond a certain point: once a women reaches the maximum rate of about one child per year, extra status and wealth will not increase her reproductive success. The same is *not* true for males, whose reproductive success is limited primarily by the number of sexual partners. By using status and wealth to attract multiple

mates, males can *vastly* increase their reproductive success. Thus, in a *matrilineal*, female-controlled group, Darwinian theory predicts that there should be very little incentive for inequality since dominant females have little to gain from it. However, in a *patrilineal*, male-controlled group the reverse is true: there is significant incentive for inequality because it can lead to tremendous differential reproductive success for dominant males.

If this theory is correct, then the origin of institutional hierarchy should coincide with a switch from *matriliny* to *patriliny*, and a corresponding increase in *polygyny* (the male practice of having many sexual partners). By piecing together evolutionary, archaeological, anthropological, and genetic evidence, a strong case can be made for this chain of events.

3.7 Evidence for the Switch from Matriliny to Patriliny

More than 100 years ago, Morgan [96] and Engels [97] proposed that human society evolved practicing matriliny, but that the rise of civilization had led to a switch to patriliny. While this theory fell out of favor during the 20th century, a variety of new evidence now supports Morgan and Engel's hypothesis.

In an evolutionary context, matriliny is the norm among mammals. For instance several other species of primate (macaques, baboons and vervets) exhibit inherited social status. However, in each of these species, status is inherited from the *mother*, not the father [98, 99]. More generally, among most social mammals it is related *females* (not males) that form the stable nucleus of social groups — at maturity (or earlier) males are forced to leave [100]. While it is surely *possible* for humans to have evolved in a patrilineal state, the evolutionary evidence suggests that it is *improbable*.

Recent genetic analysis also supports Morgan and Engel's hypothesis. Human DNA sequencing has now been used to infer historic male/female rates of migration, which can then be used to draw conclusions about historical inheritance practices. How? Matrilineal societies have very different sex-based migration patterns than patrilineal societies. In matrilineal societies, females tend to remain where they are born, while males migrate to find mates. How-

ever, in patrilineal societies the reverse is true. DNA evidence now indicates that historical sex-bias in migration is tied to the presence of agriculture. Existing hunter-gatherer societies have had much higher *male* migration rates than their agrarian counterparts [101, 102], suggesting that the former have a history of matriliny and the latter a history of patriliny.

DNA evidence also indicates that hunter-gathers (but not agrarian societies) show signs of matrilineal fertility inheritance (in which differential reproductive success is passed through the female line). This indicates that status was passed down through women [103]. Since existing hunter-gatherer societies are generally regarded as having the deepest connection with humanity's ancestral state, this DNA evidence supports the conclusion that humans evolved practicing matriliny.

The DNA evidence suggests that the loss of matriliny was tied to the spread of agriculture — a conclusion also supported by anthropological evidence. In the 1960s, David Aberle found that matriliny was most common amongst hunter-gatherers and horticultural societies [104]. More recent work indicates that it was specifically the introduction of *pastoral* agriculture (livestock) that caused the loss of matriliny [105]. To summarize, there are multiple lines of evidence suggesting that matriliny was humanity's ancestral state, and that the rise of agriculture led to a switch to patriliny.

3.8 Differential Reproductive Success: The Rise of Polygyny

We now turn to evidence for a rise in polygyny that accompanied the spread of agriculture and the loss of matriliny. Genetic researchers have recently been able to use DNA analysis to infer that the spread of agriculture led to a *massive* rise in polygyny.

The rate of polygyny has a predictable effect on the relative genetic diversity of the Y-chromosome (when compared to the diversity of maternal inherited mitochondria DNA). When rates of polygyny are high, the majority of offspring are sired by a minority of men, causing relative Y-chromosome diversity to be quite low. When rates of polygyny are low, most men sire offspring, causing

relative Y-chromosome diversity to be high. DNA analysis of a wide variety of ethnic populations now indicates that there was a roughly *tenfold decrease* in Y-chromosome diversity during the neolithic era, the period when agriculture first began [106, 107]. This suggests that the spread of agriculture led to a massive *increase* in polygyny.

In addition to the DNA evidence discussed above, many different types of anthropological evidence suggest a joint relation between the spread of agriculture, the switch to patriliny, an increase in polygyny, and the rise of social hierarchy. On the polygyny-patriliny front, societies with high rates of polygyny tend to have more male-biased systems of inheritance than societies with low rates of polygyny [108]. There is also evidence that patriliny and social hierarchy are related: analysis of modern inheritance patterns indicates that high status individuals are more likely than low status individuals to bequeath their wealth to male heirs [109].

On the polygyny-hierarchy front, L.L. Betzig has demonstrated that rates of polygyny within different traditional societies are positively correlated with the degree of social hierarchy [110]. Rates of polygyny (as measured by variance in male reproductive success) are also related to the presence/type of agriculture: societies that practice intensive and/or pastoral agriculture show much higher rates of polygyny than their hunter-gather/horticultural counterparts [111, 112]. Lastly, there is good archaeological evidence suggesting that the rise of social hierarchy coincided with the spread of agriculture [46, 113].

3.9 Big Picture Conclusions

To summarize this foray into the big picture of human inequality, the available evidence is consistent with a Darwinian explanation of the origins of social hierarchy. According to this approach, humans instinctively pursue status and power as a means for increasing reproductive success, and instinctively try to institute hereditary (rather than merit-based) systems of status as a means for increasing the reproductive success of their close relatives.

The origin of *institutional* (branching) hierarchy can then be plausibly explained in terms of the practice of kinship ranking, and a switch from matriliny to patriliny that coincided with the adoption of intensive agriculture. This new mode of production yielded dense resources that could be hoarded by dominant individuals. Under a matrilineal system, the incentive for resource hoarding was dampened because it would make little difference to female reproductive success. However, the rise of patriliny led to strong pressure for social stratification. Dominant males could use status and power to hoard resources, which then enabled them to vastly increase their reproductive success by attracting multiple (sometimes hundreds) of wives/concubines.

This Darwinian explanation suggests that institutional hierarchy formation has little to do with benefits to the wider population, and mostly to do with the self-interest of elites.[4] This provides a strong motivation for a power-based theory of income distribution. But while ancient history may be sordid — filled with despotic leaders who hoarded women, rigid caste systems dictating social status, obscene wealth concentrations in the hands of an ascribed aristocracy — many would argue that the modern era is *different*. The prevailing ethos is that we now live in a society that rewards *merit* not power. But if this is actually the case, it behooves us to test the power-income hypothesis, if only to *falsify* it.

[4] While the *motivation* for hierarchy formation may be rooted in the selfish pursuit of reproductive success, this does not preclude social hierarchy from providing benefits to the whole population. Indeed, in Darwinian fitness terms, the use of agriculture and social stratification was immensely beneficial: stratified agrarian societies were able to systematically displace/replace their hunter-gatherer counterparts. This may be related to the advantage that hierarchy gives for warfare — it allows large groups of people to operate as a cohesive unit [114, 115]. But Darwinian fitness benefits are *not* the same as increases in the material standard of living. For instance, if adult height is a reliable proxy for material affluence (a reasonable assumption under subsistence conditions), we can infer that agrarian masses (but not elites) had a significantly lower material affluence than hunter-gatherers [116].

4 A Power Theory of Personal Income Distribution

If not productivity, then what explains personal income? I propose the following hypothesis:

> **Hypothesis**: Income is most strongly determined by *social power*, as manifested by one's position within an institutional hierarchy.

In the following section I review the theoretical foundations for this approach.

4.1 Quantifying Hierarchical Power

All scientific theories require well-defined variables that can be objectively measured. One might protest that a power theory of income distribution does not meet criteria, since social power — the ability to influence or control others — is difficult to measure.

My proposed theory, however, is not about power in the *general* sense, but rather power in the *specific* context of an *institutional hierarchy*. Because of this restriction, objective measurement is made easier. The link between hierarchy and power is implicit in the etymology of the word 'hierarchy' itself, which derives from the Greek term *hierarkhs*, meaning 'sacred ruler' [117]. In essence, an institutional hierarchy is a nested set of power relations between a superior (a ruler) and subordinates (the ruled). Because of this ordered chain of command, I propose that power within a hierarchy can be quantified as follows:

> **Proposition**: Power within a social hierarchy is proportional to the number of subordinates under an individual's control.

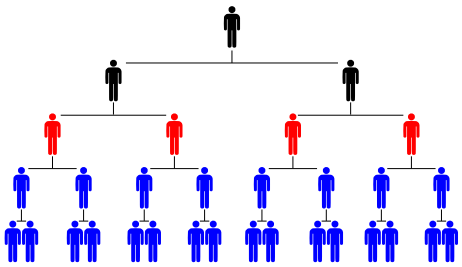
If we had access to the exact chain of command structure of an institution, we could use this definition to measure the power of each individual within a hierarchy. Unfortunately, chain of command information is rarely available.

Instead, existing case studies mostly report *aggregate* hierarchical structure only — total employment by hierarchical level. While we cannot calculate the power of *specific* individuals, we can use this data to calculate the *average* power of all individuals in a specific hierarchical level.

My implementation of this method is shown in equation 3, where $\bar{P}_h$ is the average power of individuals in hierarchical level h, and $\bar{S}_h$ is the average number of subordinates below these individuals. The 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.

$$\bar{P}_h = \bar{S}_h + 1 \tag{3}$$

The average number of subordinates $\bar{S}_h$ is equal to the sum of employment (E) in all subordinate levels, divided by employment in the level in question. Figure 4 shows a sample calculation, where red individuals occupy the level in question, and blue individuals are subordinates.



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

Figure 4: Calculating the Average Number of Subordinates

Using summation notation, we can write the following general equation for the average number of subordinates in level h:

$$\bar{S}_h = \sum_{i=1}^{h-1} \frac{E_i}{E_h} \tag{4}$$

Together, equations 3 and 4 allow us to define and measure the average power of individuals in an institutional hierarchy.

4.2 The Distribution of Power Within Firms

The principle deficiency of the productivist approach to income distribution is that observed inequalities in human productivity are systematically *too small* to account for observed levels of income inequality (Fig. 1). Does a power approach to income distribution avoid this under-explanation problem? To test if it does, we can use empirical data to estimate how unequally power is distributed within institutional hierarchies.

I assume that *business firms* are the dominant form of institutional hierarchy in capitalist societies. I have identified six firm case studies that provide adequate data to calculate average power by hierarchical level. These studies (discussed in detail in Appendix B) offer a sample of firms from the United States, Britain, the Netherlands, and Portugal. While a larger firm sample would be better, the proprietary nature of firm payroll data has proved a major obstacle to empirical research. As a result, data on firm hierarchical structure is quite limited.

The first step in the analysis is to use equations 3 and 4 to quantify average power by hierarchical level in each firm. We can then define the *distribution of power* for the entire firm by assigning each individual the average power in their respective hierarchical level. Figure 5 shows a conceptual example of this method, where the number above each individual indicates their hierarchical power.

Mathematically, the distribution of power (**P**) amounts to a vector in which average power in a specific level ($\bar{P}_h$) is repeated E_h times (the employment in that level):

$$\mathbf{P} = \left(\bar{P}_1 \overset{\times E_1)}{\cdots\cdots}, \ \bar{P}_2 \overset{\times E_2}{\cdots\cdots}, \ldots, \ \bar{P}_h \overset{\times E_h}{\cdots\cdots}\right) \tag{5}$$

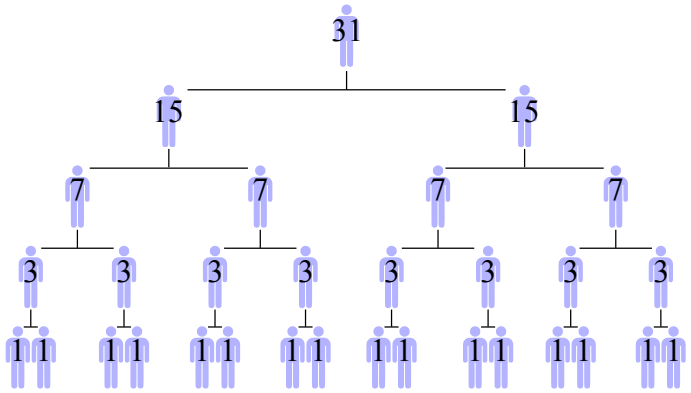


Figure 5: The Distribution of Power Within A Firm

This figure shows how power is distributed within a hypothetical firm. The number above each individual indicates their power, as defined by Eq. 3 and 4.

Given the distribution of power within a firm, we can then use the Gini index to quantify how unequally this power is distributed. In the example firm shown in Figure 5 the Gini index of power is 0.58.

I apply this method to each of the six case study firms (over all firm-year observations). Figure 6 shows the resulting distribution of power inequality within these firms and compares it to the distribution of income inequality within nation-states. Although this is a small sample size, it seems safe to conclude that hierarchical power within firms is distributed *more* unequally than income.

Thus, unlike the productivist approach, a power theory of income distribution based on hierarchy does not suffer from an under-explanation problem. Together with the historical picture discussed in section 3, this finding provides a strong motivation for the hypothesis that hierarchical power is the strongest determinant of income.

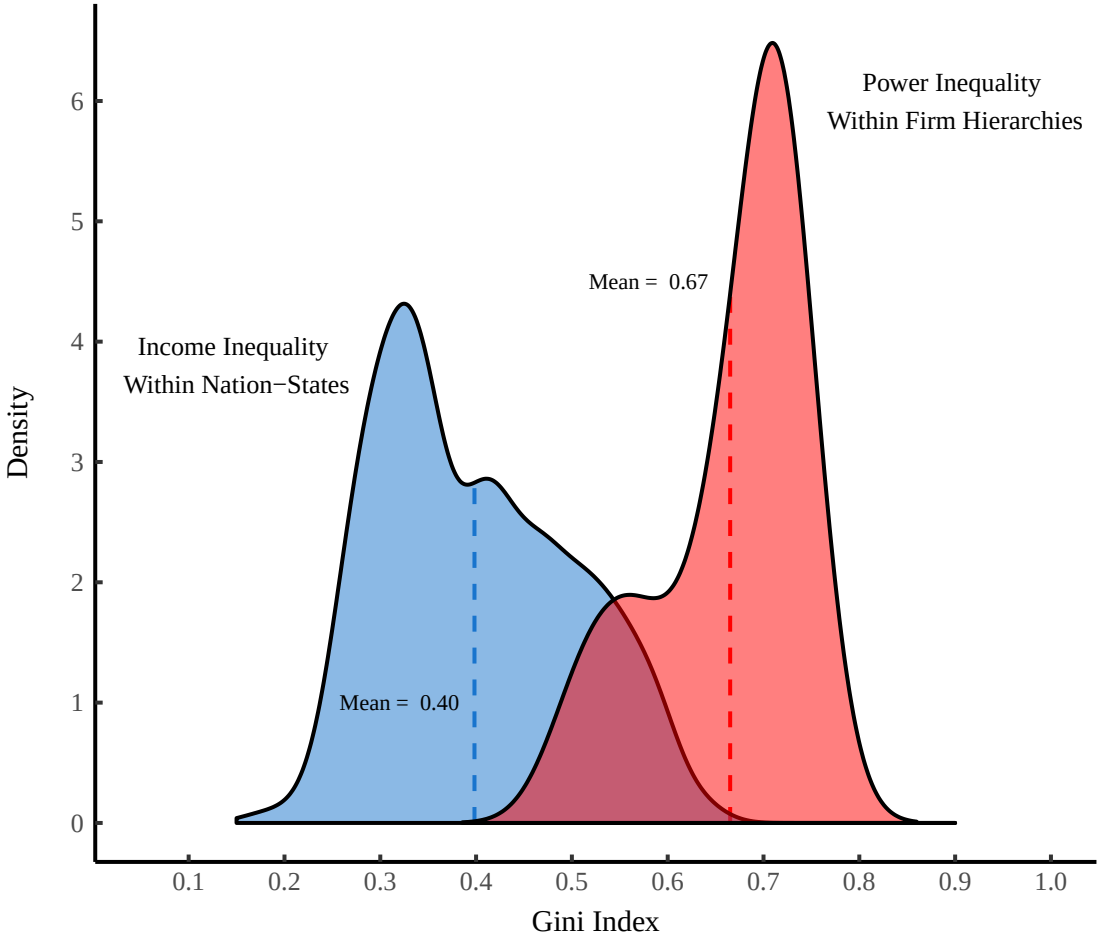


Figure 6: Income Inequality vs. Power Inequality within Firms

This figure compares the distribution of income inequality within nation states to the distribution of power inequality within six case study firms. To calculate the distribution of power, I use equations 3-5. Firm case study data comes from references [118–123] (for details, see Appendix B). The distribution of power inequality is calculated using Gini indexes for all firm-year observations. The distribution of income inequality within nation-states is calculated using all countries-year observations in the World Bank database series SI.POV.GINI.

5 Testing the Power-Income Hypothesis

To reiterate, the power-income hypothesis proposes that income is most strongly determined by *social power*, as manifested by one's position within an institutional hierarchy. To test this hypothesis, it is helpful to break it down into two parts:

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

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

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

5.1 Power-Income Correlation

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

I analyze the correlation between power and income in six case study firms — the same firms used in section 4.2 (for a detailed discussion of these studies, see Appendix B). For each firm in each observation year, I use equations 3 and 4 to calculate average power by hierarchical level. I then compare average power to average relative income by hierarchical level. In order to make comparisons across firms (and across time), I normalize all income data so that the mean income in the bottom hierarchical level is always equal to 1.

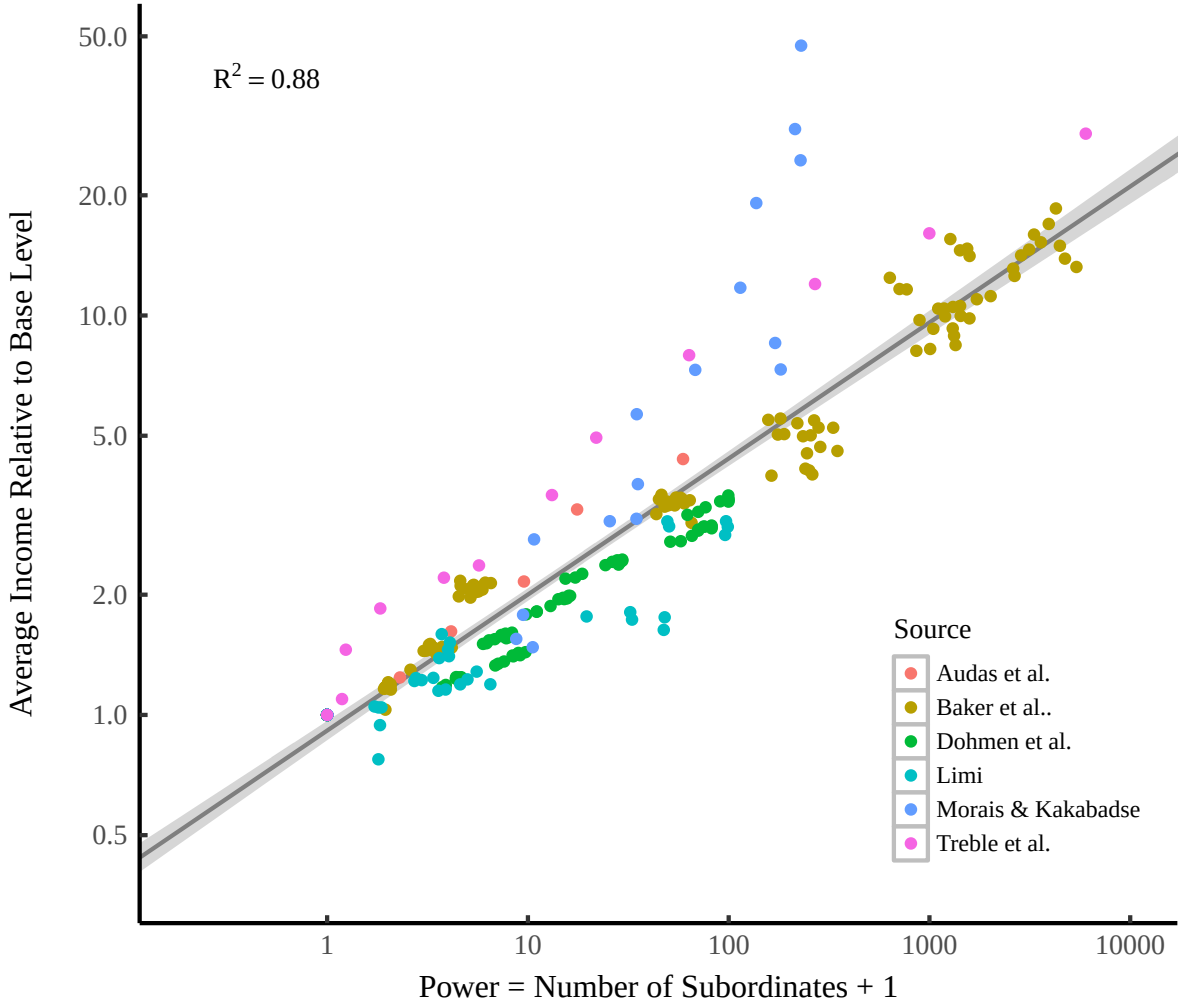


Figure 7: Average Income vs. Hierarchical Power Within Case-Study Firms

This figure shows data from six firm case studies [118–123]. The vertical axis shows average income within each hierarchical level of the firm (relative to the base level) , while the horizontal axis shows our metric for average power, which is equal to one plus the average number of subordinates below a given hierarchical level (see Eq. 3 and 4). Each point represents a single firm-year observation, and color indicates the particular case study. Grey regions around the regression indicate the 95% confidence region.

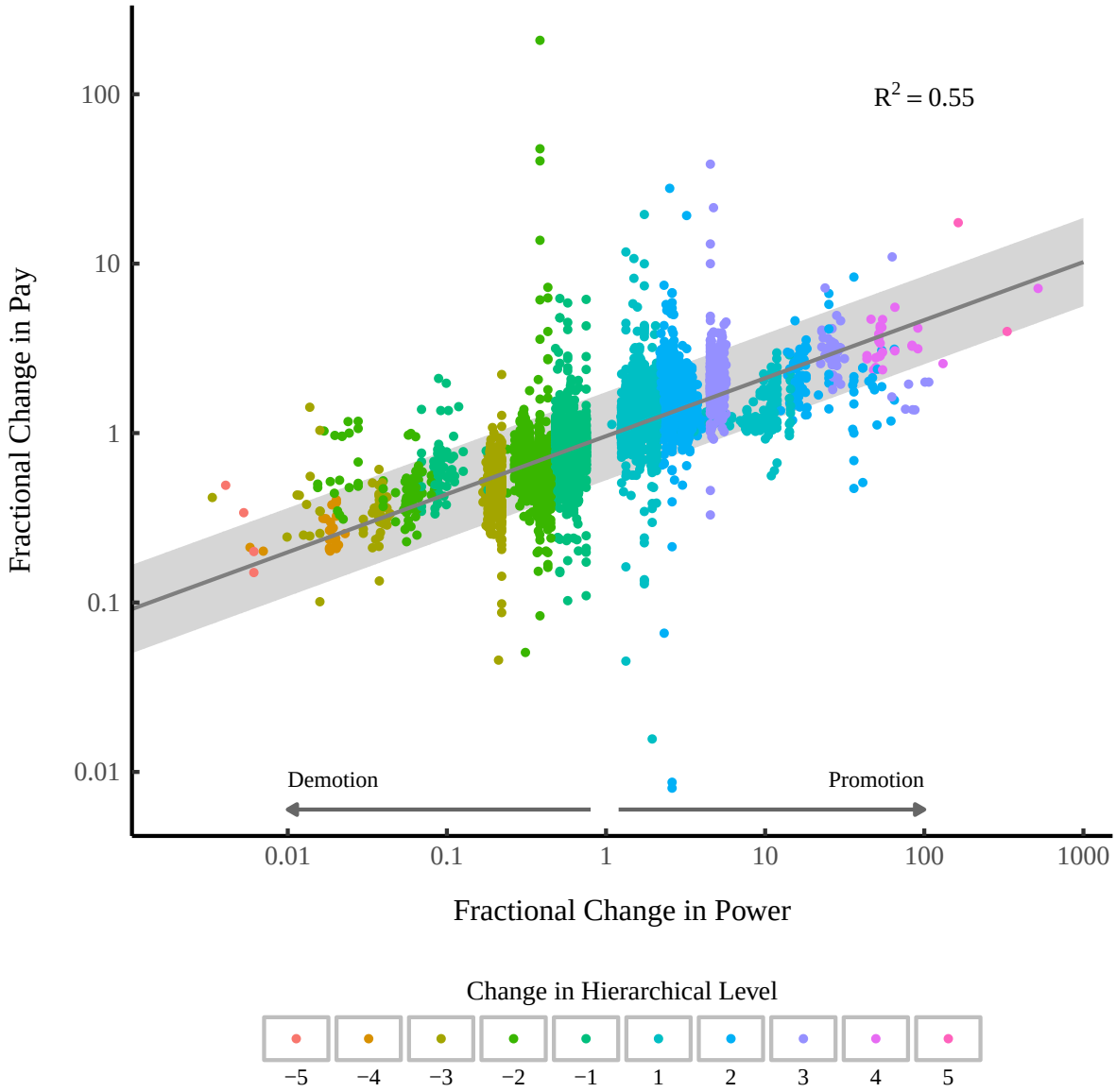


Figure 8: Changes in Power and Pay During Intra-Firm Promotions

This figure plots the fractional change in pay (Eq. 7) versus the fractional change in power (Eq. 6) for individual promotions/demotions in the Baker, Gibbs, and Holmstrom (BGH) dataset [119]. Each point represents the resulting change in pay and power of a *single* individual. Over 16,000 promotion/demotion events are plotted here. Change in hierarchical level is indicated by color. The grey region indicates the 95% prediction interval of a log-log regression. The BGH data comes from an anonymous US firm over the period 1969-1985. The dataset is available at http://faculty.chicagobooth.edu/michael.gibbs/research/index.html

The results of this analysis are shown in Figure 7. Each point represents a single firm-year observation, with the different case-study firms indicated by color. Although the firm sample is small, the evidence is conclusive: there is a strong correlation between relative income in our case-study firm hierarchies and our metric for power.

