Appendices for 'Hierarchy and the Power-Law Income Distribution Tail'

Blair Fix

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

https://osf.io/mb3ah/

The supplementary materials include:

1. Raw source data;
2. R code for all analysis;
3. Hierarchy model code.

Contents

A Sources and Methods

Sources are listed by the figure in which they appear.

Sources for Figure 4 (Modeled Income Distribution vs. US Data)

Complementary Cumulative Distribution

The US complementary cumulative distribution is calculated from data in the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015.

Cumulative Distribution

The US cumulative distribution is calculated from data in the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015.

Gini Index

I use two sources for the US Gini index. The first source is the US Current Population Survey, Table PINC-08 (available from the US Census) over the years 1994 to 2015. The second source is the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015. I estimate the Gini index by constructing a Lorenz curve from the reported cumulative frequency data. R code implementing this method is available in the Supplementary Material.

The Census and IRS data are *not* mutually consistent. IRS data is based on tax units, not individuals. The advantage of the IRS data is that it is an administrative record. Current Population Survey (CPS) data, on the other hand, is obtained by *interview*. The advantage of the CPS data is that it explicitly counts individuals. The disadvantage is that "there is a tendency in household surveys for respondents to *under report* their income" [1].

Lorenz Curve

The US Lorenz curve is calculated from data in the IRS *Individual Complete Report* (Publication 1304), Table 1.1, from 1996 to 2015.

Power Law Exponents

I estimate the power law exponent of the income distribution tail using the maximum likelihood method. US empirical data comes from the IRS Individual

Table 1: Power Law Cutoff Boundaries in US Data

Year	Percentile	α
1996	0.987	2.92
1997	0.985	2.89
1998	0.996	2.58
1999	0.996	2.58
2000	0.995	2.54
2001	0.996	2.63
2002	0.996	2.67
2003	0.996	2.65
2004	0.995	2.59
2005	0.994	2.54
2006	0.993	2.54
2007	0.993	2.54
2008	0.994	2.66
2009	0.995	2.78
2010	0.994	2.73
2011	0.994	2.74
2012	0.992	2.64
2013	0.993	2.74
2014	0.992	2.70
2015	0.991	2.72

Complete Report (Publication 1304), Table 1.1. Since this data is reported in *binned* form, I use the binned log-likelihood equation developed by Virkar and Clauset [2]:

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

Here α is the power law exponent, b_i and b_{i+1} are consecutive bin boundaries, h_i and h_{i+1} are consecutive bin counts, k is the number of bins, and n is the sum of bin counts above $b_{\min}$ (the cutoff point for the power law). The best-fit exponent α is the value that maximizes the log-likelihood function ($\mathscr{L}$). Since there is no closed-form solution to this maximization problem, I solve for α numerically. To determine the power law exponent for the top 1% of incomes in each year, I set the power law cutoff boundary ($b_{\min}$) to the empirical bin that is closest to the 99th percentile. Results are shown in Table 1.

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

$$\hat{\alpha} = 1 + n\left[\sum_{i}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} \tag{2}$$

Here $\hat{\alpha}$ is the best-fit power law exponent, x_i is the ith data point, $x_{\min}$ is the lower bound of the power law, and n is the number of data points above $x_{\min}$. To ensure compatibility with empirical power law estimates, I estimate the model's power law exponent using the *empirical* cutoff values. For each model run, I set $x_{\min}$ by randomly selecting a percentile value from Table 1.

All data and code are available in the Supplementary Material.

Probability Density Function

I estimate the normalized probability density function for US income using data from Current Population Survey Table PINC-08 (available from the US Census) over the years 1994 to 2015. This table reports binned data.

To estimate the normalized probability density function in each year, I first create a simulated income distribution ($\mathbf{I}$) using bin midpoints. Each midpoint income M_i is repeated F_i times, where F_i is the frequency count for the ith bin. I then normalize $\mathbf{I}$ by dividing all elements by the mean income $\bar{I}$.

$$\mathbf{I} = \frac{\left(M_1 \overset{\times F_1}{\cdots\cdots}, \ M_2 \overset{\times F_2}{\cdots\cdots}, \ldots, \ M_i \overset{\times F_i}{\cdots\cdots}\right)}{\bar{I}} \tag{3}$$

Lastly, I fit the simulated income distribution ($\mathbf{I}$) with a numerical density function. R code implementing this method is available in the Supplementary Material.

Top 1% Income Share

Sources for top 1% income share data are shown in Table 2.

Table 2: US Top 1% Income Share Sources

Series	Info	Source
sfainc992j	Pre-tax factor income | equal-split adults | Share | Adults | share of total (ratio)	[3]
sfainc996i	Pre-tax factor income | individuals | Share | 20 to 64 | share of total (ratio)	[3]
sfainc999i	Pre-tax factor income | individuals | Share | All Ages | share of total (ratio)	[3]
sfainc999t	Pre-tax factor income | tax unit | Share | All Ages | share of total (ratio)	[3]
sfiinc992j	Fiscal income | equal-split adults | Share | Adults | share of total (ratio)	[3]
sfiinc992t	Fiscal income | tax unit | Share | Adults | share of total (ratio)	[3]
sfiinc996i	Fiscal income | individuals | Share | 20 to 64 | share of total (ratio)	[3]
sfiinc999i	Fiscal income | individuals | Share | All Ages | share of total (ratio)	[3]
sfiinc999t	Fiscal income | tax unit | Share | All Ages | share of total (ratio)	[3]
sptinc992j	Pre-tax national income | equal-split adults | Share | Adults | share of total (ratio)	[3]
sptinc996i	Pre-tax national income | individuals | Share | 20 to 64 | share of total (ratio)	[3]
sptinc999i	Pre-tax national income | individuals | Share | All Ages | share of total (ratio)	[3]
sptinc999t	Pre-tax national income | tax unit | Share | All Ages | share of total (ratio)	[3]
sfiinc_z_US	World Top Incomes Legacy Series	[4]
lakner	Calculated from micro data	[5]
piketty_book_no_kgains	Legacy data from Capital in the 21st Century	[6]
piketty_book_with_kgains	Legacy data from Capital in the 21st Century	[6]

Sources for Figure 5 (Firm Size Distributions Associated With Top Incomes and Wealth)

Forbes 400 data is from the year 2014. Firm size data was collected by the author. For public companies, firm size data comes from Compustat. For private companies, data comes from firm websites and annual reports. The Execucomp 500 consists of the 500 top paid US executives in the Execucomp database in each year from 1992 to 2015.

B Case-Study Firms

In this section I review the case-study evidence that informs the hierarchy model. Table 3 summarizes the source data, while Figure 1 shows the hierarchical employment and pay structure of these firms. The firms remain anonymous, and are named after the authors of the case-study papers. Although the exact shapes vary, all the firms in this sample have a roughly pyramidal employment structure and inverse pyramid pay structure.

Figure 2 dissects these trends to allow further analysis. Figure 2A shows how the span of control (the employment ratio between adjacent ranks) changes as a function of hierarchical level. In these firms, the span of control is not constant, but instead tends to *increase* with hierarchical level. Similarly, Figure 2B shows the ratio of mean pay between adjacent levels. Like the span of control, the pay ratio tends to increase with hierarchical level. Lastly, Figure 2C shows income dispersion within hierarchical ranks of each firm (measured with the Gini index). Note that income dispersion within levels is quite low and there is no evidence of a trend.

In addition to case-study data of single firms, several studies have reported the aggregate hierarchical structure of a sample of firms (see Table 4 and Figure 4). The data from these firms reveals the same general trends as the case studies. However, the aggregate data is less useful because these studies capture only the top few hierarchical ranks within firms.

The case-study data plays a central role in the hierarchical model developed in this paper. From the case-study evidence, I propose the following 'stylized' facts about firm employment and pay structure:

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

The case-study evidence informs the basic structure of the model, and also some of its key parameters. The 'shape' of modeled firm hierarchies is determined from the fitted span-of-control trend shown in Figure 2A. Figure 3 shows the idealized employment hierarchy that is implied by case-study data. Error bars indicate uncertainty, calculated using the bootstrap resampling method. Parameters for intra-level income dispersion are determined from the mean of data in Figure 2C. For a detailed discussion of the model algorithm and parameter-fitting procedure, see Sections D and E.

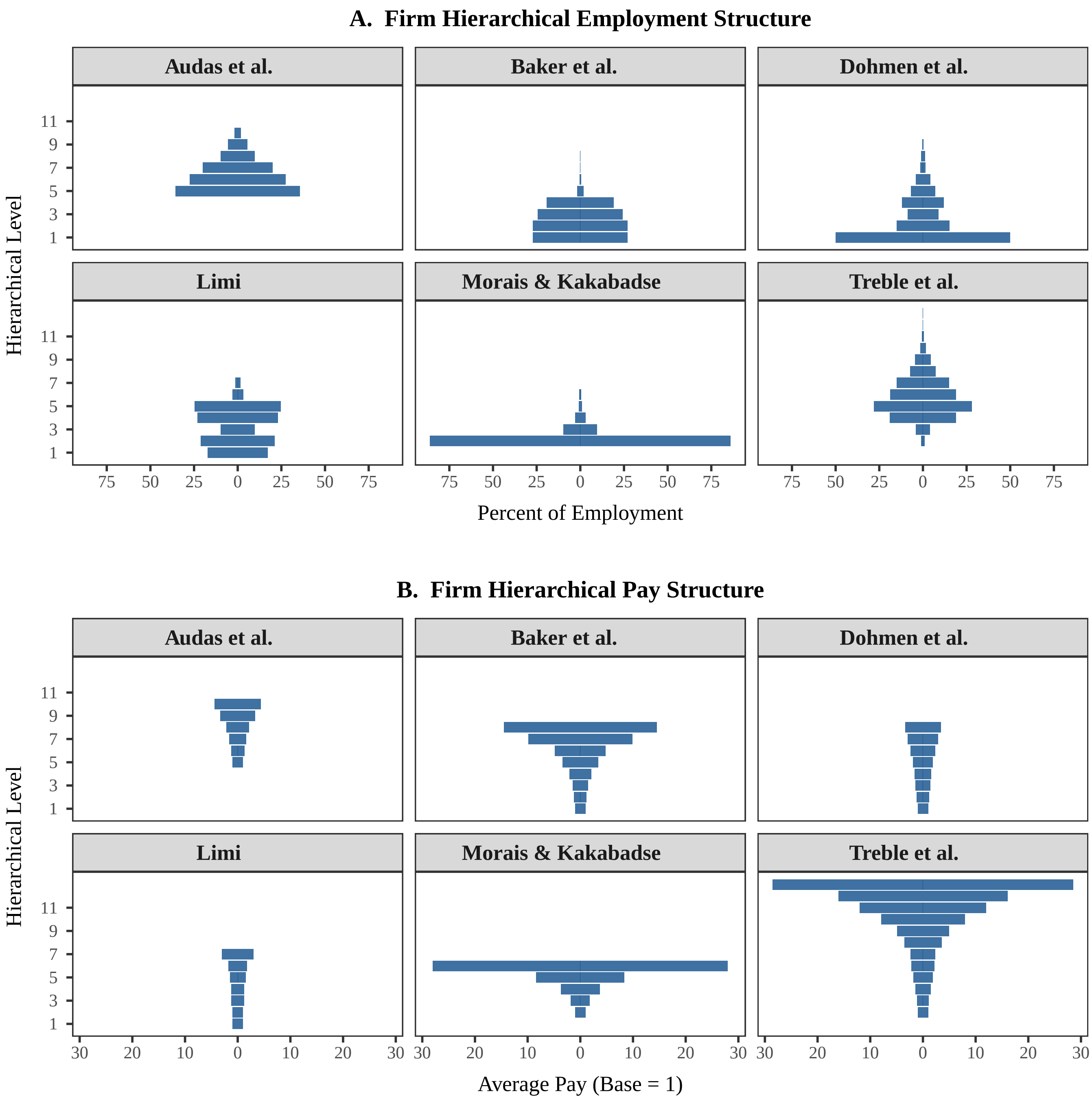

Figure 1: The Hierarchical 'Shape' of Six Different Case-Study Firms

This figure shows the hierarchical employment and pay structure of six different case-study firms. Panel A shows the hierarchical structure of employment, while Panel B shows the hierarchical pay structure.

