Journal of Computational Social Science
https://doi.org/10.1007/s42001-018-0019-8

RESEARCH ARTICLE

Hierarchy and the power-law income distribution tail

Blair Fix[1]

Received: 22 April 2018 / Accepted: 9 July 2018

Abstract

What explains the power-law distribution of top incomes? This paper tests the hypothesis that it is firm hierarchy that creates the power-law income distribution tail. Using the available case-study evidence on firm hierarchy, I create the first large-scale simulation of the hierarchical structure of the US private sector. Although not tuned to do so, this model reproduces the power-law scaling of top US incomes. I show that this is purely an effect of firm hierarchy. This raises the possibility that the ubiquity of power-law income distribution tails is due to the ubiquity of hierarchical organization in human societies.

Keywords Power law · Income distribution · Firm hierarchy · Economic modeling

Introduction

In the late nineteenth century, Pareto [1] discovered that top incomes could be modeled with a power-law distribution. This scaling behavior meant that the income distribution tail could be approximated by the simple probability function:

$$p(x) = \frac{c}{x^{\alpha}}. \tag{1}$$

Here, $p(x)$ is the probability of finding an individual (in the tail) with income x, c is a constant,[1] and α is the scaling exponent, which captures the 'fatness' of the income distribution tail. The beauty of a power law is its simplicity. The important

[1] The constant c is equal to $(\alpha - 1)/(x_{min})^{1-\alpha}$, where x_{min} is the lower bound of the power law (i.e., the beginning of the income distribution tail).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42001-018-0019-8) contains supplementary material, which is available to authorized users.

Blair Fix
blairfix@gmail.com

1 York University, Toronto, ON, Canada

Published online: 16 July 2018

properties of the distribution tail are captured by a single parameter—the power-law exponent. Since Pareto's initial discovery, the power-law scaling of top incomes has been re-confirmed many times (for a non-exhaustive list, see [2–12]).

What causes this nearly universal behavior? Is there a universal generation mechanism at work? Over the century since Pareto's landmark discovery, many power-law generation mechanisms have been suggested to explain the scaling of top incomes. While the various mechanisms (reviewed below) differ in their mathematical properties, most are united by a shared focus on the stochastic growth of individual income.

This paper investigates a very different explanation for the power-law scaling of top incomes. I test the hypothesis that it is *firm hierarchy* that creates the power-law tail. This approach was first proposed by Lydall [13], who used a simple model to show that the hierarchical structure of firms could create a power-law distribution. At the time, Lydall's work was largely speculative since little was known about the internal structure of firms. However, in the last two decades the empirical study of firm hierarchy has blossomed (for case studies, see [14–20]; for aggregate studies, see [21–29]). Enough evidence now exists that we can begin to explore the distributional consequences of hierarchy. To conduct this investigation, I use the existing case-study evidence to build a large-scale simulation of firm hierarchy. This model generalizes the trends found in case-study firms to create the first simulation of the hierarchical structure of the US private sector.

I verify the accuracy of the hierarchy model, in two ways. I first compare the model's income distribution to that of the USA. I find that the hierarchy model does a reasonably good job of reproducing the key properties of US income distribution. Importantly, the model produces (without tuning it to do so) a power-law tail that is statistically identical to US empirical data. Next, I test a key feature of the hierarchy model—that top-earning individuals should be concentrated in large firms. I find that the model's prediction is consistent with the available US evidence.

Having established the model's accuracy, I then use the model to investigate the distributional effects of hierarchy. I find that it is firm hierarchy alone (and not any of the other income dispersion factors included in the model) that is responsible for generating the power-law tail. To summarize, the hierarchy model suggests that it is firm hierarchy (and its associated properties) that creates the power-law scaling of top incomes. This finding has important implications for both the empirical and theoretical study of income distribution. On the empirical side, these results indicate that the income effects of hierarchy are significant and need to be studied in more detail. On the theoretical side, these results suggest that hierarchy is a plausible mechanism for generating the power-law scaling of top incomes. This raises the possibility that the ubiquity of power-law income distribution tails is due to the ubiquity of hierarchical organization in human societies.

The paper is laid out as follows: "Power-law generation mechanisms" reviews the different mechanisms for generating power-law distributions. "A firm hierarchy model" outlines (in non-technical terms) the basic properties of the hierarchy model. (For a technical discussion, see the Online Appendices). "Testing the hierarchy model: macro predictions" and "Testing the hierarchy model: micro predictions" test the model against empirical data. "Isolating the distributional role of firm hierarchy" demonstrates that it is firm hierarchy (alone) that is responsible for creating the model's power-law tail, and "How hierarchy generates the power-law tail analyzes how this is achieved. I conclude, in "Discussion" and "Conclusions", with a discussion of the significance of these results and propose avenues for future research.

Power-law generation mechanisms

I review here in non-technical terms the various mechanisms for generating power-law distributions, with an emphasis on those that have been applied to modeling income. For a good technical review of these mechanisms, see [30–32]. One way to generate power laws is through a stochastic, multiplicative growth process. This mechanism was identified by Gibrat [33], but was first applied to income distribution by Champernowne [34], followed by many others [35–38]. The basic idea is that individual income is subjected to stochastic, multiplicative 'shocks'. Under the condition that these multiplicative shocks are scale free (they do not depend on income size) and there is a minimum (reflective) lower bound on income, this process will produce a power-law distribution of income.

Closely related to this process is the mechanism of 'preferential attachment', sometimes called the 'rich get richer'. Developed independently by Yule [39], Simon [40], Price [41] and Barabasi and Albert [42], this process involves stochastic addition with conditional probability. It is most easily applicable to the distribution of wealth (not income). We imagine a society in which units of wealth are added at random. If the probability of an individual receiving an additional unit of wealth is proportional to his/her existing wealth, the result (after many iterations) will be a power-law distribution of wealth.

In both multiplicative growth and preferential attachment models, the source for stochastic changes in income/wealth is left unexplained. More recently, econophysicists have developed a more sophisticated class of model that attempts to explain these 'shocks' in terms of the random exchange of money between pairs of interacting agents [11, 43–51]. These 'kinetic-exchange models' draw explicitly on the statistical mechanics of gases. Agents exchange money much like gas particles exchange kinetic energy. Given certain assumptions about these interactions, kinetic-exchange models can produce a power-law distribution.

Lastly, a very simple way to produce a power law is to *exponentially* transform an *exponential* distribution. This is the mechanism by which Lydall [13] showed that firm hierarchy could create a power law. If a firm hierarchy has a constant 'span of control' (the number of subordinates controlled by each superior), then relative employment will decrease exponentially by rank. If, at the same time, income increases exponentially by rank, the result will be a power-law distribution of income. (For a technical discussion, see "How hierarchy generates the power-law tail").

