Wage Hierarchies

Natarajan Balasubramanian and Jagadeesh Sivadasan[§]

March 2026

Preliminary & Incomplete: Comments Welcome

Abstract

Using matched employer-employee data from the U.S. LEHD program (2004–2023), we examine the organization of human capital and wage hierarchies within firms by applying clustering techniques to worker-level earnings. We document six stable empirical regularities regarding these hierarchies: (i) the aggregate distribution of hierarchy sizes (i.e. the number of levels) is highly stable over time; (ii) hierarchy structure exhibits moderate sectoral heterogeneity, with systematically taller hierarchies in Public Administration and Education; (iii) hierarchy size follows a U-shaped relationship with firm employment, though average hierarchy size among the largest firms has slightly declined in recent years; (iv) wages are convex in hierarchy levels—particularly at the top of the hierarchy and within larger firms; (v) spans of control widen at higher management levels, a pattern amplified by firm size; and (vi) taller hierarchies are associated with a modest penalty to future firm growth and survival. To rationalize these findings, we develop and calibrate an extension of the Lucas (1978) span-of-control model that incorporates firm heterogeneity in productivity and span, alongside worker skill heterogeneity in the production layer.

Key Words: Human Capital, Managers, Span, Layers

JEL Codes: J31, L23, L25, J24

[§] Balasubramanian: Whitman School of Business, Syracuse University, nabalasu@syr.edu; Sivadasan: Ross School of Business, University of Michigan, jagadees@umich.edu. We thank the Michigan Census Research Data Center for their support and Clint Carter for assistance with disclosure requests.

U.S. Census Bureau Disclaimer: This research uses data from the Census Bureau's Longitudinal Employer-Household Dynamics program, which was partially supported by the following National Science Foundation grants SAS-9978093, SES-0339191, and ITR-0427889; National Institute on Aging grant AG018854; and grants from the Alfred P. Sloan Foundation. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2103. DRB Numbers: 12826.

1. Introduction

The question of how human capital and workers are organized within firms—specifically within hierarchies where managers set goals, coordinate tasks, and monitor subordinates—has long been a central focus in industrial organization, personnel economics and management. Foundational theoretical work dating back to Williamson (1967), Calvo and Wellisz (1978), and Garicano (2000) has emphasized the critical role that hierarchical structures play in determining firm productivity, wage distributions, and information flows. Understanding these internal structures is essential for explaining broader macroeconomic trends in organizational efficiency, wage inequality, and firm dynamics.

Despite the theoretical importance and broad interest in firm hierarchies, empirical research has historically been constrained by a lack of systematic data. The ideal dataset for this line of inquiry would consist of comprehensive organizational charts detailing exact reporting structures and earnings across a large, representative sample of firms over time. However, such detailed organizational data is largely unavailable at an economy-wide scale. Instead, the typical dataset in this context is a matched employer–employee dataset that tracks workers' earnings at their respective employers over time.

In this paper, we overcome this limitation by proposing a novel application of well-established statistical algorithms for determining the optimal number of clusters in a dataset to identify latent "wage clusters" within firms. We implement this approach using rich, worker-firm linked earnings data from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) dataset, covering the period from 2004 to 2023.

An extensive literature in statistics has developed numerous alternative approaches to determining the optimal number of clusters (e.g., see Milligan and Cooper (1985) discussion of 30 alternative approaches). To select our baseline methodology, we evaluated the performance of 35 alternative statistical approaches. Based on extensive simulation exercises designed to mimic varying levels of noise in wage distributions, we adopt two highly performing metrics for our baseline analysis: the "jump" in the transformed distortion measure proposed by Sugar and James (2003), and the maximum of the second difference in the Caliński and Harabasz (1974) pseudo-F index. We apply these clustering

techniques to the LEHD dataset—focusing on firms with at least 100 full-year employees to ensure meaningful organizational depth—to construct a robust, economy-wide mapping of wage hierarchies over a two-decade span.

Using this novel measurement approach, we document six empirical regularities regarding the structure of wage hierarchies, that remain highly stable over the 2004 to 2023 period (Section 4). First, we find that the aggregate distribution of hierarchy sizes (i.e. number of levels) is highly stable over time, characterized by a wide dispersion where there is higher density of firms with smaller number of layers, but also a fat tail of complex organizations that extends up to 25 layers. Second, we observe moderate structural heterogeneity across sectors; for instance, Public Administration and Education systematically feature taller hierarchies, consistent with deeper bureaucratic structures, compared to manufacturing or finance. Third, we document a surprising U-shaped relationship between the size of the hierarchy and total firm employment. Furthermore, we note a modest decline in the average hierarchy size among the largest firms in recent years, consistent with anecdotal evidence of corporate “delayering" (e.g., Rajan and Wulf 2006).

Fourth, we find a stable, convex relationship between wages and hierarchy levels. Wages increase slowly across the bottom layers of the hierarchy but rise exponentially toward the top—a convexity that is particularly pronounced within larger firms. Fifth, the span of control (the ratio of subordinates to managers) widens significantly at higher hierarchy levels, a pattern that is also heavily amplified by firm size. Finally, we find a modest but statistically significant negative correlation between the depth of a firm's hierarchy and its future performance. Firms with taller hierarchies experience a slight penalty to their five-year survival rates, as well as lower future employment and payroll growth.

To rationalize these empirical findings, we develop a structural model of firm organization (Section 5). In doing so, it is important to note that the “wage hierarchies” identified via our empirical clustering approach do not necessarily correspond perfectly to formal reporting structures in a control hierarchy. Early theoretical models, such as those by Williamson (1967) and Calvo and Wellisz (1978) posit a direct correspondence between hierarchical reporting levels and discrete wage levels. However, everyday experience suggests that wage differences within firms also arise from heterogeneity in

skill sets, even among workers operating within the traditional, lowest "production layer" posited by most models.

To account for this, we introduce a parsimonious extension of the Lucas (1978) span-of-control model, that allows for firm-level heterogeneity in both core productivity and span of control while incorporating skill heterogeneity within the lowest production layer itself. In the model, the total number of distinct production layers (or skill clusters) is determined by a flexible function of the firm's productivity draw. Furthermore, managers receive a share of the firm's profits proportional to their marginal product, dynamically linking hierarchical wage premiums to the firm's overall success.

Finally, we take this model to the data using the Simulated Method of Moments (SMM) in Section 6. We calibrate the model to target three primary stylized facts: empirical PDF of hierarchy sizes, the U-shaped relationship between hierarchy and firm size, and the convex aggregate wage gradient. The calibration achieves a strong fit across all targeted moments. By allowing smaller firms to retain a diverse collection of production skills, the flexible functional form of our production layer effectively replicates the U-shaped hierarchy-size relationship. Furthermore, the model accurately reproduces the steep, exponential rise in average wages at the top of the hierarchy, driven by the profit-sharing mechanism that scales with organizational depth and firm productivity.

1.1 Related Literature

The study of firm hierarchies and their impact on wage structures and firm dynamics has a rich history in both empirical and theoretical economics. This paper builds upon several key strands of this literature.

1.1.1. Empirical Evidence on Firm Hierarchies

Our work contributes to an emerging body of empirical research that utilizes detailed firm-level data to map internal organizational structures. Rajan and Wulf (2006) used a panel of over 300 large U.S. firms to document a "flattening" trend, characterized by an increase in the CEO's span of control and a reduction in the number of levels between the CEO and division heads. Wulf (2012) further observed that while firms are delayering, this often leads to more centralized decision-making at the top, contrary to popular belief about decentralization. Somewhat contrary to these studies, we document only modest

decline in the number of layers, though this decline is larger for bigger firms in line with the above papers.

Caliendo et al. (2015) used French manufacturing data to show that firms reorganize by adding layers as they grow, and that this reorganization is crucial for understanding wage dynamics—specifically, that adding layers can lead to lower average wages in preexisting layers while expanding without reorganization leads to higher wages. Exploiting a unique stratified panel of the entire Canadian economy, Kueng, Yang, and Hong (2017) establish three sets of stylized facts regarding firm life-cycles: dynamics are driven by startups rather than new establishments, organizational restructuring is tied to firm growth rather than age, and innovation patterns differ significantly between manufacturing and service sectors. In contrast to the U-shaped pattern we document for wage hierarchies, this paper finds an inverted U shape, with delayering of hierarchical layers beyond a size of 300 employees.

In recent work, Ewens and Giroud (2025) use 7 million resumes and job transitions to identify hierarchical layers in 3,100 U.S. public firms. They find a positive correlation between number of layers and firm size, and that "taller" firms (averaging ten layers) correlate with higher education and lower volatility but higher costs. Our LEHD-based study utilizing high-frequency wage data complements their work which exploits "structural" hierarchy via titles in resume data, and provides an "economic" mapping via wage clusters across all sectors. Our finding of a weak negative link between wage hierarchy depth to long-term firm survival and growth in our sample contrasts somewhat with the positive attributes of taller firms documented in this work.

Overall, our paper complements these studies by using detailed wage data from a significantly larger and more diverse longitudinal dataset (LEHD) covering the 2004–2023 period. While previous studies often focused on top management or specific sectors like manufacturing, our use of clustering approaches across all sectors allows us to document stable, long-term patterns in wage hierarchies, spans of control, and their relationship to firm survival and growth across the entire economy.

1.1.2. Theoretical Foundations of Firm Hierarchy and Income Distribution

Our work extends the seminal "span of control" and hierarchy models. In particular, our model builds directly on Lucas (1978), who established a foundational model for the size distribution of firms based on the distribution of managerial talent.[1] We allow for multiple layers in the organizational hierarchy with limited span forcing firms to adopt multiple management layers. Calvo and Wellisz (1978, 1979) provided early theoretical frameworks explaining how the need for supervision leads to hierarchical wage differentials and loss of control, showing that employee quality and wages naturally increase with their position in the hierarchy. In more recent work ,Garicano (2000), Caliendo and Rossi-Hansberg (2012) and Caicedo, Lucas, and Rossi-Hansberg (2016) have modeled hierarchies as a way to process knowledge and solve problems, linking organizational structure directly to productivity and the distribution of wages.

Our contribution to the modeling front is a parsimonious extension of the Lucas (1978) model that accommodates multiple hierarchical layers. By incorporating firm heterogeneity in both productivity and span, and assuming managers receive a share of profits proportional to their marginal product, we generate predictions that align with our documented empirical facts—specifically the convex relationship between wages and hierarchy levels and the U-shaped relationship between hierarchy size and firm employment. We show that incorporating heterogeneity in skills in the production layer is important for reconciling two key empirical facts: large fraction of firms with tall wage hierarchies, and allowing this heterogeneity to vary by firm productivity allows matching the U-shaped relation between firm size and hierarchy size.

2. Data

We utilize the Longitudinal Employer-Household Dynamics (LEHD) at the US Census Bureau as our primary data source. Detailed documentation of the data is available in Graham et al. (2022).[2] This program synthesizes quarterly wage data provided by

[1] Lucas (1978) models firms as two layers (One manager and multiple production workers) and suggested the need for more complex "technologies for managing managers" to capture actual business hierarchies.
[2] The documentation is available online at: https://lehd.ces.census.gov/data/lehd-snapshot-doc/latest/index.html

employers to state unemployment insurance (UI) offices, and combines it with detailed worker and employer information from other Census surveys (see Abowd et al. 2009). The underlying dataset offers robust coverage, representing about 95% of private-sector jobs and most state and local government employment across all 50 states (Card, Rothstein and Yi, 2024). The data does not track federal employees, members of the armed forces, or independent contractors/self-employed workers.

Availability of restricted access microdata for researchers is dependent on each state's participation in the LED Partnership through a Memorandum of Understanding (MOU). We utilize data from 30 states for which we have authorized access.[3]
From among the LEHD files, we use the Job History Files (JHF) to estimate annual wages at a worker-job level, and the Employer Characteristics File (ECF) to obtain industry classification for the employer. As indicated in the documentation for the Quarterly Workforce Indicators which draw on the same underlying UI data, earnings "[i]nclude gross wages and salaries, bonuses, stock options, tips, and other gratuities, and the value of meals and lodging, where supplied."[4]

For our baseline analysis, we define a firm based on the state-level employer identifier number (SEIN) in the JHF. As explained in the LEHD documentation, this identifier typically captures all of the activity of a firm within a state, and has been used as the unit of employer observation in several studies in the literature (e.g., Haltiwanger, Hyatt and McEntarfer 2018).

---

[3] The states in our sample include Alabama, Arizona, Colorado, Connecticut, District of Columbia, Delaware, Hawaii, Idaho, Illinois, Kansas, Louisiana, Maine, Maryland, Nebraska, Nevada, New Jersey, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Utah, Virginia, Washington, Wisconsin, Wyoming. The first available year for JHF data is 2004 for Arizona, and 2006 for the District of Columbia. For the other 28 states, we have data covering the 2004, 2009, 2014, 2019 and 2023 years we use in our analysis.

[4] Source: https://lehd.ces.census.gov/doc/Metadata4_QWI.htm . The documentation notes that: "earnings measures described here [include] gross wages and salaries, bonuses, stock options, tips, and other gratuities, and the value of meals and lodging, where supplied…[and does not]… include Old Age Survivor and Disability Insurance, health insurance, workers compensation, unemployment insurance, private pensions, and welfare funds.

3. Methodology, Sample Definitions and Key Variables

3.1 Selecting Optimal Number of Clusters

Our goal is to use the rich worker-firm level wage data to detect wage hierarchies (clusters) within firm. We compared the performance of 35 alternative approaches to selecting the optimal number of clusters under a number of alternative simulation scenarios, discussed in detail in Online Appendix A. Based on this evaluation, our baseline analysis uses two alternative approaches that have high scores for goodness of fit between predicted and actual number of clusters across our simulations.

The first approach is based on identifying a "jump" in the transformation of a distortion measure, as proposed by Sugar and James (2003). In our one variable context, the distortion measure is the Mahalanobis distance between X and the centers of candidate clusters:

$$d_K = \min_{c_1,\ldots,c_K} E[(X - c_X)\Gamma^{-1}(X - c_X)]$$

We estimate $\hat{d}_K$ using k-means clustering approach to the wage data. Because we use the identity matrix for Γ, the distortion is simply the mean squared error. Sugar and James (2003) propose selecting optimal number of clusters $K^*_{jumpdist}$ based on the largest "jump" in a transformation of d_K. Following suggestion in Sugar and James (2003) we use a transformation power $Y = \frac{p}{2} = \frac{1}{2}$, as we work with p=1 dimensional data. That is:

$$K^*_{\text{jumpdist}} = \arg\max_{\text{K}} J_K = \arg\max_{\text{K}} \{\hat{d}_K^{-0.5} - \hat{d}_{K-1}^{-0.5}\}$$

The second approach defines optimal number of clusters as the maximum of the second difference in the Caliński and Harabasz pseudo-F index (Caliński and Harabasz, 1974). This index is defined for data with K clusters and N observations as:

$$F_K^{CH} = \frac{\left[\frac{\text{trace}(\mathbf{B})}{K-1}\right]}{\left[\frac{\text{trace}(\boldsymbol{W})}{K-1}\right]}$$

We pick optimal number of clusters as:

$$K^*_{\text{dcalinski}} = \arg\max_{\text{K}}(d.F_{K+1}^{CH} - d.F_K^{CH}) = \arg\max_{\text{K}}\{(F_{K+1}^{CH} - F_K^{CH}) - (F_K^{CH} - F_{K-1}^{CH})\}$$

These metrics for an illustrative simulation with eight clusters is presented in Figure 1. The figures in the bottom panel show that both the approaches pick out the optimal number of clusters well.

The default approach in some statistical software (e.g., Stata) based on the Caliński and Harabasz pseudo-*F* index suggests picking the K with maximum value for the index. As Figure 1 shows, for our one-dimensional simulation exercise, the index value appears to increase with K, so that the maximum value is reached close to the largest candidate K value. This is also reflected in our simulation evaluations in Appendix A, especially for simulations with low noise (i.e., clearly separated clusters). However, the approach using the maximum of the Caliński and Harabasz pseudo-*F* performs well in high noise scenarios. For this reason, we present robustness checks using this alternative approach to picking the optimal number of clusters.

3.2 Sample Selection and Key Variables

We apply the approaches described to Section 3.1 to worker-level wage data within firms. First, we use K-means clustering to cluster the data for each firm for each year into clusters ranging 2 to 25. We then apply the baseline approaches described above to select the optimal number of clusters in the wage data. Note that for firms with employment $\leq$ 25, the optimal cluster size would trivially equal the number of employees in the firm. Thus, meaningful clustering of wages is better detected for larger firms. Hence, we limit our analysis to those firms with at least 100 full year employees. We restrict to full year employees as those that exit or enter during the year would have low annual wage numbers that could distort the clustering away from the intended goal of capturing hierarchical levels. We define full year employees as those that began their current job spell prior to the start of the year, and ended the job spell after the end of the year. To capture changes in patterns of wage hierarchies over time, we examine data for five years spanning two decades, namely 2004, 2009, 2014, 2019 and 2023.