While Figure 7 shows a correlation between *static* levels of power and pay, it is also important to test for a *dynamic* correlation. That is, we want to know if *changes* in power are related to *changes* in income when individuals are promoted/demoted within a firm. I conduct such a test using the data published by Baker, Gibbs, and Homstrom [119] — the 'BGH dataset'. This dataset contains raw personnel data for a large US firm over the years 1969-1985.

I define a promotion/demotion as any change in an individual's hierarchical level. For each such event, we define the fractional change in power ($\Delta\bar{P}$) as the ratio of power *after* versus *before* the promotion/demotion (Eq. 6). An individual's power is defined by Eq. 3. Since we do not know the exact chain of command, I assign all individuals the *average* power of their respective hierarchical level.

$$\Delta\bar{P} = \frac{\bar{P}_{\text{after}}}{\bar{P}_{\text{before}}} \tag{6}$$

For each promotion/demotion, I define the fractional change in income (ΔI) as the ratio of income *after* versus *before* the event (Eq. 7). In order to isolate the effect of the promotion from the exogenous effects of inflation and/or general wage increases, I measure all incomes *relative* to the firm mean income ($\bar{I}$) in the appropriate year.

$$\Delta I = \frac{I_{\text{after}}/\bar{I}_{\text{after}}}{I_{\text{before}}/\bar{I}_{\text{before}}} \tag{7}$$

Figure 8 show the results of this dynamic analysis. Here each plotted point represents the fractional change in pay and power for the promotion/demotion of a *single individual*. For the over 16,000 promotions/demotion events ana-

lyzed here, a highly significant correlation exists between changes in power and changes in individual income.

Interestingly, the correlation holds both for promotions and for *demotions*, the latter occurring when an individual drops hierarchical levels. The relative pay reductions accompanying these demotions are difficult to understand from a productivist approach. Do these individuals suddenly experience a drastic reduction in ability/productivity? The evidence in Figure 8 suggests a better explanation: within the BGH firm, pay is largely a function of the power of specific hierarchical *position*, irrespective of the person holding this position.

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

5.2 The Strength of the Power-Income Effect

Hypothesis B states that power affects income more strongly than *any other factor*. To test this hypothesis, I use an analysis of variance method to quantify the income effect of a wide variety of different factors. In order to make the test as thorough as possible (given data constraints), some data is *model dependent* (see the Appendix for a detailed discussion).

5.2.1 Method for Measuring Effect Size

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

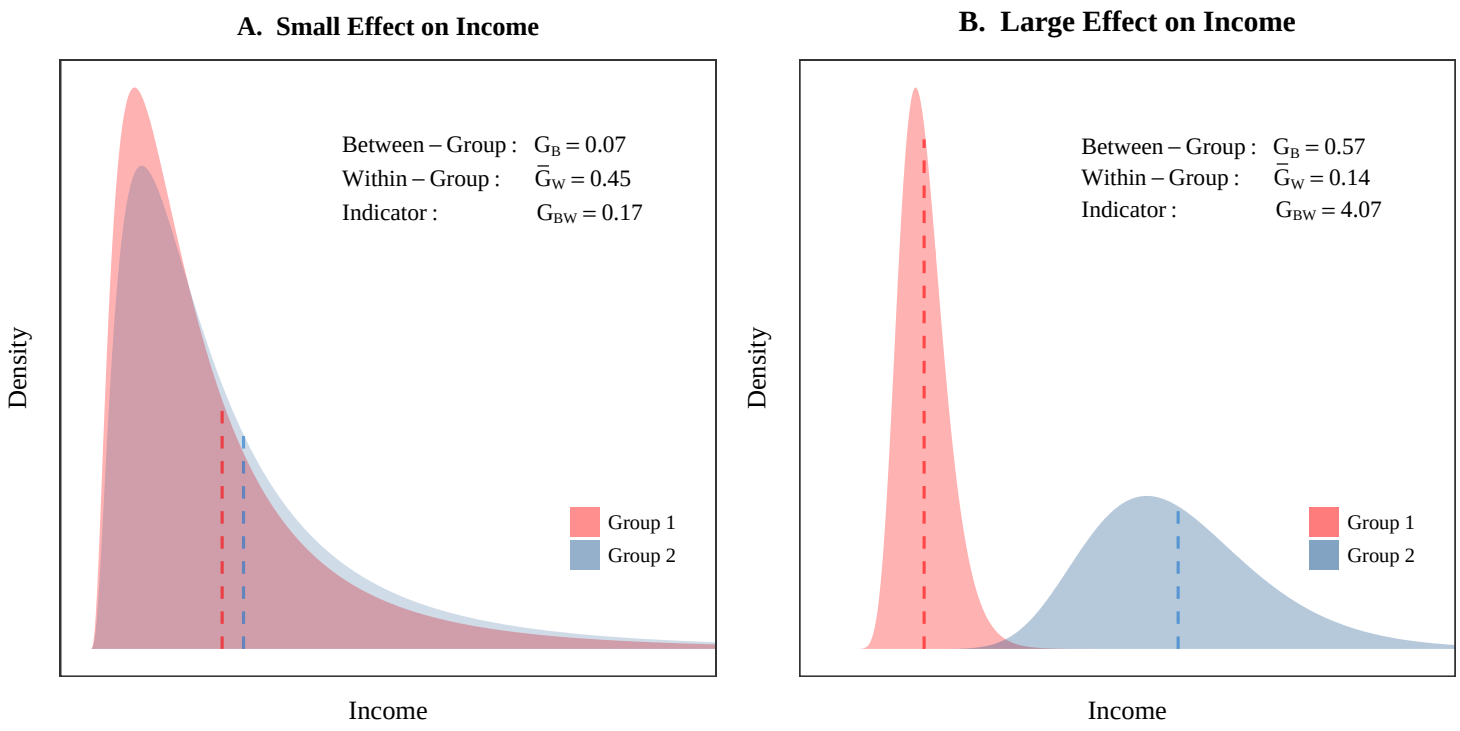


Figure 9: Analysis of Variance Using the Gini Index

This figure shows an example application of the analysis of variance method that uses the Gini index. A hypothetical 2-group variable (like 'sex') is illustrated to have a small effect on income in panel A and a large effect in panel B. The means of each distribution are indicated by a dashed line. We use equation 8 to define the between-within Gini indicator, G_{BW}. A small effect on income is indicated by a G_{BW} that is close to zero, while a large effect is indicated by a G_{BW} greater than one.

---

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

This approach is most easily understood by way of an example. Figure 9 shows a hypothetical example of how a two-group variable like 'sex' might affect income. When the separate income distributions of the two groups are plotted together, we can clearly see a *small* effect in Fig. 9A and a *large* effect

in Fig. 9B. How do we quantify the size of this effect? Most people likely judge the difference in group means against the dispersion within each group. We might call this a *signal-to-noise* ratio, where the 'signal' is the difference in group means and the 'noise' is the within-group dispersion. The larger the signal is relative to the noise, the larger the effect.

The ANOVA method allows us to generalize this concept of effect to more than two groups. The corresponding signal-to-noise ratio is often called Cohen's f^2. For this metric, the 'signal' is the dispersion *between* group means, while the 'noise' is the dispersion *within* groups (where dispersion is measure as the sum of squared differences from the mean) [124,125]. While Cohen's f^2 is a common measure of effect size, its calculation requires either raw data on individual income, or data for within-group *variance* (or *standard deviation*). Unfortunately, this type of data is difficult to obtain. Instead, what *is* readily available are aggregate statistics reporting within-group *Gini indexes*. Because of the ubiquity of the Gini index, I use it to measure effect size.

Similar to Cohen's f^2, my effect size metric is a signal-to-noise ratio (Eq. 8). However, rather than the sum of squares, I use the Gini index to measure both within-group and between-group dispersion. I call this metric the *between-within Gini ratio* (G_{BW}).

$$G_{BW} = \frac{G_B}{\overline{G}_W} \tag{8}$$

Here G_B is the between-group Gini index (the Gini index of group mean incomes), while ($\overline{G}_W$) is the average of all within-group Gini indexes. For a detailed discussion of the relation between G_{BW} and f^2 (and a more rigorous discussion of effect size) see Appendix H.

The value of G_{BW} can range from 0 to infinity, with larger values indicating a larger effect on income (see the example in Fig. 9). Of particular interest is the value $G_{BW} = 1$, which occurs when between-group dispersion is equal to within-group dispersion. Any factor that produces $G_{BW} > 1$ can be considered to have a significant impact on income, since inequality *between* groups is larger than inequality *within* groups. However the primary use of the G_{BW} metric is

not its *absolute* value, but its *relative* value when different income-affecting factors are compared.

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

$$G_n^{max} = \frac{n-1}{n} \tag{9}$$

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

To correct for this bias, I use the method proposed by George Deltas [126]. The bias-adjusted Gini index (G^{adj}) is defined by dividing the unadjusted Gini (G) by the maximum possible Gini (G_n^{max}), given the number of internal groups n:

$$G^{adj} = \frac{G}{G_n^{max}} \tag{10}$$

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

5.2.2 Grouping Individuals By Hierarchical Level

To test hypothesis B using an analysis of variance method, we must group individuals into different categories/classes of social power. My method is to group individuals by *hierarchical level* across all firms, as illustrated in Figure 10.

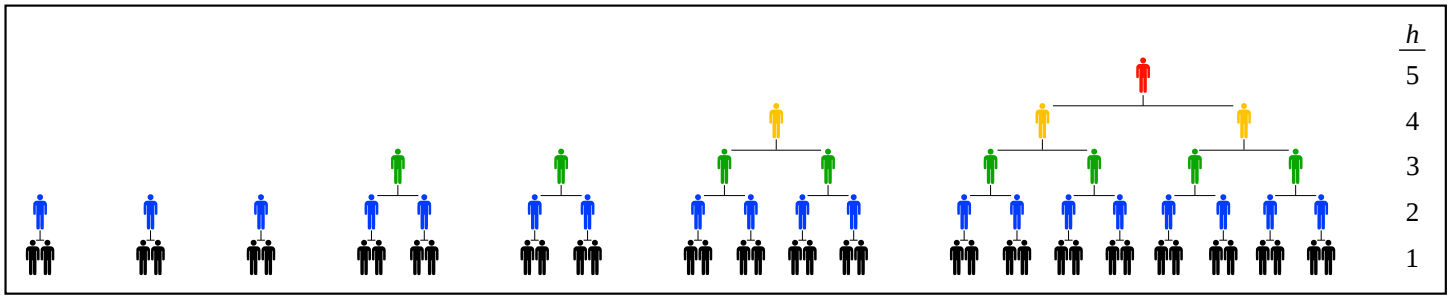


Figure 10: Grouping Power By Hierarchical Level

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

This method is theoretically attractive because hierarchical level is the principle determinant of power. If a firm has a constant 'span of control' (the number of subordinates below each superior), then power will increase *exponentially* with hierarchical level. In Figure 10, the span of control is constant both *within* and *between* firms. The result is that all individuals in each hierarchical level have the *same* power. In the real-world, we would expect this *not* to be the case. Evidence from firm case-study data suggests that the span of control varies both *within* and *between* firms (see Fig. 17 and 18 in Appendix B). As a result, we still expect that average power will increase exponentially with hierarchical level, but each hierarchical level will contain individuals with a *range* of different power.

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

5.2.3 The Data

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

Before proceeding with a discussion of the data sources used for income by hierarchical level, it is worth reviewing why I do *not* use the same case study data that was used to test hypothesis A. Testing hypothesis B requires grouping individuals by hierarchical level across a *large* number of firms. To be consistent, the firms should all be in the *same* country (ideally the United States), and the observations (that are compared) should be in the *same* year. The case study data does not meet these requirements: it is a *small* sample, with firms from many *different* countries with many *non-overlapping* years. As a result, the case study data is not useful for testing hypothesis B.

Table 1: Income-Affecting Factors Used to Test Hypothesis B

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

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

Instead, I use three different sources for estimating income distribution by hierarchical level. The first source is a seminal study by Mueller, Ouimet, and Simintzi [127] that reports income distribution by hierarchical level for 880 United Kingdom firms over the period 2004-2013. The second source is a study by Fredrik Heyman [128] that analyzes the pay distribution of the top 4 levels of management in 560 Swedish firms in the year 1995. Heyman's data comes with the caveat that it does not represent all hierarchical levels — just the top four. For this reason, I mark Heyman's results with an asterisks.

I use this non-US data because I am not aware of any equivalent US study that reports income distribution by hierarchical level over a large number of firms. While comparing US to UK/Swedish studies is not ideal, I proceed because of the lack of alternative data. If anything, the UK and Swedish data should lead to an *under*-estimate of the power-income effect in the United States. Why? Both the UK and Sweden have significantly less income inequality than the US (according to the World Bank, the most recent UK and Swedish Gini indexs are 0.33 and 0.27, while the most recent US Gini index is 0.46). If there is less *total* inequality, the potential for *between*-group inequality is diminished, resulting in a lower G_{BW} metric (see Eq. 8).

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

The idea of the Compustat Model is that firm case-study data can be used to make generalizations about the hierarchical employment and pay structure of firms. Although different firms have differently shaped hierarchies (see Fig. 17 in Appendix B), there are underlying regularities shared by all firms. The following regularities are shown in Figure 18 in Appendix B:

1. The span of control tends to *increase* with hierarchical level.
2. The ratio of average pay between adjacent hierarchical levels *increases* by level.
3. Intra-level inequality tends to be *constant* across all hierarchical levels.

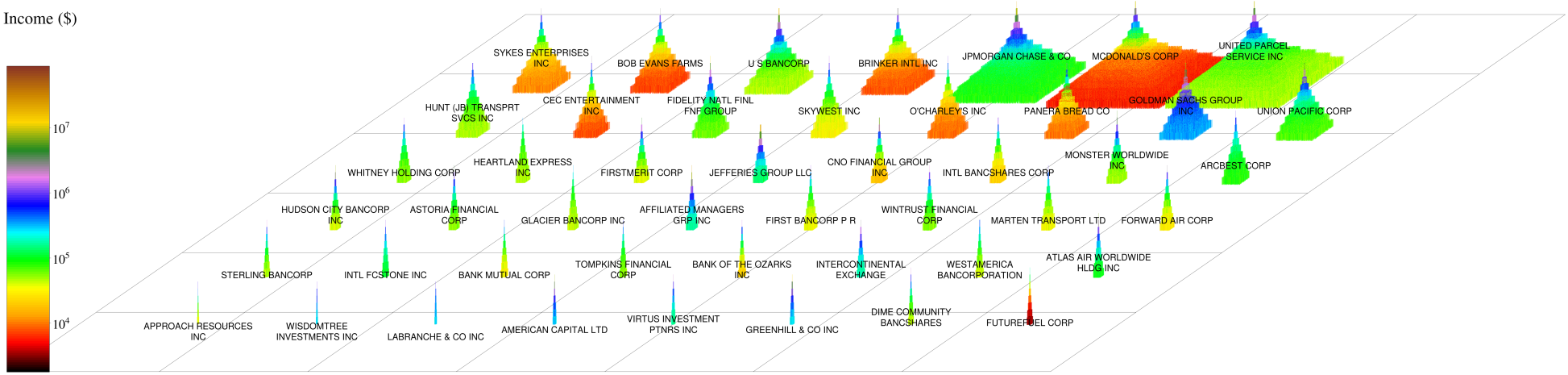


Figure 11: Visualizing the Compustat Model

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

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

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

In conjunction with case-study regressions, this Compustat data can be used to estimate the hierarchical pay structure of individual US firms (see Appendix E and F). While the details of the model are complex, the core idea is simple: since the CEO sits at the top of corporate hierarchy, his/her *relative* pay (when compared to the average pay of all employees) gives an indication of the rate at which income increases by hierarchical level. When paired with assumptions about the 'shape' of the firm (derived from case-study regressions), the model gives an unambiguous prediction about firm internal pay structure. Results of the model are visualized in Figure 11 for selected firms in 2010.

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

In addition to income distribution by hierarchical level, I also use the Compustat model to estimate the strength of the *firm*-income effect (how much

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

5.2.4 Results

The results of the analysis of variance test of hypothesis B are shown in Figures 12 and 13. Figure 12 shows the between-within Gini ratio (G_{BW}) for our 19 different income-affecting factors. For all factors except religion and cognitive score, the boxplots indicate the variation of G_{BW} over time (typically the last 20 years). For religion, the boxplot range indicates uncertainty in the G_{BW} estimate, while for cognitive score, it indicates variation between different studies.

Figure 13 shows the same data, but in a slightly different format. The G_{BW} metric consists of a *ratio* of between-to-within group income dispersion (Eq. 60). Figure 13 decomposes this ratio and shows the individual components of the metric — between-group inequality (G_B) and within-group inequality ($\bar{G}_W$). Aside from religion, density plots indicate the distribution of these values over time (for religion, density plots indicate uncertainty). The important information here is the relative position of between-group inequality relative to within-group inequality.

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

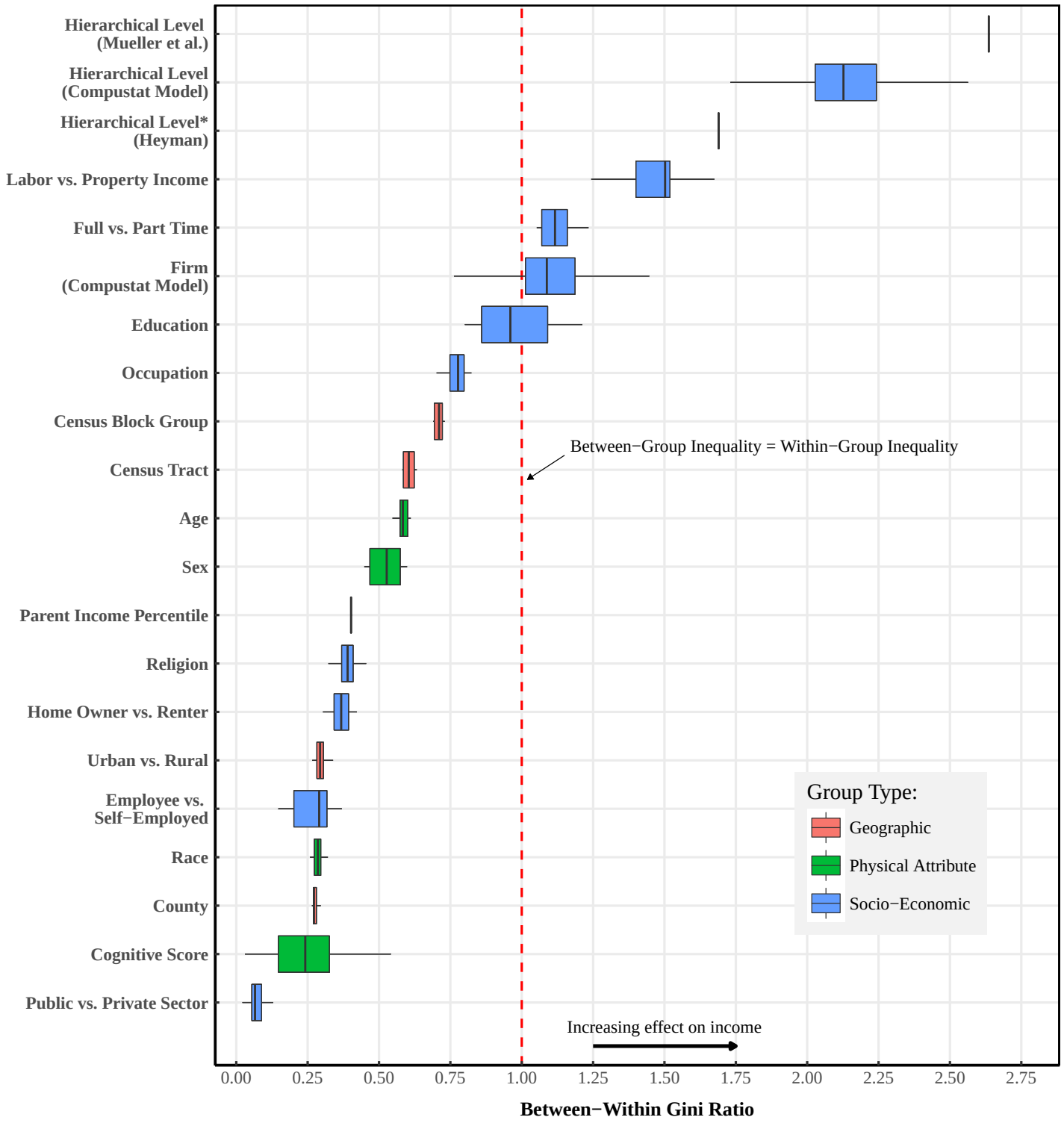


Figure 12: The G_{BW} Ratio for Different Income-Affecting Factors

This figure shows the results of an analysis of variance test of hypothesis B using the method outlined in Sec. 5.2.1. The horizontal axis shows the between-within Gini ratio (G_{BW}) defined by Eq. 8 (G_B is adjusted for bias using Eq. 10). A larger G_{BW} indicates a greater effect on income. The box plots indicate the total range (horizontal line), 25th to 75th percentile range (the box), and the median (vertical line). With the exception of hierarchical level data from Mueller et al. [127] and Heyman [128], all data is from the United States. For sources and methods, see Appendix A.

$*$ Includes only top 4 hierarchical levels

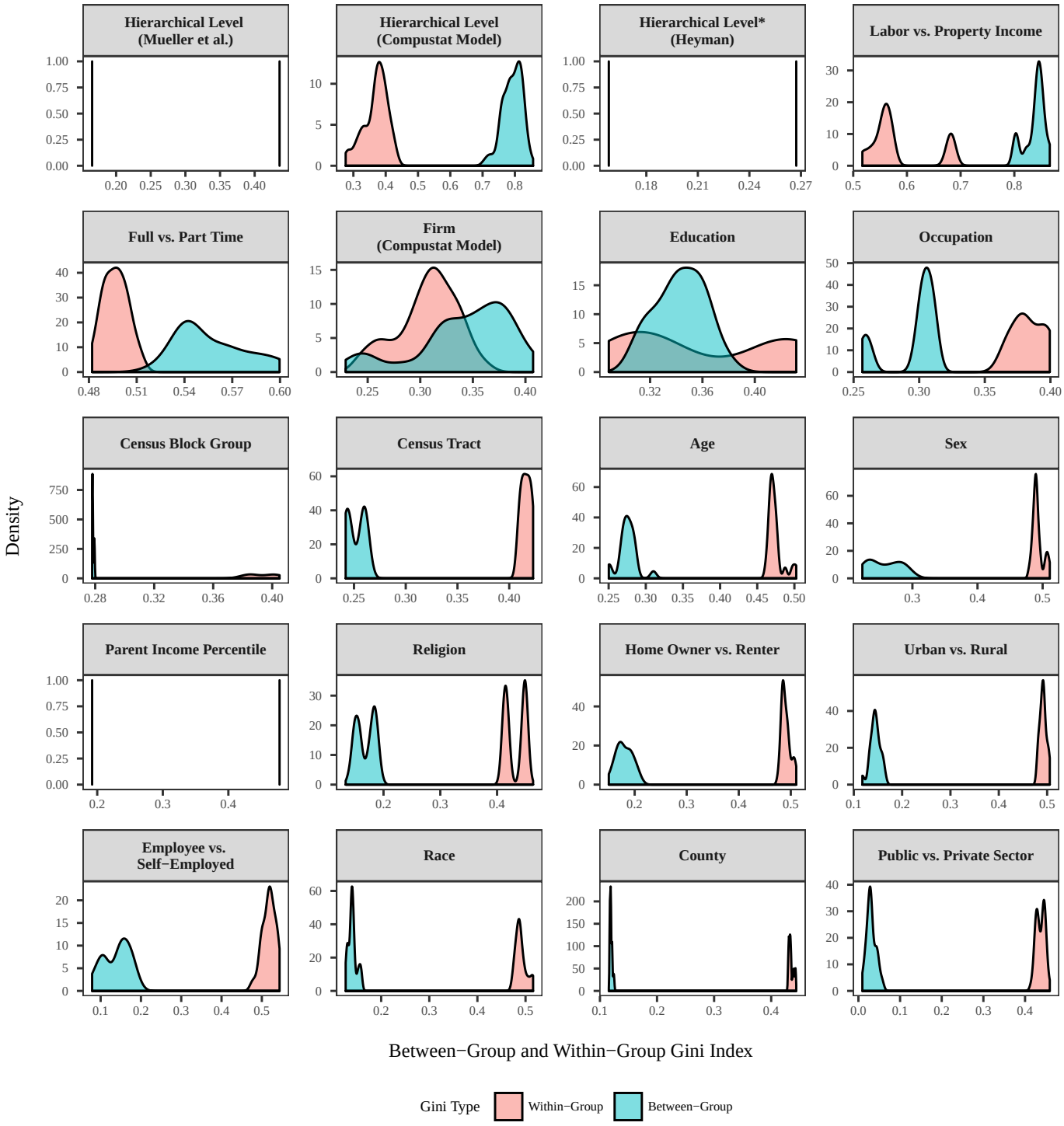


Figure 13: The G_B and G_W Index for Different Income-Effecting Factors

This figure shows distribution of *between*-group Gini indexes (G_B, the Gini index of group mean incomes, shown in blue) in relation to the distribution of *within*-group Gini indexes ($\bar{G}_W$, average within-group inequality, shown in red). Each panel plots the results for a different income-affecting factor. With the exception of 'parent income percentile' and 'religion', the density curves represent the distribution of data over different years. Panels are sorted by effect size, declining (column-wise, then row-wise) from the top left to the bottom right. With the exception of hierarchical level data from Mueller et al. [127] and Heyman [128], all data is from the United States. For sources and methods, see Appendix A.