The advantage of this hierarchy mechanism is that it ties income to *institutions*. This means that the power-law distribution of top incomes is given an explicit institutional basis—something that is important when it comes to policy discussions about how to reduce inequality. The disadvantage of the hierarchy mechanism is that relatively little is known about firm hierarchical structure. This means that the distributional consequences of hierarchy are little understood. This paper aims to remedy this situation by using the available empirical data to build a large-scale hierarchical model

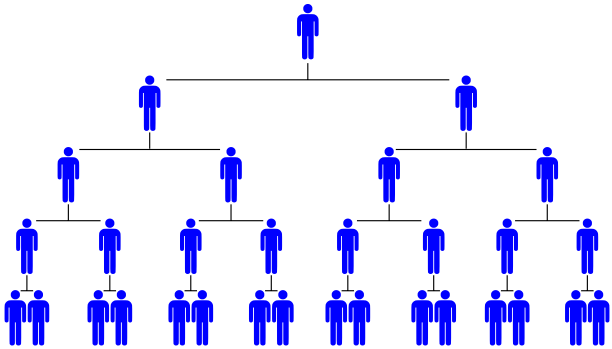

Fig. 1 A branching hierarchy. This figure shows an idealized branching hierarchy in which each superior has two subordinates. This superior/subordinate ratio—often called the span of control—can be used to mathematically describe the hierarchy. Starting from the bottom rank, each consecutive rank *decreases* in size by a factor of the span of control. Unlike employment, we expect income to *increase* with hierarchical rank

of the US private sector. This model generalizes the sparse firm hierarchy empirical data to allow the first quantitative study of the distributional effects of hierarchy.

A firm hierarchy model

The firm hierarchy model (herein the 'hierarchy model') is based on the hypothesis that human institutions are hierarchically organized, and that hierarchical rank plays a key role in determining income. The starting point for my approach is the seminal work of Simon [52] and Lydall [13]. In the late 1950s, Simon and Lydall both developed simple models that focused on the branching structure of firm hierarchies. The distinguishing feature of a branching hierarchy is that each superior has control over *multiple* subordinates (see Fig. 1).

Simon and Lydall both showed how branching hierarchical structure could explain regularities in income distribution. Simon used a simple hierarchical model of the firm to explain the observed scaling between CEO pay and firm sales [53]. Lydall showed how firm hierarchy could lead to a power-law distribution of top income (as discussed above). This paper draws on the work of Simon and Lydall, but updates their model in light of recent empirical work.

Both Simon and Lydall assumed a constant span of control within firms. (The span of control is the number of subordinates per superior). However, case-study evidence indicates that the span of control is *not* constant within firms, but instead tends to increase with hierarchical rank (see Online Appendix B). Simon and Lydall also assumed that the average income ratio between adjacent hierarchical ranks was constant. Again, case-study evidence suggests that this is not quite true. Like the span of control, the pay ratio between ranks also tends to increase with rank.

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

Modeling goals and methods

Unlike the power-law generation models discussed in "Power-law generation mechanisms", my hierarchy model is not *designed* to produce a power law. Rather, it is designed to match the available firm-level evidence, with the intention of generalizing this evidence to investigate the distributional effects of firm hierarchy. The *hope* is that the resulting model will produce a power law that matches macro-level data, but there is no guarantee that it will.

In principle, we could directly investigate the income effects of firm hierarchy using empirical data (with no need for a model). However, the available firm-level evidence is too sparse to draw conclusions about the wider distributional role of firm hierarchy. The purpose of the hierarchy model is to investigate what is *implied* by the available firm-level data. The model takes the limited firm-hierarchy evidence that does exist and fits trends and parameterized distributions to it. I then use the model algorithm (outlined in detail in Online Appendices D and E) to extrapolate these trends to create a large-scale simulation of the economy. The resulting model is *entirely* dependent on the input, firm-level data. I do not tune the model to reproduce macro-level results. Therefore, the model output is purely what is implied by generalizing the trends found in input data.

The hierarchy model is built on a tripartite income-dispersion classification scheme that allows for three sources of income dispersion (see Fig. 2):

- **Source 1**: Income dispersion *between* hierarchical levels of each firm (*inter*-hierarchical dispersion);
- **Source 2**: Income dispersion *within* hierarchical levels of each firm (*intra*-hierarchical dispersion);
- **Source 3**: Income dispersion *between* different firms (*inter*-firm dispersion).

Inter-firm and intra-hierarchical level dispersion are not explained by the model. (In the jargon of economic modeling, these dispersion sources are *exogenous*). In contrast, inter-hierarchical dispersion is *partially* explained by the model. It is explained in the sense that it is not *ex nihilo*—this dispersion does not come from nowhere. The model contains firms that have a specific hierarchical structure of employment and pay. However, the reason for this hierarchical structure is *not* explained by the model. Rather, hierarchical structure is determined from regressions on case-study data, in conjunction with firm-level data from the Compustat and Execucomp databases.

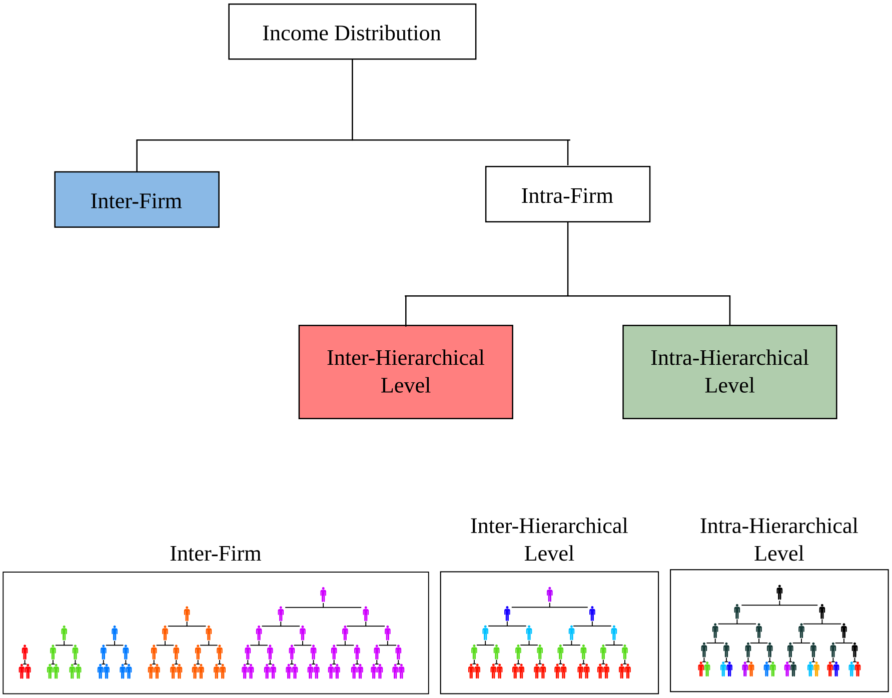

Fig. 2 A tripartite division of income distribution.This figure illustrates the income distribution grouping scheme used by the hierarchy model. The model allows for three sources of income dispersion. *Inter-firm* dispersion consists of differences in (average) pay between firms. Within each firm, there are two further sources of dispersion. *Inter-hierarchical* level dispersion consists of differences in (average) pay between hierarchical levels, while *intra-hierarchical* level dispersion consists of differences in pay within each hierarchical level

Modeling the USA

The model is designed to study the hierarchical structure of the US private sector as it was (on average) over the years 1992–2015. At the highest level of abstraction, the model has three parts. First, the model creates a firm-size distribution that dictates how many firms of a given size will exist. Second, for each firm in this distribution, the model creates a hierarchical structure. This means the model determines how many ranks will exist, and how many individuals will occupy each hierarchical rank. Lastly, the model uses each of the three dispersion sources (outlined above) to stochastically generate an income for every individual in every firm. I review here the most important elements of each step. A technical discussion can be found in the Online Appendices.