The primary variable is the annual wage of the worker, which is measured as the sum of the quarterly wages reported in the LEHD JHF files. Per the BLS documentation for the Quarterly Census of Employment and Wages (QCEW) which uses the same underlying unemployment insurance data, the reported wages "are the total compensation paid,

including bonuses, stock options, severance pay, profit distributions, the cash value of meals and lodging, tips and other gratuities, and, in some states, employer contributions to certain deferred compensation plans (such as 401(k) plans), during the calendar quarter, regardless of when the services were performed."[5]

4. Stylized Facts

We examine six characteristics of the wage hierarchy, described below. Throughout, to conform to US Census disclosure restrictions, we summarize relationships using polynomial functions. Further, to characterize relationships for firms with different hierarchy sizes, we normalize the hierarchy measure within firm-year.

4.1 Fact 1: Stable Distribution of Hierarchy Sizes

We begin by examining the distribution of hierarchy sizes in the data. Our explorations showed that the distribution across wage hierarchy sizes from 2 to 25 is well-summarized by an order 5 polynomial. The implied distributions are presented in Figure 2, and the corresponding regression results are in Online Appendix Table B1. The top panel of Figure 2 shows the overall distribution using the jump in distortion metric and the second difference of the Caliński metric. The panel on the bottom plots the distributions for each of the years.

The figure indicates a relatively flat distribution of hierarchy sizes, suggesting a large variation across firms. The patterns are similar across the two baseline metrics, and the drop-off in density at higher values suggest that our upper limit of 25 clusters is a reasonable one.[6]

Across all of our metrics, the distribution of hierarchy sizes shows remarkable stability across the years. This is noteworthy, considering that our methodology examines wage data for each firm in each year separately. There is suggestive evidence for a small decline in the density for large hierarchy sizes. While this is consistent with some of the

[5] https://www.bls.gov/opub/hom/cew/concepts.htm

[6] The robustness check using the maximum of the Caliński- metric (Appendix Figure B1) shows a greater accumulation of firms at the two tails of the distribution. As illustrated in the simulation in Figure 1 and confirmed in more extensive simulations we undertook (described in Appendix A), the Caliński metric tends to increase with the number of clusters, so that using the maximum of this metric often erroneously picks up the largest candidate number of clusters as the optimal number.

literature (e.g., Rajan and Wulf 2006, Wulf 2012), the reductions we find are very modest, suggesting there has not been a widespread impact of technological change on the clustering of wages within firms.

4.2 Fact 2: Stable Variation in Hierarchy Sizes across Sectors

In Figure 3, we present the average hierarchy size by sector, using each of the two baseline approaches, for years 2004 and 2023. These figures reveal two regularities. First, across different clustering approaches, the rank ordering of sectors by average hierarchy size is very similar, with manufacturing (NAICS 32 and 33) and Finance and Insurance (NAICS 52) being the sectors with the lowest average cluster size, while Public Administration (NAICS 92) and Education (NAICS 61) are the sectors with the highest number of clusters. The higher number of clusters for large firms in public administration is consistent with more layers of bureaucracy in this sector.

Second, consistent with Figure 2 discussed in Section 4.1 above, there is notable stability in the patterns over the nearly two-decade time frame. The inter-sector variation remains stable, and the average hierarchy size also shows little variation over time.[7]

4.3 Hierarchy Size vs Firm Size

Existing models in the literature (e.g. Caliendo & Rossi-Hansberg 2012; Williamson 1967) predict a strong positive correlation between firm size and hierarchy size. In most models, increasing the span of control reduces effectiveness, so that an increase in size of the firm necessitates additions of hierarchical layers.

We examine the cross-sectional relationship between firm size and hierarchy size in Figure 4. Two salient features stand out. One, surprisingly the relationship between hierarchy size and firm size is not monotonic. Across methodologies and time periods, the relationship is U-shaped, with larger number of clusters for smaller firms and the smallest number of clusters around the 70th to 80th percentiles of the firm size distribution, and an increase thereafter. Two, while the relationship between hierarchy size and firm size is remarkably stable across decades for the bottom half of the size distribution, there is a distinct decline in hierarchy size for the largest firms. This decline aligns with findings in

[7] Figure C2 documents similar patterns using the maximum of the Caliński metric as the methodology to select optimum number of clusters.

some of the literature (e.g., Rajan and Wulf 2006). But we note that the decline is relatively modest, at about 0.5 layers for the largest size firm corresponding to less than 5% decline in hierarchy size.[8]

4.4 Wages across the Hierarchy

A key strength of the LEHD data is the availability of rich earnings data at the individual level, allowing us to examine how the wages vary across the hierarchy level, and how this relationship has changed over the years. Because the number of hierarchy levels varies across firms, it is not straightforward to present the variation across hierarchy levels. As a baseline examination, we summarize the wage variation across levels by "standardizing" the hierarchy level variable within firms, and then regressing wages on a cubic polynomial of the standardized hierarchy level.

Figure 5a summarizes the relationship between wages and hierarchy levels across different time periods. We first estimate the average wage within each firm-year-hierarchy level. We then regress this average wage on a cubic polynomial of the "standardized" hierarchy level variable, separately for each year. The top panel of Figure 5a plots the predicted average wage from this polynomial regression, while the bottom panel presents the log of the predicted average wage. We note two regularities in the data. First, across alternative metrics and across time periods, there is a convex relationship between hierarchy levels and wage levels, consistent with the literature (e.g., Chapter 6 of Lazear and Gibbs (2014) presents a convex wage curve using internal data from one organization). Even in log terms (bottom panel), there is some evidence for a steeper slope over the top half of the hierarchy, suggesting movement up is accompanied by greater percentage increases in earnings as one moves up firm a firm's hierarchy.

Second, in the data the wage levels appear to show greater convexity in earnings in later years, suggesting that wages increase more in level terms positions at the top of hierarchies (top panel). However, the bottom panel suggests that the variation in log earnings across the hierarchy experienced largely parallel shifts over the years, consistent with similar percentage changes in earnings across the hierarchical levels, beyond the

[8] Results are similar when the maximum of the Caliński metric as the methodology to select optimum number of clusters, as documented in Figure C3.

lowest initial levels. The smaller relative shifts in log earnings at the lowest levels are noteworthy; by construction, a significant share of the firm's total employment is concentrated at the lowest levels of the hierarchy. Thus, these results are consistent with slower wage growth for a significant portion of workers in the lower parts of firm hierarchies. Nevertheless, it is noteworthy that for employees above the lowest few rungs of the hierarchy, the percentage increase in earnings have been remarkably similar.[9]

Summarizing the variation of wages across the hierarchy using the standardized hierarchy level could hide variation in the nature of this relationship by hierarchy and firm size. For example, if smaller hierarchy size is a consequence of a larger span of control of managers and if effort costs for managing a larger span are higher, we may expect manager effort and hence compensation to rise more steeply in firms with shorter hierarchies. Similarly, consistent with the documented higher wages for larger workers, it could be the case that the wage-hierarchy level relationship is steeper or more convex for larger firms.

We explore this using a regression of the average wage within firm-year-hierarchy level on a cubic polynomial of the "standardized" hierarchy level variable, with the terms of the polynomial interacted with tercile indicator variables for hierarchy size (top panel) and firm employment size (bottom panel). As surmised above, we find that the wage-hierarchy relationship is steeper for firms with shorter hierarchies (top panel), and for larger firms (bottom panel).

4.5 Span of Control across the Wage Hierarchy

In this section, we examine how the span of control – defined for layer j as the number of employees in layer $j-1$ divided by the number of employees in layer j – varies across the hierarchy. Theoretically, models may assume a constant span (e.g., Williamson 1967), while other modeling assumptions could lead to increasing or decreasing spans as we move up the hierarchy.

Figure 6a presents the variation of span across the hierarchy, for each of the five years in our sample. Following the approach in Section 4.4, we first estimate the average span within each firm-year-hierarchy level. We then regress this average span on a cubic

[9] Results using the maximum of the Caliński metric as the methodology to select optimum number of clusters are similar (Figure C4).

polynomial of the "standardized" hierarchy level variable, separately for each year. The predicted line of fit, separately for each year is plotted in Figure 6a. We find that the span increases as we move up the hierarchical levels reaching around 3.5 subordinates per employee at the highest levels.[10] There is very little variation in the average relationship between span and standardized hierarchy across the almost 20-year period between 2004 and 2023. The figures are almost identical across the two baseline alternative approaches to identify the optimal number of clusters.

Next, paralleling the analysis in Section 4.4, we explore the heterogeneity in the relationship between average span and hierarchy level by size of the hierarchy and firm size. Specifically, we first estimate the average span within each firm-year-hierarchy level. We then regress this average span on a cubic polynomial of the "standardized" hierarchy level variable, with the terms of the polynomial interacted with a hierarchy size (top panel of Figure 6b) and firm employment size (bottom panel of Figure 6b) tercile indicator variables. As could be expected, we find that span, especially beyond the median hierarchy level, is much higher for firms with shorter spans (top panel), and for larger firms (bottom panel).[11]

4.6 Correlation of Hierarchy Size with Future Firm Performance

Finally, we explore the correlation of firm hierarchy size with future firm performance using the following regression specification:

$$y_{it+5} = \alpha N_{it}^{h} + f_{it} + \epsilon_{it}$$

[10] Curiously, we find span to be less than 1 at the very lowest levels of the hierarchy. This suggests the workers in these lowest levels are not necessarily within the traditional organizational structure studied in the literatures, as suggested by our analysis of the title-linked salary data for one organization (Appendix Section A.3). The data there suggested that the lowest layer consists of support staff, who are not engaged with the primary objectives of the organizations, so the number of workers in different wage groups at the lower levels may include workers with specific support skills where the wages are determined by labor market for those skills.

[11] The need for larger spans in firms with flatter hierarchy is consistent with anecdotal evidence. For example, the HBS Case study on Lincoln Electric company notes: "Lincoln's organizational hierarchy was flat, with few levels between the bottom and the top. For example, Don Hastings, the vice president of sales, had 37 regional sales managers reporting to him." This also translates to extra effort for the managers, so that higher wages observed at higher hierarchical levels for flatter/larger firms is not surprising. A quote from Don Hastings in the Lincoln Electric Case is illustrative: "I have to work hard, there's no question about that. There are only four of us in the home office plus two secretaries. I could easily use three more people. I work every Saturday, at least half a day. Most of our regional men do too, and they like me to know it."

where y_{it+5} denotes one of three outcome variables for firm i in year t defined as follows: (i) Five-year survival dummy, defined as equal to one if the firm (specifically the SEIN) exists in the dataset in year t+5; (ii) Five-year future employment growth defined as $(E_{it+5} - E_{it})/[0.5 \times (E_{it+5} + E_{it})]$; (iii) Five-year future payroll growth defined as $(P_{it+5} - P_{it})/[0.5 \times (P_{it+5} + P_{it})]$. For the growth variables, firms that exited imputed with zero employment/payroll, so exit translates to a growth rate of negative 2. The main parameter of interest is α,the coefficient on N_{it}^{h} which denotes the f_{it} denotes a rich set of fixed effects.

In the baseline specification reported in Table1, we include interaction of state, 3 digit NAICS, and year as the fixed effect, which means we compare firms to its peers in the within the same industry and state. The results suggest a consistent and significant negative correlation between hierarchy size and future performance. While this negative correlation is consistent with theoretical models and empirical anecdotes suggesting that additional organizational layers can slow down decision making and distort information transmission within firms, we want to note that the size of the effects are very modest.[12] As the bottom two rows in Table 1 show, the estimated effect of a one standard deviation increase on number of layers (corresponding to an addition of seven layers in the wage hierarchy) are only around - 0.37% of the mean for five year survival, and about -0.15% of the mean for employment and payroll growth. The corresponding effects in terms of the standard deviation of the dependent variables range from about 1.2% (for survival) and around 1.7% to 2% for the growth variables.

One strong concern about the interpretation of the effects in table 1 relates to whether the state-industry-year effects address potential omitted variables than can impact the estimates. In particular, one concern is about how initial firm size may be an important omitted variable. As we documented above, the relation between hierarchy size and firm size is U shaped so the overall correlation to firm size is unclear. Nevertheless, it is

[12] Theoretical models going back to Williamson (1967) note a possibility of loss of control across layers of the hierarchy. Williamson (1967 p135) notes " an irreducible minimum degree of control loss results from the simple serial repro- duction distortion that occurs in communicating across successive hierarchical levels. If, in addition, goals differ be- tween hierarchical levels, the loss in control can be more extensive." Relatedly, Downs (1966, p118) noted the possible cumulative "leakage of authority" in bureaucracies as officers in each layer modify information flowing up and down the hierarchy by exercising their discretion.

known that small firms have lower survival rate, and it could be that the largest firms experience flat growth or negative growth from reversion to the mean effects (cite Haltiwanger paper AER?). To address this concern, we add interaction with size-quintile group fixed effects. Further, we include interaction with indicator variable that captures whether the firm (SEIN) comprised of multiple units, as it has been documented in the literature than multi-unit establishments experienced greater survival rates. Specifically, the regressions presented in Table C1 includes interacted fixed effects for State × NAICS 3 Digit × Employment Quintile × MultiUnit Dummy. Thus we compare firms to peers within the same initial size quintile, within the same 3-digit industry and state, with the same structure in terms of multi or single unit status. The coefficient estimates and hence the implied effect sizes are very similar to those found in Table 1.

Thus, a robust conclusion from this analysis is that there is only a very small negative correlation of firm hierarchy sizes with future firm performance, consistent with firms largely appearing to choose hierarchy sizes that are optimal for their context and consistent with their other strategic choices.

5. Theoretical Framework

In this section, we develop a theoretical model that builds on the span-of-control tradition (Lucas, 1978; Williamson, 1967) to explain some of these findings. We incorporate key modifications motivated by the empirical regularities documented in Section 4. Specifically, we posit a **Hybrid Hierarchy Model**: a combination of multi-layered management control and a production layer characterized by heterogeneous skill bundles.

5.1 Consumer Demand

We assume a representative aggregate consumer with Cobb-Douglas preferences over sectors $j \in \{0, \dots, J\}$ a la extensions of the standard Melitz (2003) model.[13]

$$U = \sum_{j=o}^{J} \beta_j \log Q_j, \qquad Q_j = \left[\int_{\omega \in \Omega_j} q_j(\omega)^{\rho} d\omega \right]^{\frac{1}{\rho}}$$

[13] The setup here draws on https://tinyurl.com/AEA2019-Trade2, Lecture 5: Some Empirical Facts on Firms and Trade (Marc Melitz), pp 42.

where: $\rho = \frac{\sigma-1}{\sigma} < 1, \sigma > 1$ is the elasticity of substitution, and $\sum_j \beta_j = 1$. Sector $j = 0$ is a homogenous sector and used as the numeraire good, i.e., has $P_o = 1$.

Given aggregate income Y, consumers on spend $E_j = \beta_j Y$ on sector j goods. Within sector j, demand is:

$$q_j(\omega) = A_j p_j(\omega)^{-\sigma}$$

Or:

$$p_j(\omega) = \left(\frac{1}{A_j}\right)^{-\frac{1}{\sigma}} q_j(\omega)^{-\frac{1}{\sigma}} = B_j q_j(\omega)^{-\mu}$$

where:

$$A_j = E_j P_j^{\sigma-1}, \quad B_j = \left(\frac{1}{A_j}\right)^{-\frac{1}{\sigma}} = (A_j)^{\mu} = E_j^{\mu} P_j^{1-\mu}, \mu = \frac{1}{\sigma} \quad P_j = \left[\int_{\omega\in\Omega_j} p_j(\omega)^{1-\sigma} d\omega\right]^{\frac{1}{1-\sigma}}$$

5.2 Firm production with hybrid hierarchy

Here we will elaborate on one sector, and for simplifying exposition, we will drop the sector subscript *j*.

Each firm corresponds to one variety ω, obtains a productivity draw φ from a distribution $F(\varphi)$. In addition to this productivity draw, the firm's output depends on an organizational hierarchy described below.

Production requires a strategic plan crafted by leaders in the topmost layer, which is a (single) CEO in this model. This plan is communicated and explained to subordinates in the next (lower) level of the hierarchy, who then communicate it further down the hierarchy.