$*$ Includes only top 4 hierarchical levels

In addition to the support for the power-income hypothesis, Figures 12 and 13 reveal a few other notable findings. Firstly, physical attributes (age, cognitive score, race, and sex) have a relatively insignificant effect on income. Geographic effects are also quite small, although they become larger as geographic area *decreases* (geographic factors ranked from largest to smallest area are: county, tract, block group).

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

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

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

To compare the income-effect of functional income type, what we really need to do is group individuals by the *proportion* of income coming from property sources. Based on the work of Piketty [1], it is reasonable to expect that

this would strongly affect income. However, I do not include such a test here because I am not aware of relevant data sources.[5]

Hypothesis B is stringent in proposing that power has the *strongest* effect on income. The evidence presented here demonstrates that hierarchical level (our grouping of power) affects income more strongly than any of the other 18 factors tested. It should be noted that no empirical analysis can ever prove definitively that some as yet *unmeasured* factor does not have a stronger effect on income. However, science proceeds on the best *available* evidence, and on this front our power-income hypothesis survives empirical testing.

6 Conclusions

This paper has proposed a new theory of personal income distribution based on the *power structure of human institutions*. I have hypothesized that social power, as measured by the number of subordinates within a hierarchy, is the strongest determinant of income.

I have tested this hypothesis in two parts. I first used firm case-study data to demonstrate both a *static* and *dynamic* correlation between income and hierarchical power. I then used an analysis of variance method to quantify the income-effect of a wide variety of different factors. To test the power-income hypothesis, I grouped individuals by hierarchical level, and found that this grouping affected income more strongly than any other factor. Given this evidence, I have concluded that there is good empirical support for a power theory of personal income distribution.

What does this theory mean for the study of income inequality? In my view, the lens of power provides a coherent way of studying income/resource distribution at *any* point in human, not just our own capitalist epoch. The drive to accumulate power and to then use power to accumulate resources

[5] In Fig. 8.10, Piketty [1]) shows how the proportion of capitalist income increases with income fractile in the United States. But this grouping is the reverse of what would be required to apply my analysis of variance method. Piketty groups individuals by income size, while the method used here would require grouping individuals by the proportion of capitalist income.

is, I believe, deeply wired into the human psyche. I have argued that this drive can plausibly be explained by Darwinian principles, as a mechanism for achieving differential reproductive success. This hypothesis is supported by a wide variety of genetic, anthropological, and archaeological evidence.

What does a power theory of income distribution mean for political economy at large? As I argued in section 2, negating productivity as a cause of income logically necessitates abandoning much of existing political economic theory, including neoclassical and Marxist theories of functional income distribution. This is not to say that my proposed theory is inconsistent with *all* political economic thought. In particular, it is consistent with seminal work of Nitzan and Bichler [21], who offer a new framework for political economy based on the hypothesis that capital is a *symbolic quantification of power*. Connecting the present theory with Nitzan and Bichler's *capital as power* framework is a fruitful path for future work.

I conclude by offering some thoughts on the ideological implications of a power theory of personal income distribution. Regardless of their scientific merit, all theories of income distribution evoke some form of human ethics that either justifies *redistribution*, or justifies the *status quo*. Productivist theories of income distribution illicit an ethics of fairness — "*To each according to what he and the instruments he owns produces*", as Milton Friedman famously put it [129]. The effect of this theory is to justify as fair any conceivable distribution of income. The result is an innate bias towards the status quo, whatever it may be.

A power theory of income distribution is very different. If we parallel Friedman's language, we might state that a power theory elicits the following ethos: '*To each according to his/her power to take*'. Few would argue that this is fair — it is the basic recipe for *despotism*. But if a power theory of income distribution is correct, then acts of income redistribution can be considered merely as *checks* on power — no different than the checks and balances that form the governmental basis of most liberal democracies.

Appendix

Supplementary materials for this paper are available at the Open Science Framework repository:

https://osf.io/ytr3b/

The supplementary materials include:

1. Data for all figures appearing in the paper;
2. Raw source data;
3. R code for all analysis;
4. Compustat model code.

A Data Sources

Age

Age mean income and within-group Gini index data is from US Census Tables PINC-02 over the years 1994-2015. Age is grouped into the following 4 categories: 18-24, 25-44, 45-64, 65 and older.

Census Blocks

Census blocks data comes from the US Census American Community Survey (ACS) over the years 2010-2014. This data is tabulated at the *household* (rather than individual) level. Neither mean household income nor household Gini index data is directly available from the ACS at the census block level. I calculate mean household income by dividing aggregate household income by the number of households.

Within group Gini indexes are estimated from binned income data using the R 'binequality' package. I construct two different estimates: one using a parametric method and the other using the midpoint method. For the parametric

method, I fit either a lognormal or gamma distribution (whichever is best) to the binned data. Gini indexes are then calculated from this fitted distribution. The midpoint method uses midpoints of the bins to estimate the Gini index. The midpoint of the upper bin (which has an open upper bound) is estimated from a best-fit power law (again, implemented in the R binequality package). Both Gini estimates are used in Figures 12 and 13. The R code implementing this method is included in the Supplementary Material.

Census Tracts

Census tract data comes from the US Census American Community Survey (ACS) over the years 2010-2015. Mean income data comes from series S1902, while intra-tract Gini indexes come from series B19083.

Cognitive Score

The between-within indicator for cognitive score is estimated using data from Figure 6 in Bowles et al. [130]. Bowles' figure presents 65 different estimates (from 24 studies between 1963 and 1992) of the relation between individual income and cognitive score. The strength of this relation is quantified using the beta coefficients (β) of a log-linear regression. This coefficient represents the slope of the regression equation shown in Eq. 11, where the logarithm of income ($\log(I)$) and cognitive score (S) have first been normalized to have a mean of 0 and standard deviation of one.

$$\log(I) = \alpha + \beta S \tag{11}$$

I use Engauge Digitizer to extract data from Bowles' graph. I then use a model to estimate the G_{BW} metric from Bowles' reported beta coefficients. The model creates a stochastic log-linear scaling relation between income and cognitive score. By adjusting the strength of this relation, we can create modeled data that has an equivalent beta coefficient to any of the points in Bowles' figure. I then use the model to calculate a G_{BW} for this beta coefficient.

The model assumes that cognitive score (S) is a normally distributed random variate with a mean of 100 and standard deviation of 15:

$$S \sim \mathcal{N}(100, 15) \tag{12}$$

We assume that the natural log of mean income ($\ln \bar{I}$) scales exponentially with cognitive score (Eq. 13). Since there is no evidence that extreme IQs lead to extreme incomes (at either the bottom or top end), I do not include them in the model. I model only those individuals with scores that are within two standard deviations of the mean ($70 < S < 130$). The parameter a determines how strongly cognitive score affects average income.

$$\ln(\bar{I}) = a(S - 70) \quad \text{for} \quad 70 < S < 130 \tag{13}$$

We assume that individual income (I) is a stochastic variable that is distributed according to a lognormal distribution defined by the location parameter μ and scale parameter σ:

$$I \sim \ln \mathcal{N}(\mu, \sigma) \tag{14}$$

Equation 15 shows how mean income $\bar{I}$ is related to μ and σ.

$$\bar{I} = e^{u + \frac{1}{2}\sigma^2} \tag{15}$$

By taking the logarithm and solving for μ, Eq. 15 can be transformed into the following:

$$\mu = \ln(\bar{I}) - \frac{1}{2}\sigma^2 \tag{16}$$

We then substitute Eq. 13 into Eq. 16 to define μ in terms of cognitive score:

$$\mu = a(S - 70) - \frac{1}{2}\sigma^2 \tag{17}$$

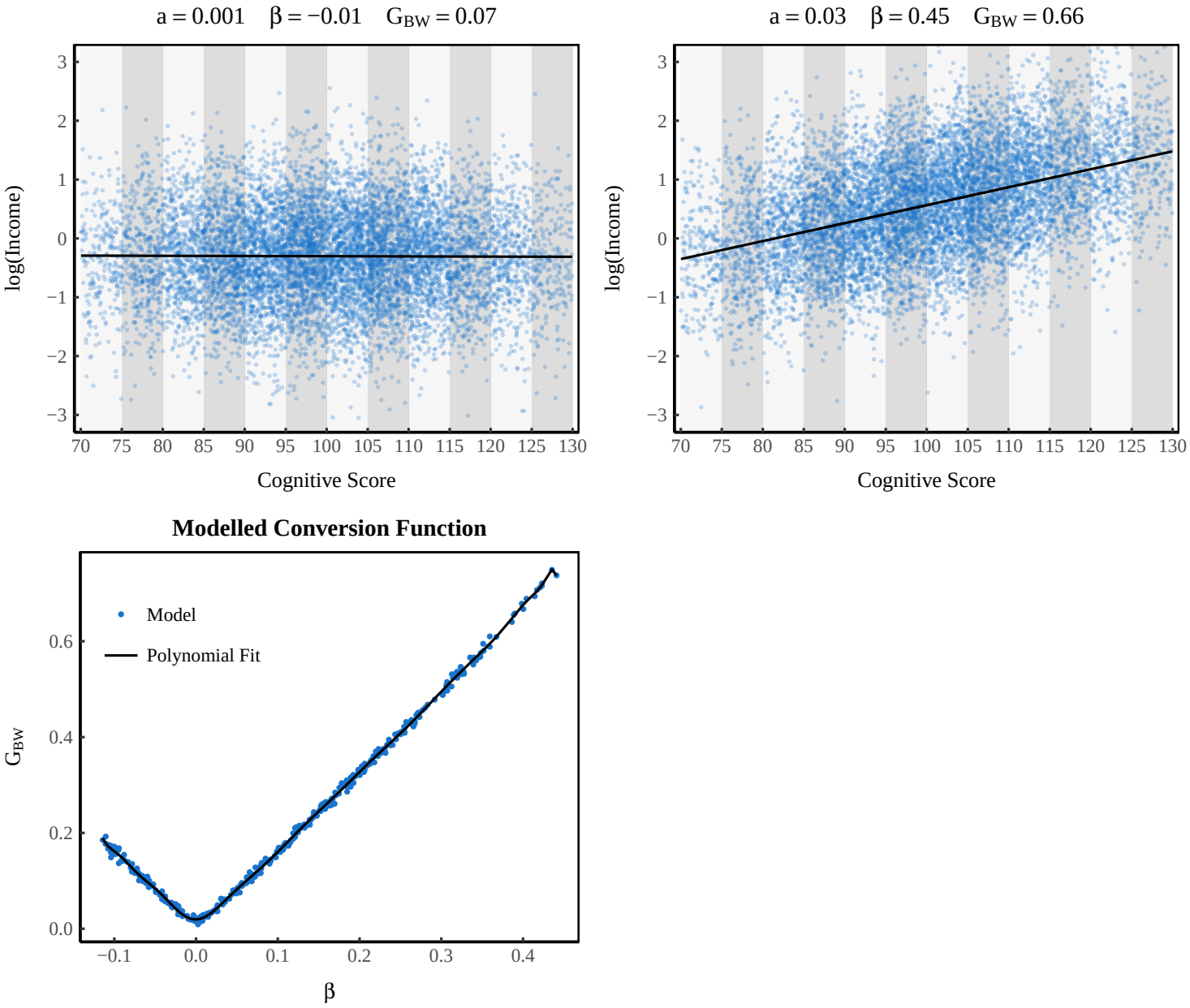


Figure 14: Cognitive Score Method — Estimating the Between-Within Indicator (G_{BW}) from Normalized Regression Coefficients (β)

This figure shows an example of the model for converting cognitive score regression data from Bowles et al. [130] to the G_{BW} indicator. Using equations 12-17, I create a stochastic scaling relation between the logarithm of individual income and cognitive score. The strength of this scaling relation is determined by the parameter a, and is quantified by the normalized regression coefficient β. The top left panel shows a weak scaling relation, while the top right shows a strong scaling relation. I then group individuals into cognitive score intervals of 5 (vertical grey bars) and calculate the G_{BW} metric. The bottom left panel shows the resulting relation between G_{BW} and β that is used to convert Bowles' data.

The algorithm for the model is as follows. We first generate a random cognitive score S, drawn from the normal distribution defined by Eq. 12. We then take this score and use Eq. 17 to define the parameter μ. Finally, we generate a random income for this cognitive score, drawn from the lognormal distribution defined by Eq. 14. This process is then repeated as many times to generate a stochastic dataset relating income to cognitive score.

The model has 2 free parameters: a and σ. Parameter a affects the rate at which income scales with cognitive score, while σ determines the amount of dispersion around the mean income $\bar{I}$. The parameter σ strongly affects the level of 'global' inequality in the model, while a has only a slight effect. For this reason, it is important to chose σ such that the model has a realistic level of inequality. I chose $\sigma = 0.8$. Over the chosen range of $-0.007 < a < 0.03$, this produces global Gini indexes that range between 0.43 and 0.47, which is roughly consistent with US data for the second half of the 20th century.

For any given value of a, the model generates a stochastic relation between cognitive score and income I. Two examples are shown in Figure 14. In Figure 14A, the small value of a produces a very weak relation between income and cognitive score. In Figure 14B, the larger value of a produces a stronger relation between income and cognitive score.

The strength of the relation is indicated by the beta coefficient β. The purpose of this model is to convert the values of β reported by Bowles et al. into the between-within Gini ratio that is used in this paper. To make this conversion, we must group individuals by their cognitive score. The bin-size of this grouping is arbitrary; I construct groupings of 5 point cognitive score intervals (indicated by the grey vertical bands in Fig. 14A-B). For each group, we calculate the mean income and within-group Gini index. The between-within Gini metric G_{BW} is then calculated by the method outlined in section 5.2.1 of the main paper.

I repeat this process for many different values of a, which produces the modeled relation between G_{BW} and β shown in Figure 14C. I then fit this relation with a high order polynomial that serves as the function for converting Bowles' β values into the G_{BW} values used in this paper.

Counties

US County data comes from the American Community survey for the years 2006-2015. County Gini indexes are from series B19083, while mean income is from series S1902.

Education

Mean income and within-group Gini indexes by educational level come from US Census tables PINC-03 over the years 1994-2014. Educational level is categorized into the following groups:

- Less Than 9th Grade
- 9th to 12th Nongrad
- High school Graduate (Incl GED)
- Some College
- Associate Degree
- Bachelor's Degree
- Master's Degree
- Professional Degree
- Doctorate Degree

Employees vs. Self-Employment

To calculate mean income and intra-group Gini indexes for employees and self-employed workers, I use US Census table PINC-07 between 1994 and 2015. This table contains three categories: *Government Wage And Salary Workers*, *Private Wage And Salary Workers*, and *Self-Employed Workers*. Table 2 shows how I have mapped these categories onto the 'employees' and 'self-employed' sectors.

Table 2: Grouping Categories of Census Table PINC-07

Employees	Self-Employed
• Government Wage And Salary Workers	• Self-Employed Workers
• Private Wage And Salary Workers	

Self-employed mean income and within-group Gini index come directly from PINC-07. To calculate the mean income of employees, I use the average of the means of government workers and private workers, weighted by the size of each group.

Since Gini indexes are not additive, I estimate the inequality among employees from binned data. I first add the binned income counts of both government and private wage/salary workers to get a binned income distribution for all 'employees'. From this binned data, I then use the the R 'binequality' package to estimate private sector Gini indexes.

I construct two different estimates: one using a parametric method and the other using the midpoint method. For the parametric method, I fit various theoretical distributions to the binned data. Gini indexes are then calculated from the best-fitting distribution. The midpoint method uses midpoints of the bins to estimate the Gini index. The midpoint of the upper bin (which has an open upper bound) is estimated from a best-fit power law (again, implemented in the R binequality package). Both Gini estimates are used in Figures 12 and 13. The R code implementing this method is included in the Supplementary Material.

Firms

Firm between-within inequality calculations use the Compustat database, and are a combination of empirical and modeled data. Firm mean income is calculated directly from Compustat data by dividing Total Staff Expenses (series XLR) by the number of employees (series EMP). Firm internal inequality is estimated using the Compustat Model. See Appendix B-G for a detailed discussion.

Full and Part Time Workers

Full and part time worker mean income and within-group inequality data comes from US Census tables PINC-05 from 1994-2015.

Parent Income Percentile

'Parent income percentile' refers to grouping individuals by the income percentile of their parents. My calculations are done using Table 1 and 2 from the online data tables of Chetty et al. [131] (a seminal study of US intergenerational mobility). For every parent income percentile x, Table 1 gives the probability $p(x,y)$ that the corresponding child will have an income in percentile y. Table 2 gives the mean income ($\bar{I}_y$) of each child percentile y.

My method for estimating group mean incomes and within-group inequality is shown in equations 18 and 19. The first step is to convert the probability $p(x,y)$ into an integer $w(x,y)$ that can be used to weight incomes. Since the probabilities in Table 1 contain 7 decimal places, I multiply $p(x,y)$ by 10^7 (Eq. 18).

$$w(x,y) = p(x,y) \times 10^7 \tag{18}$$

For each each income percentile x, we then create a vector of child incomes ($\mathbf{I_x}$) by repeating each child percentile mean income $\bar{I}_y$ by the weighting factor $w(x,y)$. Here the notation $\overset{\times w(x,y)}{\cdots\cdots}$ indicates that the value $\bar{I}_y$ is repeated $w(x,y)$ times.

$$\mathbf{I_x} = \left(\bar{I}_1 \overset{\times w(x,1)}{\cdots\cdots}, \bar{I}_2 \overset{\times w(x,2)}{\cdots\cdots}, \ldots, \bar{I}_{100} \overset{\times w(x,100)}{\cdots\cdots}\right) \tag{19}$$

We can think of $\mathbf{I_x}$ as an estimated income distribution for children of parents in income percentile x. Mean income and within-group inequality of parent group x are then estimated by calculating the mean and Gini index (respectively) of $\mathbf{I_x}$.

Note that this method neglects the income dispersion *within* each child income percentile (Chetty et al. do not provide this data). Thus, our estimated Gini index will have a slight *downward* bias. The R code implementing this method is included in the Supplementary Material.

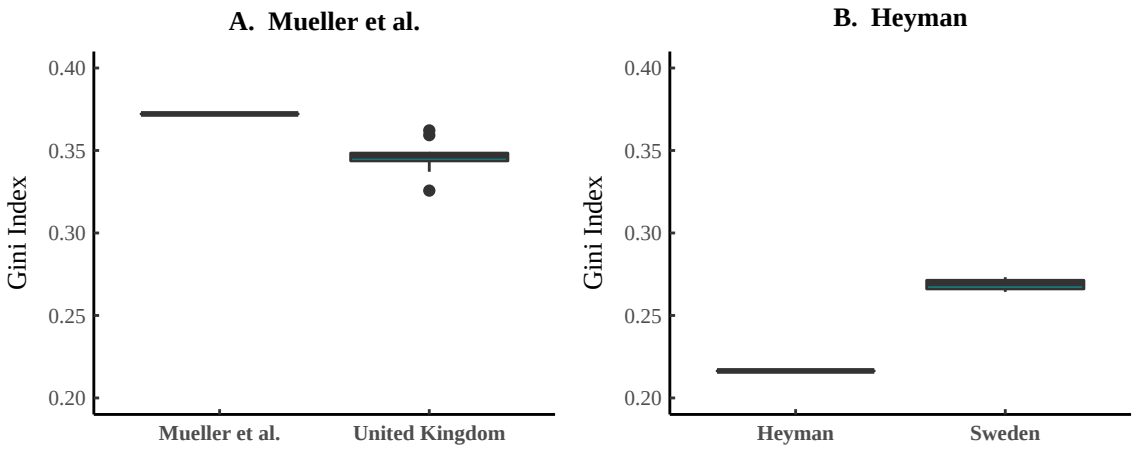


Figure 15: Aggregate Inequality Implied by Hierarchy Data

This figure compares levels of inequality implied by the Mueller et al. and Heyman firm samples against the inequality in their respective countries. UK inequality data is over the period 2004-2013, the same as covered by Mueller's data. Heyman's study covers the year 1995, while Swedish data is from 2004-2013. UK and Sweden Gini data is from the World Bank, series SI.POV.GINI.

Hierarchical Level — Heyman

This data comes from Fredrik Heyman's [128] study of 560 Swedish firms in the year 1995. His dataset includes only the top 4 levels of management. I include Heyman's results in the paper with the caveat that his data does not represent all hierarchical levels.

Heyman (Table A.1) provides the mean and standard deviation of the *logarithm* of incomes in each level. I estimate mean income ($\bar{I}$) and Gini index (G) by hierarchical level by assuming that intra-hierarchical level income is log-normally distributed. Under this assumption, the *mean* of log income is equal to the lognormal location parameter μ, while the *standard deviation* of log income is equal to the scale parameter σ. Equations 20 and 21 then define the mean income and Gini index (respectively) of each hierarchical level.

$$\bar{I} = e^{\mu+\frac{1}{2}\sigma^2} \tag{20}$$

$$G = \text{erf}\left(\frac{\sigma}{2}\right) \tag{21}$$

Figure 15 shows how the implied aggregate inequality within the Heyman's sample compares to Swedish empirical data. Heymans's sample implies a bit less inequality than the empirical data. This is not surprising, however, as Heyman's data includes only the top 4 levels of management.

Hierarchical Level — Mueller et al.

This data comes from Mueller et al. [127], who study the hierarchical pay structure of 880 United Kingdom firms over the period 2004-2013. For each hierarchical level, Mueller et al. provide the mean income as well as the 25th, 50th, and 75th income percentiles. To estimate intra-level inequality, I adapt R code written by Andrie de Vries to find the best-fit theoretical distribution for each hierarchical level. Intra-hierarchical level inequality is then calculated from the best-fit distribution.

Figure 15 shows how aggregate inequality within the Mueller et al. sample compares to UK data over the same period. Although the Mueller et al. data is slightly more unequal than the UK as a whole, it is a reasonably representative sample.

Hierarchical Level — Compustat Model

The Compustat model is discussed extensively in Appendix B-G.

Labor and Property Income

'Labor' income is defined as wages and salaries, while 'property' income is defined as the sum of interest, dividends, rents, royalties, and estates or trust

income. Mean income and within-group inequality data comes from US Census tables PINC-08 from 2003-2015.

Occupation

Data for mean income and with-group inequality by occupation comes from US Census tables PINC-06 (income by occupation of longest job) between 2007 and 2015. This table classifies occupations by major type, minor type, and detailed type. I use *detailed* categories only, which amounts to between 53 to 55 different occupation groups (depending on the year).

The US Bureau of Labor Statistics also publishes occupational wage estimates (available at https://www.bls.gov/oes/tables.htm). For the sake of completeness, I analyze this data here, but do not use it for the results published in the paper. The BLS data differs from Census data in the ways shown in Table 3.

Because the BLS does not report within-occupation Gini indexes directly, I estimate them via the reported values for 10th, 25th, 50th, 75th, and 90th income percentiles. Using an adaption of R code written by Andrie de Vries, I fit a variety of theoretical distributions to this percentile data. Within-occupation Gini indexes are calculated from the best-fit theoretical distribution.

The resulting between-within inequality indicator is shown in Figure 16A, alongside the results from Census occupation data. The two calculations differ starkly. Census data indicates that between-occupation inequality is *less* than within-occupation inequality; however, the BLS data indicate the *reverse*.

Which result is correct? The answer to this question depends on the type of income inequality we are interested in explaining. The BLS data covers only full-time, non-self-employed workers earning labor income. Census data, on the other hand, includes *all* individuals. For the purposes of this paper, the Census data is a better choice.

To demonstrate the differences between BLS and Census data, we can calculate the aggregate inequality that is implied by the data. To do this, I make the simplifying assumption that all occupations have lognormal income dis-

Table 3: Contrasting the US Census and BLS Occupational Income Data

Census Data	**BLS Data**
Includes self-employed workers	Does *not* include self-employed workers
Income for all full and part-time work	Income is for full-time equivalent workers only (hourly wage × 2080 hours).
Includes non-labor income	Does *not* include non-labor income
53-55 detailed occupation types	700-800 detailed occupational types
Reports Gini index directly	Reports 10th, 25th, 50th, 75th and 90th income percentiles

tributions. Given the mean income ($\bar{I}$) and within-group Gini index (G) of a particular occupation, we can define the lognormal location (μ) and scale (σ) parameters:

$$\sigma = 2 \cdot \text{erf}^{-1}(G) \tag{22}$$

$$\mu = \ln(\bar{I}) - \frac{1}{2}\sigma^2 \tag{23}$$

If the number of individuals engaged in this occupation is n, we can create a simulated occupational income distribution by generating n values from the lognormal distribution defined by μ and σ. We repeat this process for every occupation, and then aggregate all of the simulated occupational income distributions. The Gini index of this aggregated distribution is the level of inequality that is *implied* by the data.

The results of this analysis are shown in Figure 16B. As expected, the inequality that is implied by Census data closely matches actual levels of inequality between all individuals. However, the inequality implied by BLS data is *much* lower — a clear result of the restrictions underlying the BLS methods. For the purposes of this paper, the Census data is the correct choice.