Step 1: Create a firm-size distribution The first step of the model is to generate a distribution of firm sizes. The available evidence suggests that national firm-size distributions can be modeled by a power law [54–56]. Under this assumption, the probability of finding a firm of size x is proportional to $x^{-\alpha}$, where α is a constant. I model the US firm-size distribution with 1 million firms distributed according to a discrete power-law distribution with exponent $\alpha = 2.01$ (see Online Appendix E). This may seem like the model uses one power law (the firm size distribution) to create another (the distribution of income). However, this is not the case. Without hierarchy, the model will not create a power-law distribution (see Fig. 6).

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

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

Step 3: Endow individuals with income After each firm has a hierarchical structure, the model assigns every individual an income. Because the model has three dispersion mechanisms, this step has three components, outlined below.

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

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

Step 3B: Generate inter-firm dispersion I create inter-firm income dispersion by varying (average) pay in the bottom hierarchical level of each firm. This variation

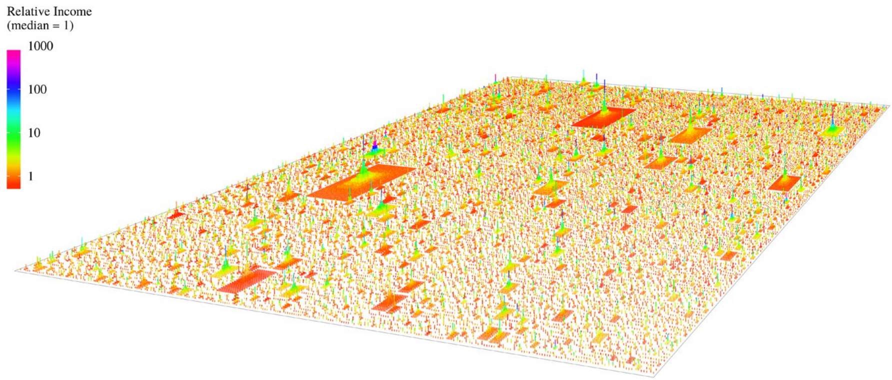

Fig. 3 A Landscape view of the hierarchy model. This figure visualizes the US hierarchy model as a landscape of three-dimensional firms. Each pyramid represents a single firm, with size indicating the number of employees and height corresponding to the number of hierarchical levels. If you look closely, you will see vertical lines corresponding to individuals. Income (relative to the median) is indicated by color. This visualization has 20,000 firms—a small sample of the actual model, which uses 1 million firms

is informed by firm-level data in the Compustat database. As discussed in Step 3A, prior to running a full-scale simulation, I fit the model to firms in the Compustat database. After having fit hierarchical pay, I use this information to estimate how base-level pay varies across these firms. This variation then informs the model's main simulation. For more details, see Online Appendices D and E.

Step 3C: Generate intra-hierarchical level dispersion The last step is to model the income dispersion *within* the hierarchical levels of each firm. The available case-study evidence suggests that income dispersion within hierarchical levels is roughly constant across all hierarchical levels (see Online Appendix B). To simplify the model, I further assume that intra-hierarchical level dispersion is constant across all firms. Informed by case-study data, I use a single parameterized distribution to randomly generate income dispersion within all hierarchical levels of every firm. For more details, see Online Appendices D and E.

Visualizing the US hierarchy model

To give an intuitive understanding of what the hierarchy model 'looks' like, Fig. 3 shows a landscape view of the model's structure. Each pyramid represents a different hierarchically organized firm. The size of each pyramid corresponds to the number of employees, height represents hierarchical level, and color represents relative income.

This figure highlights the main characteristics of the model. The firm power-law distribution is clearly visible. The vast majority of firms are small, but there are a few behemoths. Inter-firm income dispersion and inter-hierarchical level income dispersion are also visible, while intra-hierarchical level income dispersion appears negligible. Lastly, top incomes are concentrated in upper hierarchical levels, and consequently occur mostly in larger firms. These facts, which are qualitatively visible here, become more clear as we analyze the model results in quantitative terms.

Testing the hierarchy model: macro predictions

The purpose of the hierarchy model is to study the hierarchical structure of the US private sector. The first step, then, is to make sure that the model produces realistic results. To that end, Fig. 4 compares the model's aggregate income distribution to US empirical data. Although the model aims only to capture the private sector (not government), I compare the model's results to macro-level data for the entire USA. I do this because the most reliable income distribution data (from the IRS) does not differentiate between the private and public sector.

Even though the model is an extrapolation from a limited set of data, it does a reasonably accurate job of reproducing the US distribution of income. Note, though, that the model underestimates US income inequality, both in terms of the Gini index (Fig. 4a) and the income share of the top 1% (Fig. 4b). What is the source of this discrepancy? Looking at the income probability density in Fig. 4d, it appears that the US income distribution is more 'bottom heavy' than the model. The model produces too few extremely small incomes, relative to the US. This tendency is also evident in the cumulative distribution (Fig. 4f).

Why does this discrepancy occur? I demonstrate in Online Appendix F that the discrepancy can be removed by increasing the model's inter-firm income dispersion. This suggests that the model's underestimate of US inequality is due to an underestimate of inter-firm income dispersion. My guess is that this occurs because the model is based on Compustat firm data, which is not a representative sample of the US firm population. Compustat contains data for public firms only, and as a result is biased towards large firms. I suspect that a more representative firm sample would give greater inter-firm income dispersion. I include adjusted results in the Online Appendix to show that the model is *capable* of closely reproducing the important features of US income distribution. However, I do *not* use this adjusted data for any of the proceeding analysis. The purpose of the hierarchy model is to extrapolate empirical data, warts and all.

While the model slightly misrepresents the 'body' of US income distribution, it accurately reproduces the *tail*. This is evident in the complementary cumulative distribution (Fig. 4f) in the form of virtually identical model and empirical slopes in the right tail. This is important because it is the tail of the distribution (particularly, its power-law properties) that we are interested in studying. When plotted on a log–log scale (as in Fig. 4f), a power-law tail is visually evident as a straight line in the complementary cumulative distribution.