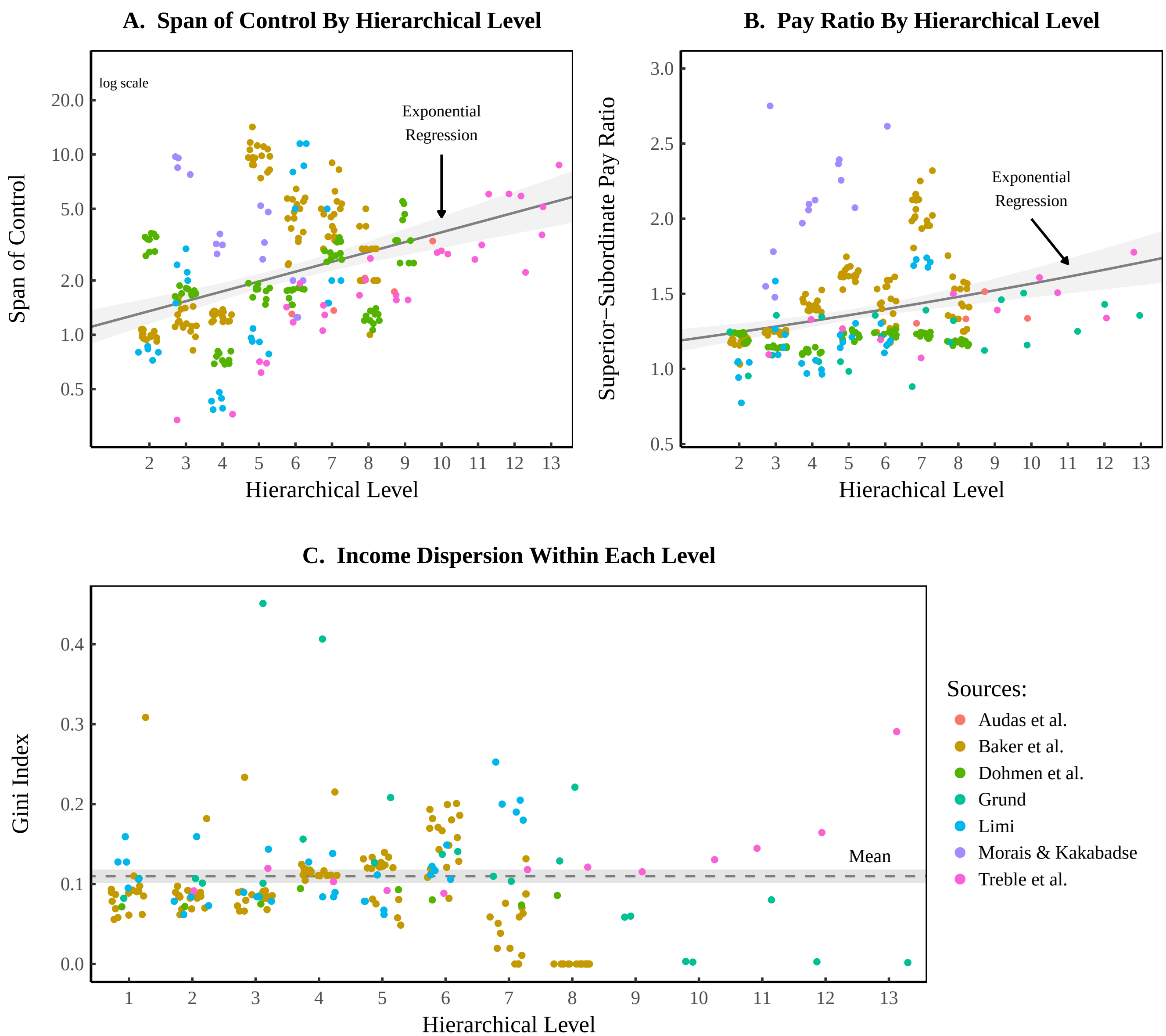

Figure 2: Analyzing the Hierarchical Structure of Case-Study Firms

This figure shows data from 7 case-study firms. Panel A shows how the span of control (the subordinate-to-superior employment ratio between adjacent levels) varies with hierarchical level. Note the log scale on the y-axis. Panel B shows how the superior-to-subordinate pay ratio varies with hierarchical level. In Panels A and B, the x-axis corresponds to the *upper* hierarchical level in each corresponding ratio. Panel C shows the Gini index of income inequality within each hierarchical level. Different case-study firms are indicated by color, with names indicating the study author. Note that horizontal 'jitter' has been introduced in all three plots in order to better visualize the data (hierarchical level is a discrete variable). The lines in Panels A and B indicate exponential regressions, while the line in Panel C shows the average Gini index. Grey regions correspond to the 95% confidence intervals.

Table 3: Summary of Firm Case Studies

Source		Years	Country	Firm Levels	Span of Control	Level Income	Level Income Dispersion
Audas	[7]	1992	Britain	All	✓	✓	
Baker	[8]	1969-1985	United States	Management	✓	✓	✓
Dohmen	[9]	1987-1996	Netherlands	All	✓	✓	✓
Grund	[10]	1995 & 1998	US and Germany	All		✓	✓
Lima	[11]	1991-1995	Portugal	All	✓	✓	✓
Morais*	[12]	2007-2010	Undisclosed	All	✓	✓	
Treble	[13]	1989-1994	Britain	All	✓	✓	✓

Notes: This table shows metadata for the firm case studies displayed in Fig. 2. The 'Firm Levels' column refers to the portion of the firm that is included in the study. 'Management' indicates that only management levels were studied. *For the analysis conducted in this paper I discard (as an outlier) the bottom hierarchical level in Morais and Kakabadse's data.

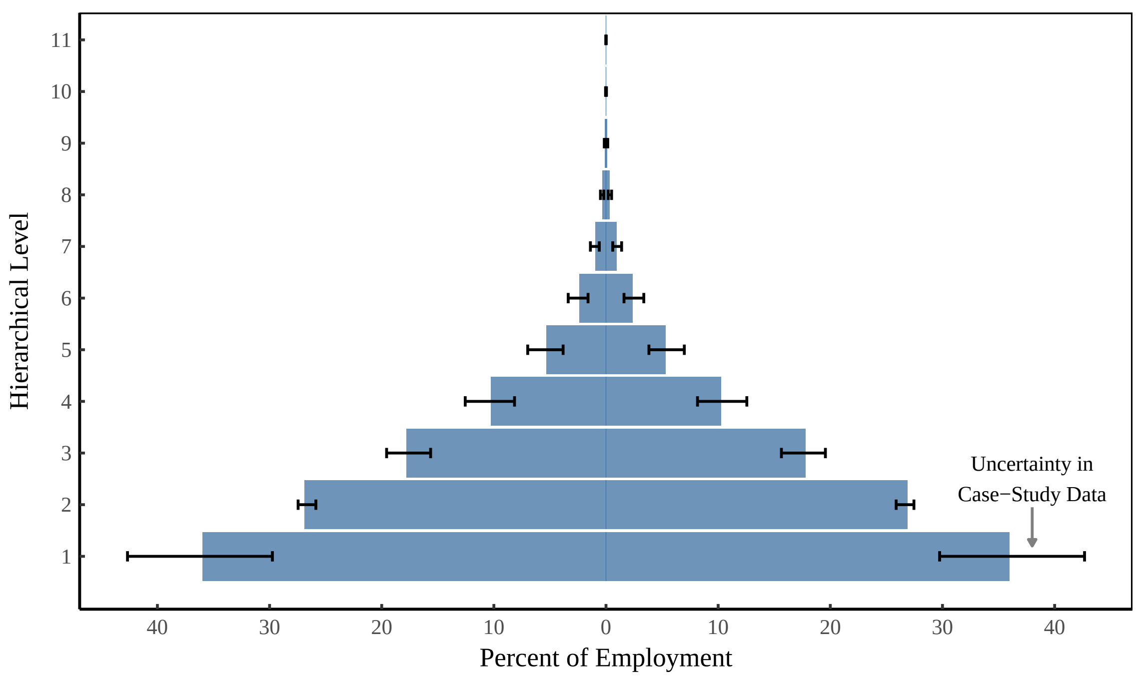

Figure 3: Idealized Firm Employment Hierarchy Implied by Case Studies

This figure shows the idealized firm hierarchy that is implied by fitting trends to case-study data (Fig. 2A). Error bars show the uncertainty in the hierarchical shape, calculated using a bootstrap resample of case-study data.