Drawing on the literature (going back to at least Williamson 1967), physical production is executed entirely by workers at the foundational layer of the firm. However, this foundational layer is not a homogenous mass; it requires an aggregation of K distinct types of workers with different sets of skill. Because these workers command different wages (as compensation for effort and time incurred in obtaining and maintaining their skills), the production layer consists of a distinct set of K wage clusters .

Thus, in terms of wage clusters, there are two separate hierarchies. One is a control hierarchy composed of n layers of managers who oversee the production workers in the

lowest level of this hierarchy. Without loss of generality, the workers in the "production hierarchy" can be ordered by wage, so that workers with the highest wages are at the top of the production hierarchy, which we describe briefly below.

1.1.3. Production Clusters/Hierarchy

To capture the complexities of production in modern firms that require workers with a diverse set of skills in a parsimonious and tractable way, we posit that heterogenous workers get combined into apparently homogenous intermediate production labor units l through a Leontief (fixed proportions) technology:

$$l = \min\left\{\frac{l_1}{\alpha_1}, \frac{l_2}{\alpha_2}, \dots, \frac{l_K}{\alpha_K}\right\}$$

Cost minimization of this Leontief function yields labor demand for particular skill types as $l_k = \alpha_k l$, and a marginal cost of the intermediate labor of $\overline{w} = \sum_{k=1}^{K} \alpha_k w_k$.

In other words, the Leontief assumptions, yields a setup where we have l actual production workers, who comprise of K types of workers, with wage for worker of type k equal to w_k. We assume this wage is function of the effort and time investment in obtaining and maintaining those skills (Caliendo and Rossi-Hansberg, 2011), so that we have:

$$w_k = w_o f(k)$$

where w_o is the wage for a worker with the baseline (minimal) skill level, and $f(k)$ captures the compensating differential for the escalating effort and knowledge required to execute higher-complexity tasks within the production process. Then the firm's total cost for the base production layer is:

$$\text{Total Production Layer Cost} = \sum_{k=1}^{K} w_k l_k = \left[\sum_{k=1}^{K} w_k(\alpha_k l)\right] = l\left[\sum_{k=1}^{K} \alpha_k w_k\right] = l.\overline{w}$$

This formulation is consistent with abstracting from the significant complexities associated with understanding different configurations of skills into teams managed by different superiors. Our setup implies that even though there are heterogenous workers, the teams allocated to individual managers in the lowest layer of the management hierarchy is homogenous. So in constructing the optimal number of layers of the control hierarchy (discussed in the next section), the production layer can be viewed as l homogenous workers, each with an average wage of w.

We allow the number of skill levels in the production hierarchy to flexibly vary with the firm's fundamental capabilities, such that $K = K(\varphi, s)$. In particular, highly productive firms or those with wider spans of control may require a broader, more specialized set of production skill levels. Because the production wage schedule is progressive, this implies that both the average production wage and the maximum production wage (the baseline for managerial compensation) are firm-specific, such that $\overline{w} = \overline{w}(\varphi, s)$.

1.1.4. Management Hierarchy

The productivity of managers cascades down the control hierarchy. Specifically, the CEO's talent θ_n impacts all of the s subordinates in layer $n-1$. Similarly, the talent of manager in level $n-1$ impacts all of her specific subordinates in layer $n-2$, and so on. The output of any single composite production worker i depends multiplicatively on the firm productivity φ, on his own productivity (θ_i^1, normalized to one and assumed to be the same across all of the composite production workers), and on the productivities of his manager, and that manager's manager and so on, up to the top most layer of the hierarchy:

$$q_i = \left[\theta_1 \prod_{m=2}^{n} \theta_m^{h_i}\right] \varphi = \left[\prod_{m=2}^{n} \theta_m^{h_i}\right] \varphi \quad (1)$$

Where h_i denotes the hierarchical path/reporting line from worker i at the bottom of the hierarchy to the top, so that $\theta_m^{h_i}$ denotes the talent/productivity of manager in layer m who is in the reporting line from worker i to the CEO.

Because manager talent is involved in translating firm productivity φ to subordinates, we can assume that higher talent makes the transmission better, but there is some loss of effectiveness across layers. That is, we assume for $2 \leq m \leq n, 0 < \theta_m \leq 1$. This assumption ensures that all else equal, adding another layer of hierarchy cannot increase the output of the production worker (e.g., Williamson 1967).

Each firm has span s, drawn from a distribution $G(s)$ which determines the maximum number of workers reporting to any manager at any level. The distribution of firm span and the firm productivity are potentially correlated, so $(\varphi, s) \sim H(s, \varphi)$.

At this stage, we assume all production workers have productivity θ^1, and that the productivity in sub-teams is symmetric.[14] That is, assume that for any layer m, the productivity of subordinate j in layer $k-1$ reporting to a manager in layer k is the same independent of the manager. Then we can write output of the firm as:

$$q = \left[\theta_n \prod_{m=n-1}^{1} \left(\sum_{j=1}^{s} \theta_{m_j}\right)\right] \varphi = \left[\prod_{m=n-1}^{2} \left(\sum_{j=1}^{s} \theta_{m_j}\right)\right] s\theta_n \varphi \quad (2)$$

For tractability, we further assume that all managers have same productivity level, i.e., $\forall m, j, \theta_{m_j} = \theta_n = \theta$. This then yields output of the firm is:

$$q = \theta^{n-1} s^{n-1} \varphi = \theta^{n-1} \varphi l = C(\theta, n) \varphi l \quad (3)$$

where the number of production workers $l = s^{n-1}$ and $C(\theta, n) = \theta^{n-1}$.

5.3 Firm profit maximization and optimal number of layers of control hierarchy

Production workers are paid an average wage $\overline{w}$. In our setup, each firm is characterized by a unique bivariate draw of the firm productivity and span (φ, s). We assume that managers earnings consist of a fixed baseline wage $(\widehat{w})$ plus a share of the profits. We assume that this baseline wage $\widehat{w}$ is set equal to the top wage in the production hierarchy so $\widehat{w} = w_K$ for a firm with K production layers. This base pay can be considered as the wage the managers would receive in case they did not get employed as managers but rather in the production layer.[15] As in Caliendo and Rossi-Hansberg (2012), we assume a per period fixed operating cost, $w_o F_o$. This yields net profit for the firm :

[14] Later, when we develop the production hierarchy, we will allow for heterogeneity in production workers.

[15] "Implicitly, we make the assumption that manager effort and knowledge match those of the workers at the top of the production hierarchy. While empirical wage profiles are visually smooth due to aggregation across heterogeneous firms and the presence of "dual career ladders" (where highly specialized individual contributors potentially overlap in compensation with middle management), we abstract from these overlaps for parsimony. By enforcing a strict sequence where the control hierarchy sits directly above the production hierarchy, the model remains tractable while capturing the fundamental transition from potentially linear, skill-based compensating differentials at the bottom to a potentially convex, span-of-control premiums at the top. Furthermore, while we assume all profits are distributed to the managers, the key results regarding the wage profile are expected to go through even if a fraction of profits were reserved for shareholders in a model extended to include outside claimants. Finally, note that because we allow K to vary with (φ, s), the baseline managerial wage also varies with firm capabilities, such that $\widehat{w} = \widehat{w}(\varphi, s)$.

$$\Pi(\mathrm{n}) = pq - \bar{w}l = Bq(\varphi, s)^{1-\mu} - \bar{w}l - \hat{w}N^m - w_o F_o$$

$$= B[\theta^{n-1}\varphi]^{1-\mu}(s^{n-1})^{1-\mu} - \bar{w}(s^{n-1}) - \hat{\mathrm{w}}\left[\frac{s^{n-1}-1}{\mathrm{s}-1}\right] - w_o F_o$$

$$= D[\theta^{n-1}]^{1-\mu}(s^{n-1})^{1-\mu} - \left[\bar{w} + \frac{\hat{w}}{s-1}\right](s^{n-1}) + \frac{\hat{\mathrm{w}}}{s-1} - w_o F_o \qquad (4)$$

where $D = B\varphi^{1-\mu} = (E^{\mu}P^{1-\mu})\varphi^{1-\mu}$, and $N^m = \left[\frac{s^{n-1}-1}{s-1}\right]$ is the number of managers in the firm. We define $w_{\text{eff}} = \left[\bar{w} + \frac{\hat{w}}{s-1}\right]$. This can be considered the effective marginal cost of adding a layer: when the firm adds a layer, the cost per worker added is the production wage $\bar{w}$, and there is an additional cost $\left(\frac{1}{s-1}\right)$ fraction of a manager required to oversee them.[16]

Comment: Our setup leads to three effects of increasing the number of layers in the hierarchy. One, which is intuitive and obvious, is that it increases the number of production workers that can be managed, which facilitates higher output and profits . A second effect here follows from our assumptions of the multiplicative accumulation of manager capabilities. This implies that, all else equal, the effective productivity of a hierarchy is affected by adding layers, as the per worker output depends on θ^n. Whether an additional layer increases or decreases productivity depends on whether $\theta > 1$ or $\theta < 1$. In the general case, we may think that there some dilution of the firm productivity as it gets transmitted down the hierarchy, so $\theta < 1$ (e.g., Williamson 1967, Downs 1966). This cumulative "loss of control" diminishes firm productive as layers are added to the hierarchy. A third effect is that adding layers increases overhead costs, from additional managers required to oversee the production. This third effect is noteworthy. Typically if managers are not involved in production, their compensation would be part of "overheads" or fixed costs that would be excluded from the optimal choice of production quantity. In our setup, production quantity is linked to number of layers, which also impacts the number of managers needed, so manager compensation factors in choice of the optimal number of layers.

[16] E.g., if the firm moves from n to $n+1$ layers, the additional number of production workers is $(s^n - s^{n-1})$, and the additional number of managers is s^{n-1}, so number of managers added per additional production worker is $\frac{s^{n-1}}{s^n - s^{n-1}} = \frac{1}{s-1}$.

FOCs with respect to n yields:

$$D\ (1-\mu)\left(\log(\theta s)\right)\left[\theta^{n^*-1}\right]^{1-\mu}(s^{n^*-1})^{1-\mu} = w_{\text{eff}}(\log s)s^{n^*-1}$$
$$\rightarrow\ w_{\text{eff}}(\log s)s^{(n^*-1)\mu} = D\ (1-\mu)\left(\log(\theta s)\right)\theta^{(n^*-1)(1-\mu)} \quad (5)$$

Taking logs on both sides allows us to solve for $n\hat{}*$:

$$n^* = 1 + \left[\frac{\log(\mathrm{D}) + \log(1-\mu) - \log(w_{\text{eff}}) + \log(\log(\theta s)) - \log(\log s)}{\mu \log s - (1-\mu)\log(\theta)}\right] \quad (6)$$

Prediction 1: Number of Layers of the Control Hierarchy: The depth (number of layers) of the hierarchy is: (i) increasing in sectoral demand shifter (E) , sector price index (P), firm productivity (φ), manager productivity (θ), worker productivity (θ_1), and elasticity of demand ($\sigma = 1/\mu$); (ii) decreasing in the effective wage (w_{eff}) and the span parameter (s)

Proof: Follows directly from (6) above (recall $D = (E^{\mu}P^{1-\mu})\varphi^{1-\mu}$). All of the results, except for that for elasticity of demand are intuitive. Increases in demand/productivity parameters or decrease in wage implies a larger optimal production labor size, which then needs more layers to be achieved, so long as s is held constant. The result for span s is intuitive as well, as larger span yields more production workers and allows firms to operate with a flatter hierarchy. The effect of increasing elasticity of demand is not intuitive, but in this model higher μ parameter reduces the optimal production labor size, so decreasing μ leads to a need for more layers (holding scope s constant). The result for span decreasing the number of layers is complicated if $\theta > 1$, but goes through with $\theta \leq 1$.

Firm size in the data is total employment, which corresponds to:

$$l_{tot} = 1 + s + \cdots s^{n-1} = \sum_{i=o}^{n-1} s^i = \frac{s^n - 1}{s-1} \quad (7)$$

From result (6), we get optimal number of production workers is given by:

$$\log l^* = (n^*-1)\log s$$
$$= \left[\frac{\log(\mathrm{D}) + \log(1-\mu) - \log(\mathrm{w}_{\text{eff}}) + \log(\log(\theta s) - \log(\log s)}{\mu \log s - (1-\mu)\log(\theta)}\right]\log s \quad (8)$$

5.4 Profit sharing and Implications for Earnings Profile over the Control Hierarchy

Extending the idea in Lucas (1978) that the manager/entrepreneur retains firm profits, we assume that a fraction λ_{π} of the profits of the firm are distributed to the

managers in the control hierarchy. A key assumption we make is that the share allocated to a manager in layer m is proportional to the marginal revenue product (MRP) of the talent of the manager.[17] This assumption is consistent with several important papers in the labor market literature on rent sharing, which find that compensation of leaders at the top of the hierarchy co-varies with firm sales and profits much more strongly than for those lower in the hierarchy (e.g., Roberts 1956, Van Reenen 1996, Bell and Van Reenen 2012).

Revenue of the firm is $R(\phi, s) = Bq^{1-\mu}$. For a manager in level m, the MRP is:

$$MRP_m = \frac{dR}{d\theta_m} = s^{m-1}\theta^{n-2} \tag{8}$$

This equation captures the idea that a manager's productivity has an impact on all of her subordinates below her chain of command. Thus, the impact of manager talent scales up along the hierarchy, with the CEO talent having the largest impact on the firm.

Total earnings of a manager in level m is given by:

$$e_m = \widehat{w} + \kappa s^{m-1}\theta^{n-2}\theta_1 \tag{9}$$

Prediction 2: Earnings profile over the control hierarchy: The earnings profile across the control hierarchy is convex.

Proof: Follows directly from (9), which yields $\frac{de_m}{dm} = \kappa\theta^{n-2}s^{m-1}\log s > 0$ and $\frac{d^2e_m}{dm^2} > 0$.

5.5 Cutoff profitability condition

The cutoff profit condition implicitly defines boundary ($\bar{\varphi}(s)$) demarcating survival:

$$\left\{B\varphi^{1-\mu}\left[\theta^{n^*-1}\right]^{1-\mu}\left(s^{n^*-1}\right)^{1-\mu}\right\} - \bar{w}s^{n^*-1} - \widehat{w}\left[\frac{s^{n^*-1}-1}{s-1}\right] - w_oF_o = 0 \tag{11}$$

5.6 Profit share received by managers

Given that the number of managers in level m is given by s^{n-m} we get from (9):

$$\text{Payroll for level m} = s^{n-m}e_m = \widehat{w}s^{n-m} + \kappa s^{n-1}\theta^{n-2}$$

[17] This echoes Caliendo and Rossi-Hansberg (2012) where higher level employees earn a wage plus compensation for the knowledge they accumulate. One difference is that managers get rewarded linearly for the length of information set z. However, the marginal knowledge can be low value as production depends on $F(Z)$. In particular, the same length of knowledge at the end point may be less valuable of $F(Z)$ approaches 1 very quickly.

Incidentally, this model yields the result that total bonus component of the payroll is the same for every layer of the hierarchy. Then we can solve for κ by setting total managerial payroll equals the profits:

$$(n^*-1)\kappa s^{n^*-1}\theta^{n^*-2} = \lambda_\pi\left[\left\{D\left[\theta^{n^*-1}\right]^{1-\mu}\left(s^{n^*-1}\right)^{1-\mu}\right\} - \bar{w}s^{n^*-1} - \hat{w}\left[\frac{s^{n-1}-1}{s-1}\right] - w_o F_o\right]$$

$$\rightarrow \kappa = \frac{\lambda_\pi\left\{\left\{D\left[\theta^{n^*-1}\right]^{1-\mu}\left(s^{n^*-1}\right)^{1-\mu}\right\} - \bar{w}s^{n^*-1} - \hat{w}\left[\frac{s^{n-1}-1}{s-1}\right] - w_o F_o\right\}}{(n^*-1)s^{n^*-1}\theta^{n^*-2}}$$

$$\rightarrow \kappa = \frac{\lambda_\pi D\theta^{1-\mu(n^*-1)}s^{-\mu(n^*-1)}}{n^*-1} - \frac{\lambda_\pi}{(n^*-1)\theta^{n^*-2}}\left(\bar{w} + \frac{\hat{w}\left(1-s^{-(n^*-1)}\right)}{s-1} + \frac{w_o F_o}{s^{n^*-1}}\right) \qquad (12)$$

5.7 Free Entry condition

We follow the models in the literature (e.g. Lippman and Rumelt 1981, Hopenhayn 1994, Melitz 2003) and assume that there is an entry cost. Specifically, identical firms (or CEOs/entrepreneurs) pay a fixed investment cost $w_o F^E$ (where F^E is the units of lowest skill labor) to enter, and upon entry draw a productivity level (φ) and a span (s) from distribution $G(\varphi, s)$.[18] If the (φ, s) combination is such that profits are below the fixed cost as described above the firms exit (and hence get zero profits).