Dating back to the work of Pareto [1], it has been common to estimate the power-law exponent by means of a linear regression on the complementary cumulative distribution. However, Clauset et al. show that this approach is inaccurate [57]. Instead, I use the more accurate maximum-likelihood method (see Online Appendix A). Estimating the power-law exponent requires making a choice about where the 'tail' of the distribution begins. I define the tail as the top 1% of

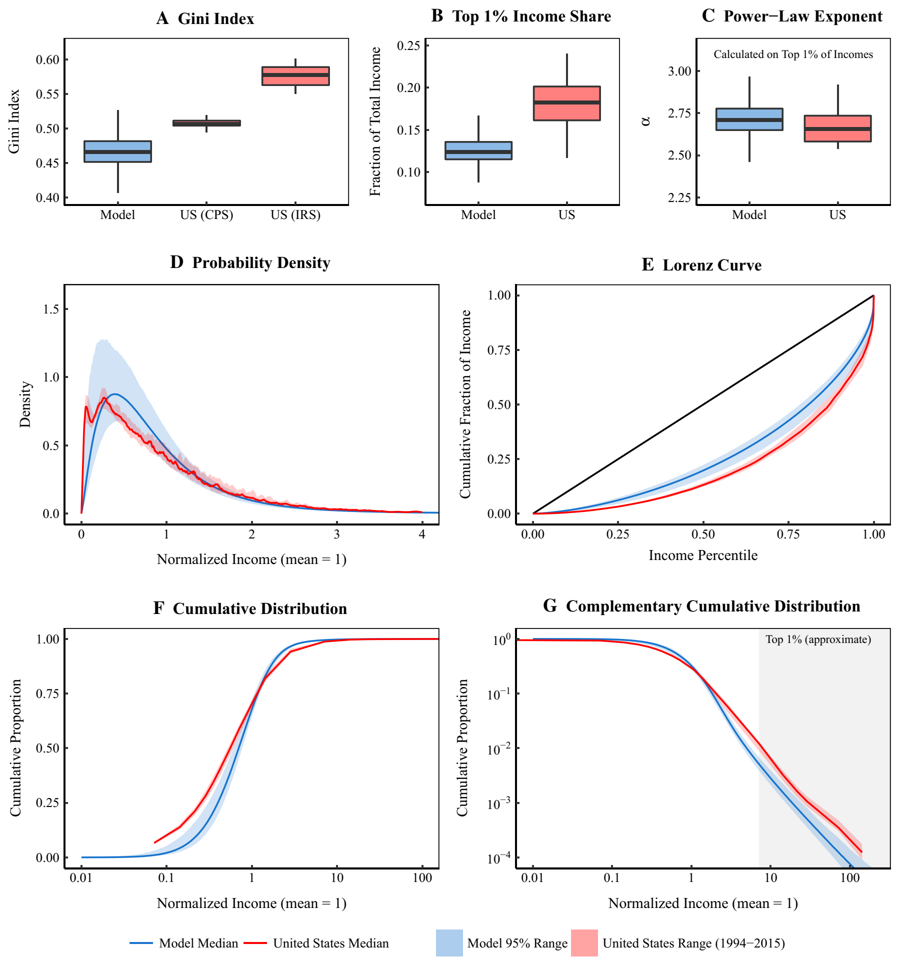

Fig. 4 Modeled income distribution vs. US data.This figure compares various aspects of the model's income distribution to US data over the years 1992–2015. **a** The Gini index, with two different US sources—the Current Population Survey (CPS) and the Internal Revenue Service (IRS). **b** The top 1% income share, using data from 17 different time series. **c** The results of fitting a power-law distribution to the top 1% of incomes (where α is the scaling exponent). **d** plots the income density curve with mean income normalized to 1 (using data from the CPS). **e**–**g** Use IRS data to construct the Lorenz curve, cumulative distribution, and complementary cumulative distribution (respectively). The cumulative distribution shows the proportion of individuals with income *less* than the given *x* value. The complementary cumulative distribution shows the proportion of individuals with income *greater* than the given *x* value. The shaded region shows the approximate threshold for the top 1% of incomes. For sources and methods, see Online Appendix A

incomes, a threshold that has been popularized by Piketty [58]. Figure 4c shows the results of fitting a power law to the top 1% of incomes (for methods, see Online Appendix A). Over many runs, the model produces a distribution tail with fitted power-law exponents that are very close to the exponents fitted to historical US data.

To summarize, the hierarchy model produces an income distribution that is roughly consistent with the US distribution of income. In particular, the model closely reproduces the tail of the US distribution, including its power-law properties.

Testing the hierarchy model: micro predictions

When discussing the model visualization shown in Fig. 3, I noted that top-earning individuals are clustered at the tops of *large* firms. This is a defining feature of the hierarchy model. It occurs because income scales strongly with hierarchical rank. As a result, top earners are found at the tops of large firms, because these firms have the most hierarchical levels. To my knowledge, this prediction is not made by any other model of income distribution. It is important, therefore, that we put it to the test.

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

I conduct this test using data from the Forbes 400 and Execucomp. The Forbes 400 list is a definitive ranking of the 400 richest Americans, and it provides the institutional source of each individual's wealth. The caveat is that this list is a ranking by *wealth*, not income. I use the Forbes 400 as a proxy for top US incomes, under the assumption that wealth and income are strongly related. I supplement the Forbes 400 data with the 'Execucomp 500', which is composed of the 500 top-paid US executives in the Execucomp database (in each year between 1992 and 2015). The advantage of the Execucomp 500 is that it is a ranking explicitly by income. The disadvantage is that we do not know if these 500 executives are actually the top-paid US individuals.

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

Results for the Fortune 400 and Execucomp 500 firm-size distributions are shown in the main panel of Fig. 5. These density plots represent the size distribution of firms associated with the richest 400 Americans and the 500 top-paid executives in the Execucomp database (respectively). To better visualize the distribution, I plot the density of the *logarithm* of firm size. Under this transformation, the null-effect result will appear as a *uniform* distribution. From the evidence shown in Fig. 5, we can immediately conclude that the null effect is false. There is definitely a relation between top incomes/wealth and firm size. But is it the relation that is predicted by the hierarchy model?

To find out, I conduct the same analysis on the model. I select the model's 500 top-paid individuals and record the size distribution of associated firms. The results are shown in Fig. 5 as the 'Model 500'. The model predicts a relation between top incomes and firm size that is very similar to the US empirical data. To be sure, the

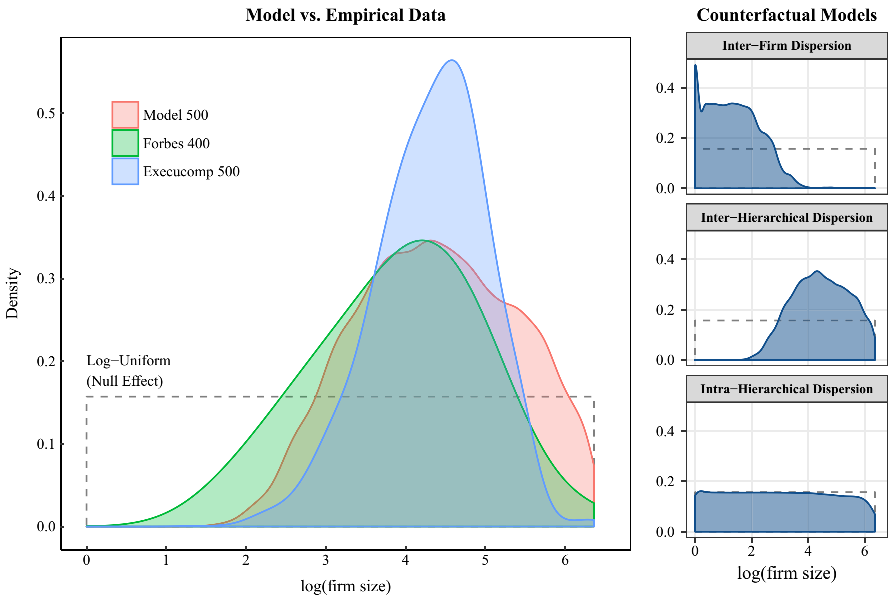

Fig. 5 Firm-size distributions associated with top incomes and wealth. This figure shows the size distribution of firms associated with top-earning individuals in the USA and in the hierarchy model (of the USA). The ‘Forbes 400’ represents the size distribution of firms associated with (owned by) the wealthiest 400 Americans in the year 2014. The ‘Execucomp 500’ represents the size distribution of firms associated with the 500 top-earning American executives (in each year from 1992 to 2015) in the Execucomp database. The ‘Model 500’ represents the size distribution of firms associated with the 500 top-earning individuals in the hierarchy model. Results for counterfactual models are shown on the right. Each counterfactual model isolates a single source of income dispersion. The top panel shows a model with inter-firm dispersion only, the middle shows a model with inter-hierarchical dispersion only, and the bottom shows a model with intra-hierarchical level dispersion only. In all plots, I also show the log-uniform distribution (dotted line), which is the null-effect prediction (i.e., no relation between firm membership and income). For sources and methods, see Online Appendix A

model results are not identical to either the Forbes 400 or the Execucomp 500 distributions. But, given the paucity of data on which the model is based (as well as the general uncertainty in the empirical analysis of top incomes), I count this result as a success. The model produces results that are roughly consistent with the US data.