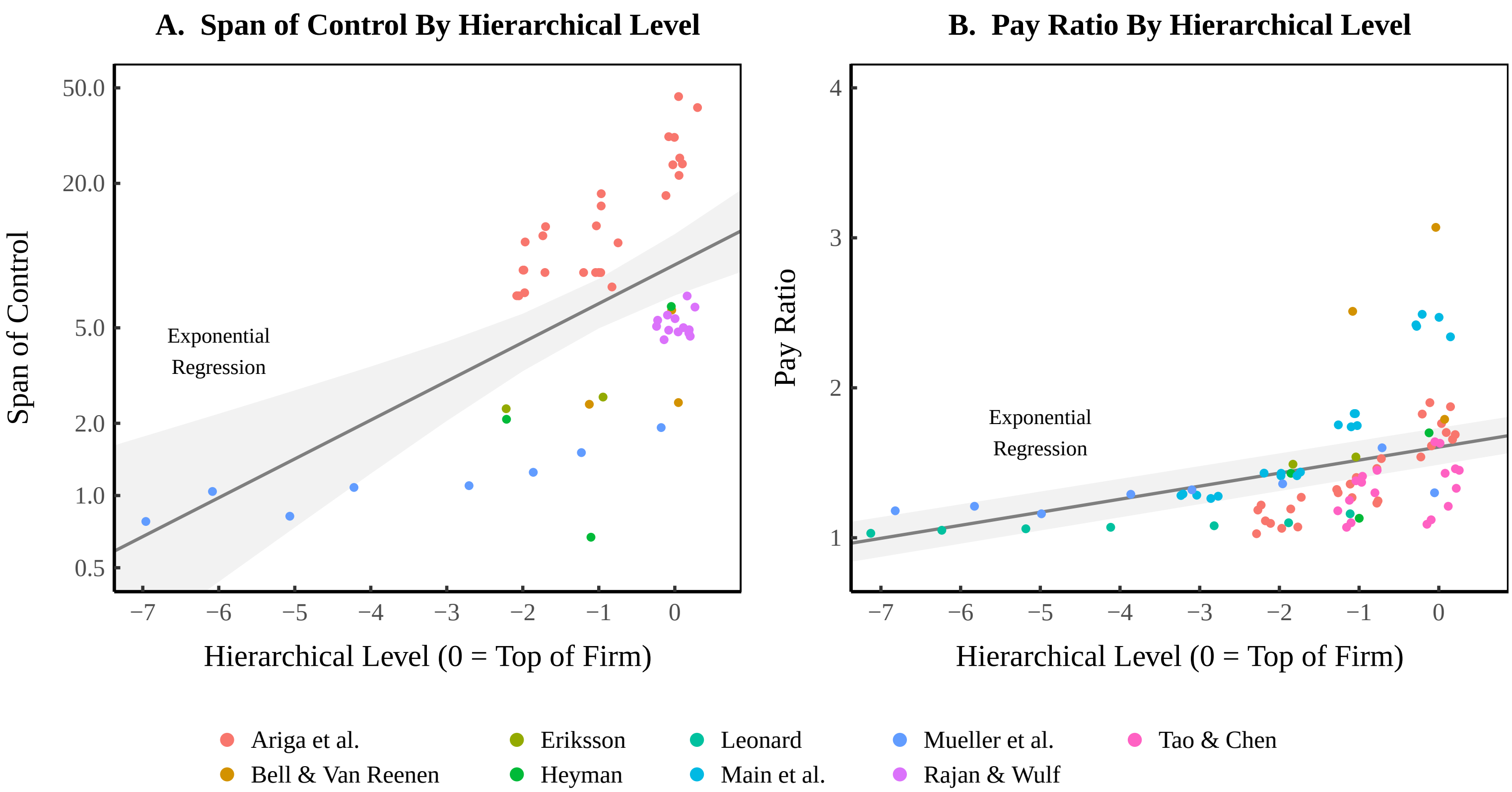

Figure 4: Aggregate Studies of Firm Hierarchical Structure

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 4: Summary of Firm Aggregate Studies

Source		Years	Number of Firms	Country	Firm Levels	Span of Control	Level Income
Ariga	[14]	1981-1989	unknown	Japan	All	✓	✓
Bell	[15]	2001-2010	552	United Kingdom	Top 3	✓	✓
Eriksson	[16]	1992-1995	210	Denmark	Management	✓	✓
Heyman	[17]	1991,1995	560	Sweden	Management	✓	✓
Leonard	[18]	1981-1985	439	United States	Top 9		✓
Main	[19]	1980–1984	200	United States	Top 4		✓
Mueller	[20]	2004-2013	880	United Kingdom	All	✓	✓
Rajan	[21]	1986-1998	261	United States	Top 2	✓	
Tao	[22]	1986-1998	8101	Taiwan	Top 2		✓

Notes: This table shows metadata for the aggregate studies displayed in Fig. 4. The 'Firm Levels' column 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).

C Compustat Data

This paper makes extensive use of the *Compustat* and *Execucomp* databases. Compustat contains data for most publicly traded US companies, while Execucomp contains data for executive compensation. Three key statistics used throughout this paper are calculated from this data: *firm mean income*, the *CEO-to-average-employee pay ratio*, and the *capitalist income fraction of executives*. I discuss the data and methods used for these calculations in the following sections.

C.1 Firm Mean Income

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

$$\text{Firm Mean Income} = \frac{\text{Total Staff Expenses}}{\text{Total Employment}} \tag{4}$$

C.2 CEO Pay Ratio

Throughout this paper, I use the term 'CEO' to refer to the executive at the top of the corporate hierarchy. I identify CEOs using the titles contained in the Execucomp series TITLEANN. Because titles vary greatly by company, identifying the top executive is not always a simple task. While a manual search would be most accurate, this is unrealistic given that the Execucomp database contains over 275 000 entries. Instead, I use the following three-step algorithm to identify the 'CEO':

1. Find all executives whose title contains one or more of the words in the 'CEO Titles' list (Table 5).
2. Of these executives, take the subset whose title does *not* contain any of the words in the 'Subordinate Titles' list (Table 5).
3. If this search returns more than one executive per firm per year, chose the executive with the highest pay.

After identifying the CEO (and matching CEO pay data with firm data contained in the Compustat database), I calculate the CEO pay ratio using the following equation:

$$\text{CEO Pay Ratio} = \frac{\text{CEO Pay}}{\text{Firm Mean Income}} \tag{5}$$

Table 5: Titles Used to Identify the 'CEO'

CEO Titles:	**Subordinate Titles**
president	vp
chairman	v-p
CEO	cfo
Chief Executive Officer	vice
chmn	chief finance officer
	president of
	coo
	division
	div
	president-
	group president
	chairmain-
	co-president
	deputy chairman
	pres.-
	Chief Financial Officer

Notes: This table shows the Execucomp titles used to identify the CEO of each company. CEOs are deemed to be those whose title contains words in the left column, but not those in the right column. Titles such as 'president-' and 'president of' are included in the subordinate list because they typically refer to a president of a division with the company: i.e. 'president of western division' or 'president-western hemisphere'.

CEO pay ratio and firm mean income data are collectively available for roughly 6000 firm-year observations over the period 1992-2016. I use this data to 'tune' my hierarchical model of the firm (see Section E) . Figure 5 shows selected summary statistics of this dataset.

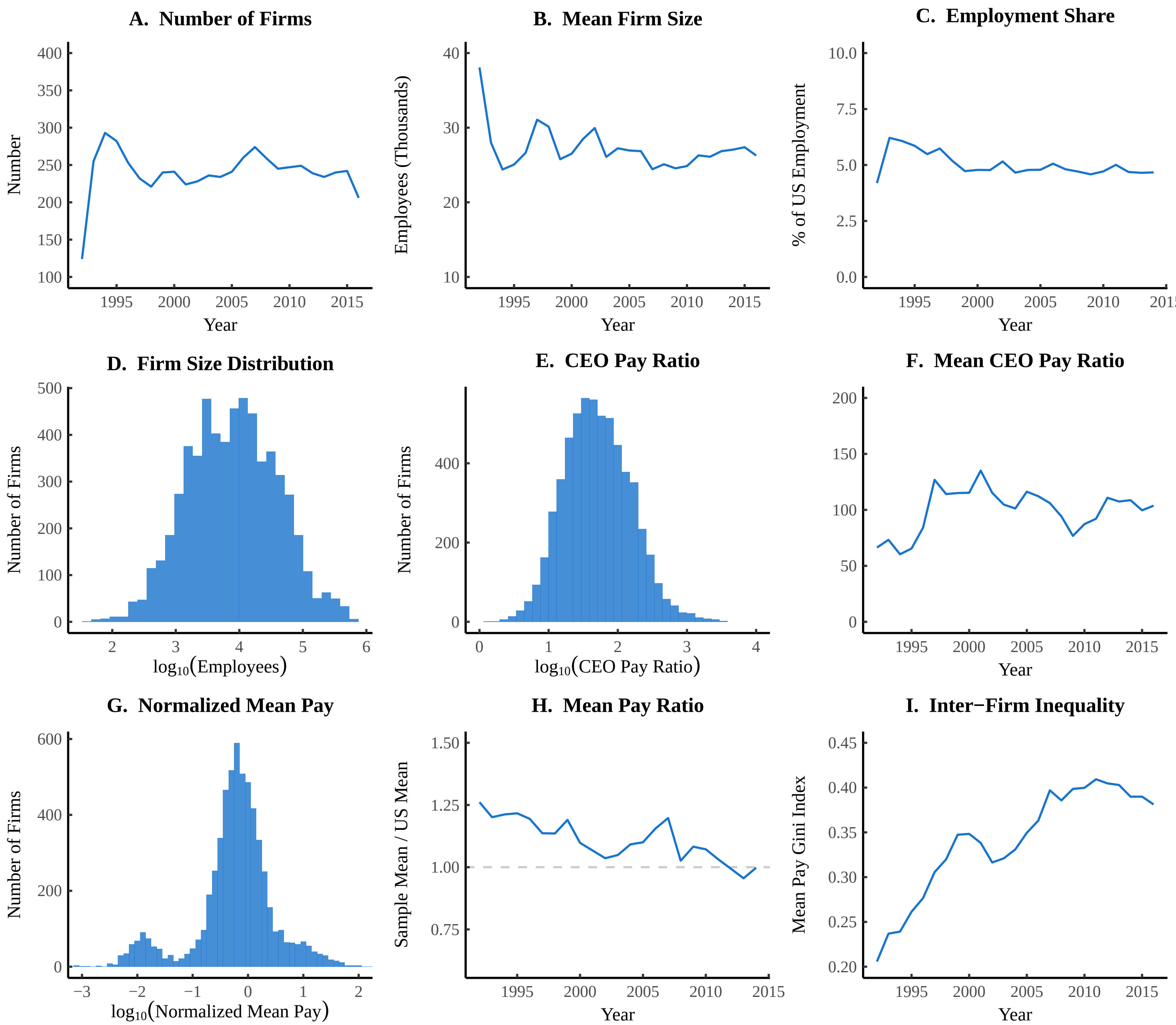

Figure 5: Selected Statistics from the Firm Sample Used for Model Tuning

This figure shows statistics for the Compustat firm sample used to tune my hierarchical model. Panel A shows the number of firms in the sample over time, Panel B the average firm size, and Panel C the share of US employment held by these firms. Panel D shows the logarithmic distribution of firm size, and Panel E shows the logarithmic distribution of the CEO pay ratio. Panel F shows the mean CEO pay ratio of all firms over time. Panel G shows the logarithmic distribution of normalized mean pay (mean pay divided by the average pay of the firm sample in each year). Panel H shows the ratio of mean pay in the Compustat sample relative to the US average (calculated from BEA Table 1.12 by dividing the sum of employee and proprietor income by the number of workers in BEA Table 6.8C-D. Panel I shows the Gini index of firm mean pay over time.

D Hierarchy Model Equations

In this section, I outline the mathematics underlying my hierarchical model of the firm. The model assumptions, outlined below, are based on the stylized facts gleaned from the real-world firm data in section B.

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

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

Table 6: Notation

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

D.1 Generating the Employment Hierarchy

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

$$s_h \equiv \begin{cases} 1 & \text{if } \ h = 1 \\ \dfrac{E_{h-1}}{E_h} & \text{if } \ h \geq 2 \end{cases} \tag{6}$$

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:

$$s_h = \begin{cases} 1 & \text{if } \ h = 1 \\ a \cdot e^{bh} & \text{if } \ h \geq 2 \end{cases} \tag{7}$$

As one moves up the hierarchy, employment in each consecutive level (E_h) *decreases* by $1/s_h$. This yields Eq. 8, a recursive method for calculating E_h. Since we want employment to be *whole* numbers, we round down to the nearest integer (notated by $\downarrow$). By repeatedly substituting Eq. 8 into itself, we can obtain a non-recursive formula (Eq. 9). In product notation, Eq. 9 can be written as Eq. 10.

$$E_h =\downarrow \frac{E_{h-1}}{s_h} \qquad \text{for} \qquad h > 1 \tag{8}$$

$$E_h =\downarrow E_1 \cdot \frac{1}{s_2} \cdot \frac{1}{s_3} \cdot \ldots \cdot \frac{1}{s_h} \tag{9}$$

$$E_h =\downarrow E_1 \prod_{i=1}^{h} \frac{1}{s_i} \tag{10}$$

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. 11, which in summation notation, becomes Eq. 12.

$$E_T = E_1 + E_2 + \ldots + E_n \tag{11}$$

$$E_T = \sum_{h=1}^{n} E_h \tag{12}$$

In practice, n is not known beforehand, so we define it using Eq. 10. 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:

$$n = \{h \quad | \quad E_h \geq 1 \quad \text{and} \quad E_{h+1} = 0\} \tag{13}$$

To summarize, the hierarchical employment structure of our model firm is determined by 3 free parameters: the span of control parameters a and b, and base-level employment E_1. Code for this hierarchy generation algorithm can be found in the C++ header files `hierarchy.h` and `exponents.h`, located in the Supplementary Material.