As in Caliendo and Rossi-Hansberg (2012), δ is the fraction of firms that exit every period), so the probability of survival is $(1-\delta)$, which yield the same expression for present value of profits.:

$$v(\varphi, s) = \max\left\{0, \frac{\Pi(\varphi, s)}{\delta}\right\}$$

Free entry drives the ex-ante expected profits to just equal the entry costs.

$$\int_o^\infty \int_0^\infty \frac{\Pi(\varphi, s)}{\delta} dG(\varphi, s) = w_o F^E \qquad (13)$$

[18] While we write the theoretical integrals over the general support $(0, \infty)$ for notational simplicity, we assume strictly positive lower bounds for both productivity and span. Thus, the joint probability density $g(\varphi, s) = 0$ for all $\varphi < \varphi_{min}$ and $s < s_{min}$, a condition we explicitly impose in our quantitative simulations.

5.8 Steady State General Equilibrium Conditions

Define probability of an entrant drawing (φ, s) that allows the firm to survive as:

$$\text{Probability of survival on entry} = \int_0^\infty \int_{\bar{\varphi}(s)}^\infty dG(\varphi, s) = 1 - \bar{G}$$

Where $\bar{\varphi}(s)$ denotes the minimum productivity threshold for a given s needed for firm to make non-zero profits. Then if M denotes the steady state mass of active firms, and M_e denotes mass of entrants, we have mass of entrants that survive equating to the mass of firms exiting every period so:

$$[1 - \bar{G}]M_e = \delta M$$

To simplify the labor market clearing condition, we follow Garicano (2000) and Caliendo and Rossi-Hansberg (2011), and posit that the economy has an overall endowment $\bar{L}$ of agents with baseline homogenous skill level with wage w_o. Then higher wages earned by skilled workers and managers are due to knowledge acquired and paid for, from teachers who are themselves at the baseline skill level. In other words, an agent earning $w' > w_0$ fundamentally utilizes $\frac{w'}{w_0}$ units of baseline labor from the economy's aggregate endowment. Then, returning to the multisector case, we get labor market clearance condition:

$$\bar{L} = \sum_j \left\{ M_j^E F_j^E + \frac{M_j}{1 - \bar{G}_j} \int_0^\infty \int_{\bar{\varphi}_j(s)}^\infty \left[l_j^{\mathrm{p}}(\varphi, s) + l_j^m(\varphi, s) + F_o \right] dG_j(\varphi, s) \right\}$$

$$= \sum_j \frac{M_j}{1 - \bar{G}_j} \left\{ \delta F_j^E + \int_0^\infty \int_{\bar{\varphi}_j(s)}^\infty \left[l_j^p(\varphi, s) + l_j^m(\varphi, s) + F_o \right] dG_j(\varphi, s) \right\} \qquad (14)$$

Where $l_j^p(\varphi, s) = \frac{\bar{w}(\varphi,s) l_j(\varphi,s)}{w_o}$ represents the firm's demand for production labor expressed in units of baseline raw labor, and $l_j^m(\varphi, s) = \frac{\hat{w}(\varphi,s)\, N_j^m(\phi,s)}{w_o}$ represents the firm's demand for management labor expressed in units of baseline raw labor, and f_o is the exogenous per-period operating fixed cost, also paid in units of baseline raw labor.

Equilibrium in the goods market implies that overall worker income equates to aggregate expenditure $Y = w_o \bar{L}$, so expenditure for each sector j is given by:

$$R_j = E_j = \beta_j Y \qquad (15)$$

which translates to the following equation for goods market clearing (aggregate revenue) in sector j:

$$R_j = \beta_j Y = \frac{M_j}{1 - \bar{G}_J} \int_0^{\infty} \int_{\bar{\varphi}(s)}^{\infty} p_j(\varphi, s) q_j(\varphi, s) dG_j(\varphi, s) \tag{16}$$

Following the approach in the literature (e.g., Helpman, Melitz, Yeaple, 2004) In our multisector model, we assume sector 1 is a perfectly competitive one with a simple linear production technology $q = l$, where l is baseline raw labor units. Then we get equilibrium price in this sector $p_1 = w_o$. We choose sector 1 good as the numeraire, which implies that $w_o = 1$.

The general equilibrium of this multi-sector economy is characterized by $3J + 1$ endogenous variables: the aggregate expenditure Y, and for each sector$j$, the price index P_j, the mass of active firms M_j, and the zero-profit productivity cutoff $\bar{\varphi}_j(s)$.These variables are pinned down by a system of $3J + 1$ equations. Because we normalize the baseline wage to one ($w_0 = 1$), the aggregate labor market clearing condition combined with the free entry conditions (zero aggregate net profits) yields $Y = \bar{L}$. Given this aggregate expenditure, the remaining sectoral variables ($P_j, M_j, \bar{\varphi}_j(s)$) are obtained by solving the system of $3J$ sectoral equations: the cutoff profit condition (Eq. 11), the free entry condition (Eq. 13), and the sectoral goods market clearing condition (Eq. 16).

Following the standard approach in the literature (e.g., Melitz 2003; Caliendo and Rossi-Hansberg 2012), we can establish the existence and uniqueness of the general equilibrium. Because firms are heterogeneous in both productivity ϕ and span s, the zero-profit condition defines a survival frontier, $\bar{\varphi}_j(s)$, rather than a single scalar cutoff. To establish the existence and uniqueness of the general equilibrium, we utilize the aggregate sectoral demand B_j. For any given s, the zero-profit frontier $\bar{\varphi}_j(s)$ is strictly decreasing in B_j, and the operating profit of every active firm is strictly increasing in B_j.[19] Consequently, the ex-ante expected value of entry is strictly monotonically increasing in B_j. Equating this expected value to the exogenous sunk entry cost F_e guarantees a unique equilibrium level

[19] Intuitively, more demand allows for more firms to survive and give surviving firms greater profits.

of sectoral demand B_j^*. This unique scalar B_j^* in turn uniquely determines the price index P_j^* and the full survival schedule $\bar{\varphi}_j^*(s)$ for each sector.

6. Calibration and Simulation

6.1 Model Setup and Exogenous Parameters

The simulation characterizes a general equilibrium environment populated by heterogeneous firms across distinct sectors. Table 3 provides a summary of parameters estimated using data moments and those calibrated or set externally. The aggregate labor force is fixed at $\bar{L} = 10000$, in baseline units of raw skill. To anchor the price level, the price of the outside sector (Sector 0) is normalized to $P_0 = 1$. The economy is divided such that Sector 0 commands a 50% expenditure share ($\beta_0 = 0.50$), while the active production sector (Sector 1, representing our empirical sample) accounts for the remaining 50% ($\beta_1 = 0.50$).

Firms face a fixed entry cost $F_e = 10000$ and a baseline per-period fixed operation cost $F_o = 120$, calibrated to generate a non-trivial selection effect where roughly \$20\%\$ of productivity and span draws fall below the survival threshold. Based on empirical survival rates for firms with greater than 100 employees, the exogenous firm exit rate is set to $\delta = 0.02$. The elasticity of substitution is fixed at $\sigma = 5$, following standard estimates in the trade and macro literature. To accommodate the tall hierarchies observed in the data, the attenuation of firm productivity across managerial layers is set to a relatively limited $\theta = 0.9$. Finally, dynamic wages across production layers are governed by \$$w_k = w_o\left(1 + \frac{\eta(k-1)}{K}\right)$, with the baseline manager-to-worker wage premium parameter set to $\eta = 2$.

Upon entry, firms draw two fundamental characteristics: an optimal span of control, s, and a core productivity level, φ. These draws are assumed to be independent. The span is drawn from a uniform distribution, $s \sim$ Uniform$[1.15, s_{max}]$, while productivity is drawn from a Pareto distribution, with the lower bound fixed at $\phi_{min} = 0.8$.

The total number of production layers a firm operates, K, is modeled as a quadratic function of its productivity draw, subject to uniform idiosyncratic noise:

$$K(\varphi) = k_o + k_1\varphi + k_2\varphi^2 + \epsilon_k$$

where $\epsilon_k \sim$Uniform$[0, 24]$.

6.2 Calibration Strategy

We utilize Simulated Method of Moments (SMM) to estimate the remaining parameters that govern the shape of the firm size and wage hierarchy distributions. Specifically, we jointly estimate the upper bound of the span distribution (s_{max}), the Pareto shape parameter for productivity (α), the profit-sharing parameter ($\lambda \in (0.01, 0.99)$), and the structural polynomial coefficients for the production layers (k_o, k_1, k_2). The SMM algorithm searches the parameter space to minimize the sum of squared errors between the simulated general equilibrium moments and our targeted empirical counterparts. We target three main moments: (i) Fact 1 – PDF of Total Hierarchy Size (Figure 2), (ii) Fact 2 – U-shaped relationship between total number of layers vs. firm size Percentile (Figure 4), and (iii) Fact 3 – The convex relationship between average wage of the hierarchy layer and (standardized) hierarchy levels (Figure 5a).

6.3 Goodness of Fit Analysis

The SMM calibration achieves a good fit across all targeted moments, as illustrated in the comparison between the simulated outputs and the empirical data in Figure 7.

Fact 1 (PDF of Total Hierarchy Size): The model captures the overall distribution of total layers reasonably well. A key characteristic of the empirical distribution is its broad dispersion, demonstrating significant support for hierarchy sizes ranging from 3 to 25 layers (the maximum allowed by our estimation strategy). The simulated distribution successfully replicates this dispersion, albeit with a few notable exceptions. Specifically, the simulation generates no firms with 3 or 4 layers and exhibits excess mass at layers 5 and 6. While the model's fit is quite accurate for firms with 7 to 22 layers, it slightly underpredicts the share of firms operating with the largest hierarchy sizes.

Fact 2 (Total Layers vs. Firm Size Percentile): The calibration matches the convex U-shaped relationship between employment size percentiles and the average total hierarchy size. The flexible quadratic functional form governing the number of production layers ($K(\varphi) = k_o + k_1\varphi + k_2\varphi^2 + \epsilon_k$) plays a critical role here. By allowing a more flexible relationship between the number of distinct skills and core firm productivity (φ), this

specification permits smaller firms to potentially retain a larger collection of skills, allowing the model to trace the empirical curve.

Fact 3 (Average Wage vs. Hierarchy): The simulation replicates the upward-sloping wage gradient observed in the data. As the standardized layer index increases, the average wage rises—slowly at first, reflecting modest earnings differentials across the lower production layers, and then accelerating rapidly at higher organizational levels. This convexity at the top of the hierarchy reflects the impact of profit sharing, which is proportional to the manager's marginal product and increases exponentially as one moves up the corporate structure.

7. Conclusions

In this paper, we leverage two decades of matched employer-employee data from the U.S. LEHD program to provide a comprehensive, economy-wide analysis of wage hierarchies. By applying unsupervised clustering algorithms to identify latent wage tiers within firms, we document six stable empirical regularities regarding the organization of human capital. These include a highly stable aggregate distribution of hierarchy sizes, systematic sectoral differences, a U-shaped relationship between organizational depth and firm size, exponential wage convexity at the top of the hierarchy, widening spans of control, and a modest negative association between hierarchy size and future firm performance.

Our approach of using large-scale administrative earnings data complements existing empirical strategies in the organizational economics literature. While much of the foundational empirical work on firm hierarchies has relied on detailed, proprietary personnel records from individual firms—or, more recently, on public resume and job-title data scraped from platforms like LinkedIn—our methodology offers a broadly representative, longitudinal perspective that spans the entire formal economy without relying on standardized job titles.

We acknowledge, however, an important conceptual distinction: the “wage clusters” identified in our empirical exercise do not necessarily map one-to-one onto formal reporting structures. In simpler, classical models of the firm, there is often a strict equivalence between hierarchical control ranks and discrete wage levels. In reality, a

worker in a lower-wage cluster may not have any direct reporting relationship to an employee in a higher-wage cluster. Nevertheless, we demonstrate that a parsimonious structural extension of the Lucas (1978) model—one that explicitly separates a heterogeneous production hierarchy (based on varying skill clusters) from a vertical control hierarchy (governed by managerial span of control and profit-sharing)—is capable of robustly fitting our key stylized facts.

Ultimately, we hope our methodology and findings will inspire further research utilizing matched employer-employee data to study the internal architecture of firms. By viewing the firm not simply as a frictionless production function, but as an active organizer of stratified human capital, there remains significant scope for future studies to deepen our understanding of organizational design, wage inequality, and firm dynamics across the broader economy.

References:

Bell, B., & van Reenen, J. (2012). *Firm performance and wages: Evidence from across the corporate hierarchy*. London School of Economics and Political Science, LSE Library.

Caicedo, S., Lucas, R. E., Jr., & Rossi-Hansberg, E. (2016). Learning, career paths, and the distribution of wages (Working Paper No. 22151). National Bureau of Economic Research.

Caliendo, L., Monte, F., & Rossi-Hansberg, E. (2015). The anatomy of French production hierarchies. *Journal of Political Economy*, 123(4), 809–852.

Caliendo, L., & Parro, F. (2015). Estimates of the Trade and Welfare Effects of NAFTA. *The Review of Economic Studies*, *82*(1), 1-44.

Caliendo, L., & Rossi-Hansberg, E. (2012). The impact of trade on organization and productivity. *The Quarterly Journal of Economics*, *127*(3), 1393-1467.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, *3*(1), 1-27.

Calvo, G. A., & Wellisz, S. (1978). Supervision, loss of control, and the optimum size of the firm. *Journal of Political Economy*, 86(5), 943–952.

Calvo, G. A., & Wellisz, S. (1979). Hierarchy, ability, and income distribution. *Journal of Political Economy*, 87(5, Part 1), 991–1010.

Card, D., Rothstein, J., & Yi, M. (2024). Industry wage differentials: A firm-based approach. *Journal of Labor Economics*, 42(S1), S11-S59.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*, 1-36.

Ewens, M., & Giroud, X. (2025). Corporate hierarchy (Working Paper). Columbia Business School and National Bureau of Economic Research. SSRN: https://ssrn.com/abstract=5391667.

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. Journal of political economy, 108(5), 874-904.

Haltiwanger, J., Hyatt, H., & McEntarfer, E. (2018). Who moves up the job ladder?. Journal of Labor Economics, 36(S1), S301-S336.

Keren, M., & Levhari, D. (1979). The optimum span of control in a pure hierarchy. *Management Science*, 25(11), 1162–1172.

Kueng, L., Yang, M.-J., & Hong, B. (2017). Sources of firm life-cycle dynamics: Size vs. age effects (Working Paper).

Lucas, R. E., Jr. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9(2), 508–523.

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), 759–773.

Sugar, C. A., & James, G. M. (2003). Finding the number of clusters in a dataset: An information-theoretic approach. *Journal of the American Statistical Association*, *98*(463), 750-763.

Van Reenen, J. (1996). The creation and capture of rents: wages and innovation in a panel of UK companies. *The Quarterly Journal of Economics*, *111*(1), 195-226.

Wulf, J. (2012). The flattened firm: Not as advertised. *California Management Review*, 55(1), 5–23.