Since the hierarchy model has three sources of income dispersion, we naturally want to know which of these sources is responsible for producing the results in Fig. 5. To answer this question, I use a counterfactual analysis. I create three different counterfactual models to supplement the original. Each counterfactual model isolates a single source of dispersion as it appears in the original model. The results of this counterfactual analysis are shown in the right-hand panels in Fig. 5. This analysis indicates that it is *exclusively* inter-hierarchical income dispersion that is responsible for associating top incomes with large institutions. The inter-hierarchical dispersion model produces results that are virtually identical to the original model. At the same time, inter-firm dispersion only and intra-hierarchical level dispersion only models produce *drastically* different results. (Note that with intra-hierarchical dispersion only, we recover the null effect. Why? In this model, firms play no part in determining income).

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

Isolating the distributional role of firm hierarchy

Having established that the hierarchy model gives credible results, I now use it to isolate the distributional effects of firm hierarchy. In particular, I am interested in determining whether or not it is hierarchy that shapes the income distribution tail. As in Fig. 5, I isolate the effects of hierarchy using a counterfactual analysis. I create three different counterfactual models of the USA, each containing only one source of income dispersion. By comparing these counterfactual models to the original model, we can determine how each dispersion source affects income distribution.

Figure 6 shows the results of this analysis. Here, I plot the income distribution (the probability density) of the original and counterfactual models. I use a log–log transformation to more clearly illustrate the distribution tail. This visualization allows us to see how each factor contributes to the original model's distribution of income. To interpret this plot, look at the vertical distance between the original and counterfactual models. The closer a particular counterfactual model comes to the original model, the more important that dispersion factor is for shaping income distribution at the point in question.

The results of this analysis are unambiguous. A clear division exists between the *body* and *tail* of the distribution. The body of the distribution is almost completely determined by *inter-firm* dispersion, while the tail of the distribution is almost completely determined by *inter-hierarchical* dispersion. Intra-hierarchical dispersion amounts to negligible noise. The inset panel in Fig. 6 shows the fitted power-law exponent for the top 1% of incomes in the original and inter-hierarchical dispersion model. This confirms what is visually obvious in the main plot—the tail of the inter-hierarchical dispersion model is virtually identical to that of the original.

To summarize, the counterfactual analysis indicates that it is inter-hierarchical pay-scaling (alone) that is responsible for generating the model's income distribution tail. This suggests that it is hierarchy that is responsible for generating the power-law tail, and that the effects of hierarchy become important in the top 1% of incomes.

How hierarchy generates the power-law tail

How does hierarchy create the (approximate) power-law distribution of top incomes? The basic mechanism was theorized by Lydall [13]. It relies on the following contrapuntal exponential tendencies of hierarchical organization:

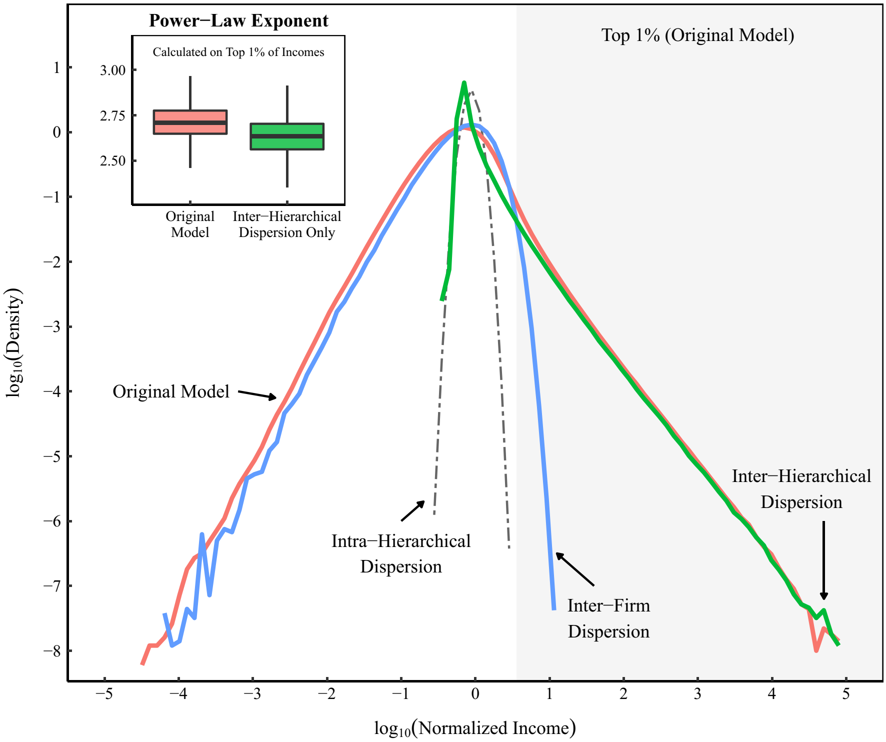

Fig. 6 Isolating the effects of hierarchy with counterfactual models. This figure compares the original hierarchy model of the USA to three different counterfactual models. Each counterfactual model contains only one of the three sources of income dispersion. The main plot shows the income probability density of each model, plotted using a log–log transformation (these results show the average distribution over many iterations). To interpret this plot, look at the vertical distance between each counterfactual model's distribution and that of the original. The smaller the distance, the greater is the distributional role played by that dispersion factor at the point in question. The shaded region indicates the top 1% of incomes (in the original model). The inset panel shows power-law exponents fitted to the top 1% of incomes in the original and inter-hierarchical dispersion model

1. Hierarchical organization causes the share of employment to *decrease* exponentially with hierarchical rank.
2. Hierarchical pay structure causes income to *increase* exponentially with rank.

These two opposing tendencies interact to produce a power-law distribution of income (in the tail). This mechanism is a specific case of a more general method. A power law will be created any time we *exponentially* transform an *exponential* distribution [32].