D.2 Generating Hierarchical Pay

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

$$p_h \equiv \begin{cases} 1 & \text{if } \; h = 1 \\ \dfrac{\bar{I}_h}{\bar{I}_{h-1}} & \text{if } \; h \geq 2 \end{cases} \tag{14}$$

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:

$$p_h = \begin{cases} 1 & \text{if } \; h = 1 \\ r^h & \text{if } \; h \geq 2 \end{cases} \tag{15}$$

Using the same logic as with employment (shown above), the mean income I_h in any hierarchical level is defined recursively by Eq. 16 and non-recursively by Eq. 17.

$$\bar{I}_h = \frac{\bar{I}_{h-1}}{p_h} \tag{16}$$

$$\bar{I}_h = \bar{I}_1 \prod_{i=1}^{h} p_i \tag{17}$$

To summarize, the hierarchical pay structure of our model firm is determined by 2 free parameters: the pay-scaling parameter r, and mean pay in the base level ($\bar{I}_1$). Code for generating hierarchical pay can be found in the C++ header files `model.h`, located in the Supplementary Material.

D.2.1 Useful Statistics

Two statistics are used repeatedly within the model: mean firm pay, and the CEO-to-average-employee pay ratio.

Mean income for all employees ($\bar{I}_T$) is equal to the average of hierarchical level mean incomes ($\bar{I}_h$) weighted by the respective hierarchical level employment (E_h):

$$\bar{I}_T = \sum_{h=1}^{n} \bar{I}_h \cdot \frac{E_h}{E_T} \tag{18}$$

To calculate the CEO pay ratio, we define the CEO as the person(s) in the top hierarchical level. Therefore, CEO pay is simply $\bar{I}_n$, average income in the top hierarchical level. The CEO pay ratio (C) is then equal to CEO pay divided by average pay:

$$C = \frac{\bar{I}_n}{\bar{I}_T} \tag{19}$$

D.3 Adding Intra-Level Pay Dispersion

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

I assume that pay dispersion within hierarchical levels is *lognormally* distributed. The lognormal distribution is defined by location parameter μ and scale parameter σ. Our empirical investigation of firm case studies indicated that pay dispersion with hierarchical levels is relatively constant (see Fig. 2C). 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.

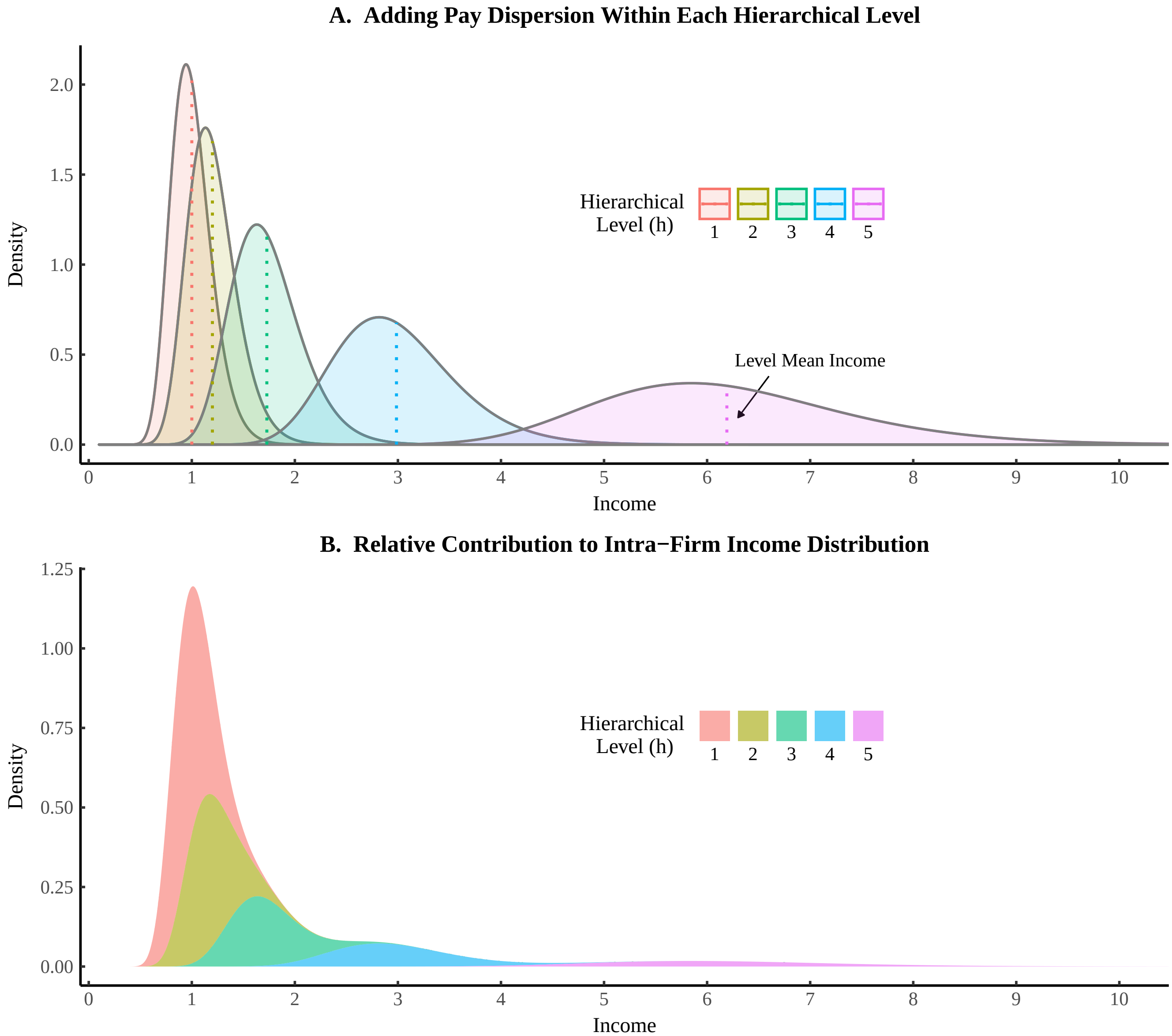

Figure 6: Adding Intra-Level Pay Dispersion to a Model Firm

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

In order to add dispersion within each hierarchical level, I multiply mean pay $\bar{I}_h$ by a lognormal random variate with an expected mean of one. Formally, this is represented by Eq. 20. Since the mean of a lognormal distribution is equal to $e^{\mu+\frac{1}{2}\sigma^2}$, I leave it to the reader to show that a mean of one requires that μ be defined by Eq. 21.

$$I_h = \bar{I}_h \cdot \ln \mathcal{N}(\mu, \sigma) \tag{20}$$

$$\mu = -\frac{1}{2}\sigma^2 \tag{21}$$

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 6. Figure 6A shows the lognormal income distributions for each hierarchical level of a 5-level firm. Figure 6B shows the size-adjusted contribution of each hierarchical level to the overall intra-firm income distribution. Lower levels have more members, and thus dominate the overall distribution. The code implementing this method can be found in the C++ header file `model.h`, located in the Supplementary Material.

Table 7: Model Parameters

Parameter	Definition	Action	Scope
α	Firm size distribution exponent	Determines the skewness of the firm size distribution	—
a, b	Span of control parameters	Determines the shape of the firm hierarchy.	Identical for all firms.
E_1	Employment in base hierarchical level	Used to build the employment hierarchy from the bottom up. Determines total employment.	Specific to each firm.
r	Pay-scaling parameter	Determines the rate at which mean income (within a firm) increases by hierarchical level.	Specific to each firm.
$\bar{I}_h$	Mean pay in base hierarchical level	Sets the base level income of the firm, which determines firm average pay.	Specific to each firm.
σ	Intra-hierarchical level pay dispersion parameter	Determines the level of inequality within hierarchical levels of a firm.	Identical for all firms.

E Restricting Parameters

As discussed in section D, the hierarchy model has many 'free' parameters. Table 7 summarizes all of the parameters used in this model. While free to take on any value, I restrict these parameters exclusively using empirical data. In the following sections, I outline the methods used for this restriction.

E.1 Firm Size Distribution

Recent studies have found that firm size distributions in the United States [23] and other G7 countries [24] can be modeled accurately with a power law. A power law has the simple form shown in Eq. 22, where the probability of observation x is inversely proportional to x raised to some exponent α:

$$p(x) \propto \frac{1}{x^{\alpha}} \tag{22}$$

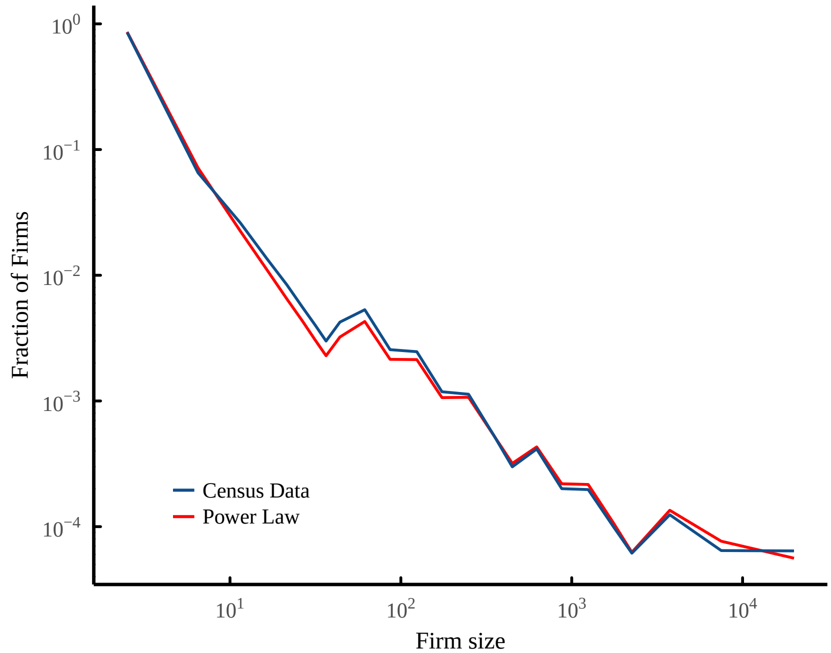

Figure 7: The United States Firm Size Distribution

This figure shows the US firm size distribution compared to a power law distribution with exponent $\alpha = 2.01$ (a simulation with 15 million firms) . The US histogram combines data for 'employer' firms with data for unincorporated self-employed workers. Data for 'employer' firms is from the US Census Bureau, Statistics of U.S. Businesses (using data for 2013). This data is augmented with Bureau of Labor Statistics data for unincorporated self-employed workers (series LNU02032185 and LNU02032192). The histogram preserves Census firm-size bins, with self-employed data added to the first bin. The last point on the histogram consists of all firms with more than 10,000 employees.