[Incomplete]

Figure 1: Alternative Approaches to Select Number of Clusters: Illustrative Simulation

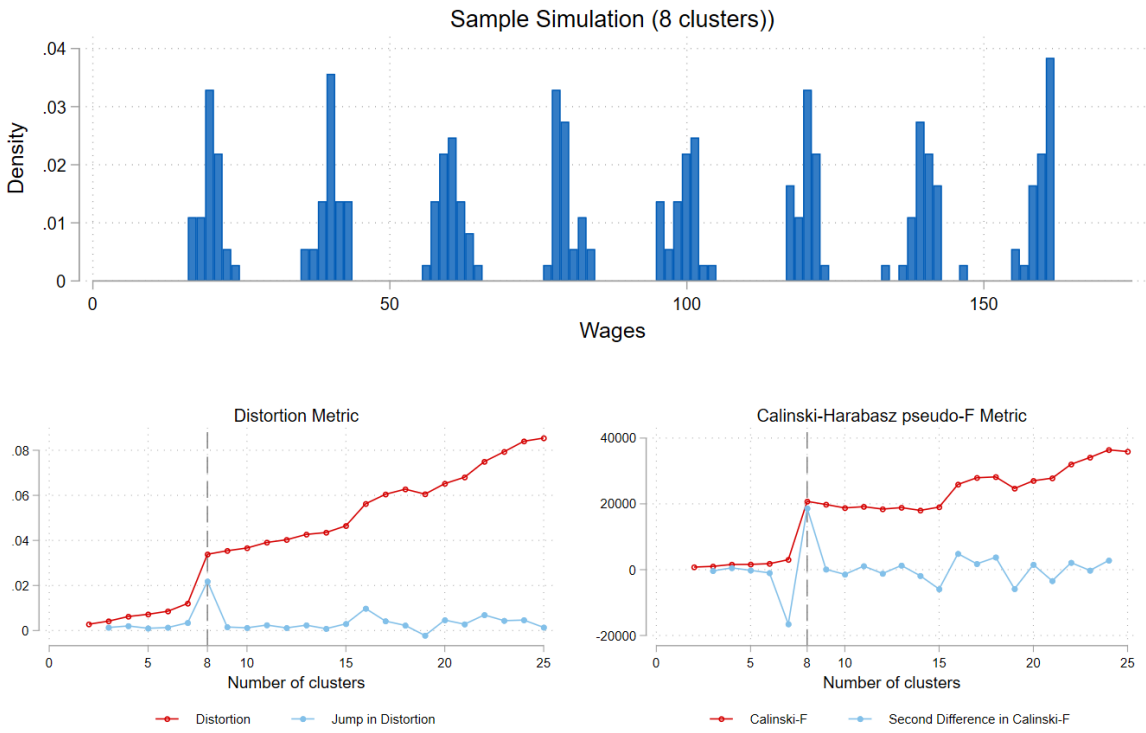
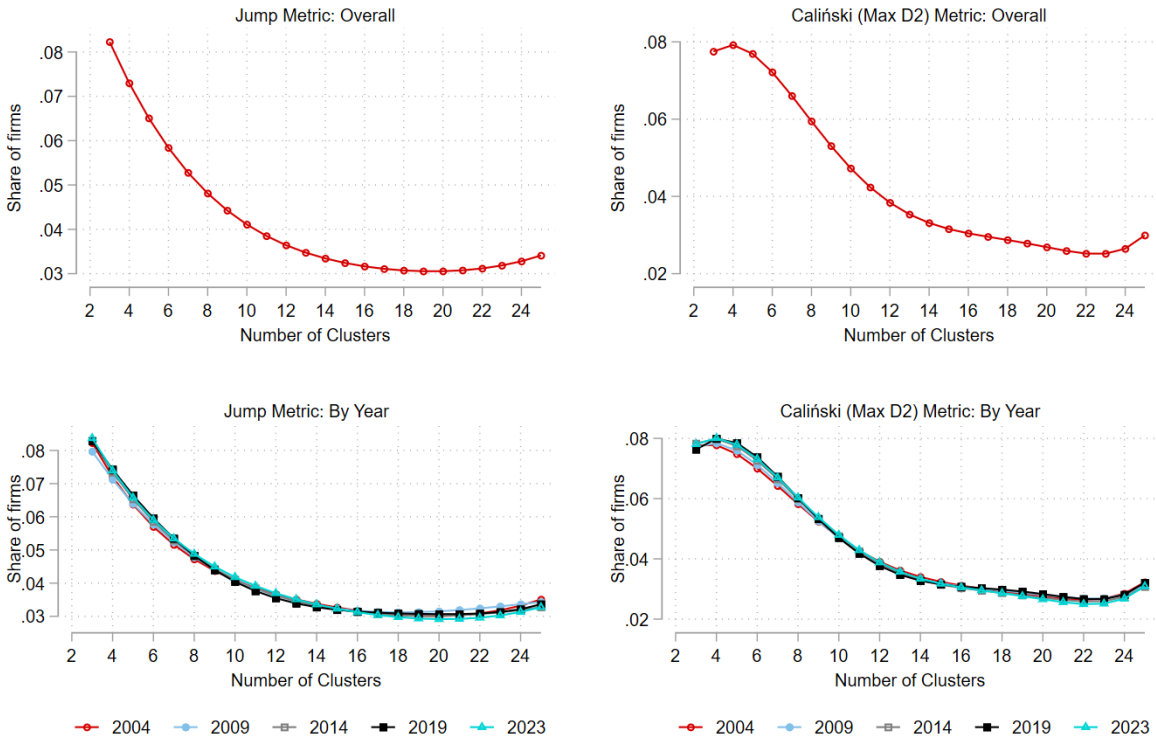


Figure 2: Distribution of Hierarchy Sizes



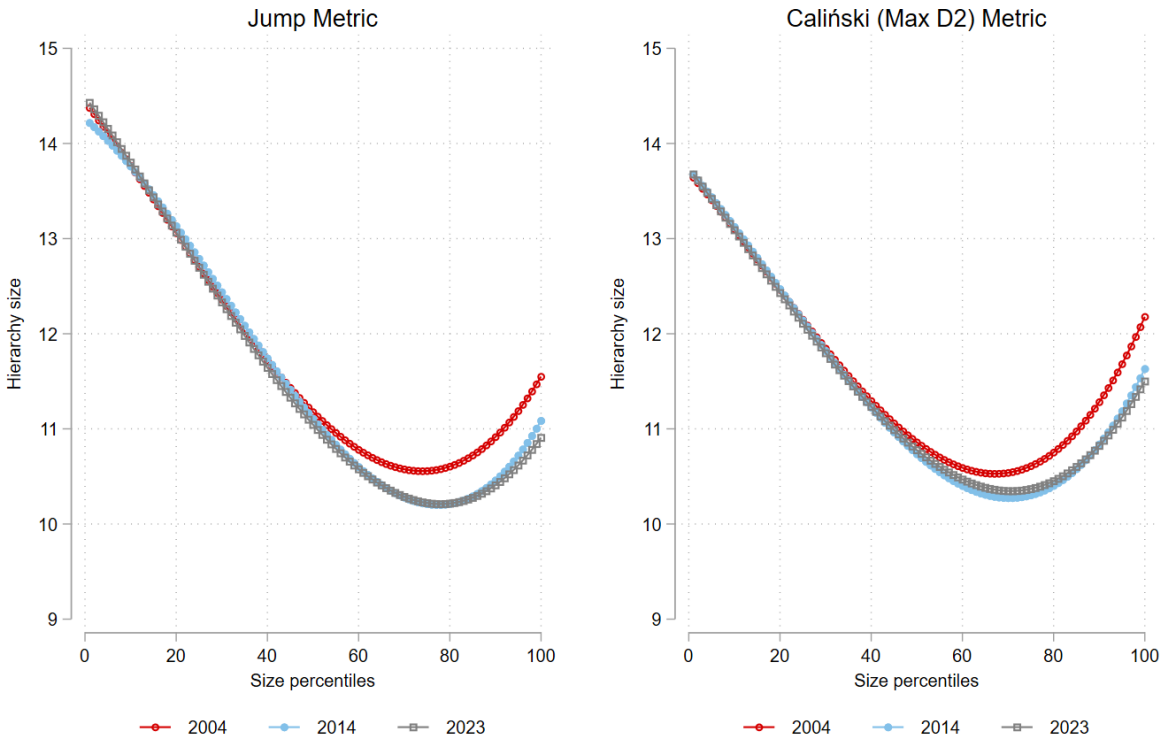
Notes: This figure presents the estimated probability density function for hierarchy size. We retain firms (SEINs) in the LEHD with at least 100 employees, and apply a K-means clustering algorithm to data on individual worker annual wages for each firm in each year. The optimal number of clusters are estimated using the maximum of two metrics: the Jump metric and the second difference of the Caliński-Harabasz metric. We fit a fifth-degree polynomial to the actual firm shares by cluster size, and the above figures present the estimates from this polynomial function (which closely approximates the actual probability distribution).

Figure 3: Hierarchy Sizes across Sectors



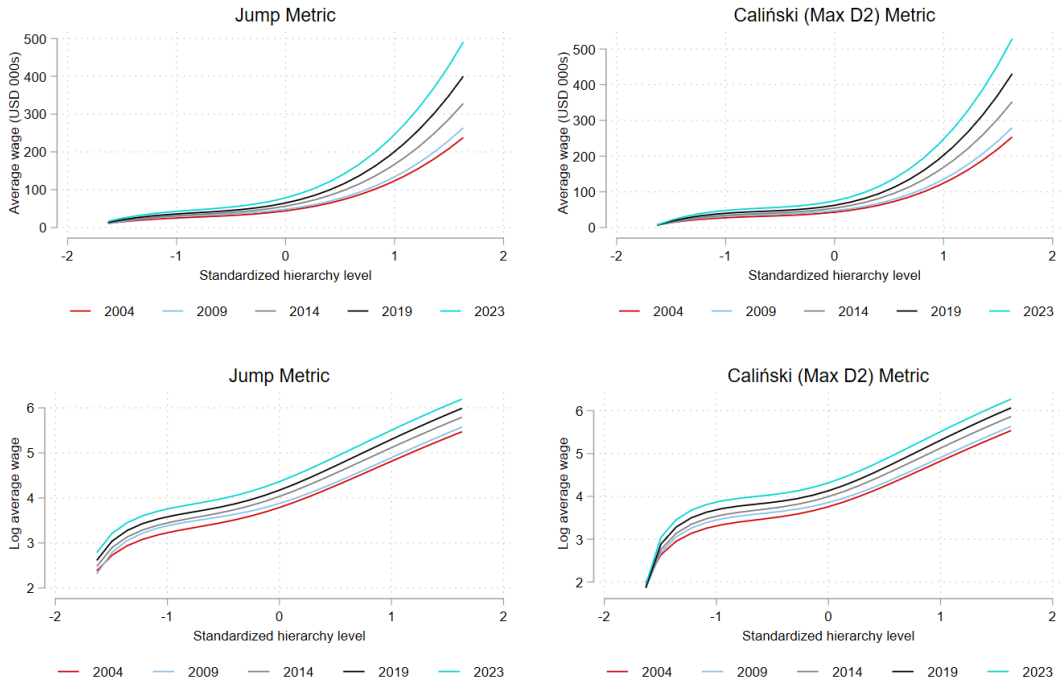
Notes: This figure presents the average number of clusters by sector for 2004 and 2023.

Figure 4: Hierarchy Size vs Firm Size



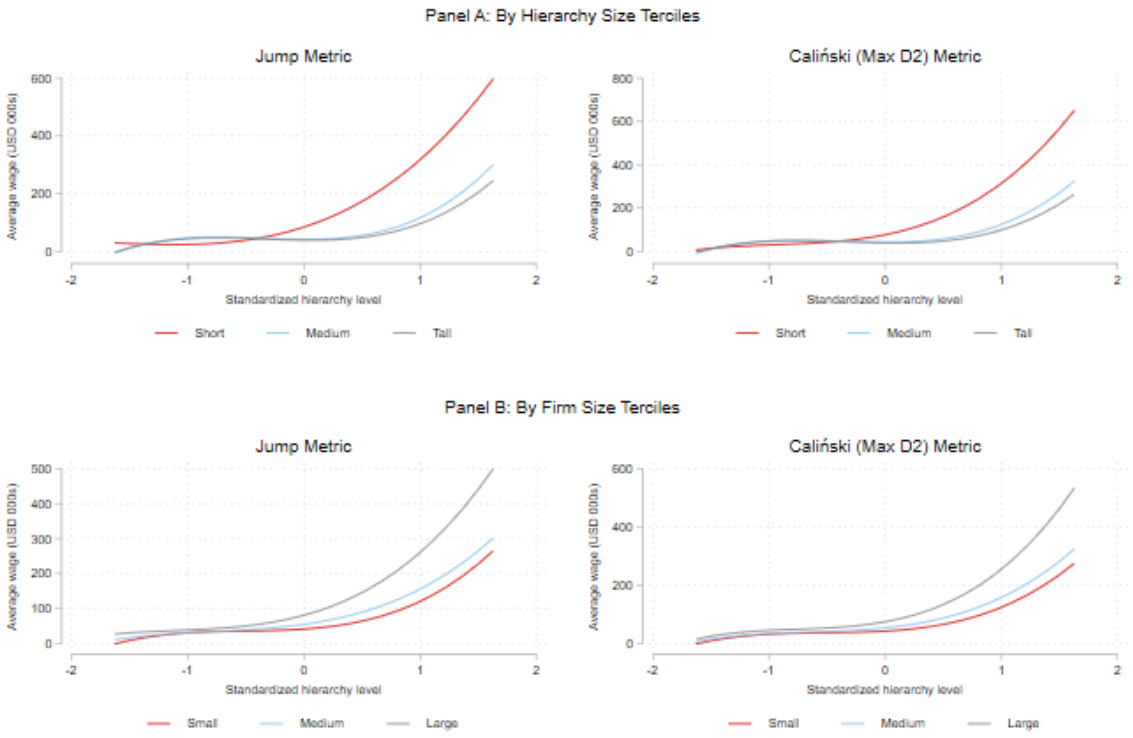
Notes: This figure presents the relationship between hierarchy size and percentiles of firm size. We first estimate the optimal number of wage clusters using the maximum of Jump metric or the maximum of the second difference of Caliński-Harabasz metric in worker-level annual earnings data for each firm-year. Then the firm hierarchy size (i.e. optimal number of clusters in the firm's worker-level wage data) was regressed on a cubic polynomial in firm size percentiles with NAICS 3-digit fixed effects.

Figure 5a: Wages across the Hierarchy



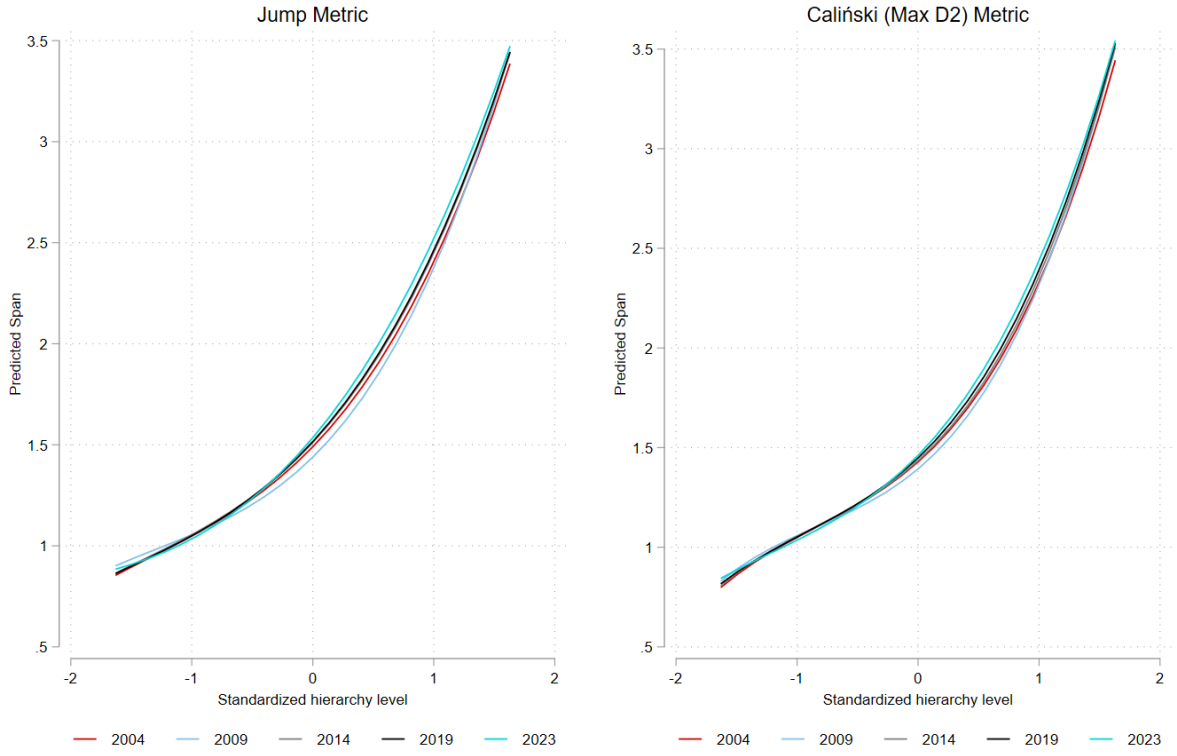
Notes: This figure presents the relationship between wages and hierarchy level, where optimal number of wage clusters in the hierarchy are identified using the maximum of Jump metric or the maximum of the second difference of Caliński-Harabasz metric in worker-level annual earnings data for each firm-year. We estimate the average wage within each firm-year-hierarchy level. We then regress this average wage on a cubic polynomial of the "standardized" hierarchy level variable, separately for each year. The top panel presents the predicted average wage from the cubic polynomial regression. The bottom panel presents the log of the predicted average wage from the cubic polynomial regression. All regressions included NAICS 3-digit fixed effects.