The proof works as follows. Suppose we have some quantity y that is exponentially distributed (here a is a negative constant):

$$p(y) \sim e^{ay}. \tag{2}$$

In the case of hierarchical class structure, this would be the probability of finding someone with a hierarchical rank y. Suppose that we have another variable, x, that is *also* exponentially related to y:

$$x = e^{by}. \tag{3}$$

In the context of hierarchical organization, x would be income, which *increases* exponentially with rank. We want to know how income (x) is distributed. To find out, we use the change of variable formula to get f_x, the density function of x:

$$f_x = f_y\big(\, y(x)\,\big) \cdot \big|\, y'(x)\,\big|. \tag{4}$$

We let $f_y = e^{ay}$. Since $x = e^{by}$, we note that $y(x) = \frac{1}{b}\ln x$ and $y'(x) = 1/bx$. Substituting into the change of variable formula gives:

$$f_x \;=\; e^{\frac{a}{b}\ln x} \cdot \frac{1}{bx} \;=\; \frac{1}{b}x^{a/b-1}. \tag{5}$$

Thus, the variate x (income) has a power-law distribution with exponent $\alpha = a/b - 1$.

To reiterate, hierarchical organization creates a power-law distribution because of two contrapuntal, exponential tendencies: (1) employment tends to *decrease* exponentially with rank; and (2) income tends to *increase* exponentially with rank. Figure 7 illustrates this contrapuntal behavior in the hierarchy model. Figure 7a shows the aggregate hierarchical employment structure of the model. As expected, the hierarchical employment distribution has a bottom-heavy pyramid shape. The vast majority of people occupy low ranks and only a tiny elite have high rank. The inset panel highlights the *exponential* nature of this distribution. Figure 7b shows the model's aggregate hierarchical *pay* structure. As expected, hierarchical pay has an inverted pyramid shape. The average income at the top of the hierarchy dwarfs that at the bottom. Again, the inset plot highlights the exponential nature of this relation.

Note that neither employment nor pay has a *purely* exponential relation with rank. This is a design feature of the model, stemming from case-study evidence. In the case-study data, income tends to *increase* supra-exponentially (faster than an exponential) with rank. Conversely, employment tends to *decrease* supra-exponentially with rank (see Online Appendix B for details). In any case, when we combine these two supra-exponential tendencies, the result still seems to be (roughly) a power-law distribution of income in the model's tail.

While the above derivation highlights the basic power-law generation mechanism, the hierarchy model's inner workings involve some added complexity. First, the above derivation assumes that rank (y) is a continuous variable. In the model, rank is a *discrete* variable, which would result in a discontinuous distribution of pay (x) in Eq. 5. Lydall noted this in his original derivation, and posited that a process of 'blurring' would occur (due to stochastic differences in pay between firms) that would make the resulting distribution continuous [13]. In this regard, Lydall's intuition appears to be correct.

Figure 8 shows how the various discrete hierarchical ranks contribute to produce the continuous power-law tail. Each panel shows the distribution of income

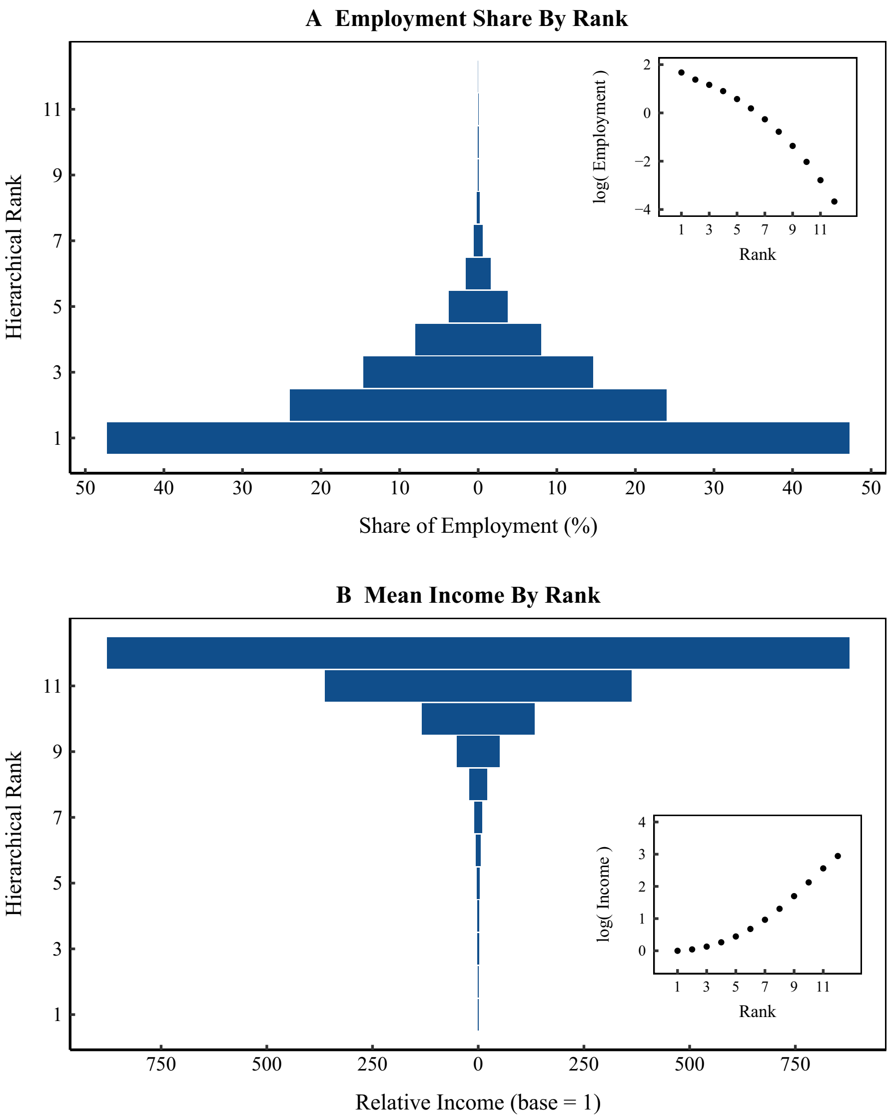

Fig. 7 The hierarchy model's contrapuntal exponential tendencies. This figure shows the two contrapuntal exponential tendencies associated with the hierarchy model's class structure. **a** The model's aggregate distribution of employment by hierarchical rank. The bottom-heavy shape results from firm hierarchical structure (in conjunction with the firm-size distribution). The inset graph shows the logarithm of employment share, plotted against rank. A pure exponential function would appear as a straight line. The curve in this relation indicates that employment declines with rank slightly faster than an exponential function. **b** The model's mean pay by hierarchical rank (normalized so that the base level =1). The inset graph shows the logarithm of income plotted against rank. The curve in this relation indicates that income increases with rank slightly faster than an exponential function