Figure 7 compares the US firm size distribution with a power law of exponent $\alpha = 2.01$. Although not perfect, the fit is good enough for modeling purposes. I assume that the firm sizes can be modeled with a discrete power law random variate. I model the US firm size distribution with $\alpha = 2.01$.

A characteristic property of power law distributions is that as α approaches 2, the mean becomes *undefined*. In the present context, this means that the model can produce firm sizes that are extremely large — far beyond anything that exists in the real world. To deal with this difficulty, I *truncate* the power law distribution at a maximum firm size of 2.3 million. This happens to be the present size of Walmart, the largest US firm in existence.

Code for the discrete power law random number generator can be found in

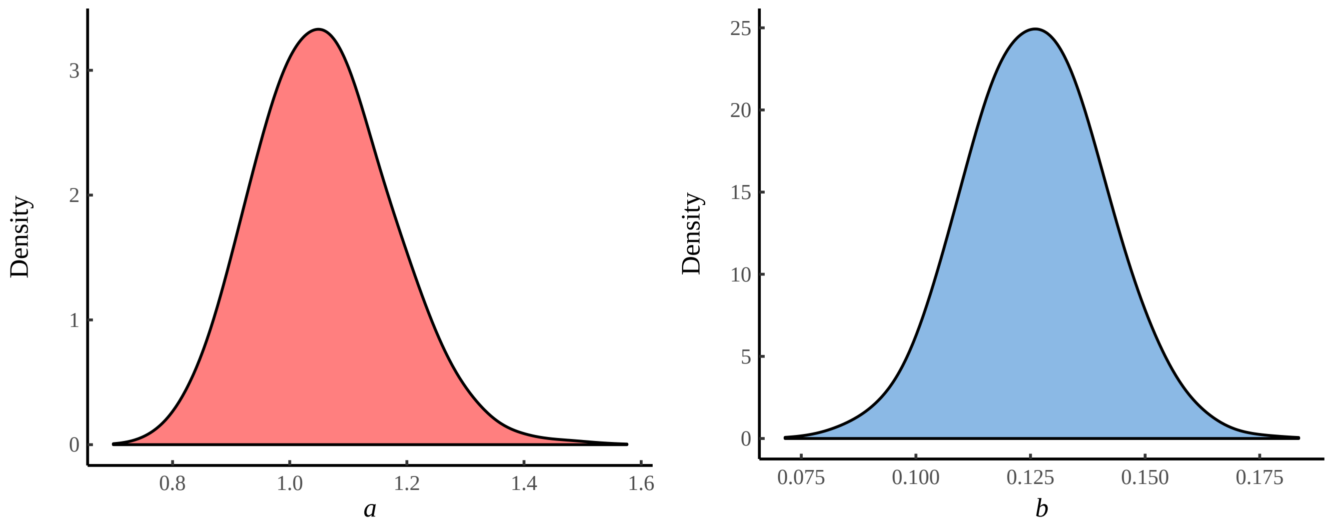

Figure 8: Density Estimates for Span of Control Parameters

This figure shows density estimates for the parameters a and b, which together determine the 'shape' of the firm hierarchy. These parameters are determined from regressions on firm case-study data (Fig. 2). The density functions are estimated using a bootstrap analysis, which involves resampling (with replacement) the case study data many times, and calculating the parameters a and b for each resample.

---

the C++ header file `rpld.h`, located in the Supplementary Material. This code is an adaption of Collin Gillespie's [25] discrete power law generator found in the R poweRlaw package (which is, in turn, an adaption of the algorithm outline by Clauset [26]).

E.2 Span of Control Parameters

The parameters a and b together determine the shape of firm employment hierarchy. These parameters are estimated from an exponential regression on case study data (Fig. 2A). The model proceeds on the assumption that these parameters are *constant* across all firms.

Because the case-study sample size is small, there is considerable uncertainty in these values. I incorporate this uncertainty into the model using the *bootstrap* method [27], which involves repeatedly resampling the case-study data (with replacement) and then estimating the parameters a and b from this resample. Figure 8 shows the probability density distribution resulting from this bootstrap analysis. I run the model many times, each time with a and b determined by a bootstrap resample of case-study data.

Code implementing this bootstrap can be found in the C++ header file `boot_span.h`.

E.3 Base Level Employment

Given span of control parameters a and b, each firm hierarchy is constructed from the bottom hierarchical level up. Thus, we must know base level employment. In practice, however, we don't know this value — instead we are given *total* employment for a particular firm. While it may be possible to use the equations in section D to define an analytic function relating total employment to base level employment, this is beyond my mathematical abilities.

Instead, I use the model to reverse engineer the problem. I input a range of different base employment values into equations 7, 10, and 12 and calculate total employment for each value. The result is a discrete mapping relating base-level employment to total employment. I then use the C++ Armadillo interpolation function to linearly interpolate between these discrete values. This allows us to predict base level E_1, given total employment E_T. Code implementing this method can be found in the C++ header file `base_fit.h`, located in the Supplementary Material.

E.4 Pay-Scaling Parameter

The pay-scaling ratio r determines the rate at which mean pay increases by hierarchical level. Unlike the span of control parameters, the pay-scaling parameter is allowed to vary between firms. But how should it vary? I restrict the variation of this parameter in a two-step process. I first 'tune' the model to Compustat data. This results in a distribution of pay-scaling parameters specific to Compustat firms. I then fit this data with a parameterized distribution, from which simulation parameters are randomly chosen.

E.4.1 Fitting Compustat Pay-Scaling Parameters

I fit the pay-scaling parameter r to Compustat firms using the CEO-to-average-employee pay ratio (C). The first step of this process is to build the employment hierarchy for each Compustat firm using parameters a, b, and E_1 (the latter is determined from total employment). Given this hierarchical employment structure, the CEO pay ratio in the modeled firm is uniquely determined by the parameter r. Thus, we simply choose r such that the model produces a CEO pay ratio that is equivalent to the empirical ratio.

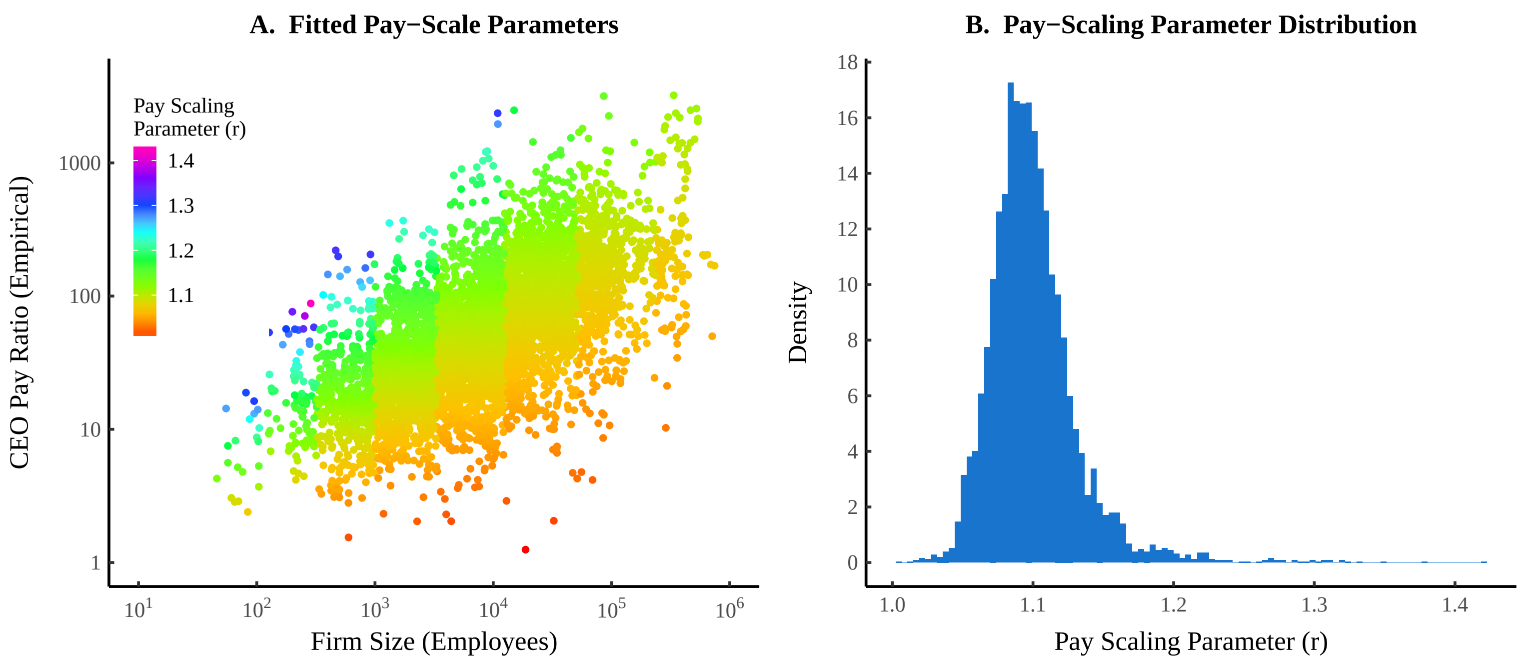

Figure 9: Fitting Compustat Firms with a Pay-Scaling Parameter

This figure shows the fitted pay-scaling parameters (r) for all Compustat firms. Panel A shows the relation between the CEO pay ratio and firm size, with the fitted pay-scaling parameter indicated by color. The discrete changes in color (evident as vertical lines) correspond to changes in the number of hierarchical levels within firms. The pay-scaling parameter distribution for all firms (and years) is shown in panel B.

To solve for this r value, I use numerical optimization (the bisection method) to minimize the error function shown in Eq. 23. Here $C_{\text{Compustat}}$ and C_{model} are Compustat and modeled CEO pay ratios, respectively.

$$\epsilon(r) = \Big| C_{\text{model}} - C_{\text{Compustat}} \Big| \tag{23}$$

For each firm, the fitted value of r minimizes this error function. To ensure that there are no large errors, I discard Compustat firms for which the best-fit r parameter produces an error that is larger than $\epsilon = 0.01$). Fitted results for r are shown in Figure 9. Code implementing this method can be found in the C++ header file `fit_model.h`, located in the Supplementary Material.