Figure 5b: Wages across the Hierarchy: By Hierarchy and Firm Size



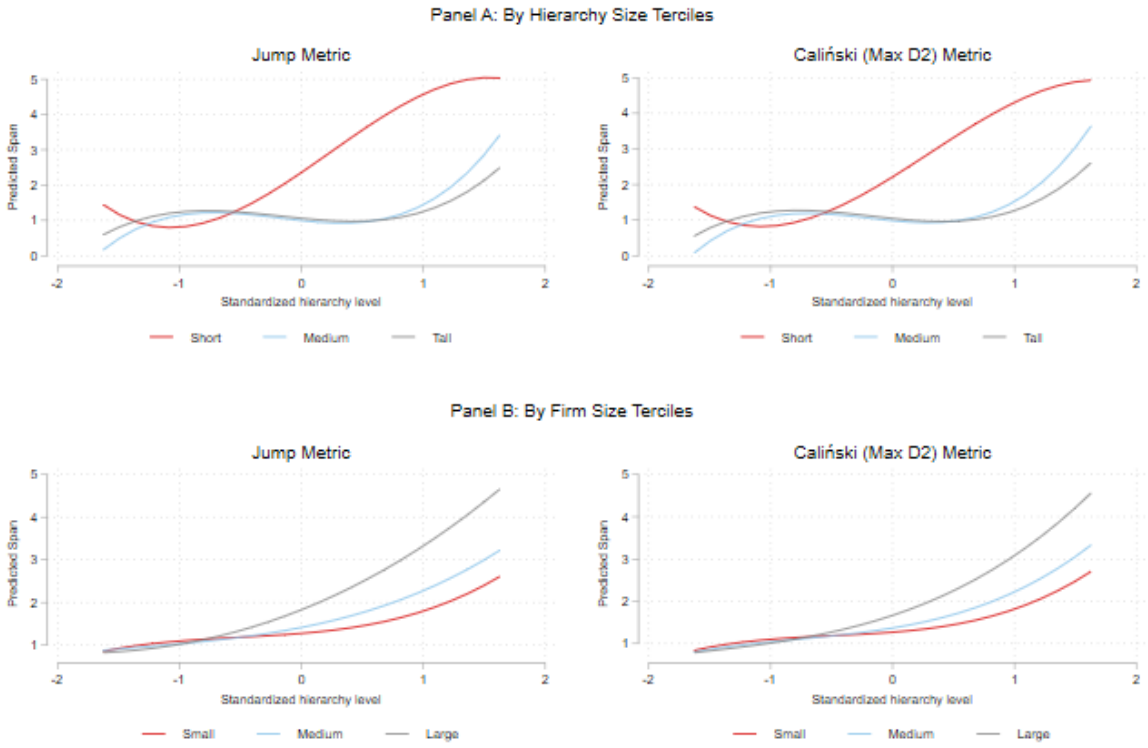
Notes: This figure presents the relationship between wages and hierarchy level, where optimal number of wage clusters in the hierarchy are identified using the maximum of Jump metric or the maximum of the second difference of Caliński-Harabasz metric in worker-level annual earnings data for each firm-year. We estimate the average wage within each firm-year-hierarchy level. We then regress this average wage on a cubic polynomial of the "standardized" hierarchy level variable, with the terms of the polynomial interacted with a hierarchy size (top panel) and firm employment size (bottom panel) tercile indicator variables. The top panel presents the predicted average wage from the cubic polynomial regression, separately for terciles of the hierarchy size. The bottom panel presents the predicted average wage from the cubic polynomial regression, separately for terciles of the firm employment size. All regressions included NAICS 3-digit fixed effects.

Figure 6a: Span across the Hierarchy



Notes: This figure presents the relationship between span and hierarchy level, where optimal number of wage clusters in the hierarchy are identified using the maximum of Jump metric or the maximum of the second difference of Caliński-Harabasz metric in worker-level annual earnings data for each firm-year. We estimate the average span within each firm-year-hierarchy level. We then regress this average span on a cubic polynomial of the "standardized" hierarchy level variable, separately for each year. The figure presents the predicted average span from the cubic polynomial regression. All regressions included NAICS 3-digit fixed effects.

Figure 6b: Span across the Hierarchy: By Hierarchy and Firm Size



Notes: This figure presents the relationship between span and hierarchy level. We first estimate the average span within each firm-year-hierarchy level. We then regress this average span on a cubic polynomial of the "standardized" hierarchy level variable, with the terms of the polynomial interacted with a hierarchy size (top panel) and firm employment size (bottom panel) tercile indicator variables. The top panel presents the predicted average span from the cubic polynomial regression, separately for terciles of the hierarchy size. The bottom panel presents the predicted average span from the cubic polynomial regression separately for terciles of the firm employment size. All regressions included NAICS 3-digit fixed effects.

Figure 7: Fit of Simulated Data to Aggregate Moments

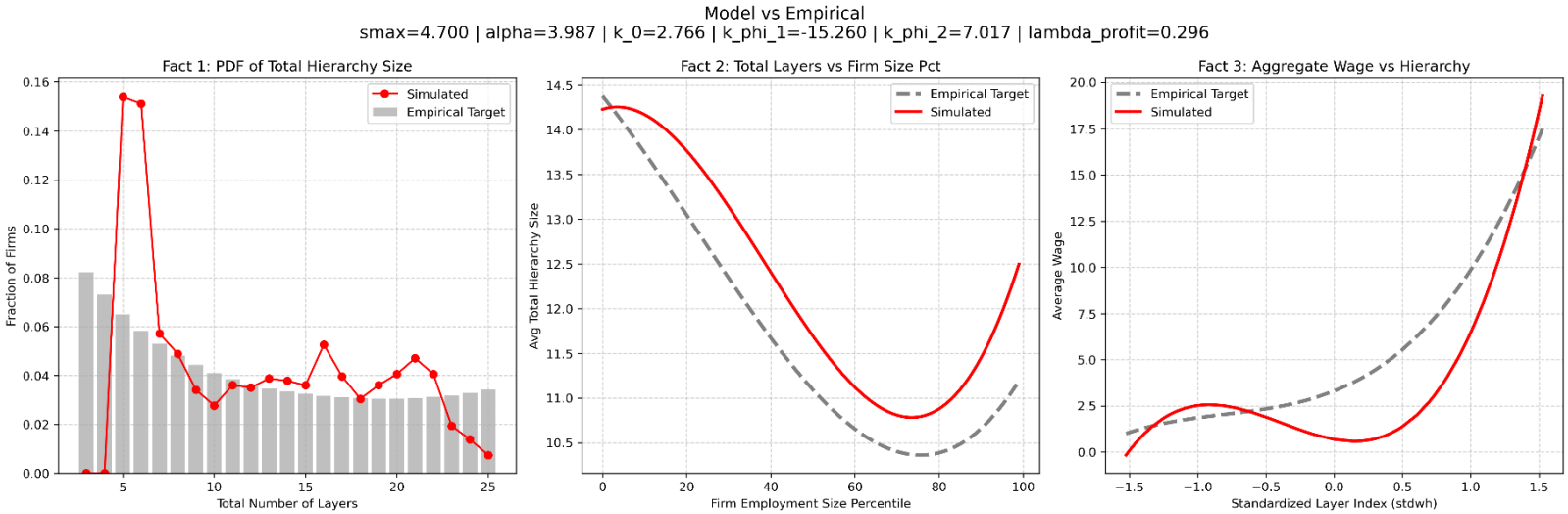


Table 1: Hierarchy Size and Future Firm Performance

	Five Year Survival Dummy		Five Year Employment Growth		Five Year Payroll Growth	
	(1)	(2)	(3)	(4)	(5)	(6)
(Number of Clusters by Jump Metric)/100	-0.03938*		-0.208***		-0.1917**	
	(0.01881)		(0.06044)		(0.05912)	
(Number of Clusters by Caliński (Max D2) metric)/100		-0.0485**		-0.1833***		-0.1855***
		(0.01452)		(0.04351)		(0.0433)
Adjusted R-square	0.045	0.045	0.065	0.065	0.066	0.066
Number of Observations (Rounded)	243000	243000	243000	243000	243000	243000
Fixed Effects	State-NAICS3-Year		State-NAICS3-Year		State-NAICS3-Year	
Mean of DV	0.9193	0.9193	9.728	9.728	9.852	9.852
SD of DV	0.2724	0.2724	0.7215	0.7215	0.7621	0.7621
Mean of Number of Clusters/100	0.1281	0.1175	0.1281	0.1175	0.1281	0.1175
SD of Number of Clusters/100	0.06996	0.06931	0.06996	0.06931	0.06996	0.06931
Effect Size (1SD RHS as proportion of DV mean)	-0.30%	-0.37%	-0.15%	-0.13%	-0.14%	-0.13%
Effect Size (1SD RHS as proportion of DV SD)	-1.01%	-1.23%	-2.02%	-1.76%	-1.76%	-1.69%

Table 2: Summary of Key Assumptions

Parameter	Assumption	Comment
σ (Demand elasticity)	$\sigma > 1$	Standard in the literature.
(φ, s) (Firm productivity, scale)	$\sim G(\varphi, s)$	Not necessarily correlated, which is consistent with variation in number of layers and/or span within size categories. Correlation can be disciplined by the data.
θ_m (Talent/productivity of manager in layer m)	$0 < \theta_m \leq 1$	Captures idea that some of the firm capability gets diluted as plans are communicated down the layers.

Table 3: Simulation parameters

Parameter	Value/Assumptions	Comment
Parameters estimated using Simulated Method of Moments (SMM)		
$s_{\max}, \alpha$	s ~Uniform$[1.15, s_{\max}]$ φ is distributed pareto $F(\varphi) = 1 - \left(\frac{\varphi_{min}}{\varphi}\right)^{\alpha}$ Set $\phi_{min} = 0.8$.	$G(\varphi, s)$is a mixture of uniform distribution for the span and Pareto for firm productivity Assume s and φ are uncorrelated
λ_{π}	Allowed to vary between (0.01, 0.99)	Share of profits distributed to managers
k_o, k_1, k_2	Bounds : {(0.1, 5), (-30,10), (-5, 15)}	Number of production layers $K = k_o + k_1\varphi + k_2\varphi^2 + \epsilon_k$ where ϵ_k~Uniform$[0, 24]$
Externally Calibrated/Normalized Parameters		
σ	5	Head and Mayer (2014), Simonovska and Waugh (2014)
θ	0.95	Attenuation of firm productivity across layers is fairly limited, to allow for tall hierarchies.
$\bar{L}$	10000	Size of the labor force, in units of baseline skill level
P_0	1	Sector 0 has price normalized to 1
β_0, β_1	0.5, 0.5	Equal expenditure shares for numeraire sector 0 and sector 1 (which corresponds to aggregate empirical data)
δ	0.02	In this sample (of firms with >100 employees), mean 5 years survival is about 92%, so an annual exit rate of 2% is reasonable.
F_e	10000	Allow for sufficient average equilibrium profits
η	2	In $w_k = w_o f(k)$, we set $w_k = w_o\left(1 + \frac{\eta(k-1)}{K}\right)$ and accordingly η specifies baseline manager wage to raw worker wage ($\widehat{w}/w_o$)
F_o	120	Set to generate non-trivial cutoff (~20% firms (φ, s) draws below cutoff)

ONLINE APPENDIX

Wage Hierarchies

Natarajan Balasubramanian and Jagadeesh Sivadasan

Appendix A: Alternative Approaches to Select Optimal Number of Clusters

We examine a total of 35 alternative approaches to estimating the optimal number of clusters. The approaches, and the source code for each are summarized in table A.1 at the bottom of this appendix. The first 8 approaches were coded by the authors in Stata, the next 26 uses the NBClust R package developed by Charrad et al (2014), while the final approach uses the log likelihood obtained from the Scikit-learn python package (Pedregosa et al 2011). In all cases, we evaluate the possible number of clusters ranging from 2 to 25.

A.1. Simulation outline

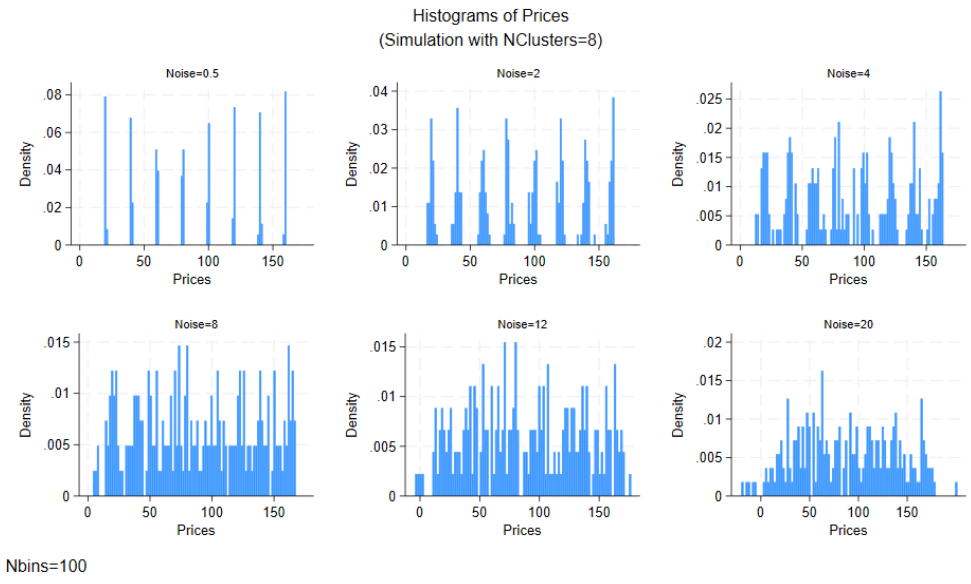
We simulate data on wages using the following equation:

$$\text{Wage} = (\text{mod}(_n, g) + 1) * 20 + \omega * \text{invnorm}(\text{uniform}())$$

where _n is the observation number, g is the number of intended clusters, and ω is a measure of noisiness. We consider 7 alternative cases for g {3, 5, 8, 11, 14, 17, 20}, and 6 different noise levels {0.5, 2, 4, 8, 12, 20}.

Histograms from alternative simulations across different noise levels are summarized in Figure A1 below, for a scenario with g=8 as the intended actual number of clusters.

Figure A1: Histograms across Alternative Simulations



Note that with higher levels of noisiness, the number of clusters in the raw data becomes unnoticeable as there is significant overlap across clusters. Thus, we may not

expect optimal stopping rules to discern clusters consistent with what we are nominally simulating. Hence, when we analyze goodness of fit, we examine performance separately for low vs high noise simulations.

A.2. Evaluation of Alternative Selection Rules

We present goodness of fit analysis for the 35 approaches using a sample size of 250. We examine two alternative approaches to measure the goodness of fit.

1. Mean Absolute Demeaned Error: The first metric we used for goodness of fit is the mean absolute demeaned error, defined as:

$$MADE = abs[(\hat{g} - g) - ME]$$

Where $ME = \frac{1}{7}\sum_{g\in\{3,5,8,11,14,17,20\}}(\hat{g} - g)$ is the mean error across the seven alternative cluster simulations within each noise level and each random sample seed.[20] (Recall we have five random sample draws within each noise and number of cluster simulations. We use the same five seeds (so the same invnorm(uniform()) draws) to generate these five samples within each of the combination of noise and number of cluster scenarios.).

Using a simple mean absolute error (defined as $MAE =$ |Predicted Ncluster − Simulated Ncluster| $= abs[(\hat{g} - g)]$) as a goodness of fit criterion has a weakness: a method that always picks a number close to the middle of the tested range (e.g., one that always picks 10 as the optimal number of clusters) would have a low error, relative to a method that always picks an extreme value (say always zero or always 20), whereas both are arguably equally poor estimates. The $MADE$ measure adjusts for this by first demeaning the error. This adjusted (demeaned error, $[(\hat{g} - g) - ME]$) would have the same zero mean across all methodologies. Then the absolute value of this demeaned error gives better sense how much the predictions really co-move with the simulated number of clusters. (This is more directly examined in a second metric below.)

The results summarized in a bar plot in Figure A2a and in a scatter plot in Figure A2b. Notably, the three information criterion (*aic*, *bic*, *dic*) and the *logrss* approach have remarkably low (zero) error across all 105 simulations where the noise is low. However,

[20] The simplest metric we considered was the mean absolute error, defined as $MAE =$ |Predicted Ncluster − Simulated Ncluster|.

there is no clearly superior approach per this metric in noisy simulations. Considering the sum of the average of the metric in high noise and low noise simulations (points highlighted in red in the scatter plot), the *jumpdist* and *dcalinski* metrics are among the best performers.