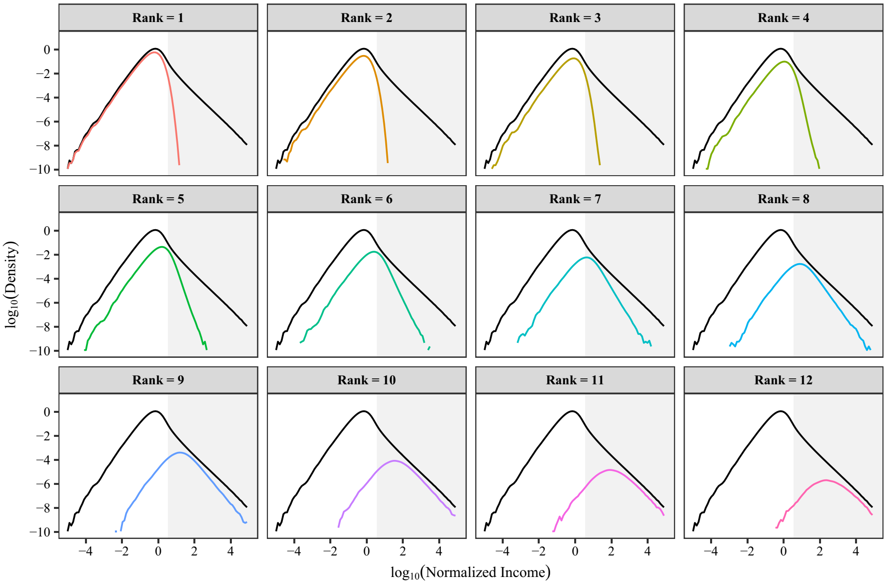

Fig. 8 The model's distribution of income by hierarchal rank. This figure shows the distribution of income for each hierarchical rank in the hierarchy model. In each panel, a rank-specific income distribution (color) is compared to the model's aggregate income distribution (black). The rank-specific distributions are normalized so that cumulative density of all ranks sums to one. The shaded region indicates the top 1% of incomes (in the aggregate model distribution). To interpret this plot, look at how closely each rank-specific distribution comes to the aggregate distribution. The closer the two are, the greater is the rank's contribution to income distribution at that point. The power-law right tail (evident as the straight line in the aggregate distribution) is jointly created by ranks five and up

of a *specific* hierarchical rank in relation to the model's aggregate income distribution. (The rank-specific distributions are normalized so that the cumulative density of all ranks sums to one.) In this plot, the exponential growth of income with rank appears as a horizontal shift in the income distribution of each rank. At the same time, each successive rank has exponentially fewer members, which appears as a downward shift in the income distribution. When the contributions of all ranks are summed, the result is an approximate power-law distribution of top incomes. As Lydall suspected, a complex blurring process occurs (between ranks) that smooths out what would otherwise be a discontinuous distribution.

Discussion

Whenever two or more theories describe the same phenomenon, we need to determine if they are consistent with one another, or if they are mutually inconsistent. Thus, we should ask—is the hierarchy model's explanation of the power-law distribution of top incomes at odds with the stochastic growth models described in "Power-law generation mechanisms"? Or are the two approaches mutually consistent?

The primary difference between the two approaches is that the hierarchy model is *static*, while the stochastic models are *dynamic*. The hierarchy model begins with the observation that firms have a hierarchical structure, and that this (static) structure could explain the distribution of income at a point in time. The hierarchy model says nothing about the dynamics of individual income, but instead focuses on institutional structure. In contrast, the stochastic approach begins with the observation that individual incomes *change* over time. Since income distribution represents a snapshot of these changing incomes, it must be possible to explain income distribution in terms of the dynamics of individual income. The static and dynamic approaches explain the power-law distribution of top incomes from very different angles. Therefore, I see no fundamental clash between the hierarchy model and exogenous stochastic growth models in the tradition of Champernowne [34]. More research is needed to determine how the two approaches are related.

That being said, the stochastic growth and firm hierarchy models each have very different implications for how we should study (and potentially alleviate) inequality. Stochastic growth models put the focus on isolated individuals. This makes it difficult to connect inequality to the wider political and socioeconomic setting (the search for such a connection is a major goal of many economists and sociologists [59–71]). In contrast, the hierarchy model suggests that the income distribution power-law tail is an outcome of the internal compensation policies of firms. This puts the focus squarely on firms and how they remunerate their employees as a function of hierarchical rank. This perspective opens the door to future research that connects the internal pay policies of firms to the wider distribution of income (and potentially to government policy).

Conclusions

In 1959, when Lydall [13] first proposed that firm hierarchy could create a power-law distribution of income, his hypothesis was largely speculative. At the time, little was known about the internal pay structure of firms. Nearly 60 years later, data on firm hierarchical structure is still scarce, but enough evidence exists that we can begin to investigate the distributional effects of firm hierarchy. This paper has presented a first attempt at doing so.

The key finding is that the empirically informed hierarchy model is capable of reproducing the power-law scaling of top US incomes, while at the same time accurately connecting top-earning individuals to large firms. Importantly, the model indicates that it is hierarchical pay-scaling alone that is responsible for these results. Of course, the hierarchy model's results are contingent on the input data, which is limited. While I have made every effort to incorporate uncertainty in the underlying case-study data, results may change when new data comes along. Further research is needed to verify these results and see if they can be replicated in other countries.

Uncertainty aside, the hierarchy model suggests that the ubiquitous power-law scaling of top incomes may be a result of the ubiquitous use of hierarchical organization in human societies. This implies that when we study the tail of the distribution of income, we may be studying the effects of social hierarchy.

Acknowledgements I would like to thank the Social Sciences and Humanities Resource Council of Canada (Grant no. 767-2015-1015) for its support. I thank also Jonathan Nitzan, who has offered feedback on aspects of this paper.

References

1. Pareto, V. (1897). *Cours d'economie politique* (Vol. 1). Geneva: Librairie Droz.
2. Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*, *84*(334), 129.
3. Aoyama, H., Souma, W., Nagahara, Y., Okazaki, M. P., Takayasu, H., & Takayasu, M. (2000). Pareto's law for income of individuals and debt of bankrupt companies. *Fractals*, *8*(03), 293.
4. Coelho, R., Richmond, P., Barry, J., & Hutzler, S. (2008). Double power laws in income and wealth distributions. *Physica A: Statistical Mechanics and its Applications*, *387*(15), 3847.
5. Clementi, F., & Gallegati, M. (2005). Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In A. Chatterjee, S. Yarlagadda, & B. K. Chakrabarti (Eds.), *Econophysics of wealth distributions* (pp. 3–14). Berlin: Springer.
6. Clementi, F., & Gallegati, M. (2005). Power law tails in the Italian personal income distribution. *Physica A: Statistical Mechanics and Its Applications*, *350*(2–4), 427.
7. Di Guilmi, C., Gaffeo, E., & Gallegati, M. (2003). Power law scaling in world income distribution. *Economics Bulletin*, *15*, 1–7.
8. Drgulescu, A., & Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A: Statistical Mechanics and its Applications*, *299*(1–2), 213.
9. Nirei, M. (2009). Pareto distributions in economic growth models. IIR working paper WP#09-05
10. Toda, A. A. (2012). The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior & Organization*, *84*(1), 364.
11. Silva, A. C., & Yakovenko, V. M. (2004). Temporal evolution of the thermal and superthermal income classes in the USA during 19832001. *EPL (Europhysics Letters)*, *69*(2), 304.
12. Souma, W. (2001). Universal structure of the personal income distribution. *Fractals*, *9*(04), 463.
13. Lydall, H. F. (1959). The distribution of employment incomes. *Econometrica: Journal of the Econometric Society*, *27*(1), 110.
14. Audas, R., Barmby, T., & Treble, J. (2004). Luck, effort, and reward in an organizational hierarchy. *Journal of Labor Economics*, *22*(2), 379.
15. Baker, G., Gibbs, M., & Holmstrom, B. (1993). Hierarchies and compensation: A case study. *European Economic Review*, *37*(2–3), 366.
16. Dohmen, T. J., Kriechel, B., & Pfann, G. A. (2004). Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. *Journal of Population Economics*, *17*(2), 193.
17. Grund, C. (2005). 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.
18. Lima, F. (2000). Internal labor markets: A case study. FEUNL working paper, vol. 378.
19. Morais, F., & Kakabadse, N. K. (2014). 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.
20. Treble, J., Van Gameren, E., Bridges, S., & Barmby, T. (2001). The internal economics of the firm: Further evidence from personnel data. *Labour Economics*, *8*(5), 531.
21. Ariga, K., Brunello, G., Ohkusa, Y., & Nishiyama, Y. (1992). Corporate hierarchy, promotion, and firm growth: Japanese internal labor market in transition. *Journal of the Japanese and International Economies*, *6*(4), 440.
22. Bell, B., & Van Reenen, J. (2012). Firm performance and wages: Evidence from across the corporate hierarchy. CEP discussion paper, vol. 1088.