E.4.2 Generating a Pay Scaling Distribution

Once we have generated r parameters for every Compustat firm, the next step is to fit a parameterized distribution to this data. For Compustat firms, the dis-

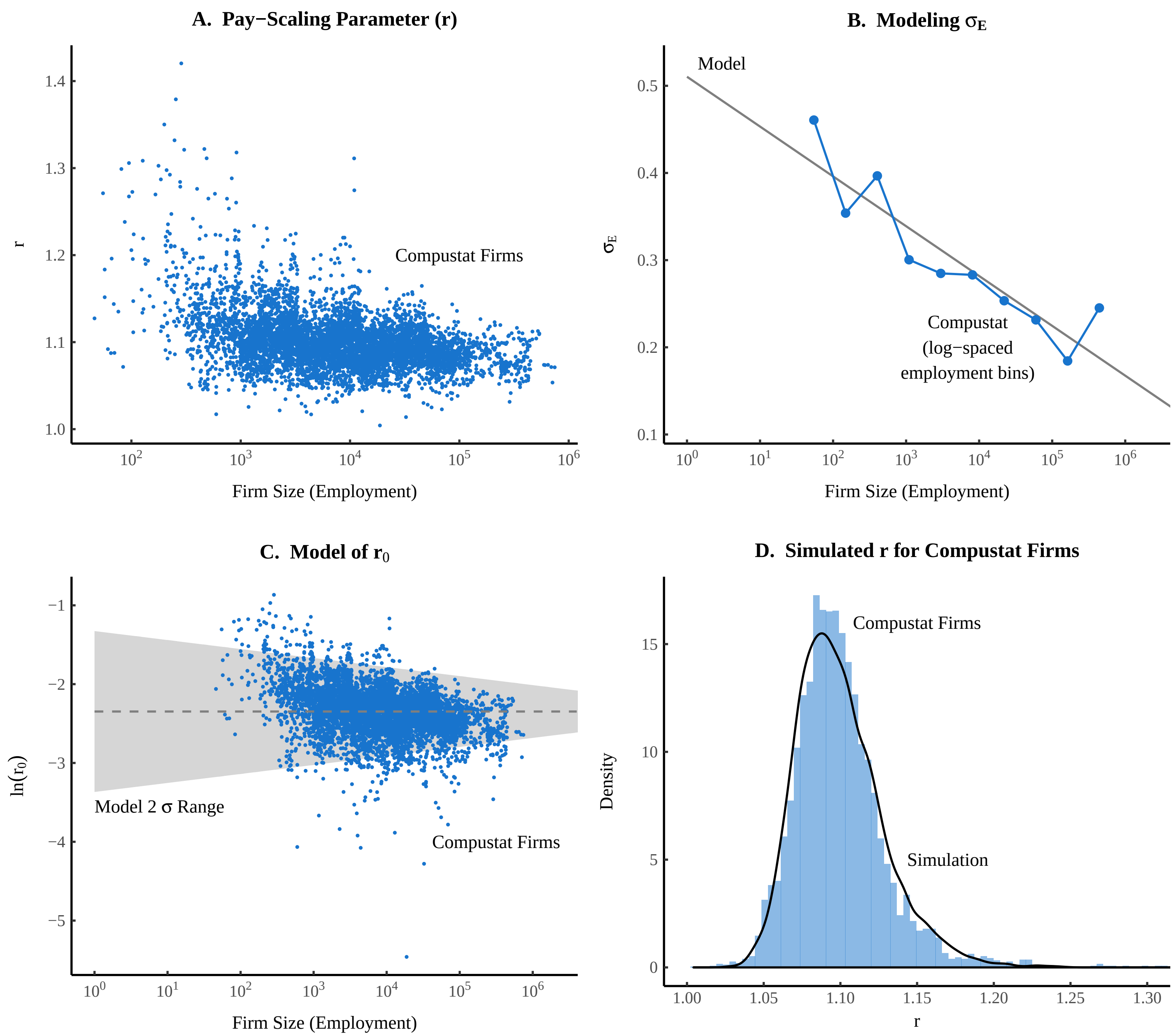

Figure 10: Modeling the Firm Pay Scaling Distribution

This figure visualizes the model used to simulate firm pay-scaling parameters (r). Panel A shows the relation between r and firm employment for Compustat firms. For the simulation, the distribution of r is modeled with the lognormal variate r_0. Panel B shows how the lognormal scale parameter σ_E (defined by Eq. 28) changes with firm size. The straight line indicates the modeled relation. Panel C shows how the modeled dispersion of $\ln(r_0)$ declines with firm size, and how this relates to Compustat r_0 data. The 2σ range indicates 2 standard deviations from the mean (on log-transformed data). Panel D shows how the distribution of r for Compustat firms compares to the *simulated* distribution achieved by applying the model to the same Compustat firms.

persion of r is approximately lognormal, and tends to decline with firm size (see Figure 10A). I model r as a shifted function of the lognormal variate r_0:

$$r = 1 + \ln\mathcal{N}(r_0) \tag{24}$$

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

$$r_0(E) = \ln\mathcal{N}(r_0; \mu, \sigma_E) \tag{25}$$

I use the tuned Compustat data to solve for the parameters μ and σ. We first transform Compustat r values using Eq. 26 to get the Compustat distribution of r_0:

$$r_0 = r - 1 \tag{26}$$

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

$$\mu = \overline{\ln(r_0)} \tag{27}$$

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

$$\sigma_E = \mathrm{SD}\left[\ \ln(r_0)\ \right]_E \tag{28}$$

Figure 10B plots σ_E vs. E for logarithmically spaced size groupings of Compustat firms. I model this relation using a log-linear regression. Figure 10C shows how the modeled dispersion in r_0 varies with firm size, and how this compares to Compustat data.

Once we have fitted the parameters μ and σ to the tuned Compustat data, we can generate r values for simulated firms using equations 24 and 25. Although the model is simple, it produces reasonably accurate results. To test this accuracy, we can apply the model to the same Compustat firms for which it is 'tuned'. For each Compustat firm, we use the method outlined above to stochastically generate a pay-scaling value r. As Figure 10D shows, the resulting simulated distribution of r fairly accurately reproduces the original data.

When we move from simulating Compustat firms to a real-world distribution of firms, this model involves significant extrapolations for small firms. Why?

The Compustat firm sample has very few observations for firms smaller than 100. And those small firms that are included in the sample are likely *not* representative of the wider population, since they are small *public* firms. In the real world, virtually all small firms are *private*. As with all extrapolations, we simply do the best with the data that is available, while noting that better data might render the extrapolation moot. The code implementing this model can be found in the C++ header file `r_sim.h`, located in the Supplementary Material.

E.5 Base-Level Mean Pay

As with the pay-scaling parameter, base level mean pay varies across firms. How should it vary? Again, I restrict the variation of this parameter in a two-step process. I first 'tune' the model to Compustat data. This results in a distribution of base pay specific to Compustat firms. I then fit this data with a parameterized distribution, from which simulation parameters are randomly chosen.

E.5.1 Fitting Compustat Base Level Pay

Having already fitted a hierarchical pay structure to each Compustat firm (in the process of finding r), we can use this data to estimate base pay for each firm. To do this, we set up a ratio between base level pay ($\bar{I}_1$) and firm mean pay ($\bar{I}_T$) for both the model and Compustat data:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{\bar{I}_1^{\text{model}}}{\bar{I}_T^{\text{model}}} \tag{29}$$

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. 17 and 18). If we run the model with $\bar{I}_1^{\text{model}} = 1$, then Eq. 29 reduces to:

$$\frac{\bar{I}_1^{\text{Compustat}}}{\bar{I}_T^{\text{Compustat}}} = \frac{1}{\bar{I}_T^{\text{model}}} \tag{30}$$

We can then rearrange Eq. 30 to solve for an estimated base pay for each Compustat firm ($\bar{I}_1^{\text{Compustat}}$):

$$\bar{I}_1^{\text{Compustat}} = \frac{\bar{I}_T^{\text{Compustat}}}{\bar{I}_T^{\text{model}}} \tag{31}$$

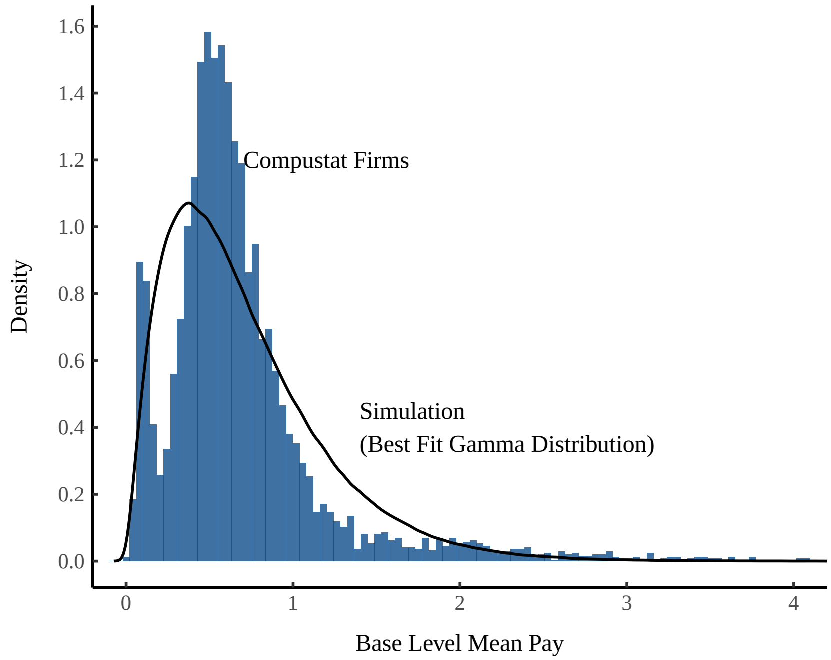

Figure 11: Modeling Firm Base Level Mean Pay

This figure shows the distribution of fitted base-level mean pay for Compustat firms. I model this data with a gamma distribution, from which simulated firm base-level mean pay is randomly drawn. Note that fitting the unimodal gamma distribution to the bi-modal Compustat data means that the fit is not great. (The gamma distribution does fit the data better than other skewed distributions such as the Weibull or lognorma). The lower mode in the Compustat data is likely not representative of the general firm population. This lower mode is made up almost entirely of chain restaurants, which seem to be over-represented in this sample.

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

E.5.2 Generating a Base Pay Distribution

Once each Compustat firm has a fitted value for base-level mean pay, we fit this data with a parametric distribution which is then used to stochastically generate base-level mean pay for the simulation. Since Compustat data is comprised of observations over *multiple* years, in order to aggregate this data into a single distribution, we must account for inflation. Rather than use a price index like the GDP deflator, I divide all firm mean pay data by the average Compustat mean pay in the appropriate year. Since our simulation is concerned only with relative incomes (rather than absolute incomes) no pertinent information is lost in this process.

I model the Compustat firm base pay distribution with a gamma distribution (Fig. 11). Note that because the Compustat data has a bimodal structure (that I do not aim to replicate), the gamma distribution is not a particularly strong fit. Nonetheless the gamma model closely replicates the inequality of firm base pay (which has a Gini index of roughly 0.35). Code implementing this model can be found in the C++ header file `base_pay_sim.h` (in the Supplementary Material).

E.6 Intra-Hierarchical Level Income Dispersion

Intra-hierarchical level income dispersion is modeled with a lognormal distribution, with the amount of inequality determined by the scale parameter σ. I estimate σ from the case-study data shown in Figure 2C. 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. 32 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: $G = \text{erf}(\sigma/2)$.

$$\sigma = 2 \cdot \text{erf}^{-1}(\bar{G}) \tag{32}$$

The model proceeds on the assumption that σ is constant for all hierarchical levels within all firms. Because the case-study sample size is small, there is considerable uncertainty in these values. I quantify this uncertainty using the *bootstrap* method [27], which involves repeatedly resampling the case-study data (with replacement) and then estimating the parameter σ from this resampled data.

Figure 12 shows the probability density distribution resulting from this bootstrap analysis. In order to incorporate this uncertainty, I run the model many times, with each run using a different bootstrapped value for σ. Code implementing this method can be found in the C++ header file `boot_sigma.h`, located in the Supplementary Material.