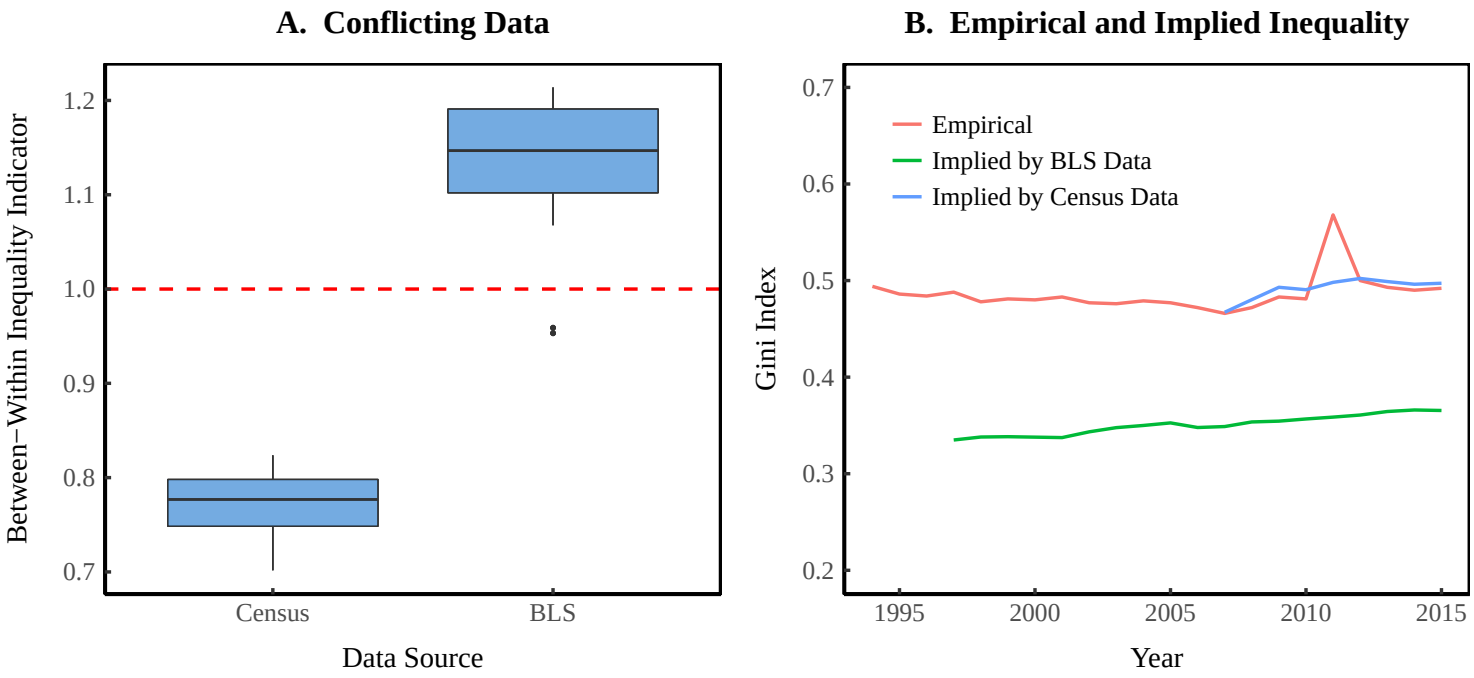


Figure 16: Inequality by Occupation — Data Discrepancies

This figure shows differences in the occupation income data published by the US Census versus that published by the US Bureau of Labor Statistics (BLS). Panel A shows calculations of the between-within indicator (G_{BW}) for both BLS and Census data. The BLS data gives a much higher G_{BW} value, meaning between-occupation inequality is far greater (relative to within-occupation inequality) in BLS data than it is in the Census data. Why? The two datasets imply very different levels of aggregate (society-wide) inequality, as shown in panel B. This is because the BLS data includes only full-time wage/salary earners, while the Census data includes all individuals. The level of aggregate inequality implied by the Census data closely matches actual levels. I use Census data only in this paper.

Owner vs. Renter

Mean income and intra-group Gini indexes by home-ownership status come from US Census table PINC-01 between 1994 and 2015. I use the following two categories: (1) Owner Occupied; and (2) Renter Occupied.

Public vs. Private Sector

To calculate mean income and intra-group Gini indexes for public and private sector workers, I use US Census table PINC-07 between 1994 and 2015. This table contains three categories: *Government Wage And Salary Workers*, *Private Wage And Salary Workers*, and *Self-Employed Workers*. Table 4 shows how I have mapped these categories onto the 'public' and 'private' sectors.

Table 4: Grouping Categories of Census Table PINC-07

Public Sector	**Private Sector**
• Government Wage And Salary Workers	• Private Wage And Salary Workers
	• Self-Employed Workers

The mean income and Gini index of the public sector is thus equivalent to the values for government wage/salary workers. Private sector mean income is calculated as the average of the means of private wage/salary worker income and self-employed worker income, weighted by the size of each group.

Since Gini indexes are not additive, I estimate the inequality of private sector income from binned data. I first add the binned income counts of both private wage/salary workers and self-employed workers to get a binned income distribution for the private sector. From this binned data, I then use the R 'binequality' package to estimate private sector Gini indexes.

I construct two different estimates: one using a parametric method and the other using the midpoint method. For the parametric method, I fit various theoretical distributions to the binned data. Gini indexes are then calculated from the best-fitting distribution. The midpoint method uses midpoints of the bins to estimate the Gini index. The midpoint of the upper bin (which has an open upper bound) is estimated from a best-fit power law (again, implemented in the R binequality package). Both Gini estimates are used in Figures 12 and 13. The R code implementing this method is included in the Supplementary Material.

Race

Data for mean income and within-group inequality by race comes from US Census tables PINC-01 between 1994 and 2015. Data for 2002-2015 contain the following four categories: Asian, Black, Hispanic, and White. Data for 1994-2001 contains only three categories: Black, Hispanic, and White.

Religion

Religion income data comes from the Pew Research Center 2007 U.S. Religious Landscape Survey (RLS). I use the following groups:

- Agnostic
- Atheist
- Baptist
- Buddhist
- Church of Christ, or Disciples of Christ
- Congregational or United Church of Christ
- Episcopalian or Anglican
- Hindu
- Holiness (Nazarenes, Wesleyan Church, Salvation Army)
- Jewish
- Lutheran
- Methodist
- Mormon
- Muslim
- Nondenominational or Independent Church
- Nothing in particular
- Orthodox
- Pentecostal
- Presbyterian
- Reformed (include Reformed Church in America; Christian Reformed; Calvinist)
- Roman Catholic

The RLS reports the binned income of each respondent. I use the R 'binequality' package to estimate group mean income and Gini indexes (using the midpoint method). Because some religions have a very small sample size, I use the bootstrap method [132] to estimate a plausible range of values for group mean incomes and intra-group income inequality.

Sex

Data for mean income and within-group inequality by sex (male/female only) comes from US Census tables PINC-01 between 1994 and 2015.

Urban vs. Rural

Data for urban/rural mean income and intra-group Gini index comes from US Census tables PINC-01 between 1994 and 2015. I define 'urban' as individuals inside metropolitan statistical areas, and 'rural' as individuals outside these areas.

B Hierarchical Structure and Pay Within Case-Study Firms

Purely based on worldly experience, most people would agree that firms are hierarchically organized, and that pay tends to increase as one moves up the hierarchy. But the exact structure of this hierarchy has not been widely studied.

Figure 17 shows the hierarchical employment and pay structure of six different firms whose data has been made available to social scientists. The firms remain anonymous, and are named after the authors of the case-study papers (see Table 7 for details). By and large, these studies confirm our basic intuition about firm structure. Although the exact shapes vary, all of the firms in Figure 17 have a roughly pyramidal employment structure and inverse pyramid pay structure.

To analyze the structure of these firms in further detail, I define and calculate the three metrics shown in Table 5. Results are shown in Figure 18. Figure 18A 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 18B 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 18C shows the *intra-level Gini index* as a function of hierarchical level. Unlike the other two quantities, intra-level in-

Table 5: Metrics of Firm Hierarchical Employment and Pay Structure

Name	Definition
Span of Control	Employment ratio between adjacent hierarchical levels.
Inter-Level Pay Ratio:	Ratio of mean pay between adjacent hierarchical levels.
Intra-Level Gini Index	The Gini index of income inequality within a specific hierarchical level of a firm.

come inequality seems to be more-or-less constant across all hierarchical levels (a linear regression reveals no significant trend).

As well as single-firm case studies, a handful of studies exist that have analyzed the hierarchical structure of multiple firms. Although these aggregate studies offer more scope than case studies, they have one major shortcoming: they rarely study the structure of *entire* firms. Instead, these aggregate studies typically focus on the span of control and pay in the *top* hierarchical levels of a firm (the CEO and adjacent upper management levels). But this is a problem. I have conceptualized firm hierarchy as a *bottom-up* ranking (see Fig. 10). Under this definition, a CEO in a small firm will be in a very different hierarchical level than a CEO in a large firm. Thus, when we compare the span of control between a CEO and his subordinates across firms of different size, we are likely comparing very different hierarchical levels.

As a result of this shortcoming, the aggregate studies summarized in Table 8 are less useful than the case studies in Table 7. However, keeping in mind their shortcomings, these aggregate studies still reveal the same trends as our case study data. Figure 19 shows the analysis of these aggregate studies. Note that hierarchical level is counted from the *top down*, where level 0 is the CEO. Figure 19A and B (respectively) indicate that there is still a tendency for the span of control and inter-level pay ratio to increase with hierarchical level.

From this evidence, I propose the following 'stylized' facts (Table 6) about firm employment and pay structure:

Table 6: Stylized Facts About Firm Employment and Pay

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

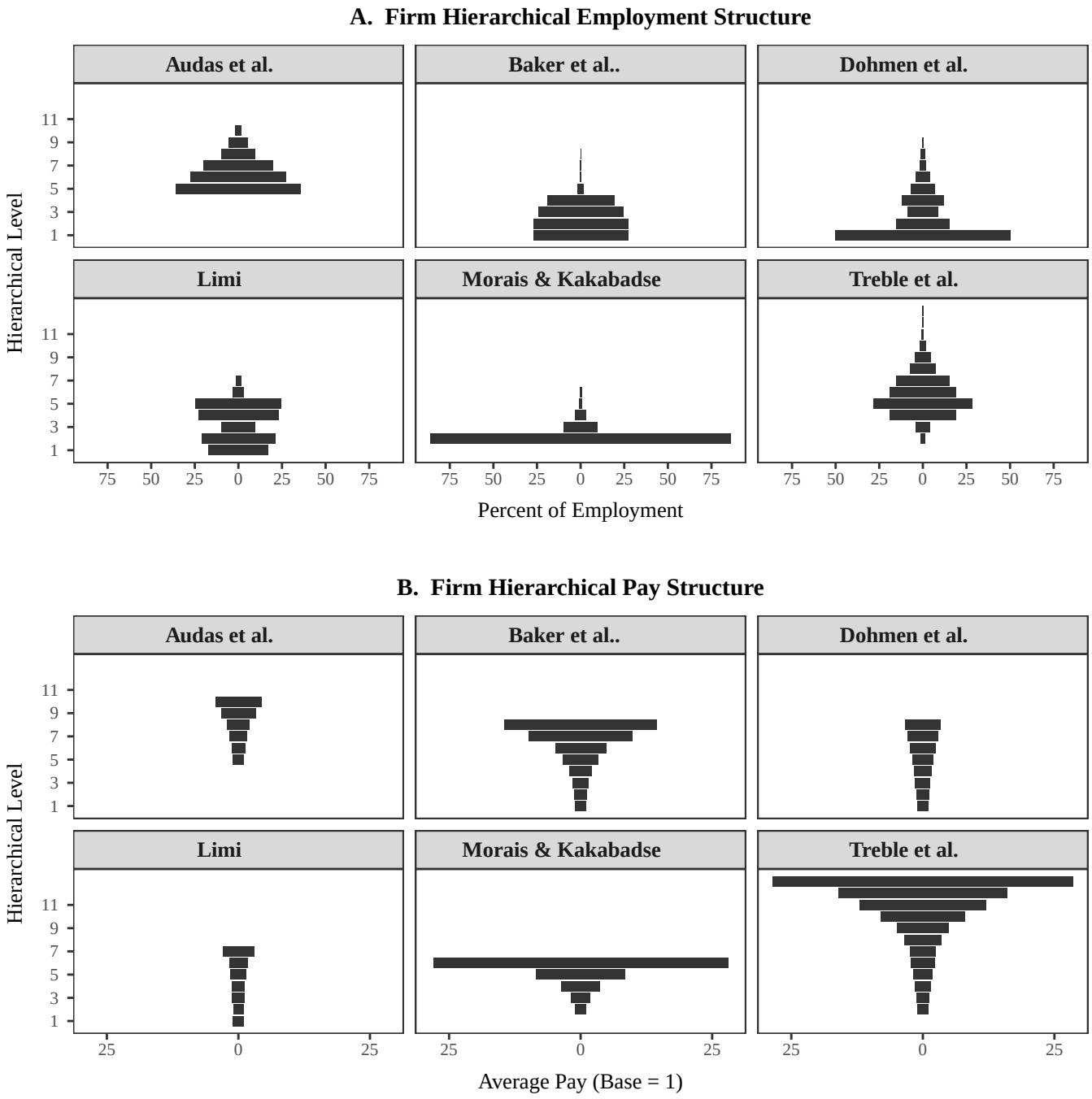


Figure 17: The Hierarchical Employment and Pay Structure of Six Different Firms

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.

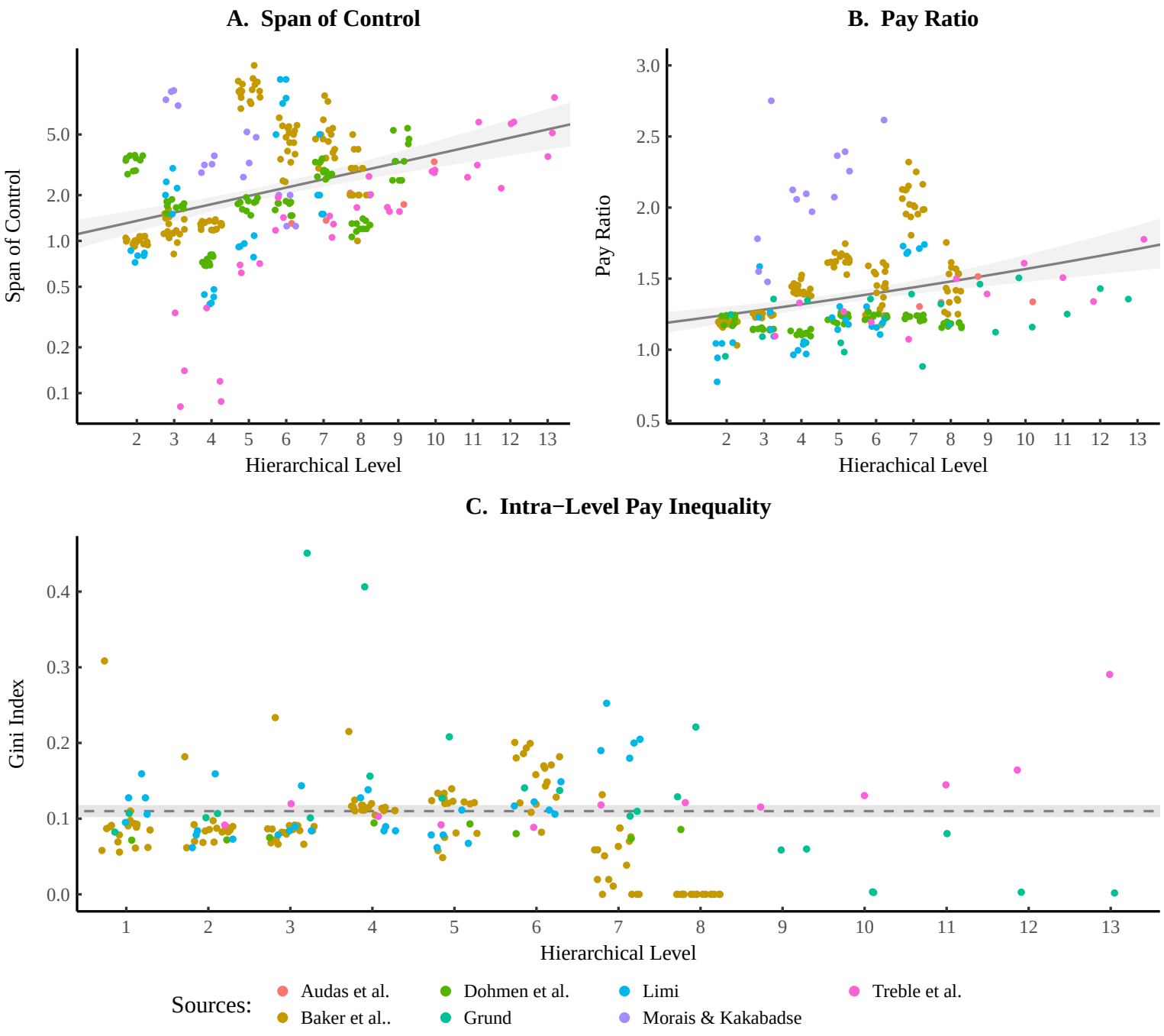


Figure 18: Case Studies of Firm Hierarchical Structure

This figure shows data from 7 different single-firm case studies. Panel A shows how the span of control (the employment ratio between adjacent levels) relates to hierarchical level. Panel B shows how the pay ratio between adjacent levels varies with hierarchical level. In these two panels, span of control and pay ratios between two hierarchical levels, h and $h-1$, are plotted on the x-axis at level h. Panel C shows levels of income inequality *within* individual hierarchical levels of each firm. Note that horizontal 'jitter' has been introduced in all three plots in order to better visualize the data (hierarchical level is a discrete variable). Grey regions correspond to the 95% confidence interval for regressions (or in panel C, the mean).

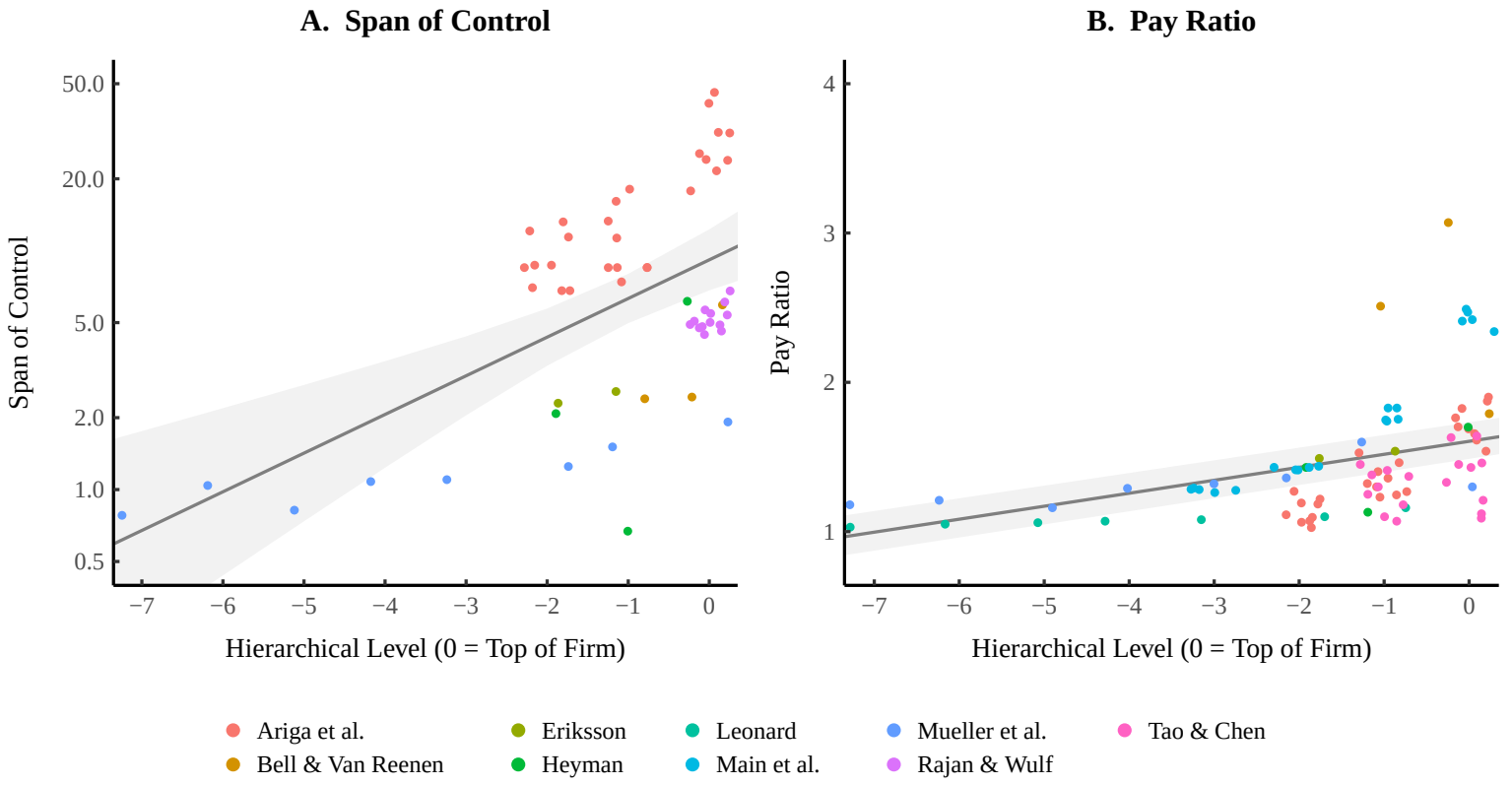


Figure 19: Aggregate Studies of Firm Hierarchical Structure

This figure shows data from 9 different aggregate firm studies. Most of these studies only survey the top several hierarchical levels in each firm. Because of this, I order hierarchical levels from the top down, where the CEO is level 0, the level below is -1, etc. Panel A shows how the span of control (the employment ratio between adjacent levels) relates to hierarchical level. Panel B shows how the pay ratio between adjacent levels varies with hierarchical level. In both plots, horizontal 'jitter' has been introduced in order to better visualize the data (hierarchical level is a discrete variable). Grey regions correspond to the 95% confidence interval for regressions.

Table 7: Firm Case Studies

Source		Years	Country	Firm Levels	Span of Control	Level Income	Level Income Dispersion
Audas	[118]	1992	Britain	All	✓	✓	
Baker	[119]	1969-1985	United States	Management	✓	✓	✓
Dohmen	[120]	1987-1996	Netherlands	All	✓	✓	✓
Grund	[133]	1995 & 1998	US and Germany	All		✓	✓
Lima	[121]	1991-1995	Portugal	All	✓	✓	✓
Morais	[122]	2007-2010	Undisclosed	All	✓	✓	
Treble	[123]	1989-1994	Britain	All	✓	✓	✓

Notes: This table shows metadata for the firm case studies displayed in Fig. 18. 'Firm Levels' refers to the portion of the firm that is included in the study. 'Management' indicates that only management levels were studied.

Table 8: Firm Aggregate Studies

Source		Years	Number of Firms	Country	Firm Levels	Span of Control	Level Income
Ariga	[134]	1981-1989	unknown	Japan	All	✓	✓
Bell	[135]	2001-2010	552	United Kingdom	Top 3	✓	✓
Eriksson	[136]	1992-1995	210	Denmark	Management	✓	✓
Heyman	[128]	1991,1995	560	Sweden	Management	✓	✓
Leonard	[137]	1981-1985	439	United States	Top 9		✓
Main	[138]	19801984	200	United States	Top 4		✓
Mueller	[127]	2004-2013	880	United Kingdom	All	✓	✓
Rajan	[139]	1986-1998	261	United States	Top 2	✓	
Tao	[140]	1986-1998	8101	Taiwan	Top 2		✓

Notes: This table shows metadata for the aggregate studies displayed in Fig. 19. 'Firm Level's refers to the portion of the firm that is included in the study. 'Top 2', 'Top 3', etc. indicates that only the top n levels were included in the study (where the top level is the CEO).

Table 9: Income Inequality Within Case Study Firms

Source		Years	Mean Gini Index
Baker et al.	[119]	1969-1985	0.32
Dohmen et al.	[120]	1991	0.18
Lima	[121]	1991-1995	0.15
Morais and Kakabadse	[122]	2007-2010	0.23
Treble et al.	[123]	1989-1997	0.26

B.1 Inequality Within Case Study Firms

I report here my estimates for inequality within the case study firms. Of the seven case studies summarized in Table 7, only one (Morais and Kakabadse) directly reports a firm Gini index. However, four other studies — Baker et al., Dohmen et al., Lima, and Treble et al. — provide enough data to allow estimates of firm internal inequality. I outline my calculation methods below. The resulting Gini estimates are shown in Table 9.

Baker et al.

Baker et al. have made their raw personnel data publicly available at the site below. I use this raw data to calculate the firm internal Gini index.

http://faculty.chicagobooth.edu/michael.gibbs/research/index.html

Dohmen et al.

Dohmen et al. report the following data that I use to estimate the firm Gini index:

1. Fraction of employment by hierarchical level (Tbl. 1);

2. Density plots of income distribution by hierarchical level (Fig. 5).

I use the *Engauge Digitizer* program to digitize and pull data from the density plots. I then use the resulting numerical density functions to estimate the firm Gini index.

We define $f_h(x)$ as the income density function for hierarchical level h. The income density function of the entire firm $f_T(x)$ is then defined by Eq. 24 – the sum of the density functions for each hierarchical level, weighted by the fraction of total employment (E_h/E_T).

$$f_T(x) = \sum_{h=1}^{n} \frac{E_h}{E_T} \cdot f_h(x) \tag{24}$$

The firm Gini index is then defined by Eq. 25-27. Equation 25 defines the mean income of the firm $(\bar{I})$, while equation 26 defines the cumulative income distribution function $F(x)$. Equation 27 then defines the Gini index (G). I use numerical integration implemented in R to evaluate these integrals.

$$\bar{I} = \int_0^\infty x \cdot f_T(x)\, dx \tag{25}$$

$$F(x) = \int_0^\infty f_T\, dx \tag{26}$$

$$G = \frac{1}{\bar{I}} \int_0^\infty F(x)(1 - F(x))dx \tag{27}$$

Grund

Grund [133] does not provide enough information to calculate firm-wide inequality. However, I am able to calculate intra-level income dispersion, (which appears in Fig. 18C). I use data from Grund's Fig. 1, which shows mean income by level, as well as what I assume to be 5th and 95th percentiles. After

digitizing this data, I use the best-fit theoretical distribution to estimate the Gini index.

Lima

Lima provides the following summary statistics, which I use to estimate a firm Gini index:

1. Employment within each hierarchical level (Tbl. 1);
2. Mean pay within each hierarchical level (Fig. 2);
3. Wage coefficient of variation by hierarchical level (Tbl. 6).