23. Eriksson, T. (1999). Executive compensation and tournament theory: Empirical tests on Danish data. *Journal of Labor Economics*, *17*(2), 262.
24. Heyman, F. (2005). Pay inequality and firm performance: Evidence from matched employer–employee data. *Applied Economics*, *37*(11), 1313.
25. Leonard, J. S. (1990). Executive pay and firm performance. *Industrial and Labor Relations Review*, *43*(3), 13.
26. Main, B. G., O'Reilly, C. A, I. I. I., & Wade, J. (1993). Top executive pay: Tournament or teamwork? *Journal of Labor Economics*, *11*(4), 606.
27. Mueller, H. M., Ouimet, P. P., & Simintzi, E. (2016). Within-firm pay inequality. SSRN working paper.
28. Rajan, R. G., & Wulf, J. (2006). The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. *The Review of Economics and Statistics*, *88*(4), 759.
29. Tao, H. L., & Chen, I. T. (2009). The level of technology employed and the internal hierarchical wage structure. *Applied Economics Letters*, *16*(7), 739.
30. Kumamoto, S. I., & Kamihigashi, T. (2018). Power laws in stochastic processes for social phenomena: An introductory review. *Frontiers in Physics*, *6*, 20.
31. Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, *1*(2), 226.
32. Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323.
33. Gibrat, R. (1931). *Les inegalites economiques*. Paris: Recueil Sirey.
34. Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, *63*(250), 318.
35. Gabaix, X., Lasry, J. M., Lions, P. L., & Moll, B. (2016). The dynamics of inequality. *Econometrica*, *84*(6), 2071.
36. Nirei, M., & Souma, W. (2007). A two factor model of income distribution dynamics. *Review of Income and Wealth*, *53*(3), 440.
37. Rutherford, R. (1955). Income distributions: A new model. *Econometrica: Journal of the Econometric Society*, *23*(3), 277.
38. Wold, H. O., & Whittle, P. (1957). A model explaining the Pareto distribution of wealth. *Econometrica, Journal of the Econometric Society*, *25*(4), 591.
39. Yule, G. U, I. I. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. *Philosophical Transactions of the Royal Society B*, *213*(402–410), 21.
40. Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425.
41. Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology*, *27*(5), 292.
42. Barabasi, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509.
43. Angle, J. (1986). The surplus theory of social stratification and the size distribution of personal wealth. *Social Forces*, *65*(2), 293.
44. Angle, J. (2006). The inequality process as a wealth maximizing process. *Physica A: Statistical Mechanics and Its Applications*, *367*(15), 388.
45. Chatterjee, A., Chakrabarti, B. K., & Chakraborti, A. (2007). *Econophysics and sociophysics: Trends and perspectives*. Milan: Wiley.
46. Chatterjee, A., & Chakrabarti, B. K. (2007). Kinetic exchange models for income and wealth distributions. *The European Physical Journal B*, *60*(2), 135.
47. Hegyi, G., Neda, Z., & Santos, M. A. (2007). Wealth distribution and Pareto's law in the Hungarian medieval society. *Physica A: Statistical Mechanics and its Applications*, *380*(1), 271.
48. Kitov, I. O. (2009). Mechanical model of personal income distribution (pp. 1–220). arXiv preprint arXiv:0903.0203
49. Pianegonda, S., Iglesias, J. R., Abramson, G., & Vega, J. L. (2003). Wealth redistribution with conservative exchanges. *Physica A: Statistical Mechanics and its Applications*, *322*, 667.
50. Scheffer, M., Bavel, B. V., Leemput, I. A. V. D., Nes, E. H. V. (2017) Inequality in nature and society. *Proceedings of the National Academy of Sciences*, *114*, 201706412. https://doi.org/10.1073/pnas.1706412114
51. Yakovenko, V. M., & Rosser, J. B, Jr. (2009). Colloquium: Statistical mechanics of money, wealth, and income. *Reviews of Modern Physics*, *81*(4), 1703.
52. Simon, H. A. (1957). The compensation of executives. *Sociometry*, *20*(1), 32.

53. Roberts, D. R. (1956). A general theory of executive compensation based on statistically tested propositions. *The Quarterly Journal of Economics*, *70*(2), 270.
54. Axtell, R. L. (2001). Zipf distribution of US firm sizes. *Science*, *293*, 1818.
55. Fix, B. (2017). Energy and institution size. *PLoS ONE*, *12*(2), e0171823. https://doi.org/10.1371/journal.pone.0171823.
56. Gaffeo, E., Gallegati, M., & Palestrini, A. (2003). On the size distribution of firms: Additional evidence from the G7 countries. *Physica A: Statistical Mechanics and its Applications*, *324*(12), 117. https://doi.org/10.1016/S0378-4371(02)01890-3.
57. Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661.
58. Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge: Harvard University Press.
59. Brown, C. (1988). In Mangum, G. & P. Philips (Eds), *Three worlds of labor economics* (Vol. 51 pp. 515–530)
60. Brown, C. (2005). Is there an institutional theory of distribution? *Journal of Economic Issues*, *39*(4), 915.
61. Dahrendorf, R. (1959). *Class and class conflict in industrial society*. Stanford: Stanford University Press.
62. Huber, E., Huo, J., & Stephens, J. D. (2017). Power, policy, and top income shares. *Socio-Economic Review*. https://doi.org/10.1093/ser/mwx027.
63. Kalecki, M. (1971). *Selected essays on the dynamics of the capitalist economy 1933–1970*. Cambridge: Cambridge University Press.
64. Lenski, G. E. (1966). *Power and privilege: A theory of social stratification*. Chapel Hill: UNC Press Books.
65. Marx, K. (1867). *Capital* (Vol. I). Penguin/New Left Review: Harmondsworth.
66. Mills, C. W. (1956). *The power elite*. Oxford: Oxford University Press.
67. Nitzan, J., & Bichler, S. (2009). *Capital as power: A study of order and creorder*. New York: Routledge.
68. Peach, J. T. (1987). Distribution and economic progress. *Journal of Economic Issues*, *21*(4), 1495.
69. Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge: Cambridge University Press.
70. Weber, M. (1978). *Economy and society: An outline of interpretive sociology*. Berkeley: University of California Press.
71. Wright, E. O. (1979). *Class structure and income determination* (Vol. 2). New York: Academic Press.