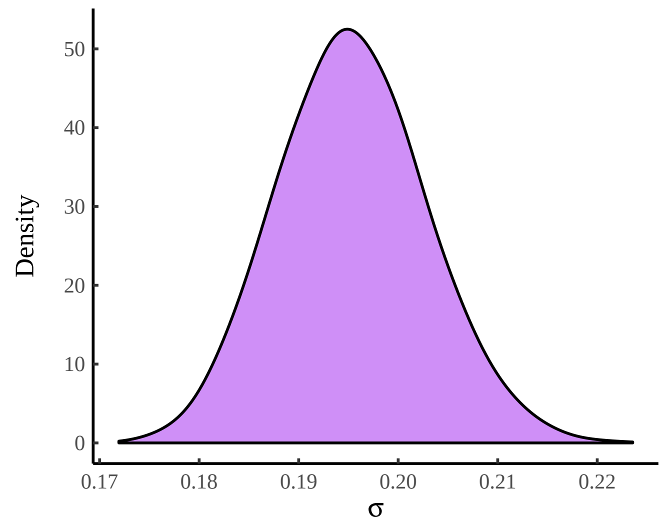

Figure 12: Density estimates for Intra-Hierarchical Level Pay Dispersion Parameter σ

This figure shows the distribution of the lognormal scale parameter σ, which determines pay dispersion within all hierarchical levels of all firms. The distribution is calculated using the bootstrap method.

E.7 Counterfactual Models

To isolate the distributional effects of hierarchy, I create three counterfactual models, each with only one income-dispersion source. This is achieved as follows:

Inter-firm dispersion only: To create this model, I set the hierarchical pay-scaling parameter (r) to 1 for all firms (removing hierarchical pay-scaling) and set the intra-hierarchical dispersion parameter (σ) to zero (removing dispersion within hierarchical levels).

Inter-hierarchical dispersion only: To create this model, I set base-level pay ($\bar{I}_1$) in all firms to an identical constant (removing dispersion between firms), and set the intra-hierarchical dispersion parameter (σ) to zero (removing dispersion within hierarchical levels).

Intra-hierarchical dispersion only: To create this model, I set base-level pay ($\bar{I}_1$) in all firms to an identical constant (removing dispersion between firms), set the hierarchical pay-scaling parameter (r) to 1 for all firms (removing hierarchical pay-scaling).

E.8 Summary of Model Structure

The model is implemented in C++ using a modular design. Each major task is carried out by a separate function that is defined in a corresponding header file. Table 8 summarizes this structure sequentially in the order that functions are called. In each step, I briefly summarize the action that is performed, giving reference to the section where this action is described in detail.

Table 8: Model High-Level Structure

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

Notes: Model code makes extensive use of Armadillo, an open-source C++ linear algebra library [28].

F The Adjusted Hierarchy Model

The hierarchy model tends to underestimate US income inequality. This may be caused by the model's reliance on Compustat Firm data (see Appendix E), which is biased towards large firms. The result is that the model likely has too little inter-firm income dispersion. Here I present the results of an *adjusted* model in which inter-firm income dispersion is increased so that the model closely reproduces US macro-level data.

As outlined in Appendix E, inter-firm income dispersion is modeled by fitting a gamma distribution to Compustat firm data. The gamma distribution has the following probability density function:

$$p(x) = \frac{1}{\Gamma(k)\theta^k} \cdot x^{k-1} \cdot e^{-k/\theta} \tag{33}$$

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

$$k_{\text{adjust}} = c \cdot k_{\text{fit}} \tag{34}$$

The adjusted model then uses the parameter k_{adjust} instead of k_{fit}. All of the model's other parameters remain constant. Note that for $c > 1$, inter-firm dispersion is *decreased* (relative to the original model). For $c < 1$, inter-firm dispersion is *increased*. I choose the value c so that the adjusted model produces the best match to US data. Model results for $c = 0.5$ are shown in Figure 13. By increasing inter-firm dispersion, we significantly improve model's fit to the body of the US distribution of income. Note that the adjusted model's Gini index is significantly higher than in the original model, and now better matches US data. Results in the tail remain virtually unchanged. (This is expected, since hierarchy shapes the tail).

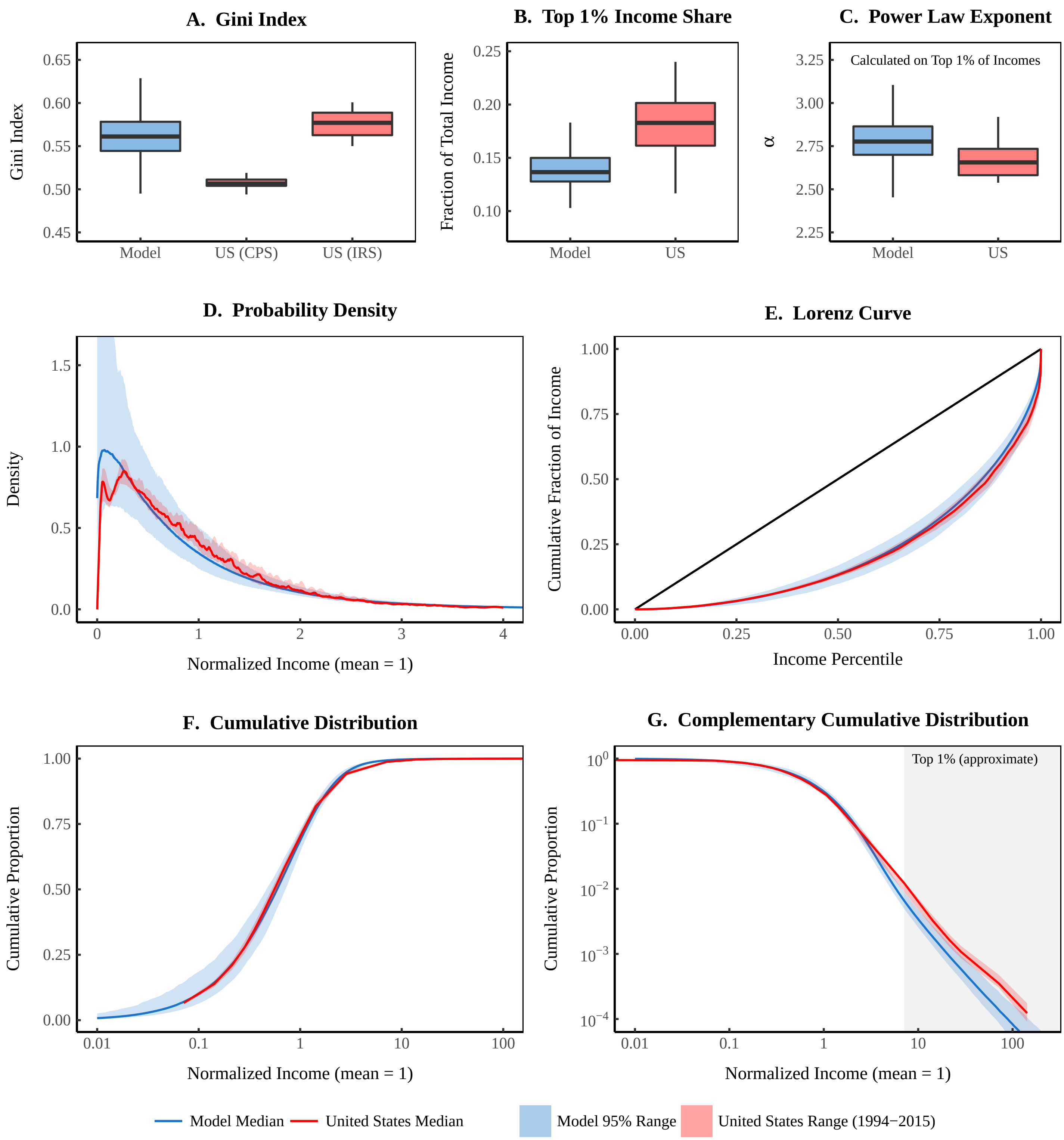

Figure 13: Adjusted Model Income Distribution vs. US Data

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

G A Null-Effect Model for US Top Incomes and Firm Size

A key prediction of the hierarchy model is that top incomes should be concentrated at the top of large institutions. To test this prediction, I look at the size distribution of firms associated with top incomes. Here I develop a null-effect model, which is what we would expect to find if there is absolutely no relation between firm membership and income. In the null-effect case, we should find that the size distribution of firms associated with *top earners* is exactly the same as the size distribution of firms associated with the general population.

To determine the null-effect we must find the size distribution of firms associated with the general population. Before doing so, some clarification is in order. What we are talking about is the size distribution of firms associated with *individuals*. As shown in Figure 14, this is quite different from the firm-size distribution. To determine the firm-size distribution, each firm is counted *once*. However, when we map firm size to individuals, each firm is weighted by the number of individuals within it. When we do this, we are really looking at the distribution of *employment* by firm size. So what is this distribution?

If we randomly select an individual from the private sector population, let $p(i_x)$ be the probability that this individual is associated with a firm of size x. This probability will determine the size distribution of firms associated with a random sample of individuals. Let $p(x)$ be the probability of randomly selecting a firm of size x from the firm population. Using Figure 14 for guidance, we can see that $p(i_x)$ is given by:

$$p(i_x) \sim x \cdot p(x) \tag{35}$$

If we know $p(x)$ — the probability distribution of firms — we can use Eq. 35 to predict the firm-size distribution associated with a random sample of individuals. Let's do so for the United States. The US firm-size distribution can be approximated by the power-law distribution $p(x) \sim x^{-2}$ (see Appendix E). Substituting this into Eq. 35 gives:

$$p(i_x) \sim x^{-1} \tag{36}$$

Because firm sizes generally span many orders of magnitude, it is more convenient to look at the log transformation of Eq. 36. Therefore, we want to know

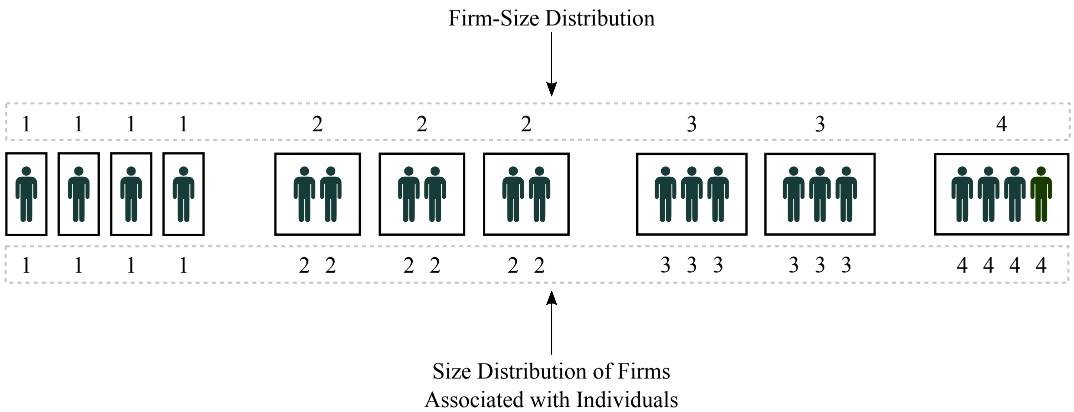

Figure 14: Mapping Firm Sizes to Individuals

This figure illustrates the mapping of firm size to individuals. Each box represents a firm, with size indicated above. The mapping of firm size to individuals appears below each firm. Let $p(x)$ be the probability of randomly selecting a *firm* of size x from the firm population. Let $p(i_x)$ be the probability of randomly selecting an *individual* associated with a firm of size x (from the individual population). Noting that each firm size x appears x times in the individual-to-firm mapping, we can state that $p(i_x) \propto x \cdot p(x)$.