I use the *Engauge Digitizer* program to digitize and pull data from Fig. 2. To calculate the firm Gini index, I assume income within each hierarchical level is lognormally distributed. For each hierarchical level h, I then use equation 28 to define the lognormal scale parameter σ that produces a distribution with an equivalent coefficient of variation, c_v:

$$\sigma_h = \sqrt{\ln(c_v^2 + 1)} \tag{28}$$

Once we have σ_h, we use equation 29 to calculate the lognormal location parameter μ for each hierarchical level . Here $\bar{I}_h$ is the mean pay in hierarchical level h (which Lima reports directly).

$$\mu_h = \ln(\bar{I}_h) - \frac{1}{2}\sigma_h^2 \tag{29}$$

Once we have the appropriate lognormal parameters for each hierarchical level, we use these distributions to create a simulated payroll. To do this, we draw E_h numbers (employment in level h) from each lognormal distribution $\ln \mathcal{N}(\mu_h, \sigma_h)$. I then calculate the Gini index from this simulated payroll.

Treble et al.

Treble et al. report the following summary statistics, which I use to estimate a firm Gini index:

1. Employment within each hierarchical level (Fig. 2);
2. Mean pay within each hierarchical level (Fig. 3);
3. 5th and 95th wage percentile by hierarchical level (Fig. 4).

Again, I use *Engauge Digitizer* to pull data from all graphs. To estimate the intra-level Gini index, I adapt code writtent by Andrie de Vries to fit a parameterized distribution to the mean and 5th/95th percentiles.

C A Hierarchical Model of the Firm

Table 10: Notation

Symbol	Definition
a	span of control parameter 1
b	span of control parameter 2
c_v	coefficient of variation
C	CEO to average employee pay ratio
E	employment
f	either (1) a generic function; or (2) a probability density function
F	cumulative distribution function
G	Gini index of inequality
h	hierarchical level
$\bar{I}$	average income
μ	lognormal location parameter
n	number of hierarchical levels in a firm
p	pay ratio between adjacent hierarchical levels
r	pay-scaling parameter
s	span of control
σ	lognormal scale parameter
T	total for firm
$\downarrow$	round down to nearest integer
$\prod$	product of a sequence of numbers
$\sum$	sum of a sequence of numbers

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.

Model Assumptions

1. Firms are hierarchically structured, with a span of control that increases *exponentially* with hierarchical level.
2. The ratio of mean pay between adjacent hierarchical levels increases *exponentially* with hierarchical level.
3. Intra-hierarchical-level income is lognormally distribute and constant across all levels.

Using these assumptions, I first develop an algorithm that describes the hierarchical employment within a model firm, followed by an algorithm that describes the hierarchical pay structure.

C.1 Generating the Employment Hierarchy

To generate the hierarchical structure of a firm, we begin by defining the span of control (s) as the ratio of employment (E) between two consecutive hierarchical levels (h), where $h = 1$ is the *bottom* hierarchical level. It simplifies later calculations if we define the span of control in level 1 as $s = 1$. This leads to the following piecewise function:

$$s_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \dfrac{E_h}{E_{h-1}} & \text{if } h \geq 2 \end{cases} \tag{30}$$

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:

$$s_h = \begin{cases} 1 & \text{if } h = 1 \\ a \cdot e^{bh} & \text{if } h \geq 2 \end{cases} \tag{31}$$

As one moves up the hierarchy, employment in each consecutive level (E_h) *decreases* by $1/s_h$. This yields Eq. 32, a recursive method for calculating E_h. In this model, we want employment to be *whole* numbers. To accomplish this I have included the $\downarrow$ symbol to indicate that the last step is to round *down* to the nearest whole number. By repeatedly substituting Eq. 32 into itself, we can obtain a non-recursive formula (Eq. 33). In product notation, Eq. 33 can be written as Eq. 34.

$$E_h = \downarrow \frac{E_{h-1}}{s_h} \qquad \text{for} \qquad h > 1 \tag{32}$$

$$E_h = \downarrow E_1 \cdot \frac{1}{s_2} \cdot \frac{1}{s_3} \cdot \ldots \cdot \frac{1}{s_h} \tag{33}$$

$$E_h = \downarrow E_1 \prod_{i=1}^{h} \frac{1}{s_i} \tag{34}$$

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. 35, which in summation notation, becomes Eq. 36.

$$E_T = E_1 + E_2 + \ldots + E_n \tag{35}$$

$$E_T = \sum_{h=1}^{n} E_h \tag{36}$$

In practice, n is not known beforehand, so we define it using Eq. 34. 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:

$$n = \{h \quad | \quad E_h \geq 1 \quad \text{and} \quad E_{h+1} = 0\} \tag{37}$$

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.

C.2 Generating Hierarchical Pay

To model the hierarchical pay structure of a firm, we begin by defining the inter-hierarchical pay-ratio (p_h) as the ratio of mean income ($\bar{I}$) between adjacent hierarchical levels. Again, it is helpful to use a piecewise function so that we can define a pay-ratio for hierarchical level 1:

$$p_h \equiv \begin{cases} 1 & \text{if } h = 1 \\ \dfrac{\bar{I}_h}{\bar{I}_{h-1}} & \text{if } h \geq 2 \end{cases} \tag{38}$$

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:

$$p_h = \begin{cases} 1 & \text{if } h = 1 \\ r^h & \text{if } h \geq 2 \end{cases} \tag{39}$$

Using the same logic as with employment (shown above), the mean income I_h in any hierarchical level is defined recursively by Eq. 40 and non-recursively by Eq. 41.

$$\bar{I}_h = \frac{\bar{I}_{h-1}}{p_h} \tag{40}$$

$$\bar{I}_h = \bar{I}_1 \prod_{i=1}^{h} p_i \tag{41}$$

Mean income for all employees ($\bar{I}_T$) is then the weighted average of hierarchical level mean income ($\bar{I}_h$) and hierarchical level employment (E_h):

$$\bar{I}_T = \sum_{h=1}^{n} \bar{I}_h \cdot \frac{E_h}{E_T} \tag{42}$$

We define the CEO as the person(s) in the top hierarchical level. Therefore, CEO pay is simply $\bar{I}_n$, average income in the top hierarchical level. The *CEO-to-average-employee pay ratio* is given by the equation below. For succinctness, I refer to this ratio as the 'CEO pay ratio' C:

$$C = \frac{\bar{I}_n}{\bar{I}_T} \tag{43}$$

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 ($\bar{I}_1$)

C.3 Adding Intra-Level Pay Dispersion

Up to this point, we have modelled only the *mean* income within each hierarchical level of a firm. The last step in the modelling process is to make the firm more realistic by adding pay *dispersion* within each hierarchical level.

For this model, I assume that pay dispersion within hierarchical levels is *lognormally* distributed. This means that income (I) in hierarchical level h is described by the probability density function $\ln \mathcal{N}(I_h; \mu, \sigma)$, where μ and σ are the *location* and *scale* parameters, respectively:

$$\ln \mathcal{N}(I_h; \mu, \sigma) = \frac{1}{I_h \cdot \sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln I_h - \mu)^2}{2\sigma^2}\right] \tag{44}$$

Our empirical investigation of firm case studies indicated that pay dispersion with hierarchical levels is relatively constant (see Fig. 18C). Given this finding, I assume *identical inequality* within all hierarchical levels. This means that the lognormal scale parameter σ is the same for all hierarchical levels.

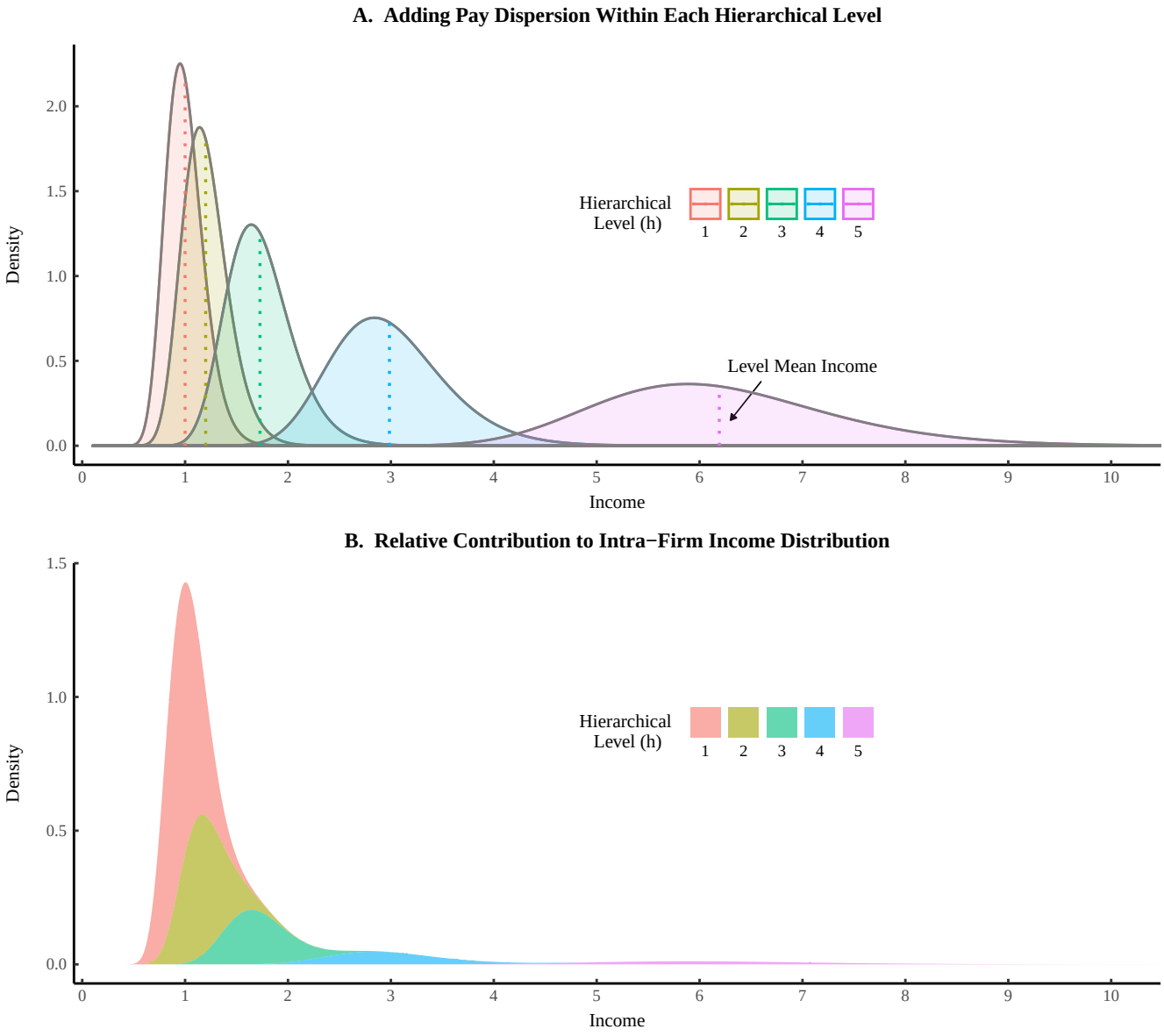


Figure 20: Adding Intra-Level Pay Dispersion to a Model Firm

This illustrates a model firm with lognormal pay dispersion in each hierarchical level. The model firm has a pay-scaling parameter of $r = 1.2$ and an intra-level Gini index of 0.13. Panel A shows the separate distributions for each level, with mean income indicated by a dashed vertical line. Panel B shows contribution of each hierarchical level to the resulting income distribution for the whole firm (income density functions are summed while weighting for their respective employment. Span of control parameters are identical to those used Table 11.

To define μ, I use Eq. 45, the formula for the mean income ($\bar{I}_h$) of our lognormal distribution. Solving for μ gives Eq. 46.

$$\bar{I}_h = e^{\mu+\frac{1}{2}\sigma^2} \tag{45}$$

$$\mu = \ln(\bar{I}_h) - \frac{1}{2}\sigma^2 \tag{46}$$

Given a value for σ (which is a free parameter), we can define the pay distribution within any hierarchical level of a firm. This process is shown graphically in Figure 20. Figure 20A shows the lognormal income distributions for each hierarchical level of a 5-level firm with pay-scaling parameter $r = 1.2$. Figure 20B shows the size-adjusted contribution of each hierarchical level to the overall intra-firm income distribution. Lower levels have more members, and thus dominate the overall distribution.

Once we have defined the probability distributions governing income in each hierarchical level, the last step is to simulate individual pay, and ultimately construct a firm payroll. We do this by defining income as a random lognormal variable:

$$I_h \sim \ln \mathcal{N}(\mu_h, \sigma) \tag{47}$$

We construct a completed firm payroll by drawing E_h random numbers for each level h, and combining them all in the payroll vector I. Using subscripts to denote the hierarchical level and superscripts to denote the individual in that level (ranging from 1 to E_h) we get:

$$I = \{I_1^1, I_1^2, ..., I_1^{E_1}, I_2^1, I_2^2, ..., I_2^{E_2}, ..., \bar{I}_n\} \tag{48}$$

Note that the last entry, the CEO pay $\bar{I}_n$, is *not* a random variable. In order to preserve the CEO pay ratio (dictated by the Compustat dataset) I do not allow this value to vary stochastically.

C.4 Example of the Model Algorithm

We begin by choosing the arbitrary values of a, b, E_1, r, and $\bar{I}_1$ shown in Table 11. We then input these values into the hierarchy-building algorithm. Column A shows the hierarchical levels of the firm (h), where $h = 1$ is the base level. Using parameters a and b, we first calculate the span of control (column B), which defines the employment-ratio between adjacent hierarchical levels. In column C, we begin with the base level and use Eq. 32 to calculate employment in each hierarchical level. In column D, we calculate the pay ratio p_h using the pay-scaling parameter r. Finally, in column E, we calculate mean income $\bar{I}_h$ in each hierarchical level

Once we have this table of values, we can calculate aggregate statics like total employment (the sum of column C) and mean pay (the mean of column E, weighted by column C). We can also calculate the CEO pay ratio. These results are shown at the bottom of Table 11

The last step of the model is to generate a simulated payroll by adding lognormal dispersion to each hierarchical level. For large firms, this involves drawing many random numbers from a lognormal distribution. For example purposes, it is convenient to choose a small firm. Table 12 shows a firm with the same span of control parameters as in Table 11, but with a base size of 10. As before, we use the model algorithm to calculate mean pay in each level. We then use Eq. 46 to calculate the lognormal location parameter in each level. The last step is to create the simulated payroll. For each hierarchical level h, we draw E_h random numbers from a lognormal distribution with parameters σ and μ_h. Note that we do not let income in the top hierarchical level vary stochastically — this preserves the CEO pay ratio on which the model is based.

Once we have the simulated payroll, we can calculate the firm's income inequality. The resulting Gini index will vary randomly, due to the stochastic nature of the model. For large firms (more than 1000 employees) this variation is negligible. For small firms, if we wish to know the 'true' Gini index that is predicted from the sum of the lognormal density functions, we can do two things:

Table 11: Example of the Model Algorithm

Parameters

a	b	E_1	r	$\bar{I}_1$
1	0.2	10 000	1.15	1

A	**B**	**C**	**D**	**E**
Hierarchical Level	Span of Control	Employment	Pay Ratio	Mean Income
h	$s_h = e^{0.2h}$	$E_h = \downarrow \frac{E_{h-1}}{s_h}$	$p_h = 1.15^h$	$\bar{I}_h = \bar{I}_{h-1} \cdot p_h$
10	7.39	0	–	–
9	6.05	1	3.52	468.5
8	4.95	8	3.06	133.2
7	4.06	44	2.66	43.5
6	3.32	182	2.31	16.4
5	2.72	607	2.01	7.1
4	2.23	1652	1.75	3.5
3	1.82	3678	1.52	2.0
2	1.49	6703	1.32	1.3
1	–	10000	–	1

Results

Total Employment	Mean Pay	CEO Pay	CEO Pay Ratio
$E_T = \sum_{h=1}^{n} E_h$	$\bar{I}_T = \sum_{h=1}^{n} \bar{I}_h \cdot \frac{E_h}{E_T}$	$\bar{I}_n$	$C = \bar{I}_n / \bar{I}_T$
22 875	1.87	468.5	250

Table 12: Adding Intra-Level Pay Dispersion to a Firm

Parameters

a	b	E_1	r	$\bar{I}_1$	σ
1	0.2	10	1.2	1	0.24

Hierarchical Level	Pay Ratio	Mean Pay	Scale Parameter	Location Paramter
h	$p_h = 1.2^h$	$\bar{I}_h = \bar{I}_{h-1} \cdot p_h$	σ	$\mu_h = \ln(\bar{I}_h) - \frac{1}{2}\sigma^2$
4	1.73	2.99	0.24	1.07
3	1.44	1.73	0.24	0.52
2	1.20	1.20	0.24	0.15
1	–	1.00	0.24	-0.03

Generating a Simulated Payroll

Hierarchical Level	Employment	Simulated Payroll
h	E_h	$I_h \sim \ln \mathcal{N}(\mu_h, \sigma)$
4	1	$\{2.99\}$
3	3	$\{1.62, 1.88, 1.16\}$
2	6	$\{1.11, 0.94, 1.08, 1.15, 1.07, 2.13\}$
1	10	$\{0.75, 0.65, 1.04, 1.09, 0.96, 0.95, 0.97, 1.09, 0.75, 0.87\}$

1. Run the model many times and take the mean of resulting sample of Gini indexes;
2. Multiply all hierarchical employment E_h by a large, constant factor.

Mathematically, these two approaches produce identical results. However, method 2 is computationally faster. In our example, the Gini of the simulated payroll is $G = 0.206$. The Gini predicted from the sum of lognormal density functions is $G = 0.217$.

D The Compustat Data

To model US intra-firm income distribution, I use the Compustat data series shown in Table 13. Selected statistics from this dataset are shown in Figure 21.

Employment, staff expense, and executive compensation data available are for roughly 300 firms per year over the period 1992–2015 (Fig. 21A). Although this is a small firm sample, the firms themselves are very large, with mean sizes of between 20,000 and 30,000 employees (Fig. 21B). As a result, this firm sample accounts for roughly 5% of US employment (Fig. 21C). Unlike the total US firm size distribution, which has a power-law shape [141], this Compustat firm sample is lognormally distributed (Fig. 21D).

From the three data series shown in Table 13, we can calculate the following:

$$\text{Employee Mean Income} = \frac{\text{Total Staff Expenses}}{\text{Employees}} \tag{49}$$

$$\text{CEO Pay Ratio} = \frac{\text{Top Exec Pay}}{\text{Employee Mean Income}} \tag{50}$$

Table 13: Compustat Data Series

Database	Series ID	Description
ExecuComp	TDC1	Executive Total Compensation
Fundamentals Annual	XLR	Total Staff Expenses
Fundamentals Annual	EMP	Employees

Notes: Executive compensation series TDC1 = Salary + Bonus + Other Annual + Restricted StockGrants + LTIPPayouts + All Other + Value of Option Grants

Note that 'CEO pay' is defined as the income of the top-paid executive in a given firm. The CEO pay ratio of this sample is lognormally distributed (Fig. 21E) with an average of between 50 and 150 (Fig. 21F). The normalized firm mean pay distribution is shown in Figure 21G (each firm's pay is divided by average of the annual sample). The resulting log distribution has bimodal structure. Figure 21H shows how the mean pay of the Compustat sample compares to mean pay for all US workers. Employees in these Compustat firms earn slightly more than the US population at large — a result that is consistent with the well-known firm size-wage gap [142].

Figure 21I shows inter-firm income inequality — the Gini index of firm mean pay in our Compustat sample. Inter-firm income inequality in this sample tended to increase over the time period in question.

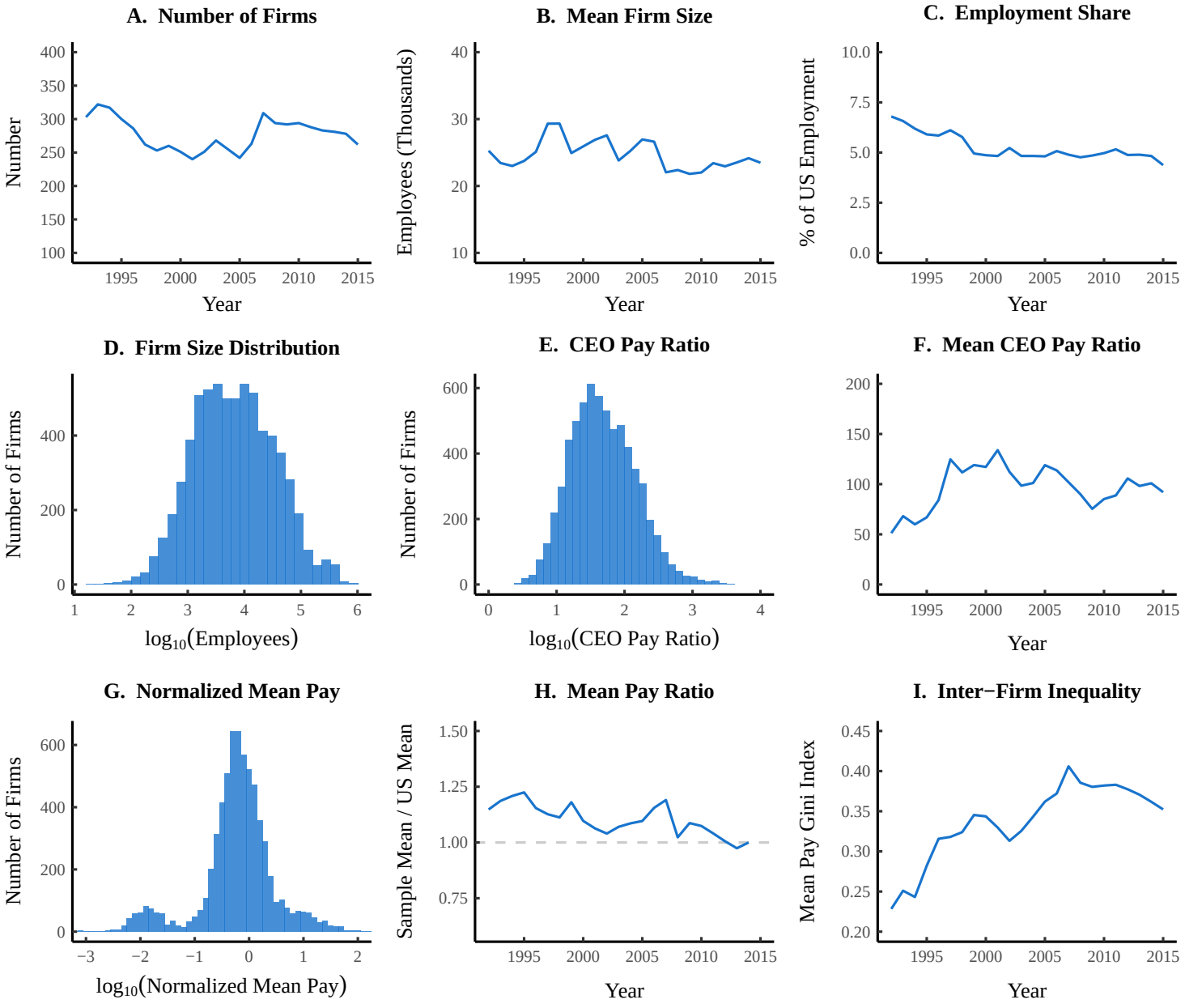


Figure 21: Selected Statistics from the Compustat Firm Sample

This figure shows statistics for the Compustat firm sample, which consists of US firms for which the data series in Table 13 are available. In panel H, US mean income per worker is calculated from national accounts (BEA Table 1.12, National Income by Type of Income) by dividing the sum of employee and proprietor income by the number of workers (BEA Table 6.8C-D, persons engaged in production).

Table 14: Compustat Model Parameters

Parameter	Definition	Estimation Method
a, b	Span of control parameters	Estimated from case study data (Fig. 18A). *Values are fixed for all modeled firms*.
σ	Intra-hierarchical level pay dispersion parameter	Estimated from case study data (Fig. 18C). *Value is fixed for all modeled firms*.
E_1	Employment in base hierarchical level	Estimated numerically, given a, b, and Compustat firm employment E_T.
r	Pay-scaling parameter	Estimated from Compustat CEO pay ratio, given a, b, and E_1.
$\bar{I}_h$	Mean pay in base hierarchical level	Estimated from Compustat firm mean pay $\bar{I}_T$, given a, b, E_1, and r.

E Estimating Compustat Model Parameters

We now apply the algorithm developed in Appendix C to model intra-firm income distribution within the Compustat firm sample (Sec. D). The Compustat model is characterized by the 6 parameters shown in Table 14. In order to model the internal income distribution of Compustat firms, we need to estimate these 6 parameters for each firm. My methods are summarized in Table 14 and discussed in detail in the following sections.