Figure A2a: Average of Mean Absolute Demeaned Error

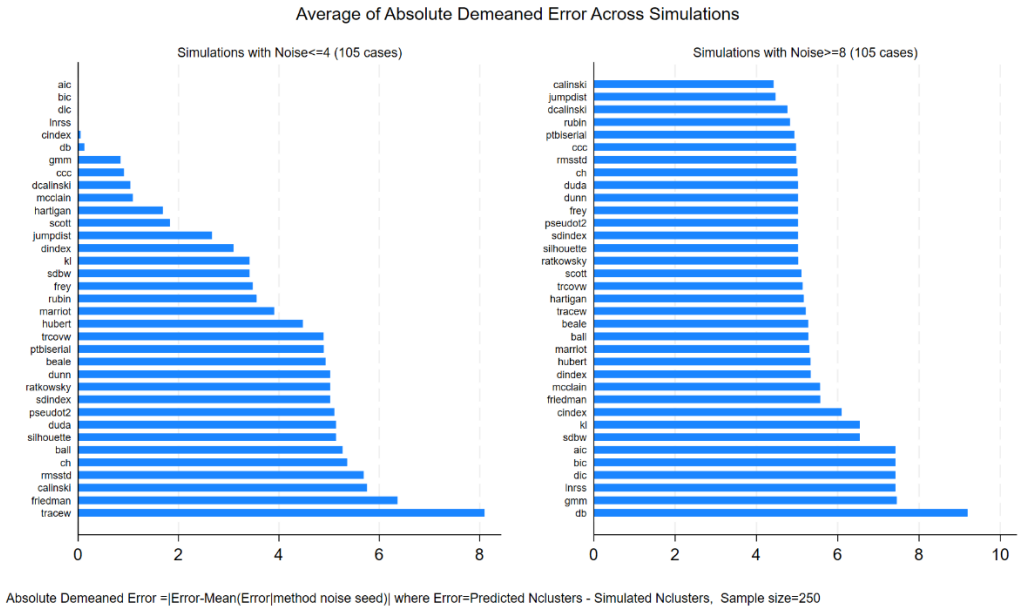
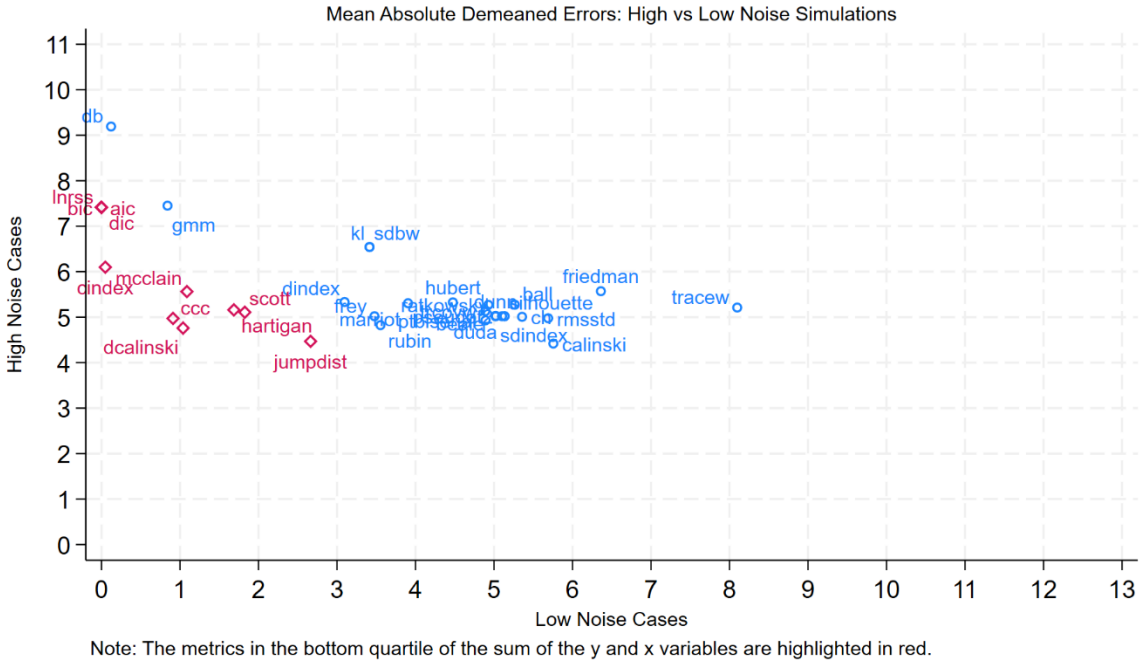


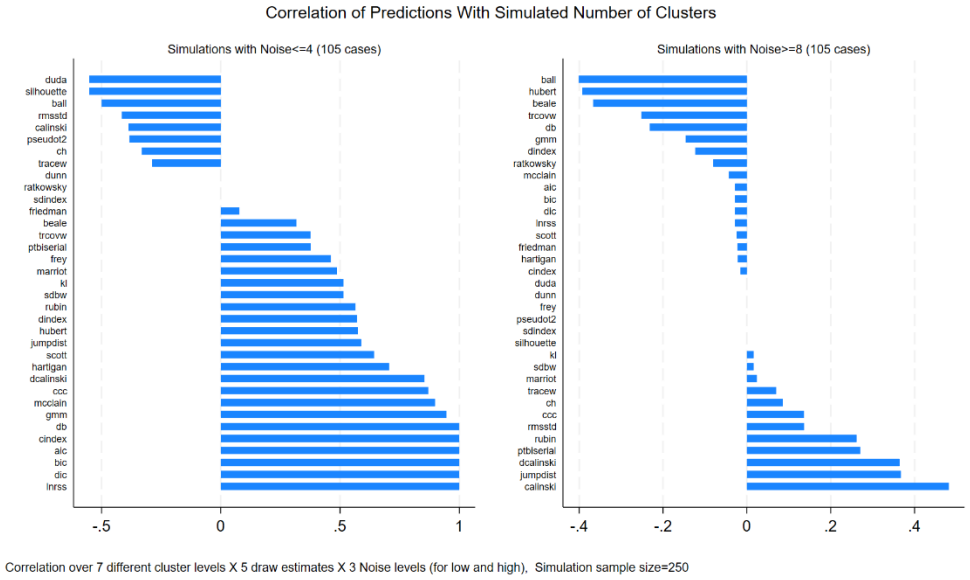
Figure A2b: Average of Mean Absolute Demeaned Error: High vs Low Noise Simulations



Note: The metrics in the bottom quartile of the sum of the y and x variables are highlighted in red.

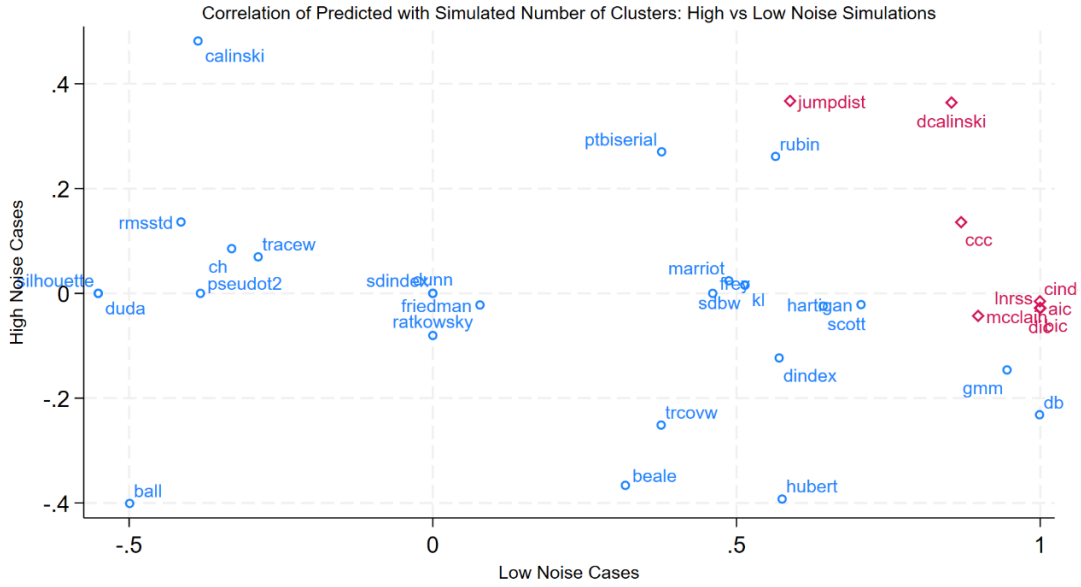
2. Correlation of Prediction with Actual Number of Clusters: A method that is consistently under (or over) the mark is arguably superior if the variations of the predictions across datasets are strongly correlated with the variation in the actual number of clusters. Hence as an alternate measure, we examine (for given noise level choice), the correlation of the predicted number of clusters with the actual number of clusters. In particular, for each of the 6 noise levels, we have 7 clusters × 5 random samples = 35 datasets, so we measure the correlation across a 35 observations sample for each noise level. The results are summarized in the bar chart in Figure A3a and the scatterplot in Figure A3b below, where we take average over the 3 low and 3 high noise scenarios. Considering the sum of the average correlations in high noise and low noise simulations (points highlighted in red), the *jumpdist* and *dcalinski* metrics are again among the best performers (showing highest correlations with the true/simulated number of clusters across simulations).

Figure A3a: Correlation of Predicted and Actual Number of Clusters



Correlation over 7 different cluster levels X 5 draw estimates X 3 Noise levels (for low and high), Simulation sample size=250

Figure A3b: Correlation of Predicted and Actual Number of Clusters: High vs Low Noise Simulations



Note: The metrics in the top quartile of the sum of the y and x variables are highlighted in red.

Table A1. Alternative Approaches to Determine Optimal Number of Clusters

N	Short Name	Full Name	Source	Decision Rule	Code
1	calinski	Caliński and Harabasz	Caliński and Harabasz, 1974	Maximum value of the index (default approach under the Stata "cluster stop" command)	Stata in-built with kmeans
2	dcalinski	Caliński and Harabasz	Caliński and Harabasz, 1974	Minimum of second differences between levels of the index	Stata in-built with kmeans
3	jumpdist	Jump Method, or Information-theoretic approach	Determining the number of clusters in a data set - Wikipedia, Sugar and James (2003)	Minimum of second differences between levels of the index	Authors' Stata code
4	aic	Akaike Information Criterion	Akaike information criterion - Wikipedia	Maximum of second differences between levels of the index	Authors' Stata code
5	bic	Bayesian Information Criterion (Gaussian case)	Bayesian information criterion - Wikipedia	Maximum of second differences between levels of the index	Authors' Stata code
6	dic	Deviance Information Criterion	Deviance information criterion - Wikipedia	Maximum of second differences between levels of the index	Authors' Stata code
7	lnrss	log(Residual Sum of Squares), a sub-component of the Akaike Information Criterion	Akaike information criterion - Wikipedia	Maximum of second differences between levels of the index	Authors' Stata code
8	rmsstd	Root Mean Square Standard Deviation Index	Wilkinson et al, (Chapter 4 of Systat manual)	Maximum of second differences between levels of the index	Authors' Stata code

N	Short Name	Full Name	Source	Decision Rule	Code
9	ch	Calinski and Harabasz	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package
10	duda	Je(2)/Je(1)	Charrad et al (2014)	Smallest nc such that index > criticalValue	Charrad et al (2014) NbClust R package
11	pseudot2	Pseudo t^2	Charrad et al (2014)	Smallest nc such that index < criticalValue	Charrad et al (2014) NbClust R package
12	cindex	C-index	Charrad et al (2014)	Minimum value of the index	Charrad et al (2014) NbClust R package
13	beale	Beale	Charrad et al (2014)	nc such that critical value of the index >= alpha	Charrad et al (2014) NbClust R package
14	ccc	Cubic Clustering Criterion (CCC)	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package
15	ptbiserial	Point-Biserial	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package
16	db	Davies and Bouldin	Charrad et al (2014)	Minimum value of the index	Charrad et al (2014) NbClust R package
17	frey	Frey and Van Groenewoud	Charrad et al (2014)	the cluster level before that index value < 1.00	Charrad et al (2014) NbClust R package
18	hartigan	Hartigan	Charrad et al (2014)	Maximum difference between hierarchy levels of the index	Charrad et al (2014) NbClust R package
19	ratkowsky	c/k^1.5	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package
20	scott	n log (|T|/|W|)	Charrad et al (2014)	Maximum difference between hierarchy levels of the index	Charrad et al (2014) NbClust R package

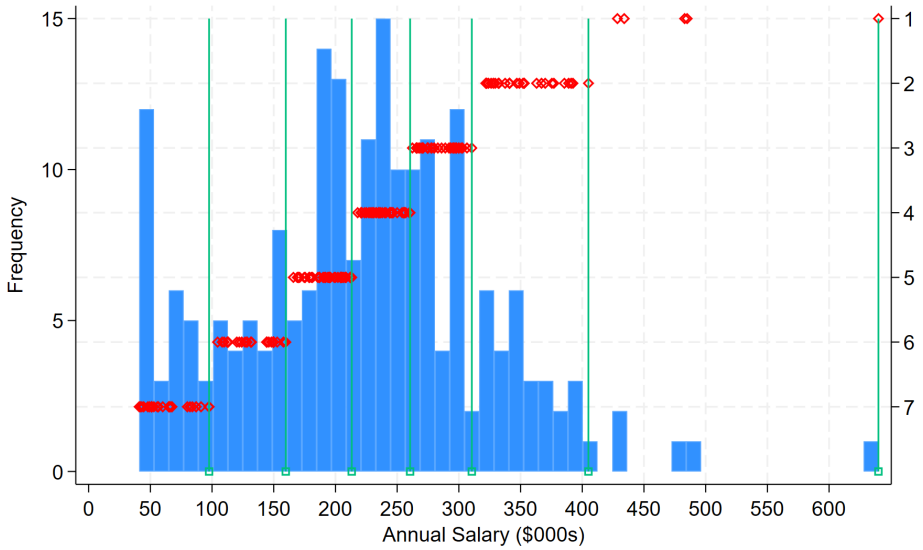
N	Short Name	Full Name	Source	Decision Rule	Code
21	marriot	k^2 W	Charrad et al (2014)	Max. value of second differences between levels of the index	Charrad et al (2014) NbClust R package
22	ball	Ball and Hall	Charrad et al (2014)	Maximum difference between hierarchy	Charrad et al (2014) NbClust R package
23	trcovw	Trace Cov W	Charrad et al (2014)	Maximum difference between hierarchy levels of the index	Charrad et al (2014) NbClust R package
24	tracew	Trace W	Charrad et al (2014)	Maximum value of absolute second differences between levels of the index	Charrad et al (2014) NbClust R package
25	friedman	Trace W^-1 B	Charrad et al (2014)	Maximum difference between hierarchy levels of the index	Charrad et al (2014) NbClust R package
26	mcclain	McClain and Rao	Charrad et al (2014)	Minimum value of the index	Charrad et al (2014) NbClust R package
27	rubin	|T|/|W|	Charrad et al (2014)	Minimum value of second differences between levels of the index	Charrad et al (2014) NbClust R package
28	kl	KL	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package
29	silhouette	Silhouette	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package
30	dindex	D	Charrad et al (2014)	Graphical method	Charrad et al (2014) NbClust R package
31	dunn	Dunn	Charrad et al (2014)	Maximum value of the index	Charrad et al (2014) NbClust R package

N	Short Name	Full Name	Source	Decision Rule	Code
32	hubert	Modified Statistic of Hubert	Charrad et al (2014)	Graphical method	Charrad et al (2014) NbClust R package
33	sdindex	SD	Charrad et al (2014)	Minimum value of the index	Charrad et al (2014) NbClust R package
34	sdbw	SDbw	Charrad et al (2014)	Minimum value of the index	Charrad et al (2014) NbClust R package
35	gmm	Gaussian Mixture Model	Dempster, Laird, and Rubin (1977)	Minimum of second differences between levels of the index	Scikit-learn, Python (Pedregosa at al, 2011)

A.3 Validation using title-linked salary data

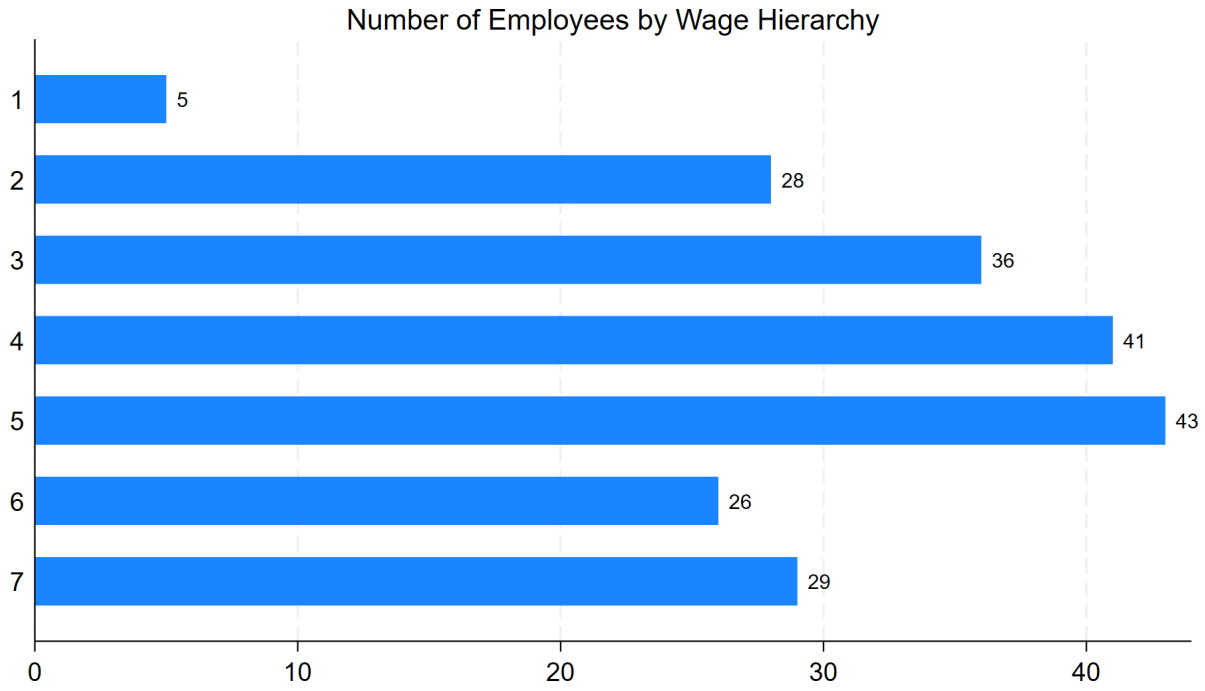
We undertake analysis of wage clusters in a dataset for one year of salary data from a business school of a midwestern public university (with 208 employees), using the ***dcalinski approach*** for picking optimal number of clusters (given the superior performance of the metric per the tests above).