---

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

$$f_y = f_x\big(\ x(y)\ \big) \cdot \big|\ x'(y)\ \big| \tag{37}$$

We let $f_y = p(\ln i_x)$ and $f_x = c \cdot x^{-1}$ (where c is constant). The transformation function is $y = \ln x$. We then note that $x(y) = e^y$ and $x'(y) = e^y$. Substituting into Eq. 37 gives:

$$f_y = c \cdot (e^y)^{-1} \cdot e^y = c \tag{38}$$

Since $f_y = p(\ln i_x)$, we can state that $p(\ln i_x) = c$, the uniform distribution. If we randomly draw a sample of individuals from the US private sector, we predict that their associated firm-size distribution will be *log-uniform*. This is the null-effect. If there is absolutely no relation between income and firm membership, we should find that the size distribution of firms associated with top incomes (in the US) is log-uniformly distributed.

H The Effect of Hierarchy on Inequality

An interesting question to ask is — what effect does hierarchy have on income inequality? In this section, I isolate the inequality effects of hierarchy using the counterfactual models of the United States. Each model contains only *one* of the three sources of income dispersion used in the original model. By comparing these counterfactual models to the original model, we can determine how each dispersion source affects income inequality.

The results in Figure 15 indicate that hierarchy's effect on inequality depends on *how* we measure inequality. When using the Gini index (Figure 15A), we see that the model with inter-firm dispersion has inequality that is closest to the original model. (The model with inter-hierarchical dispersion comes a distant second). This suggests that hierarchy does not have a particularly strong effect on inequality. However, things change drastically when we switch to measuring inequality in terms of the income share of the top 1% (Fig. 15B). Now we find that the model with inter-hierarchical dispersion has inequality that is nearly identical to the original model. The other two sources of dispersion are inconsequential. How can this be?[1]

To understand this apparent contradiction, we can look at the Lorenz curves for each model (Fig. 15C). The Lorenz curve offers a convenient way to visualize the 'shape' of inequality. The curve traces the cumulative fraction of income held by all individuals below a given income percentile. The Gini index and the top 1% income share are both intimately related to the Lorenz curve. The Gini index is proportional to the area between the Lorenz curve and the line of perfect equality (the black line in Fig. 15C). The income share of the top 1% is

[1] Some readers may note that I am using non-decomposable metrics to measure inequality. Since neither the Gini index nor the top 1% income share is decomposable, the inequality of the counterfactual models will *not* sum to the inequality of the original model. Thus we cannot quantify exactly 'how much' each factor contributes to income inequality. Although there are inequality metrics that are decomposable (such as the Theil index, or simply the variance), I choose not to use them here. For starters, such measures are generally far less intuitive than the Gini index or top income shares. Second, decomposable measures merely give *a* decomposition of inequality — not *the* decomposition. Decomposition requires deciding how to weight the number of incomes of a given size against the size of the income. Since there are many ways to do this, there are many equally valid decompositions of inequality. Anthony Shorrocks [29] summarizes the problem nicely: "Inequality comparisons are invariably sensitive to the choice of inequality index used since alternative measures tend to emphasize inequality at different points in the distribution. Replacing one index by another will therefore almost always change the relative significance of the between- and with-group terms".

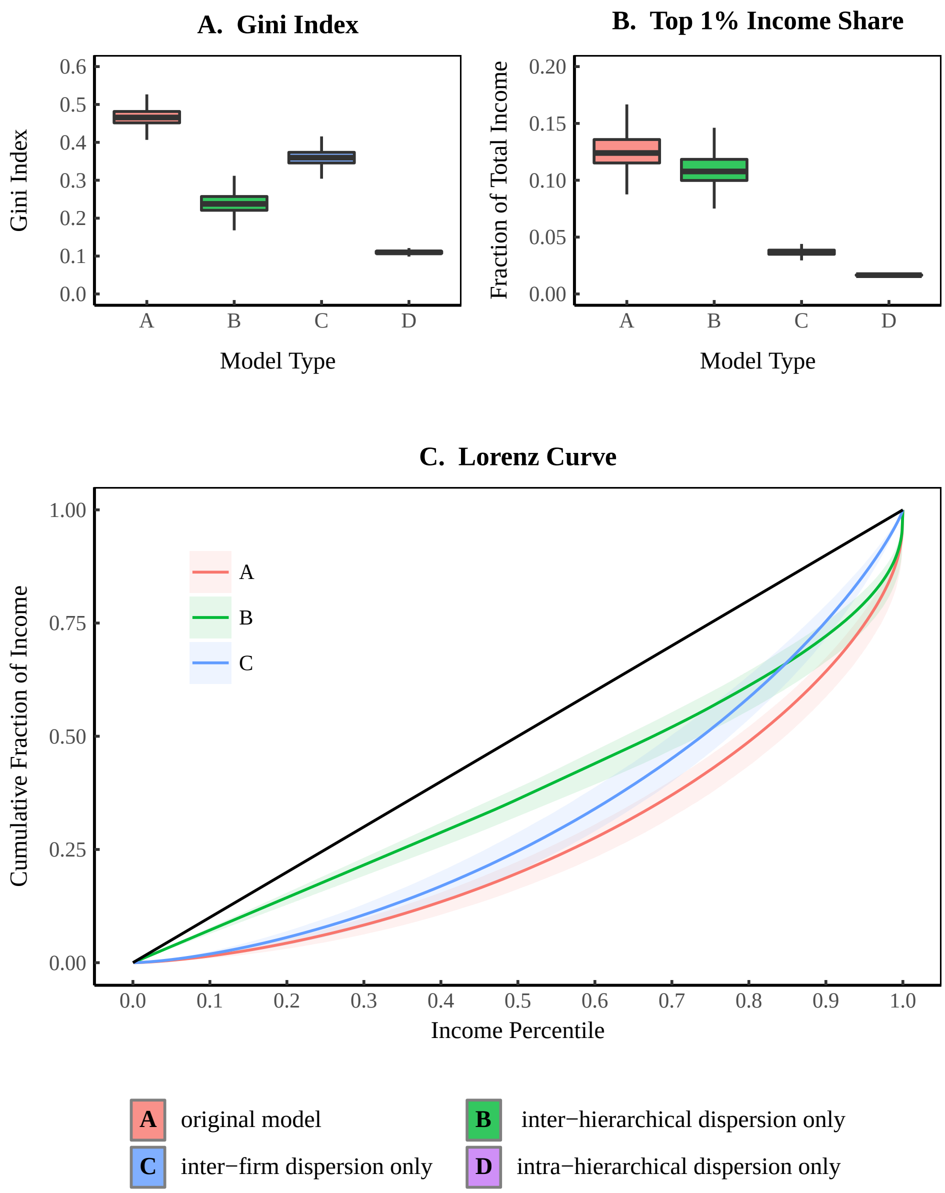

Figure 15: How Hierarchy Affects Inequality

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

equal to the vertical distance between the Lorenz curve and $y = 1$, at the point $x = 0.99$.

The apparent contradiction between the Gini and top 1% results is now easy to understand. It is caused by an *intersection* between the inter-firm Lorenz curve and the inter-hierarchical level Lorenz curve. For incomes *below* this intersection, inter-firm dispersion plays the most important role in shaping inequality. However, for incomes *above* the intersection, hierarchy plays the most important role in shaping inequality.

These results reinforce those in the main paper. Hierarchy is important for shaping the tail of the distribution (the top 1%), while dispersion between firms shapes the rest of the distribution. These results also demonstrate the pitfalls of using a single metric to quantify inequality. No single metric can capture all of the information in a Lorenz curve. The Gini index places an emphasis on the body of the distribution, while top income fraction metrics capture the dynamics of the tail. The hierarchy model suggest that when we study top income shares, we are studying the effects of firm hierarchy.

References

1. Current Population Survey. Annual Social and Economic Supplement; 2017.
2. Virkar Y, Clauset A. Power-law distributions in binned empirical data. The Annals of Applied Statistics. 2014;8(1):89–119.
3. Alvaredo F, Atkinson AB, Piketty T, Saez E, Zucman G. The world wealth and income database. Website: http://www wid world. 2016;.
4. Alvaredo F, Atkinson T, Piketty T, Saez E. The World Top Incomes Database; 2013. Available from: http://topincomes.parisschoolofeconomics.eu/#.
5. Atkinson A, Lakner C. Capital and labor: the factor income composition of top incomes in the United States, 1962-2006. World Bank Policy Research Working Paper. 2017;8268.
6. Piketty T. Capital in the Twenty-first Century. Cambridge: Harvard University Press; 2014.
7. Audas R, Barmby T, Treble J. Luck, effort, and reward in an organizational hierarchy. Journal of Labor Economics. 2004;22(2):379–395.
8. Baker G, Gibbs M, Holmstrom B. Hierarchies and compensation: A case study. European Economic Review. 1993;37(2-3):366–378.
9. Dohmen TJ, Kriechel B, Pfann GA. Monkey bars and ladders: The importance of lateral and vertical job mobility in internal labor market careers. Journal of Population Economics. 2004;17(2):193–228.
10. Grund C. The wage policy of firms: comparative evidence for the US and Germany from personnel data. The International Journal of Human Resource Management. 2005;16(1):104–119.
11. Lima F. Internal Labor Markets: A Case Study. FEUNL Working Paper. 2000;378.
12. Morais F, Kakabadse NK. The Corporate Gini Index (CGI) determinants and advantages: Lessons from a multinational retail company case study. International Journal of Disclosure and Governance. 2014;11(4):380–397.
13. Treble J, Van Gameren E, Bridges S, Barmby T. The internal economics of the firm: further evidence from personnel data. Labour Economics. 2001;8(5):531–552.

14. Ariga K, Brunello G, Ohkusa Y, Nishiyama Y. Corporate hierarchy, promotion, and firm growth: Japanese internal labor market in transition. Journal of the Japanese and International Economies. 1992;6(4):440–471.

15. Bell B, Van Reenen J. Firm performance and wages: evidence from across the corporate hierarchy. CEP Discussion paper. 2012;1088.

16. Eriksson T. Executive compensation and tournament theory: Empirical tests on Danish data. Journal of Labor Economics. 1999;17(2):262–280.

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

18. Leonard JS. Executive pay and firm performance. Industrial and Labor Relations Review. 1990;43(3):13–28.

19. Main BG, O'Reilly III CA, Wade J. Top executive pay: Tournament or teamwork? Journal of Labor Economics. 1993;11(4):606–628.

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

21. Rajan RG, Wulf J. The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. The Review of Economics and Statistics. 2006;88(4):759–773.

22. Tao HL, Chen IT. The level of technology employed and the internal hierarchical wage structure. Applied Economics Letters. 2009;16(7):739–744.

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

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

25. Gillespie CS. Fitting heavy tailed distributions: the poweRlaw package. arXiv preprint arXiv:14073492. 2014;.

26. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM review. 2009;51(4):661–703.

27. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: CRC press; 1994.

28. Sanderson C, Curtin R. Armadillo: a template-based C++ library for linear algebra. Journal of Open Source Software. 2016;.

29. Shorrocks AF. The class of additively decomposable inequality measures. Econometrica: Journal of the Econometric Society. 1980;48(3):613–625.