E.1 Parameters Derived from Case-Study Data

The parameters a, b, and σ are estimated from the firm case-study data shown in Figure 18 and assumed to be fixed for all Compustat firms. The parameters a and b, which together determine how the span of control changes with hierarchical level, come from an exponential regression on data in Fig. 18A. The

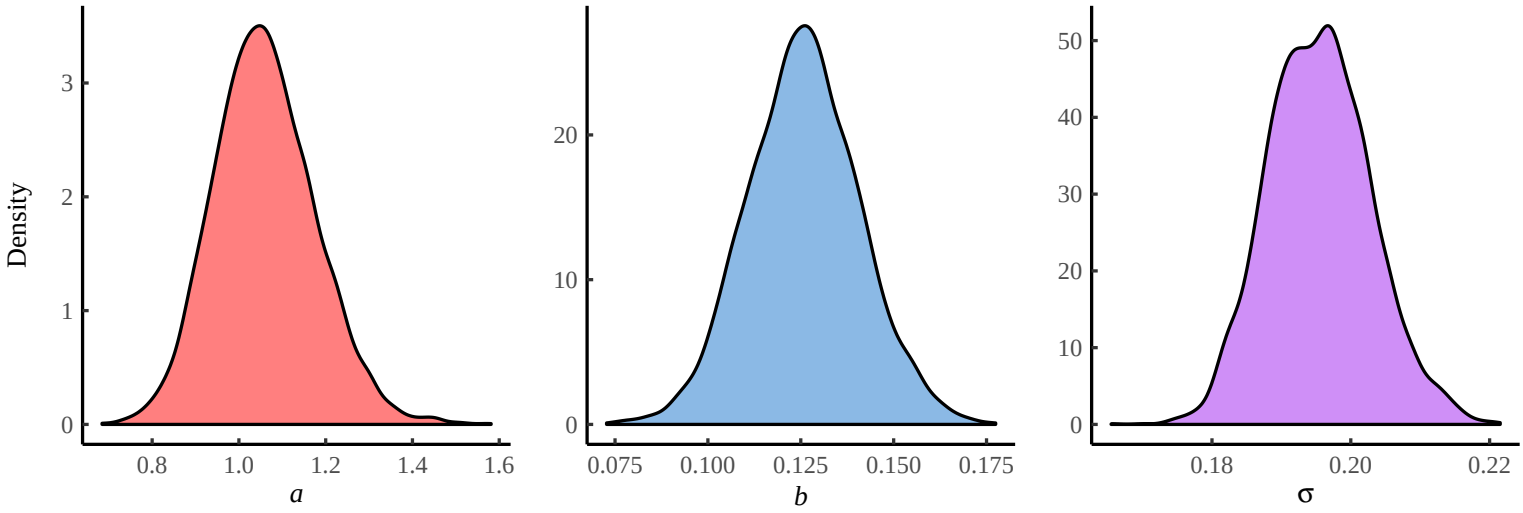


Figure 22: Density Estimates for Case-Study Derived Parameters

This figure shows density estimates for the parameters a, b, and σ. Parameters a and b together determine the 'shape' of the firm hierarchy. The parameter σ determines the amount of income dispersion within each hierarchical level. Each parameter is determined from regressions on firm case-study data (Fig. 18). The density functions are estimated using a bootstrap analysis, which involves resampling (with replacement) the case study data many times, and calculating the parameters a, b, and σ for each resample.

---

parameter σ determines the amount of pay dispersion within each hierarchical level of a firm (all levels are assumed to have the same amount of dispersion). This dispersion is modeled with a lognormal distribution, and σ is the 'scale' parameter (see Eq. 44).

We estimate σ from the case-study data shown in 18C. Note that this data uses the Gini index as the metric for dispersion. To estimate σ, we first calculate the mean Gini index of all data ($\bar{G}$). We then use Eq. 51 to calculate the value σ, which corresponds to the lognormal scale parameter that would produce a lognormal distribution with an equivalent Gini index. This equation

is derived from the definition of the Gini index of a lognormal distribution — $G = \text{erf}(\sigma/2)$.

$$\sigma = 2 \cdot \text{erf}^{-1}(\bar{G}) \tag{51}$$

Once a, b, and σ are estimated from case-study data, the model proceeds on the assumption that these parameters are *constant* across all Compustat firms. While real-world Compustat firms likely have parameters that vary widely, the hope is that our case-study regression provides a reasonable mid-point estimate.

Regressions on the case-study data provide a single best-fit estimate of the values of a, b, and σ. However, because the case-study sample size is small, there is considerable uncertainty in these values. This uncertainty can be quantified using the *bootstrap* method [132], which involves repeatedly resampling the data (with replacement) and then estimating the parameters a, b, and σ from this resampled data. Figure 22 shows the probability density distributions resulting from this bootstrap analysis.

To incorporate this uncertainty into the model, I run the model many times — once for each bootstrapped estimate of a, b, and σ. In each iteration, we first resample the case-study data and calculate values of a, b, and σ. We then use these values (particularly a and b) to calculate all other model parameters. The results shown in this paper are based on 5000 bootstrap runs of the model.

E.2 Base Level Employment

Having estimated the span of control parameters a and b, the next step is to calculate base-level employment E_1 for each Compustat firm. We do this by using data for total employment E_T.

The modeled-relation between total employment E_T and base-level employment E_1 is determined by equations 31, 34, and 36 (see Appendix C). Given values for a and b, these equations produce a unique relation between E_1 and E_T.

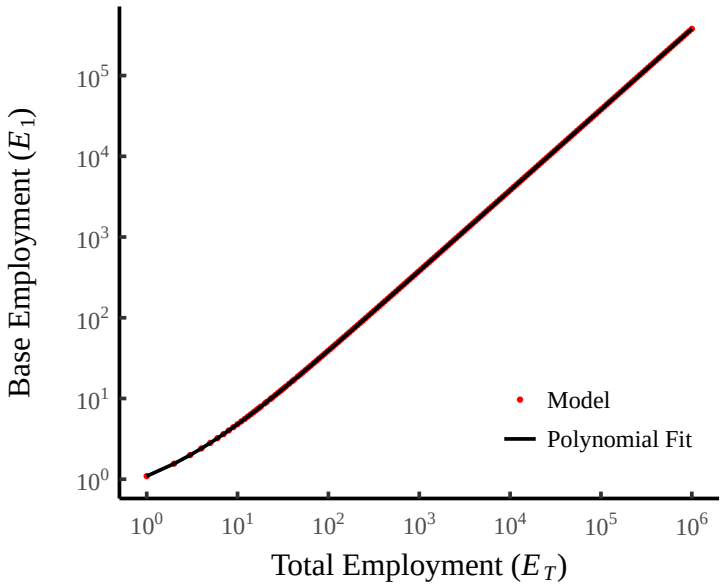


Figure 23: Finding Base Level Employment From Total Employment

This figure shows the modeled relation between total employment in a firm (E_T) and base-level employment (E_1). This relation, defined by Eq. 31, 34, and 36, depends on the span of control parameters a and b (here $a = 1.05$ and $b = 0.13$). I fit this numerical mapping with a high-order polynomial to allow fast (but accurate) estimation of E_1 from E_T. The R code for this procedure is available in the Supplementary Material.

$$E_T = f_{a,b}(E_1) \tag{52}$$

What we want is an inverse function that gives E_1 from E_T:

$$E_1 = f_{a,b}^{-1}(E_T) \tag{53}$$

Although there may be a way to define this inverse function analytically, it is beyond my mathematical abilities. Instead, I use the model to reverse engineer the problem. I define $f_{a,b}$ numerically by inputting a range of different values for E_1 into equations 31, 34, and 36 and calculating E_T for each value.

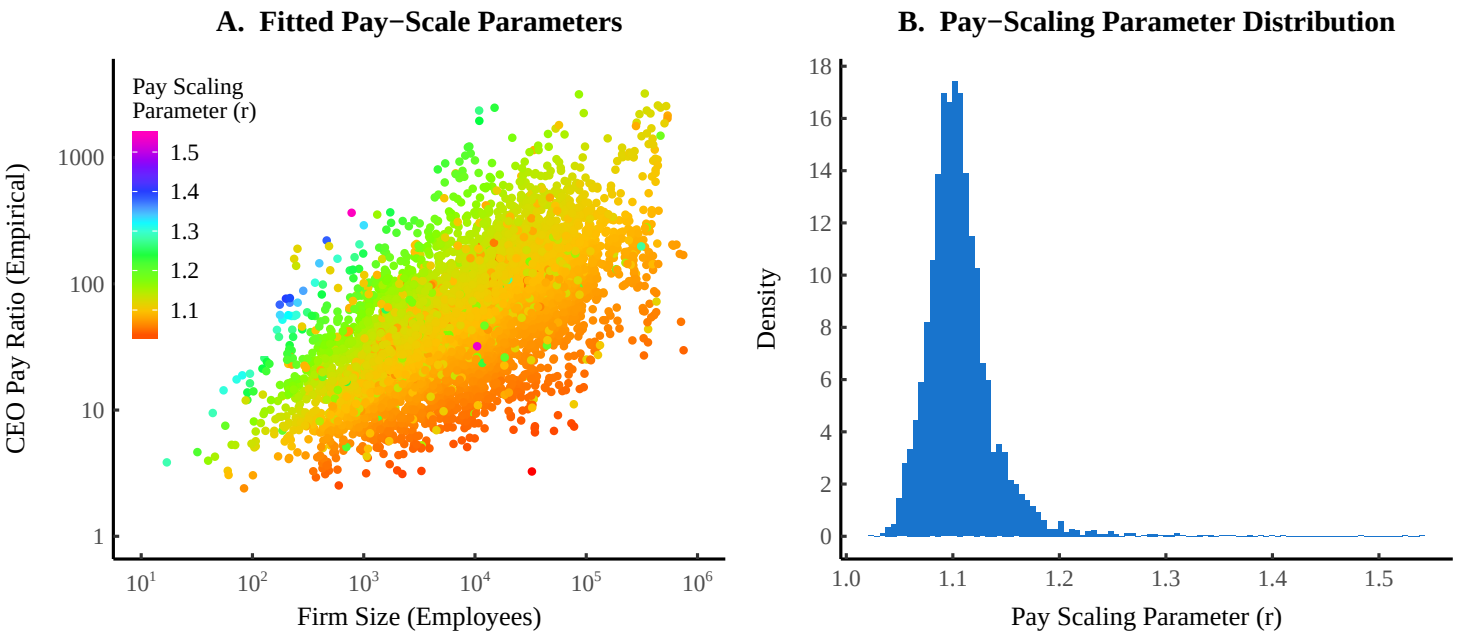


Figure 24: Fitting Compustat Firms with a Pay-Scaling Parameter

This figure shows the fitted pay-scaling parameters (r) for all Compustat firms. Panel A shows the relation between the CEO pay ratio and firm size, with the fitted pay-scaling parameter indicated by color. The pay-scaling parameter distribution for all firms (and years) is shown in panel B. These results show the average of 5000 model runs, each with different bootstrapped parameters a, b, and σ.

The result is a discrete mapping relating base-level employment to total employment (see Fig. 23). I then fit this mapping with a high-order polynomial, which then serves as an approximation to the inverse function f^{-1}. This polynomial can then be used to quickly and accurately calculate E_1 from E_T for every Compustat firm.

E.3 Pay-Scaling Parameter

Once we have calculated base-level employment (E_1) for all Compustat firms, we can estimate their respective pay-scaling ratios (r) using the CEO-to-average-

employee pay ratio (C). The pay-scaling ratio r determines the rate at which mean pay increases by hierarchical level.

Having estimated a, b, and E_1 for each Compustat firm, the model (specifically equations 34, 39, 41, 42,and 43) produces a CEO pay ratio (C_{model}) that is a unique function of the pay-scaling parameter r:

$$C_{\text{model}} = f_{a,b,E_1}(r) \tag{54}$$

As with base-employment, I am not aware of an analytical method for defining the inverse function f^{-1}_{a,b,E_1}. Instead I use a numerical optimization method to solve for r. I define an error function $\epsilon(r)$ that quantifies the error between the actual value of a firm's CEO pay ratio ($C_{empirical}$) and the value predicted by the model (C_{model}) for a given value of r:

$$\epsilon(r) = \Big|\, C_{\text{model}} - C_{\text{empirical}} \Big| \tag{55}$$

For each firm, the correct value of r is that which *minimizes* this error function. I use the R non-linear optimization function 'nlminb' to solve this minimization problem. To ensure that there are no large errors, I discard Compustat firms for which the best-fit r parameter produces an error that is larger than 5% of $C_{\text{empirical}}$. Fitted results for r are shown in Figure 24.

E.4 Base-Level Pay

Once we have the pay-scaling parameter r, we can estimate base-level pay for each Compustat firm. To do this, we set up a ratio between base level pay ($\bar{I}_1$) and firm mean pay ($\bar{I}_T$) for both the model and Compustat data:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{\bar{I}_1^{\text{model}}}{\bar{I}_T^{\text{model}}} \tag{56}$$

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. 41 and 42). If we run the model with $\bar{I}_1^{\text{model}} = 1$, then Eq. 56 reduces to:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{1}{\bar{I}_T^{\text{model}}} \tag{57}$$

We can then rearrange Eq. 57 to solve for an estimated base pay for each Compustat firm ($\bar{I}_1^{\text{Compustat}}$):

$$\bar{I}_1^{\text{Compustat}} = \frac{\bar{I}_T^{\text{Compustat}}}{\bar{I}_T^{\text{model}}} \tag{58}$$

F Compustat Model Results

I review here the results of the Compustat model that are not discussed in the main paper. All results are generated using 5000 bootstrap model runs over different values for the parameters a, b, and σ. From the data generated by the model, many different calculations are possible. I review here the following: (1) estimates for income inequality within Compustat firms; (2) estimates for income by hierarchical level; and (3) aggregate inequality of all firms in the model.

F.1 Inequality Within Compustat Firms

Figure 25 shows estimate of income inequality within Compustat firms. In Figure 25A, I illustrate how firm Gini indexes are related to both the CEO Pay ratio and firm size. Note that the CEO pay ratio is a reliable indicator of firm inequality only for firms of the *same size*. A general feature of a hierarchical firm model is that when internal inequality is held constant, the CEO pay ratio nonetheless tends to increase with firm size (a feature first demonstrated by Herbert Simon [52]). In Figure 25A, this feature is evident as color contours of constant firm inequality that scale with both firm size and the CEO pay ratio.

Figure 25B shows the overall distribution of all firm Gini indexes. According to our model, 90% of Compustat firms have internal Gini indexes between 0.2 and 0.5. Note that the distribution is right-skewed — a small minority of firms have extremely unequal pay.

In Figure 25C I compare firm inequality in the Compustat model to inequality within the case-study firms discussed in Appendix B. The results indicate that Compustat firms are slightly more unequal than the case study firms. However, because the case-study sample size is small, this difference is not statistically significant. A Kolmogorov-Smirnov test gives a p-value of 0.20, indicating that there is a reasonable (20%) probability that the two firm samples (model and case study) come from the *same* distribution. Thus, under the

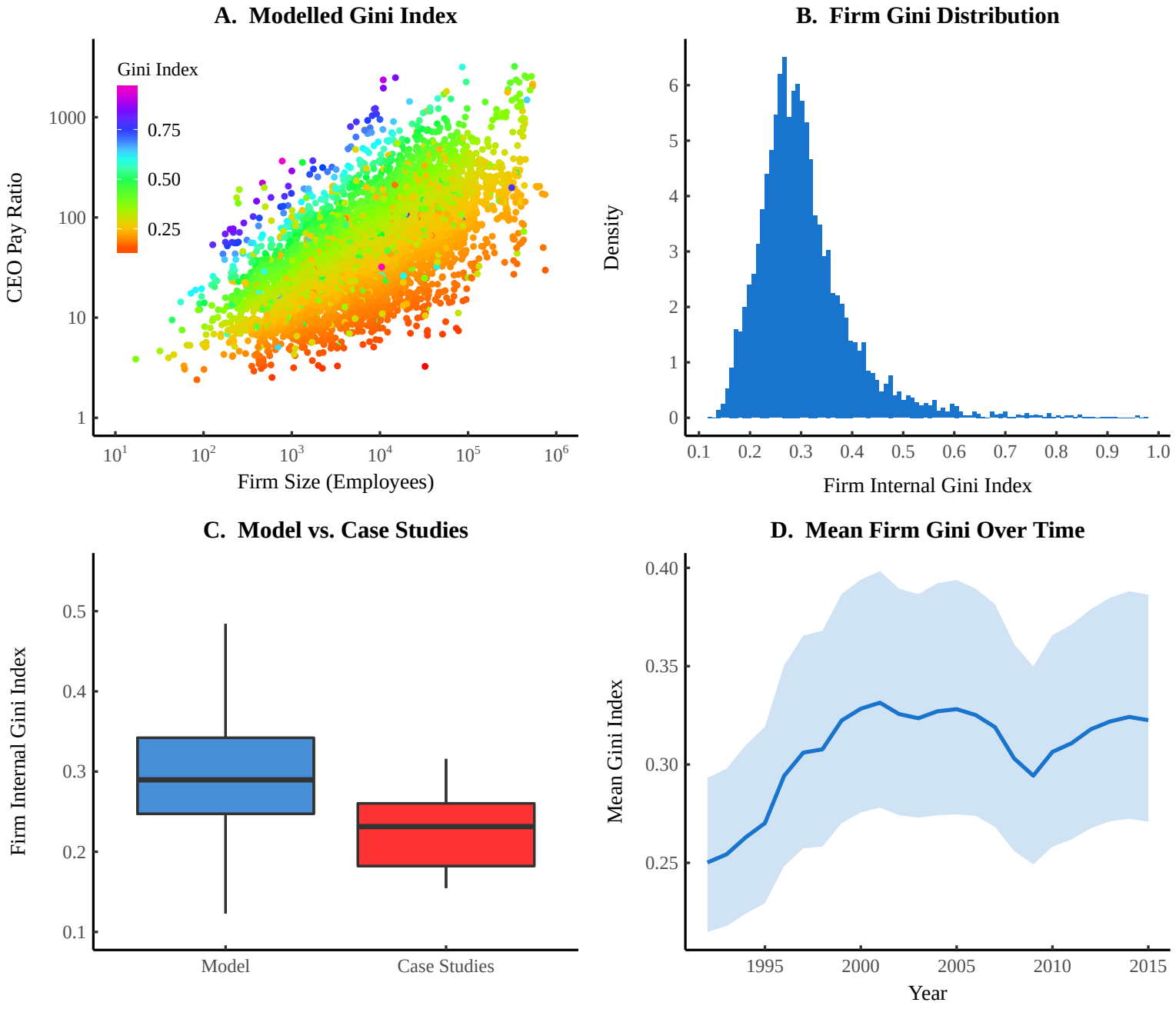


Figure 25: Compustat Model Results for Intra-Firm Inequality

This figure shows the firm internal Gini index results of the Compustat model. Panel A shows how firm internal inequality (indicated by color) is related to the CEO pay ratio and firm size. Panel B shows the distribution of modeled Gini indexes for all firms. Panel C compares model results to the Gini index of case study firms (see section B.1 for case study methods). Panel D shows time evolution of the average Gini index of all modeled firms. The shaded region indicates the 95% confidence interval. All results are computed from 5000 model runs, each with different bootstrapped parameters a, b, and σ.

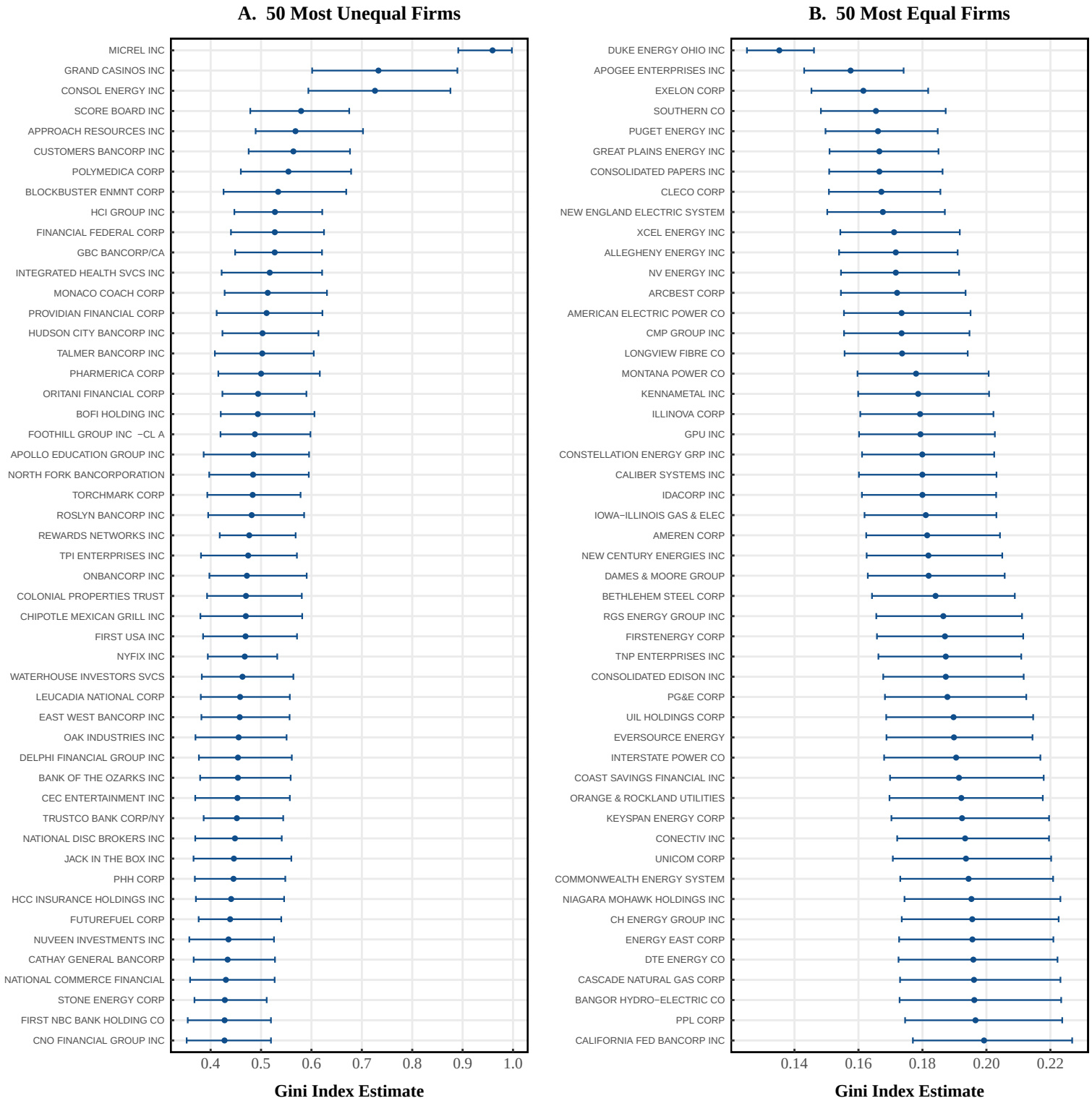


Figure 26: The Most Equal and Unequal Compustat Firms

This figure shows the 50 most unequal (panel A) and 50 most equal firms (panel B). Points indicate the mean Gini index for each firm, while the error bars show the 95% confidence interval calculated from 5000 bootstrap model runs.

conventional 5% significance level, we cannot reject the null-hypothesis that these samples come from the same distribution.

But if the case study data and Compustat model produce firm internal Gini distributions that are statistically indistinguishable, why not simply use case study data for the test of hypothesis B (hierarchical power has the strongest effect on income)? There are several reasons the case study data cannot be used. Firstly, the case study sample size is extremely small. Secondly, the firms cover many different countries (not just the US, the desired country). Thirdly, the observation years often do not overlap. To test hypothesis B, we need a *large* firm sample from a *single* country in a *single* year. While model dependent, the results inferred from Compustat data satisfy these conditions, while case study data does not.

Figure 25D shows the time-evolution of average inequality within Compustat firms. During the late 1990s inequality rapidly increased, followed by relative stability from 2000 onward. While the *trend* is clear, there is significant uncertainty in the *absolute* level of inequality (as indicated by the shaded region). This uncertainty is due to the small case-study sample size on which key model parameters are based (see Appendix E).

Finally, Figure 26 shows Gini index estimates for the 50 most equal and 50 most unequal firms. What is most interesting about these results is the *sectoral composition* of the 50 most equal firms. The vast majority (80%) are energy/utility companies. In the United States, firms in the utility sector are highly regulated, which leads to far more scrutiny over executive pay. Previous studies have found similar results — executives in regulated firms earn far less than those in unregulated firms [143]. This finding has important implications for a power theory of income distribution. It suggests that government regulation serves as a *check* on power, limiting the degree to which elites are able to use their status to amass wealth.

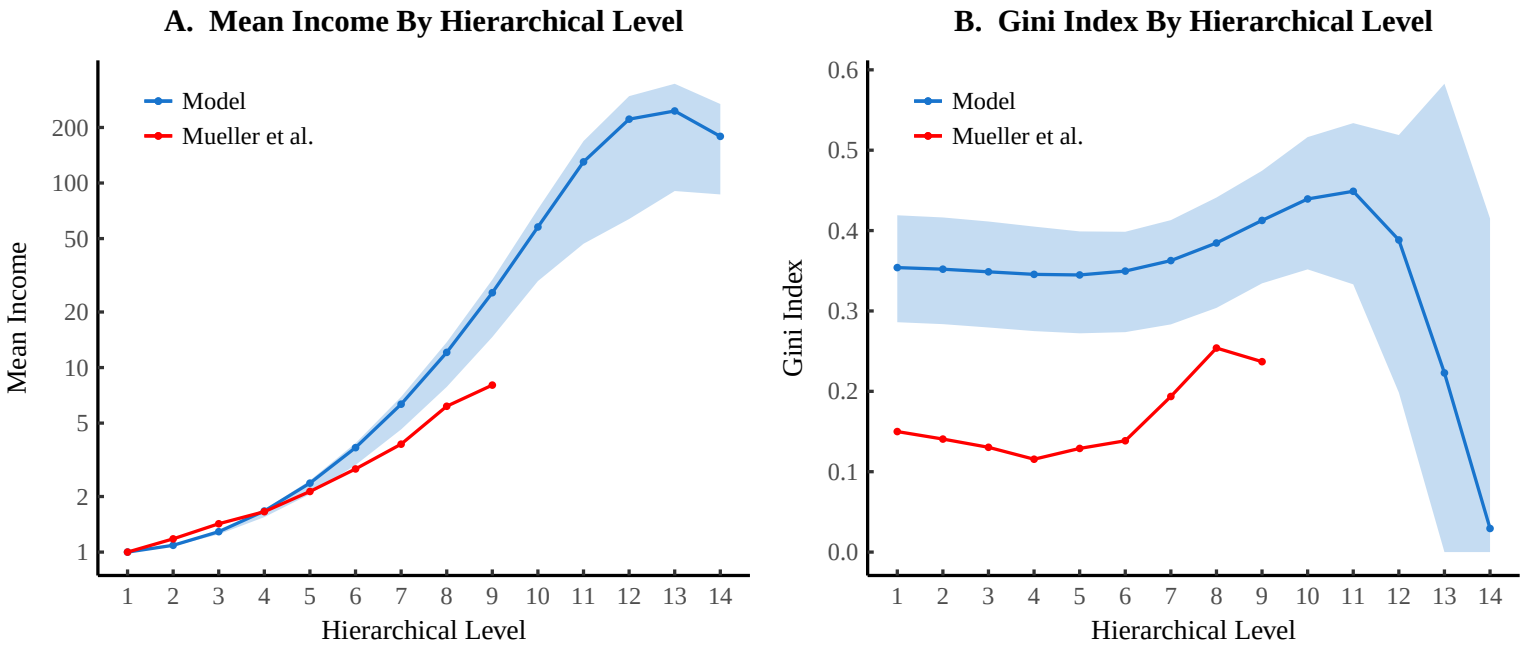


Figure 27: Compustat Model Results for Income by Hierarchical Level

This figure compares the results of the Compustat model to the UK data from Mueller et al. [127]. Panel A shows average income by hierarchical level (across all firms) indexed to pay in level 1. Panel B shows how intra-level inequality changes by hierarchical level. Shaded regions indicate the 95% confidence region of the model, estimated from 5000 bootstrap runs (see Appendix E).

F.2 Income By Hierarchical Level

Besides estimating firm internal inequality, I use the Compustat model to estimate income and inequality by hierarchical level. To do this, we group all individuals by their hierarchical level, regardless of firm membership (see Fig. 10).

Model results are shown in Figure 25, and are compared to the UK data documented by Mueller et al. [127]. Figure 25A shows how mean income changes by hierarchical level. In both the Compustat model and Mueller's data, mean

income increases super-exponentially with hierarchical level — that is, it increases *faster* than an exponential function, which would appear as a straight line on the log-linear scale. Figure 25B shows how intra-level income inequality changes by hierarchical level. For hierarchical levels 1-10, both the Compustat model and Mueller's data show similar trends.

The similarities between the model and Mueller's data lend credence to the model. However, what explains the differences? One key factor is that the United States has much greater income inequality than the United Kingdom, and the Compustat firm sample comes from the former and Mueller's sample the latter. As it turns out, both the Compustat model and Mueller's data imply aggregate levels of inequality that are consistent with their respective national Gini indexes (see Fig. 15 and 28).

In this light, the results in Figure 25A make sense — in the more unequal United States, income scales more rapidly with hierarchical level than in the United Kingdom. The results in Figure 25B can be similarly explained — in the more unequal United States, intra-hierarchical level income dispersion is greater than in the UK.

Another interesting result in Figure 25 is the conspicuous change in model trends for hierarchical levels above 11. Above this level, mean income no longer increases with hierarchical level, and intra-level inequality declines precipitously. The former result may simply be an artifact of the particular firm sample. Going back to Figure 25A, note that the four largest firms have particularly low CEO pay ratios. Given the model's assumptions, only the very largest firms will have more than 11 hierarchical levels. Since the 4 largest firms have particularly low CEO pay ratios, resulting mean income in hierarchical levels 12-14 will be relatively low.

The precipitous drop in intra-level inequality for hierarchical levels 12-14 is likely due to the convergence to a size of one. This is because there is often only one firm with 12 or more hierarchical levels, and the top level of this firm will contain only one individual. By definition, there is zero inequality in a sample size of one.

F.3 Aggregate Inequality

An important test of the Compustat model is to see if it produces aggregate levels of inequality that are comparable to US empirical data. Figure 28 shows the results of such a test. Here I plot the time-series trends in both US historical inequality and aggregate inequality in the Compustat model. This latter metric is calculated by aggregating (by year) all individuals in the model into a single sample, and then calculating the inequality of the resulting income distribution.

Figure 28A compares the model's aggregate Gini index against three different types of data published by the US Census: Gini by *individual*, *family*, and *household*. Two findings are evident. Firstly, the model is roughly consistent with the US empirical data over the period 2000-2015. However, the model produces too little inequality during the 1990s. Secondly, the US empirical data shows *contradictory* trends — roughly *constant* inequality among individuals, but secularly *increasing* inequality among families and households. The model reproduces the secular trend. But which empirical data should we believe? My vote is that the secular increase is the correct trend.

Largely in response to his dissatisfaction with official inequality statistics, Thomas Piketty [1] has focused on measuring inequality in the tail of the income distribution. Figure 28B and C show Piketty's series for the top 10% and 1% income share in the United States. Both series show secularly increasing inequality over the period in question. The model reproduces these *trends* quite accurately, but at a lower absolute level of inequality.

How can it be that the model more or less matches US Gini index data, but gives much less inequality than Piketty's metrics? A plausible explanation is that official data simply *underestimates* inequality. However, the validity (or lack their of) of official inequality statistics is not something that this paper is concerned with. Rather, I simply take official data as a given, and use it to test my power-income hypothesis. As such, the important take-home finding here is that the Compustat model produces a level of inequality that is consistent

with official US data. This means that it is fair to compare the model's results to other results derived from official data.

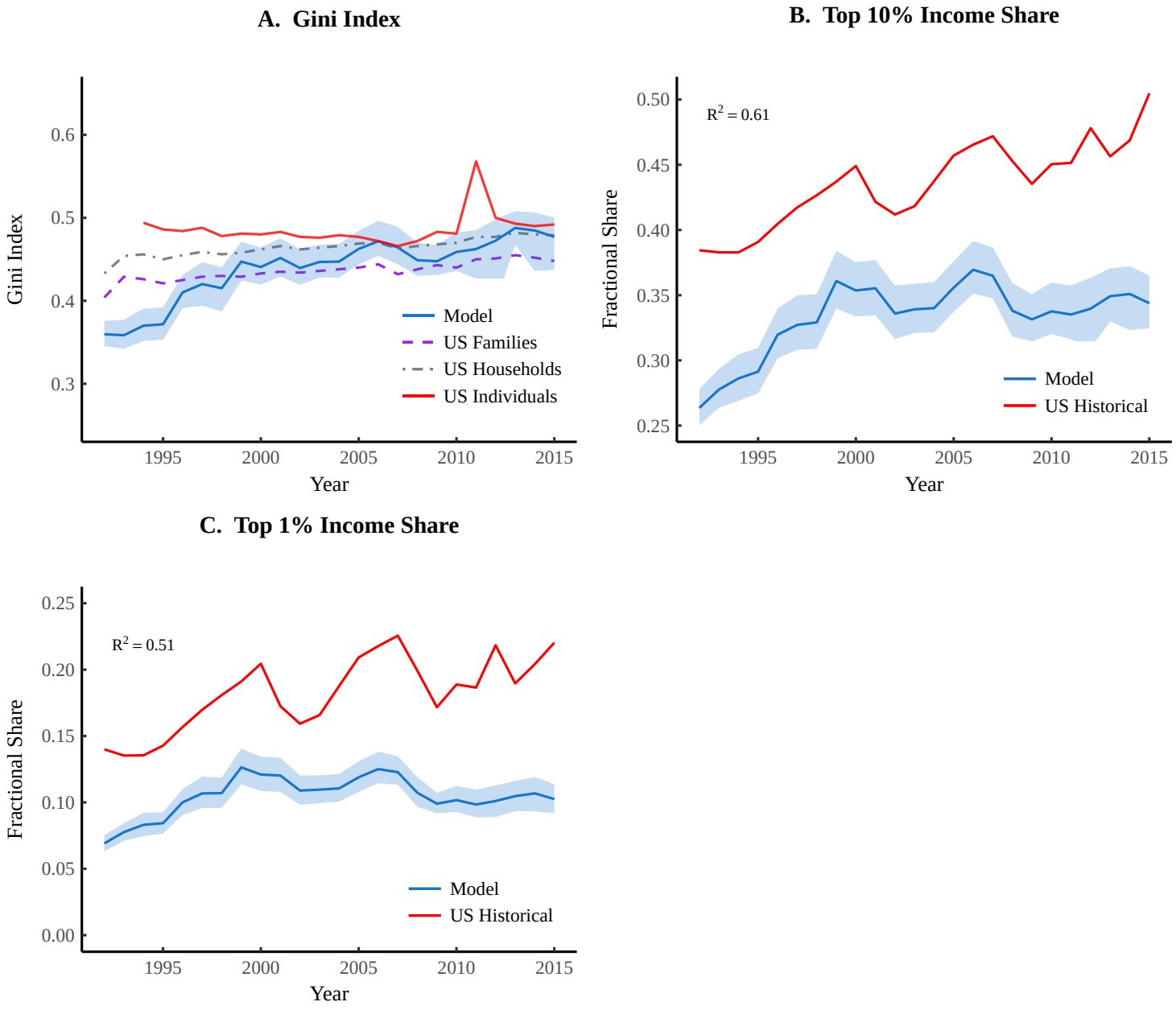


Figure 28: Compustat Model Aggregate Inequality vs. US Historical Data

This figure compares estimates of aggregate income inequality in the Compustat model to US historical data. Panel A compares the model GIni index to three different US measures (the Gini of individuals, families and households). Panel B shows the income share of the top 10%, while panel C shows the top 1%. The shaded regions indicate the 95% confidence interval of the model, estimated over 5000 bootstrap runs. US Gini index data is for individuals, and comes from US Census table PINC-05. The 2011 outlier in US data is likely a statistical error. Families and Household Gini indexes are from the Federal Reserve Bank, series GINIALLRF and GINIALLRH, respectively. US top 10% and top 1% share data is from the World Wealth and Income Database, series *sptinc992j*.

G A Sensitivity Analysis of the Compustat Model

The Compustat model relies on three parameters — a, b, and σ — that are determined from regressions on firm case study data (see Appendix E). Parameters a and b define the span of control, and ultimately determine the 'shape' of a firm's hierarchy. The parameter σ determines the level of income inequality within each hierarchical level of a firm.

Unfortunately, our case study analysis contains only seven firms, and we have no way of knowing if it is a representative sample on which to base our model. Given this ambiguity, it is important to understand the 'sensitivity' of our model results to changes in the parameters a, b, and σ.

Recall that the model results presented in the paper are based on 5000 bootstrap runs of the model — each run uses a different value of a, b, and σ generated by running regressions on resampled (with replacement) case-study data. I use this bootstrapped data to analyze how each parameter affects the following metrics.

1. The G_{BW} metric for hierarchical levels;
2. The G_{BW} metric for firms;
3. Aggregate levels of inequality within the entire model

Figure 29 shows the results of this analysis. We can immediately conclude that the model is not sensitive to the value of σ, which has virtually no effect on any of the above metrics. However, the model appears to be highly sensitive to the parameters a and b. This sensitivity is least pronounced for hierarchical level results. But for firm G_{BW} and aggregate inequality, changes in a and b have a strong effect on model results.

This is an important finding. It suggests that our hierarchical level results (used to test the power-income hypothesis) are relatively robust. A different firm case-study sample would likely not lead to significant changes in our findings. Our firm results, however, are less robust. A different firm case-study sample could lead to very different results.

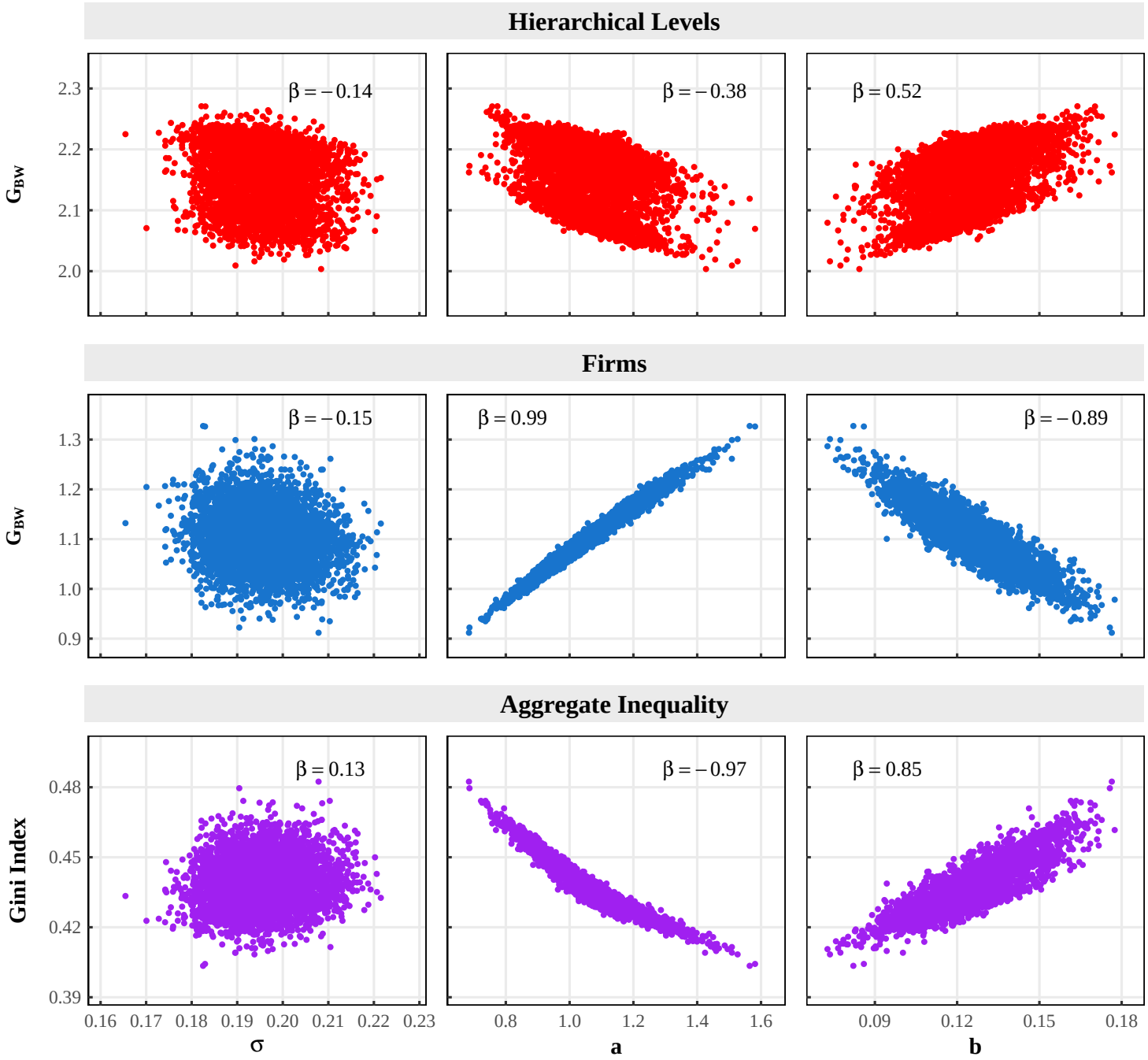


Figure 29: A Sensitivity Analysis of Model Parameters

This figure shows the results of a sensitivity analysis of Compustat model parameters. The top row shows the effect that each parameter has on the G_{BW} metric (the between-group vs. within-group Gini index) for hierarchical levels. The middle row shows the same for firms, and the bottom row for aggregate inequality in the model. Each plotted point indicates a different parameter combination. The normalized regression coefficient β (which can range from -1 to 1) quantifies the model sensitivity.

H The Between-Within Gini Metric and Effect Size

In this section, I discuss what is meant by 'effect size' (in the context of this paper) and how my between-within Gini metric relates to more standard measures of effect size.

H.1 What is Meant By 'Effect Size'

In the context of income distribution, there are two possible ways that we might define effect size:

Definition A: How much a factor effects *total inequality*.
Definition B: How much a factor effects *individual income*.

Effect size definition A refers to what we might call 'inequality accounting'. For instance, we might ask: how much do differences in pay between two groups contribute to total inequality? The point is that this definition attempts to measure how a given factor effects *total inequality*. In general, inequality accounting depends crucially on the *size* of the various groups.

Figure 30 shows this phenomena. In both panels, groups A and B have equal differences in mean income and equal within-group income dispersion. However the total inequality obtained by merging the two groups varies dramatically depending on the relative size of A to B. In Fig. 30A, the two groups are of equal size, while in Fig. 30B, group B is 50 times smaller than group A. The resulting merger of A and B produces much more inequality when the two groups are equal size than when they are not.

Effect size definition B is concerned only with the effect on *individual income*, not on accounting for total inequality. The key difference is that for definition A, we care about group size, while for definition B we do not. In more technical terms, effect size definition B should be calculated by drawing *equal sized samples* from each group. Perhaps the simplest and most intuitive metric of effect size definition B is *Cohen's d* (Eq. 59), defined as the difference in

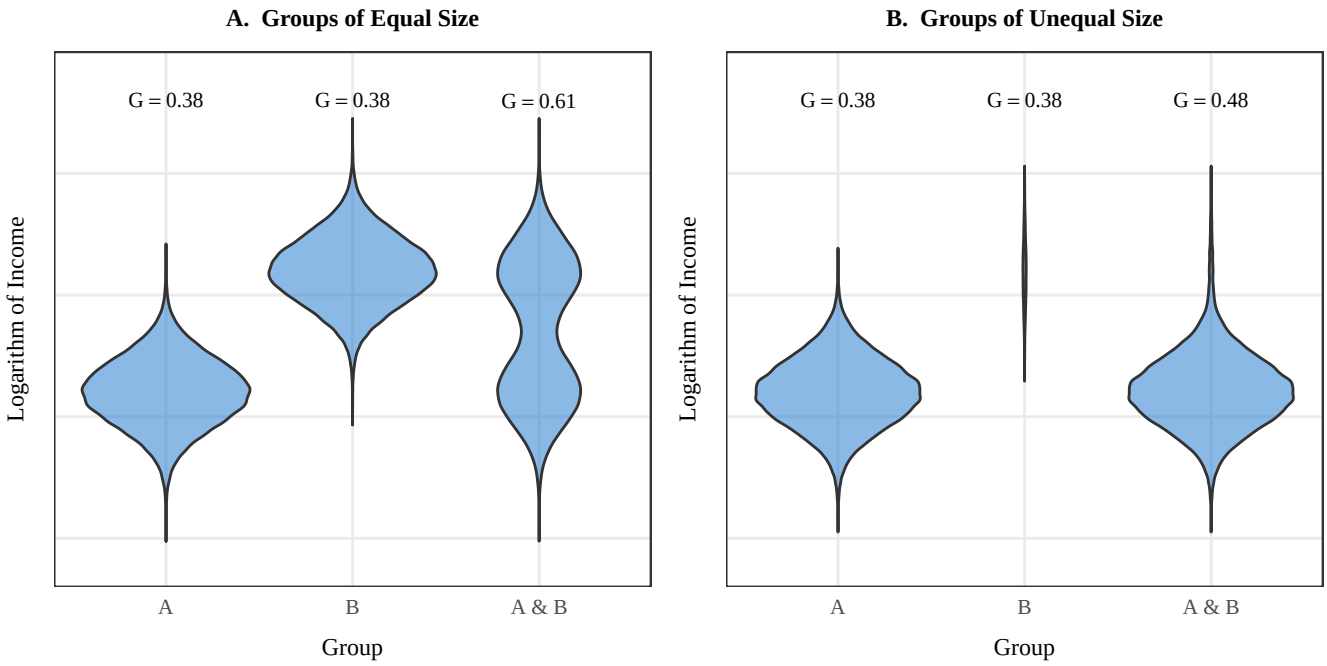


Figure 30: How Group Size Affects Inequality Accounting

This figure shows how differences in group size effect total inequality. In both panels, the income distribution of two different groups (A and B) are shown. The distributions are displayed as 'violin' plots, where the thickness of the violin indicates the number of individuals with that income. In both panels, groups A and B have identical differences in mean income, and identical within-group income dispersion (Gini indexes are shown above each violin). In the left panel, both groups have the same size. In the right panel, group B is 50 times smaller than group A. The rightmost violin plot in each panel shows the income distribution produced by merging groups A and B. Far more inequality is produced when the two groups are of equal size than when there are large differences in size.

means ($\bar{x}$) between two group samples (A and B), divided by the within-group standard deviation (s_W).

$$d = \frac{\bar{x}_B - \bar{x}_A}{s_W} \tag{59}$$

Cohen's d can be interpreted as a *signal-to-noise ratio*. The 'signal' is the difference in means (the effect we want to measure), while the 'noise' is the dispersion within groups, as measured by the standard deviation. In the case of income, the size of the signal-to-noise ratio indicates how accurately we can predict someone's income based only on knowledge of their group membership (either A or B). The larger the signal-to-noise ratio, the more accurate the prediction.

In the example shown in Figure 30, group A and B have the same difference in means and the same within-group standard deviation in both the left and right panel. Therefore Cohen's d would measure an identical effect size. To be clear, this is the effect on *individual income* (definition B), not the effect on inequality (definition A).

H.2 Measuring Effect Size

In this paper, I am concerned only with effect size definition B — the effect on individual income. I have proposed the between-within Gini metric (G_{BW}) as a measure of this type of effect size. This metric is defined by equation 60, where G_B is the Gini index of group means and $\bar{G}_W$ is the mean of all within-group Gini indexes:

$$G_{BW} = \frac{G_B}{\bar{G}_W} \tag{60}$$

How does this metric relate to more standard measures of effect size? It amounts to a signal-to-noise ratio that is similar to Cohen's f^2 measure, the latter of which is a generalization of Cohen's d to many different groups. Cohen's d uses the *difference* between means in the numerator. In order to generalize to many groups, f^2 uses the *sum of squared differences* (SS). To obtain f^2

(Eq. 61), we divide the sum of squares between-groups (SS_B) by the sum of squares within groups (SS_W). See Fleishman [124] and Steiger [125] for a more detailed discussion of the f^2 metric.[6]

$$f^2 = \frac{SS_B}{SS_W} \tag{61}$$

To be clear, SS_B is the sum of squared differences between each group mean ($\bar{x}_i$) and the grand mean ($\bar{x}_{GM}$), multiplied by group size n. Similarly, SS_W is the sum of squared differences between each observation (x_{ij}) and its group mean ($\bar{x}_i$). This double sum operates over each of the k groups and n observations within each group. Lastly, i indexes groups, and j indexes observations within each group.

$$SS_B = n\sum_{i=1}^{k}(\bar{x}_i - \bar{x}_{GM})^2 \tag{62}$$

$$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2 \tag{63}$$

Like Cohen's d, the f^2 metric is a signal-to-noise ratio. The 'signal' is the sum of squares between groups, while the 'noise' is the sum of squares within groups. When applied to income, the size of f^2 indicates the accuracy with which we can predict individual income from group membership.

Comparing the form of f^2 and G_{BW}, we see that the two measures of effect size are very similar. Both are signal-to-noise ratios, consisting of a ratio of between-group dispersion to within-group dispersion. The difference is that f^2 uses the sum of squares to measure dispersion, while G_{BW} uses the Gini index. Given the similarity between G_{BW} and f^2, there should be some relation between the two measures.

[6] A more common formula for this metric is $f^2 = \eta^2/(1-\eta^2)$, where $\eta = SS_B/SS_T$, the sum of squares between groups divided by the total sum of squares. Since $SS_T = SS_B + SS_W$, simple algebraic substitution can prove that the two definitions of f^2 are equivalent.

Rather than attempt to show this similarity analytically, I use simulated data. I build a model based on the following assumptions:

1. Income within groups is lognormally distributed.
2. Within-group income dispersion is the same for all groups (but can vary over different model iterations).
3. Mean income between groups is lognormally distributed (and can vary between iterations).
4. Total inequality is (roughly) constant for all iterations.
5. The size of each group is constant.
6. The number of groups varies (between iterations) from 2 to 100.

For each iteration of the model, we define the mean income of each group by drawing randomly from a lognormal distribution. We then simulate individuals within each group by drawing randomly from (a different) lognormal distribution. The model has 2 key parameters: the lognormal scale parameter that defines the dispersion *between* groups, and the lognormal scale parameter that determines dispersion *within* groups. Varying these parameters changes the size of the group-income effect. For consistency, I use only parameter combinations that produce roughly the same level of total inequality (a Gini index of 0.5).

For each set of simulated data, I calculate both G_{BW} and f^2. Because analysis of variance typically assumes that within-group data is normally distributed, I calculate f^2 using the *logarithm* of income. The results are shown in Figure 31. As expected, there is an extremely strong relation between the two effect-size measures. This indicates that an f^2 test of the power-income effect would likely give very similar results to the G_{BW} findings shown in Figure 12. To reiterate, I do not conduct such an f^2 test in this paper because the relevant data is not available.

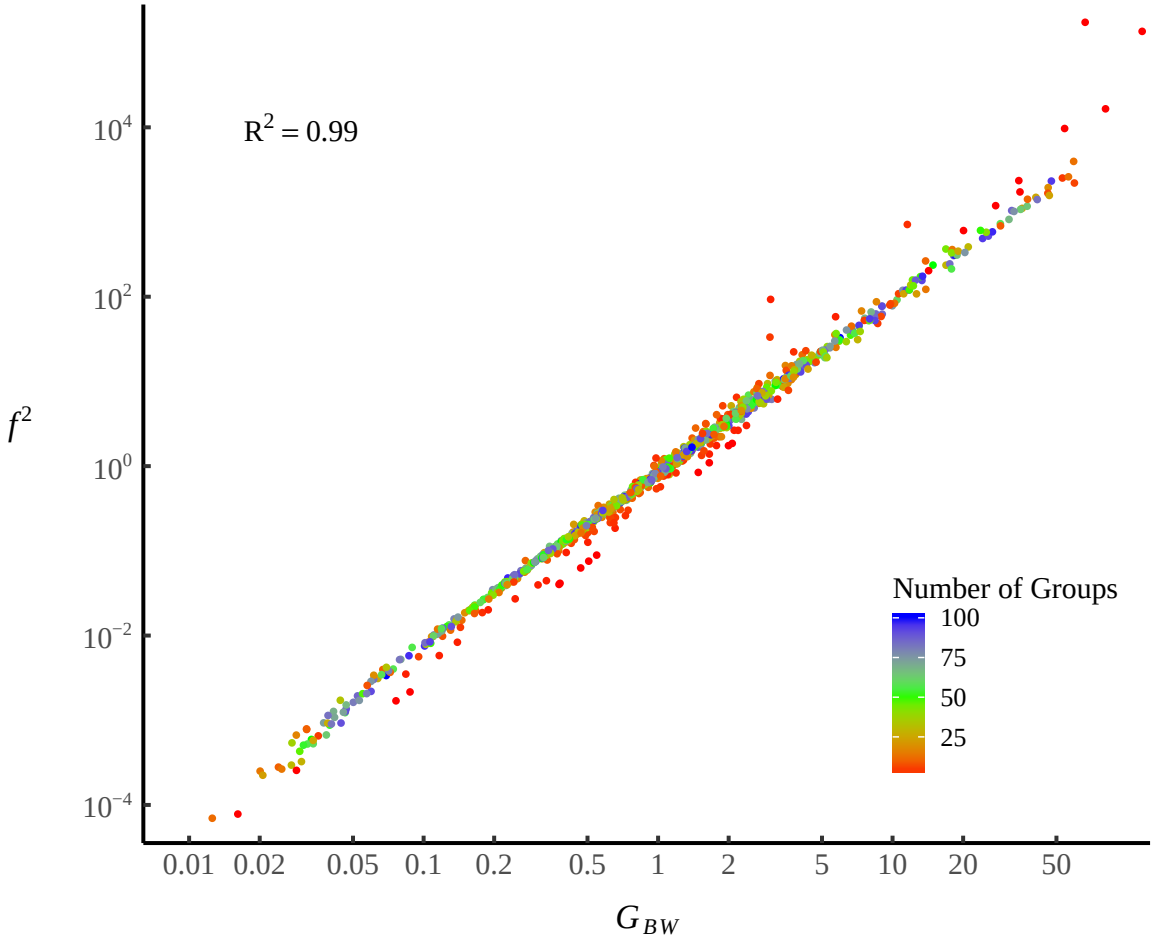


Figure 31: Standard Effect Size Measure f^2 vs. the Gini Metric G_{BW}

This figure compares Cohen's f^2 metric of effect size (Eq. 61) to my between-within Gini metric, G_{BW} (Eq. 60). The comparison uses simulated data, and each data point represents different parameter combinations (see model assumptions above). Color indicates the number of groups used in each iteration. R code for the model is available in the Supplementary Material.

H.3 A Note on Group Size

The simulation shown in Figure 31 shows a special case where all groups have the same size. However, for the vast majority of the income-affecting factors studied in this paper, the groups do *not* have the same size. For instance, there are *vastly* more people in lower hierarchical levels than in upper hierarchical levels. How do we deal with this situation?

The key ingredient of effect-size definition B is that it weights each different group *equally* (rather than weighting a group by its size). One way to achieve this equal weighting is to draw *equal-sized samples* from each group and calculate Cohen's f^2 using equations 61-63. But what if we do not have raw data? What if we have only summary statistics such as the mean and standard deviation of each group?

We can proceed by noting that an alternative way to define Cohen's f^2 is as the ratio of between-group variance σ_B^2 and mean within-group variance $\bar{\sigma_W^2}$:

$$f^2 = \frac{\sigma_B^2}{\bar{\sigma_W^2}} \tag{64}$$

$$\sigma_B^2 = \sum_{i=1}^{k} \frac{(\bar{x}_i - \bar{x}_{GM})^2}{k} \tag{65}$$

$$\bar{\sigma_W^2} = \sum_{i=1}^{k}\sum_{j=1}^{n} \frac{(x_{ij} - \bar{x}_i)^2}{nk} = \sum_{i=1}^{k} \frac{\sigma_i^2}{k} \tag{66}$$

Here, σ_B^2 and $\bar{\sigma_W^2}$ are derived by dividing SS_B and SS_W (respectively) by nk. Given group means ($\bar{x}_i$) and within-group standard deviations(σ_i), we can use this alternative formula to calculate f^2.

What equation 64 does is give *identical weight* to each group's summary statistics. This accomplishes the same thing as if we took equal sized samples from raw data and used Eq. 61-63 to calculate f^2. This same logic applies to my construction of the G_{BW} metric: it calculates a signal-to-noise ratio from summary statistics by giving equal weight to each group.

References

[1] Thomas Piketty. *Capital in the Twenty-first Century*. Harvard University Press, Cambridge, 2014.

[2] Thomas Piketty and Emmanuel Saez. Income Inequality in the United States, 1913-1998 (series updated to 2000 available). Technical report, National Bureau of Economic Research, 2001.

[3] Thomas Piketty and Emmanuel Saez. The evolution of top incomes: a historical and international perspective. Technical report, National Bureau of Economic Research, 2006.

[4] Anthony Barnes Atkinson and Thomas Piketty. *Top incomes: A global perspective*. Oxford University Press, 2010.

[5] Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. The top 1 percent in international and historical perspective. *The Journal of Economic Perspectives*, 27(3):3–20, 2013.

[6] Francis Galton. *Hereditary genius: An inquiry into its laws and consequences*, volume 27. Macmillan, 1869.

[7] Vilfredo Pareto. *Cours d'economie politique*, volume 1. Librairie Droz, 1897.

[8] Gian Singh Sahota. Theories of personal income distribution: a survey. *Journal of Economic Literature*, 16(1):1–55, 1978.

[9] David G. Champernowne. A model of income distribution. *The Economic Journal*, 63(250):318–351, 1953.

[10] Robert Gibrat. *Les inegalites economiques*. Recueil Sirey, 1931.

[11] R.S.G. Rutherford. Income distributions: a new model. *Econometrica: Journal of the Econometric Society*, 23(3):277–294, 1955.

[12] Herbert A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[13] Herbert A. Simon and Charles P. Bonini. The size distribution of business firms. *The American Economic Review*, 48(4):607–617, 1958.

[14] George Kingsley Zipf. *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge, 1949.

[15] John Angle. The surplus theory of social stratification and the size distribution of personal wealth. *Social Forces*, 65(2):293–326, 1986.

[16] Arnab Chatterjee and Bikas K. Chakrabarti. Kinetic exchange models for income and wealth distributions. *The European Physical Journal B*, 60(2):135–149, 2007.

[17] Arnab Chatterjee, Bikas K. Chakrabarti, and Anirban Chakraborti. *Econophysics and sociophysics: trends and perspectives*. John Wiley & Sons, Milan, 2007.

[18] Salete Pianegonda, Jose Rroberto Iglesias, G. Abramson, and J. L. Vega. Wealth redistribution with conservative exchanges. *Physica A: Statistical Mechanics and its Applications*, 322:667–675, 2003.

[19] Mauro Gallegati, Steve Keen, Thomas Lux, and Paul Ormerod. Worrying trends in econophysics. *Physica A: Statistical Mechanics and its Applications*, 370(1):1–6, 2006.

[20] David Ricardo. *Principles of political economy and taxation*. G. Bell and Sons, London, 1817.

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

[22] Gary S. Becker. Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49, 1962.

[23] Jacob Mincer. Investment in human capital and personal income distribution. *The journal of political economy*, 66(4):281–302, 1958.

[24] Theodore W. Schultz. Investment in human capital. *The American Economic Review*, 51(1):1–17, 1961.

[25] Isaak Ilich Rubin. *Essays on Marx's theory of value*. Black Rose Books Ltd., 1973.

[26] C. Harry Boissevain. Distribution of abilities depending upon two or more independent factors. *Metron*, 13(4):49–58, 1939.

[27] Benoit Mandelbrot. The Pareto-Levy law and the distribution of income. *International Economic Review*, 1(2):79–106, 1960.

[28] John E. Hunter, Frank L. Schmidt, and Michael K. Judiesch. Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75(1):28, 1990.

[29] Avi J. Cohen and Geoffrey C. Harcourt. Retrospectives: Whatever happened to the Cambridge capital theory controversies? *Journal of Economic Perspectives*, 17(1):199–214, 2003.

[30] Joan Robinson. The production function and the theory of capital. *The Review of Economic Studies*, 21(2):81–106, 1953.

[31] John Pullen. *The marginal productivity theory of distribution: a critical history*. Routledge, London, 2009.

[32] John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

[33] John C. Haltiwanger, Julia I. Lane, and James R. Spletzer. Productivity differences across employers: The roles of employer size, age, and human capital. *The American Economic Review*, 89(2):94–98, 1999.

[34] Jonathan Haskel, Denise D. Hawkes, and Sonia C. Pereira. Skills, human capital and the plant productivity gap: UK evidence from matched plant, worker and workforce data. *CEPR Discussion Paper*, (5334), 2005.

[35] Judith K. Hellerstein, David Neumark, and Kenneth R. Troske. Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. Technical Report Working Paper No. 5626, National Bureau of Economic Research, 1996.

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

[37] Susana Iranzo, Fabiano Schivardi, and Elisa Tosetti. Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data. *Journal of Labor Economics*, 26(2):247–285, 2008.

[38] Nicholas Oulton. Competition and the dispersion of labour productivity amongst UK companies. *Oxford Economic Papers*, 50(1):23–38, 1998.

[39] Gerhard E. Lenski. *Power and privilege: A theory of social stratification*. UNC Press Books, 1966.

[40] Jean-Jacques Hublin, Abdelouahed Ben-Ncer, Shara E. Bailey, Sarah E. Freidline, Simon Neubauer, Matthew M. Skinner, Inga Bergmann, Adeline Le Cabec, Stefano Benazzi, Katerina Harvati, and Philipp Gunz. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature*, 546(7657):289–292, June 2017.

[41] Daniel Richter, Rainer Grün, Renaud Joannes-Boyau, Teresa E. Steele, Fethi Amani, Mathieu Rué, Paul Fernandes, Jean-Paul Raynal, Denis Geraads, Abdelouahed Ben-Ncer, Jean-Jacques Hublin, and Shannon P. McPherron. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature*, 546(7657):293–296, June 2017.

[42] Kenneth M. Ames. On the evolution of the human capacity for inequality and/or egalitarianism. In *Pathways to power*, pages 15–44. Springer, 2010.

[43] Sergey Gavrilets. On the evolutionary origins of the egalitarian syndrome. *Proceedings of the National Academy of Sciences*, 109(35):14069–14074, 2012.

[44] Christopher Boehm, Harold B. Barclay, Robert Knox Dentan, Marie-Claude Dupre, Jonathan D. Hill, Susan Kent, Bruce M. Knauft, Keith F. Otterbein, and Steve Rayner. Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. *Current Anthropology*, 34(3):227–254, 1993.

[45] Carles Boix. Origins and persistence of economic inequality. *Annual Review of Political Science*, 13:489–516, 2010.

[46] T. Douglas Price and Ofer Bar-Yosef. Traces of inequality at the origins of agriculture in the ancient Near East. In *Pathways to Power*, pages 147–168. Springer, 2010.

[47] T. Douglas Price and Gary M. Feinman. Social inequality and the evolution of human social organization. In *Pathways to power*, pages 1–14. Springer, 2010.

[48] Gary M. Feinman. The emergence of inequality. In *Foundations of Social Inequality*, pages 255–279. Springer, 1995.

[49] C. Wright Mills. *The power elite*. Oxford University Press, 1956.

[50] Jim Sidanius and Felicia Pratto. *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press, 2001.

[51] Max Weber. *Economy and society: An outline of interpretive sociology*. Univ of California Press, 1978.

[52] Herbert A. Simon. The compensation of executives. *Sociometry*, 20(1):32–35, 1957.

[53] Harold F. Lydall. The distribution of employment incomes. *Econometrica: Journal of the Econometric Society*, 27(1):110–115, 1959.

[54] F. G. Barroso, Concepción L. Alados, and J. Boza. Social hierarchy in the domestic goat: effect on food habits and production. *Applied Animal Behaviour Science*, 69(1):35–53, 2000.

[55] A. M. Guhl, N. E. Collias, and W. C. Allee. Mating behavior and the social hierarchy in small flocks of white leghorns. *Physiological Zoology*, 18(4):365–390, 1945.

[56] S. Kondo and J. F. Hurnik. Stabilization of social hierarchy in dairy cows. *Applied Animal Behaviour Science*, 27(4):287–297, 1990.

[57] G. B. Meese and R. Ewbank. The establishment and nature of the dominance hierarchy in the domesticated pig. *Animal Behaviour*, 21(2):326–334, 1973.

[58] Robert M. Sapolsky. The influence of social hierarchy on primate health. *Science*, 308(5722):648–652, 2005.

[59] Jacob Uhrich. The social hierarchy in albino mice. *Journal of Comparative Psychology*, 25(2):373, 1938.

[60] Brenda J. Bradley, Martha M. Robbins, Elizabeth A. Williamson, H. Dieter Steklis, Netzin Gerald Steklis, Nadin Eckhardt, Christophe Boesch, and Linda Vigilant. 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*, 102(26):9418–9423, 2005.

[61] Michael P. Haley, Charles J. Deutsch, and Burney J. Le Boeuf. Size, dominance and copulatory success in male northern elephant seals, Mirounga angustirostris. *Animal Behaviour*, 48(6):1249–1260, 1994.

[62] Derek J. Girman, M. G. L. Mills, Eli Geffen, and Robert K. Wayne. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (Lycaon pictus). *Behavioral Ecology and Sociobiology*, 40(3):187–198, 1997.

[63] Ulrike Gerloff, Bianka Hartung, Barbara Fruth, Gottfried Hohmann, and Diethard Tautz. 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*, 266(1424):1189–1195, 1999.

[64] Emily E. Wroblewski, Carson M. Murray, Brandon F. Keele, Joann C. Schumacher-Stankey, Beatrice H. Hahn, and Anne E. Pusey. Male dominance rank and reproductive success in chimpanzees, Pan troglodytes schweinfurthii. *Animal Behaviour*, 77(4):873–885, 2009.

[65] Daniel G. Frankel and Tali Arbel. Group formation by two-year olds. *International Journal of Behavioral Development*, 3(3):287–298, 1980.

[66] Ritch C. Savin-Williams. Dominance hierarchies in groups of middle to late adolescent males. *Journal of Youth and Adolescence*, 9(1):75–85, 1980.

[67] Floyd F. Strayer and M. Trudel. Developmental changes in the nature and function of social dominance among young children. *Ethology and Sociobiology*, 5(4):279–295, 1984.

[68] Rosemary L. Hopcroft. Sex, status, and reproductive success in the contemporary United States. *Evolution and Human Behavior*, 27(2):104–120, 2006.

[69] Talcott Parsons. An analytical approach to the theory of social stratification. *American Journal of Sociology*, 45(6):841–862, 1940.

[70] Kingsley Davis and Wilbert E. Moore. Some principles of stratification. *American sociological review*, 10(2):242–249, 1945.

[71] Ralf Dahrendorf. *Class and class conflict in industrial society*. Stanford University Press, 1959.

[72] K. Marx. *Capital, Volume I*. Harmondsworth: Penguin/New Left Review, 1867.

[73] Antonio Gilman. The development of social stratification in Bronze Age Europe. *Current anthropology*, 22(1):1–23, 1981.

[74] Dusan Boric. Social dimensions of mortuary practices in the Neolithic: a case study. *Starinar*, (47):67–83, 1996.

[75] Paul Halstead. Spondylus shell ornaments from late Neolithic Dimini, Greece: specialized manufacture or unequal accumulation? *Antiquity*, 67(256):603–609, 1993.

[76] P. Van der Velde. Banderamik social inequalitya case study. *Germania*, 68(1):19–38, 1990.

[77] Gonzalo Aranda and Fernando Molina. Wealth and Power in the Bronze Age of the South-East of the Iberian Peninsula: The Funerary Record of Cerro De La Encina. *Oxford Journal of Archaeology*, 25(1):47–59, February 2006.

[78] Gonzalo Aranda-Jiménez, Sandra Montón-Subías, and Silvia Jiménez-Brobeil. Conflicting evidence? Weapons and skeletons in the Bronze Age of south-east Iberia. *Antiquity*, 83(322):1038–1051, 2009.

[79] Giampaolo Graziadio. The process of social stratification at Mycenae in the Shaft Grave period: a comparative examination of the evidence. *American journal of archaeology*, 95(3):403–440, 1991.

[80] Kristian Kristiansen. *Europe before history*. Cambridge University Press, New York, 2000.

[81] Robin Harding. Data shift to lift US economy 3%. *Financial Times*, 2013.

[82] Michael Dee, David Wengrow, Andrew Shortland, Alice Stevenson, Fiona Brock, Linus Girdland Flink, and Christopher Bronk Ramsey. An absolute chronology for early Egypt using radiocarbon dating and Bayesian statistical modelling. *Proc. R. Soc. A*, 469:20130395, 2013.

[83] Julian Reade. Assyrian king-lists, the royal tombs of Ur, and Indus origins. *Journal of Near Eastern Studies*, 60(1):1–29, 2001.

[84] Zhiyu Guo, Kexin Liu, Xiangyang Lu, Hongji Ma, Kun Li, Sixun Yuan, and Xiaohong Wu. The use of AMS radiocarbon dating for XiaShangZhou chronology. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 172(1):724–731, 2000.

[85] Gaetano Mosca. *On the Ruling Class*. McGraw Hill, New York, 1939.

[86] Richard Dawkins. *The Selfish Gene*. Oxford University Press, 1976.

[87] William D. Hamilton. The genetical evolution of social behaviour. II. *Journal of theoretical biology*, 7(1):17–52, 1964.

[88] Philip Drucker. Rank, wealth, and kinship in Northwest Coast society. *American Anthropologist*, 41(1):55–65, 1939.

[89] Raymond Firth. We, the Tikopia: kinship in primitive Polynesia. *Boston: Beacon*, 1936.

[90] Edward Winslow Gifford. *Tongan society*. Bernice P. Bishop Museum, 1929.

[91] Per Hage and Frank Harary. *Island networks: communication, kinship, and classification structures in Oceania*. Number 11. Cambridge University Press, 1996.

[92] Paul Kirchhoff. *The principles of clanship in human society*. Bobbs-Merrill, 1955.

[93] Edmund R. Leach. *Political systems of highland Burma: a study of Kachin social structure*. G. Bell and Sons, London, 1954.

[94] Kalervo Oberg. Types of social structure among the lowland tribes of South and Central America. *American anthropologist*, 57(3):472–487, 1955.

[95] Marshall D. Sahlins. Poor man, rich man, big-man, chief: political types in Melanesia and Polynesia. *Comparative studies in society and history*, 5(03):285–303, 1963.

[96] Lewis Henry Morgan. *Ancient society; or, researches in the lines of human progress from savagery, through barbarism to civilization*. Charles H. Kerr & Company, Chicago, 1877.

[97] Friedrich Engels and Tristram Hunt. *The origin of the family, private property and the state*. Verlag der Schweizerischen Volksbuchhandlung, Zurich, 1884.

[98] Joan B. Silk. Kin selection in primate groups. *International Journal of Primatology*, 23(4):849–875, 2002.

[99] Joan B. Silk. Nepotistic cooperation in non-human primate groups. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1533):3243–3254, 2009.

[100] Paul J. Greenwood. Mating systems, philopatry and dispersal in birds and mammals. *Animal behaviour*, 28(4):1140–1162, 1980.

[101] Giovanni Destro-Bisol, Francesco Donati, Valentina Coia, Ilaria Boschi, Fabio Verginelli, Alessandra Caglia, Sergio Tofanelli, Gabriella Spedini, and Cristian Capelli. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Molecular Biology and Evolution*, 21(9):1673–1682, 2004.

[102] Paul Verdu, Noémie SA Becker, Alain Froment, Myriam Georges, Viola Grugni, Lluis Quintana-Murci, Jean-Marie Hombert, Lolke Van der Veen,

Sylvie Le Bomin, Serge Bahuchet, and others. Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African Pygmies and non-Pygmies. *Molecular biology and evolution*, 30(4):918–937, 2013.

[103] Michael GB Blum, Evelyne Heyer, Olivier François, and Fréderic Austerlitz. Matrilineal Fertility Inheritance Detected in HunterGatherer Populations Using the Imbalance of Gene Genealogies. *PLoS Genetics*, 2(8):e122, 2006.

[104] David F. Aberle. Matrilineal descent in cross-cultural perspective. In *Matrilineal kinship*, pages 655–727. 1961.

[105] Clare Janaki Holden and Ruth Mace. Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1532):2425–2433, 2003.

[106] Monika Karmin, Lauri Saag, Mário Vicente, Melissa A. Wilson Sayres, Mari Järve, Ulvi Gerst Talas, Siiri Rootsi, Anne-Mai Ilumäe, Reedik Mägi, Mario Mitt, and others. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome research*, 25(4):459–466, 2015.

[107] Shao-Qing Wen, Xin-Zhu Tong, and Hui Li. Y-chromosome-based genetic pattern in East Asia affected by Neolithic transition. *Quaternary International*, 426:50–55, 2016.

[108] John Hartung, Mildred Dickemann, Umberto Melotti, Leopold Pospisil, Eugenie C. Scott, John Maynard Smith, and William D. Wilder. Polygyny and inheritance of wealth [and comments and replies]. *Current anthropology*, 23(1):1–12, 1982.

[109] Martin S. Smith, Bradley J. Kish, and Charles B. Crawford. Inheritance of wealth as human kin investment. *Ethology and Sociobiology*, 8(3):171–182, 1987.

[110] Laura L. Betzig. Despotism and differential reproduction: A cross-cultural correlation of conflict asymmetry, hierarchy, and degree of polygyny. *Ethology and Sociobiology*, 3(4):209–221, 1982.

[111] Clare Janaki Holden, Rebecca Sear, and Ruth Mace. Matriliny as daughter-biased investment. *Evolution and Human Behavior*, 24(2):99–112, 2003.

[112] Laura Betzig. Means, variances, and ranges in reproductive success: comparative evidence. *Evolution and Human Behavior*, 33(4):309–317, 2012.

[113] T. Douglas Price. Social Inequality at the Origins of Agriculture. In *Foundations of Social Inequality*. Plenum Press, New York, 1995.

[114] Peter Turchin and Sergey Gavrilets. Evolution of complex hierarchical societies. *Social Evolution and History*, 8(2):167–198, 2009.

[115] Peter Turchin. Warfare and the evolution of social complexity: a multilevel-selection approach. *Structure and Dynamics*, 4(3), 2010.

[116] Carles Boix and Frances Rosenbluth. Bones of contention: the political economy of height inequality. *American Political Science Review*, 108(01):1–22, 2014.

[117] Nicolas Verdier. Hierarchy: a short history of a word in Western thought. In *Hierarchy in natural and social sciences*, pages 13–37. Springer, 2006.

[118] Rick Audas, Tim Barmby, and John Treble. Luck, effort, and reward in an organizational hierarchy. *Journal of Labor Economics*, 22(2):379–395, 2004.

[119] George Baker, Michael Gibbs, and Bengt Holmstrom. Hierarchies and compensation: A case study. *European Economic Review*, 37(2-3):366–378, 1993.

[120] Thomas J. Dohmen, Ben Kriechel, and Gerard A. Pfann. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. *Journal of Population Economics*, 17(2):193–228, 2004.

[121] Francisco Lima. Internal Labor Markets: A Case Study. *FEUNL Working Paper*, 378, 2000.

[122] Filipe Morais and Nada K. Kakabadse. The Corporate Gini Index (CGI) determinants and advantages: Lessons from a multinational retail company case study. *International Journal of Disclosure and Governance*, 11(4):380–397, 2014.

[123] John Treble, Edwin Van Gameren, Sarah Bridges, and Tim Barmby. The internal economics of the firm: further evidence from personnel data. *Labour Economics*, 8(5):531–552, 2001.

[124] Allen I. Fleishman. Confidence intervals for correlation ratios. *Educational and Psychological Measurement*, 40(3):659–670, 1980.

[125] James H. Steiger. Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological methods*, 9(2):164–182, 2004.

[126] George Deltas. The small-sample bias of the Gini coefficient: results and implications for empirical research. *Review of economics and statistics*, 85(1):226–234, 2003.

[127] Holger M. Mueller, Paige Parker Ouimet, and Elena Simintzi. Within-Firm Pay Inequality. *SSRN Working Paper*, 2016.

[128] Fredrik Heyman. Pay inequality and firm performance: evidence from matched employeremployee data. *Applied Economics*, 37(11):1313–1327, 2005.

[129] Milton Friedman. *Capitalism and Freedom*. University of Chicago press, Chicago, 1962.

[130] Samuel Bowles, Herbert Gintis, and Melissa Osborne. The determinants of earnings: A behavioral approach. *Journal of economic literature*, 39(4):1137–1176, 2001.

[131] Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623, 2014.

[132] Bradley Efron and Robert J. Tibshirani. *An introduction to the bootstrap*. CRC press, London, 1994.

[133] Christian Grund. The wage policy of firms: comparative evidence for the US and Germany from personnel data. *The International Journal of Human Resource Management*, 16(1):104–119, 2005.

[134] Kenn Ariga, Giorgio Brunello, Yasushi Ohkusa, and Yoshihiko Nishiyama. Corporate hierarchy, promotion, and firm growth: Japanese internal labor market in transition. *Journal of the Japanese and International Economies*, 6(4):440–471, 1992.

[135] Brian Bell and John Van Reenen. Firm performance and wages: evidence from across the corporate hierarchy. *CEP Discussion paper*, 1088, 2012.

[136] Tor Eriksson. Executive compensation and tournament theory: Empirical tests on Danish data. *Journal of Labor Economics*, 17(2):262–280, 1999.

[137] Jonathan S. Leonard. Executive pay and firm performance. *Industrial and Labor Relations Review*, 43(3):13–28, 1990.

[138] Brian GM Main, Charles A. O'Reilly III, and James Wade. Top executive pay: Tournament or teamwork? *Journal of Labor Economics*, 11(4):606–628, 1993.

[139] Raghuram G. Rajan and Julie Wulf. The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. *The Review of Economics and Statistics*, 88(4):759–773, 2006.

[140] Hung-Lin Tao and I.-Ting Chen. The level of technology employed and the internal hierarchical wage structure. *Applied Economics Letters*, 16(7):739–744, 2009.

[141] Robert L. Axtell. Zipf distribution of US firm sizes. *Science*, 293:1818–1820, 2001.

[142] Charles Brown and James Medoff. The employer size-wage effect. *Journal of political Economy*, 97(5):1027–1059, 1989.

[143] Paul Joskow, Nancy Rose, Andrea Shepard, John R. Meyer, and Sam Peltzman. Regulatory constraints on CEO compensation. *Brookings Papers on Economic Activity. Microeconomics*, 1993(1):1–72, 1993.