(a) Density function and clusters: First, we present a histogram of the wage data, with the cutoffs for the optimal number of clusters. The dcalinski approach suggests that there are 7 wage hierarchies in this organization.



As can be noted, the dcalinski approach appears to pick cutoffs for the clusters around big dips in the probability density. In the lowest hierarchy level (level 7) are employees with wages less than 97.6K, the top (level 1) are employees with salary in excess of $401K.

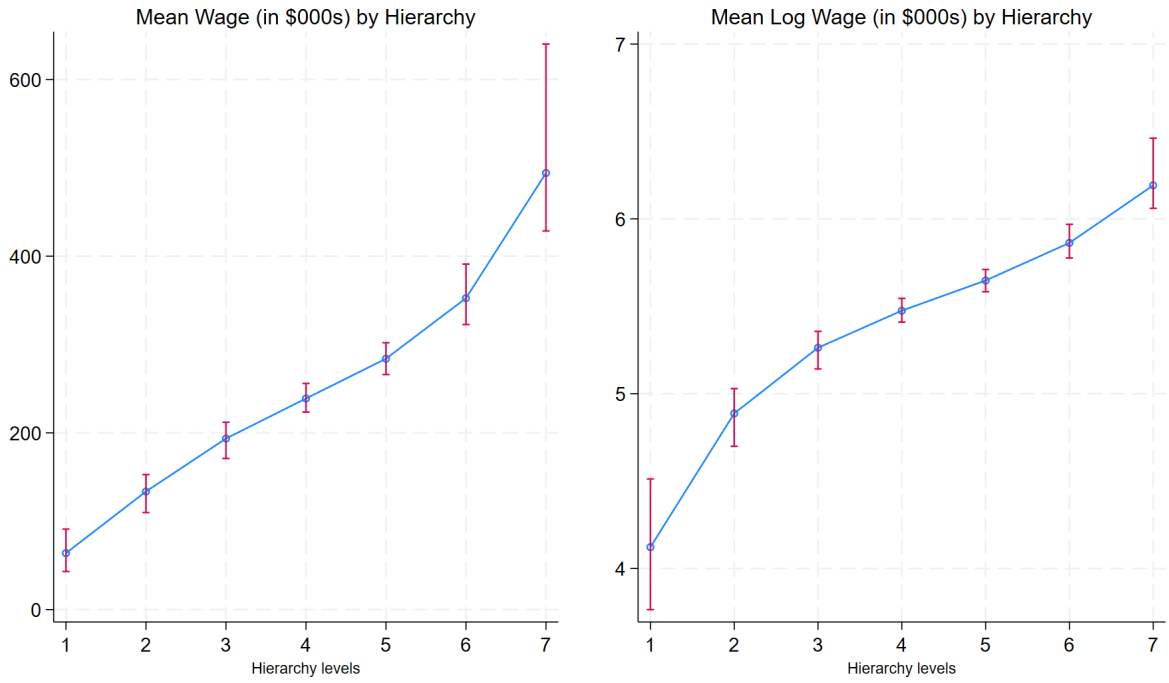
(b) Shape of wage hierarchy: We find that the structure of the wage hierarchies is not pyramidal; the largest size is in level 5 (43 employees), while two of the lower levels have smaller numbers.



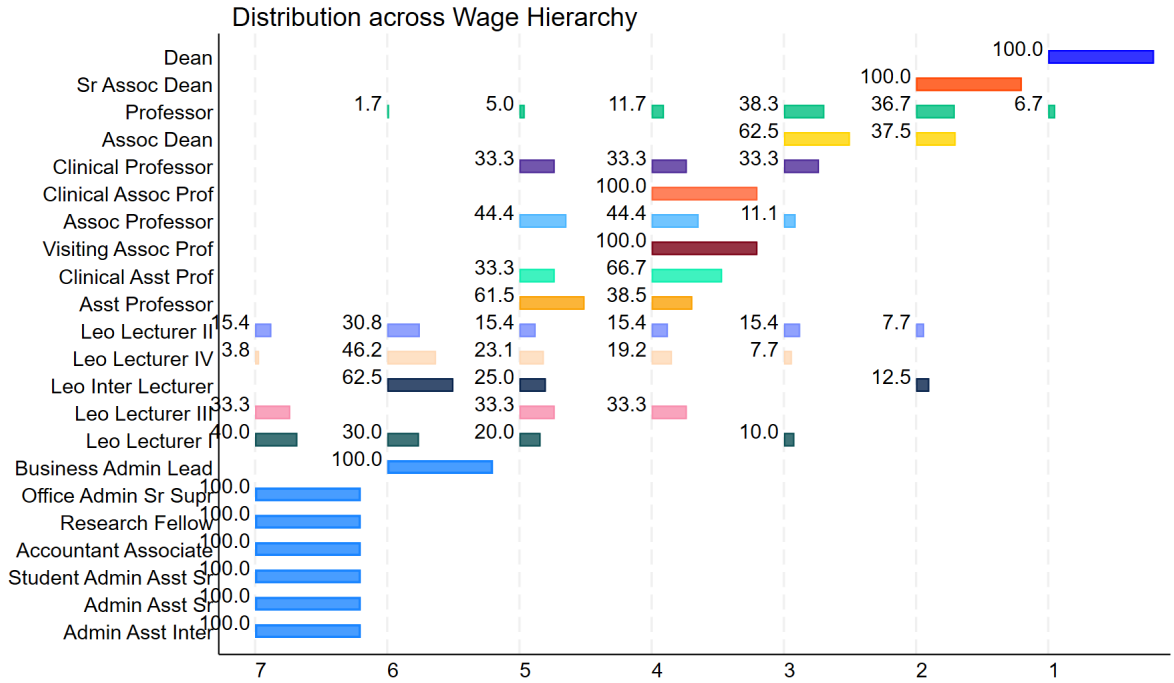
(c) Wage distribution across levels: The figure below presents the average wage in each level. The figure suggests convexity in wage across layers, especially at higher wage levels (consistent with Figure 11.1 in Lazear and Gibbs (2014)).

The log wage figure (on the right) suggests an initial concavity, with a large percentage increase in wage between the lowest and the next higher levels, and drops in percentage wage differences across until about the third level.

Notably, the convex part of the wage figure corresponds to the part of the wage hierarchy (levels 4 through 1) which have the traditional pyramidal shape expected in traditional hierarchy structures. (As in Figure 11.1 in Lazear and Gibbs (2014), there is some evidence of increase in spread at higher hierarchy levels.)



(d) Job titles across levels: The following figures present information on the composition of the wage hierarchy.



The figures illustrate three points. One, the wage hierarchies appear to pick up traditional title hierarchies, especially at higher wage hierarchy levels. E.g., Assistant professors are confined to levels 5 and 4, whereas full professors are largely in layers 3 and 2. In particular, the topmost layer picks up the Dean and (likely senior) professors.

Two, the consistency of wage hierarchies with title hierarchies is imperfect. E.g., there is considerable overlap between assistant and associate professors, with both the groups distributed across levels 5 and 4. This likely reflects heterogeneity in wages across subgroups of the business school, driven by market factors. In particular, new junior faculty in some fields (e.g., finance) can get paid significantly more than those in other fields, so that junior faculty (assistant professors) in higher wage groups may have significant overlap with associate professors in other groups. This feature may not be unique to this organization; it is plausible that titles have heterogenous meanings across different sub-groups in organizations (e.g., director in information technology group vs director in human resources).

Three, the bottom rungs of the hierarchy appear to be dominated by support staff, whose roles only indirectly support the core educational objectives of organization. Again, this could be a more general feature, in that the bottom or particular layers in a wage hierarchy may include employees engaged in non-core or support activities. To the extent that theories of organizational structure typically consider only employees engaged in core activities, the presence of non-core employees in wage data could imply that the empirical distributions (e.g., of wages or number of employees across layers) may deviate from predictions of traditional models.

Appendix B: Model Results

First, the consumer allocates expenditure at the composite product level (upper tier). Utility maximization subject to income constraint yields:

$$\max_{X_j} \sum_{j=0}^{J} \beta_j \log Q_j + \lambda \left(Y - \sum_j P_j Q_j \right)$$

FOC for Q_j yields:

$$\beta_j = \lambda P_j Q_j$$

Then the assumption that $\sum_j \beta_j = 1$, and assuming consumers use up all income so $Y = \sum_j P_j Q_j$ implies that $\lambda = \frac{1}{Y}$, and in equilibrium:

$$E_j = P_j Q_j = \beta_j Y$$

At the lower level, within product Q_j, consumer has to allocate among goods $q_j(\omega)$. Cost minimization for the CES utility function $\left[\int_{\omega \in \Omega_j} q_j(\omega)^\rho d\omega\right]^{\frac{1}{\rho}}$ yield P_j, the minimum cost of purchasing one unit of the composite good $Q_j$is given (from standard CES utility function theory): as:

$$P_j = \left[\int_{\omega \in \Omega_j} p_j(\omega)^{1-\sigma} d\omega\right]^{\frac{1}{1-\sigma}}$$

Note $\sigma = \frac{1}{1-\rho}$, and $\rho = \frac{\sigma-1}{\sigma}$. The derivation of this price index is as follows. To find the minimum expenditure $E_j$required to purchase one unit of Q_j, we solve:

$$\min_{q_j(\omega)} \int_\omega p_j(\omega) q_j(\omega) d\omega\,, \quad \text{subject to: } \bar{Q}_j^\rho = \left[\int_{\omega \in \Omega_j} q_j(\omega)^\rho d\omega\right]$$

Taking the derivative of the lagrangian wrrt $q_j(\omega)$ and setting to zero yields:

$$p_j(\omega) = \lambda \rho q_j(\omega)^{\rho-1}$$

$$\rightarrow\ q_j(\omega) = \left[\frac{p_j(\omega)}{\lambda\rho}\right]^{\frac{1}{\rho-1}} = (\lambda\rho)^\sigma p_j(\omega)^{-\sigma}$$

Then we have:

$$\bar{Q}_j^{\rho} = \int_{\omega\in\Omega_j} q_j(\omega)^{\rho} d\omega = (\lambda\rho)^{\sigma\rho} \int_{\omega\in\Omega_j} \left(p_j(\omega)^{-\sigma}\right)^{\rho} d\omega = (\lambda\rho)^{\sigma\rho} \int_{\omega\in\Omega_j} \left(p_j(\omega)^{1-\sigma}\right) d\omega$$

$$= (\lambda\rho)^{\sigma\rho} P_j^{1-\sigma}$$

Where we simply define: $P_j^{1-\sigma} = \int_{\omega\in\Omega_j} \left(p_j(\omega)^{1-\sigma}\right) d\omega$.

Now:

$$E_j = \int_{\omega} p_j(\omega) q_j(\omega) d\omega = \int_{\omega} p_j(\omega) (\lambda\rho)^{\sigma} p_j(\omega)^{-\sigma} d\omega = (\lambda\rho)^{\sigma} \int_{\omega} p_j(\omega)^{1-\sigma} d\omega$$

$$\rightarrow E_j = \frac{\bar{Q}_j}{P_j^{\frac{1-\sigma}{\rho}}} P_j^{1-\sigma} = \bar{Q}_j P_j$$

$$\rightarrow P_j = \frac{E_j}{\bar{Q}_j}$$

Appendix C: Additional Figures and Tables

Figure C1: Distribution of Hierarchy Sizes: Maximum of Caliński Metric

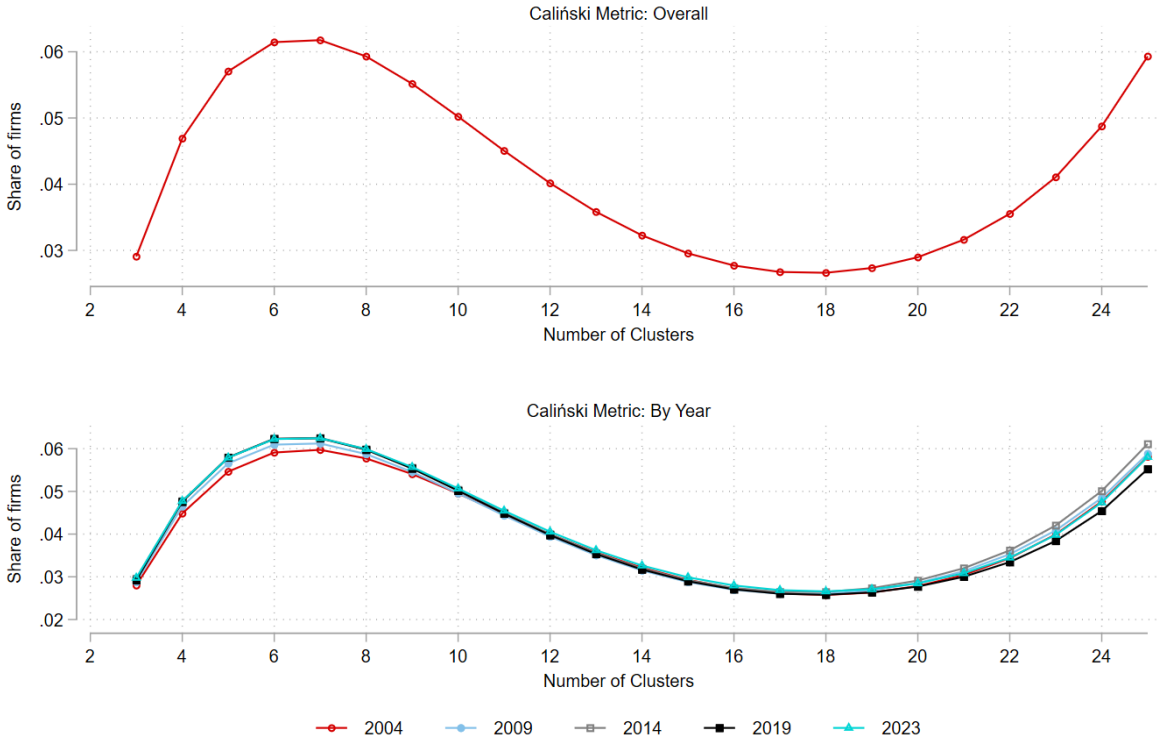


Figure C2: Hierarchy Sizes across Sectors: Maximum of Caliński Metric

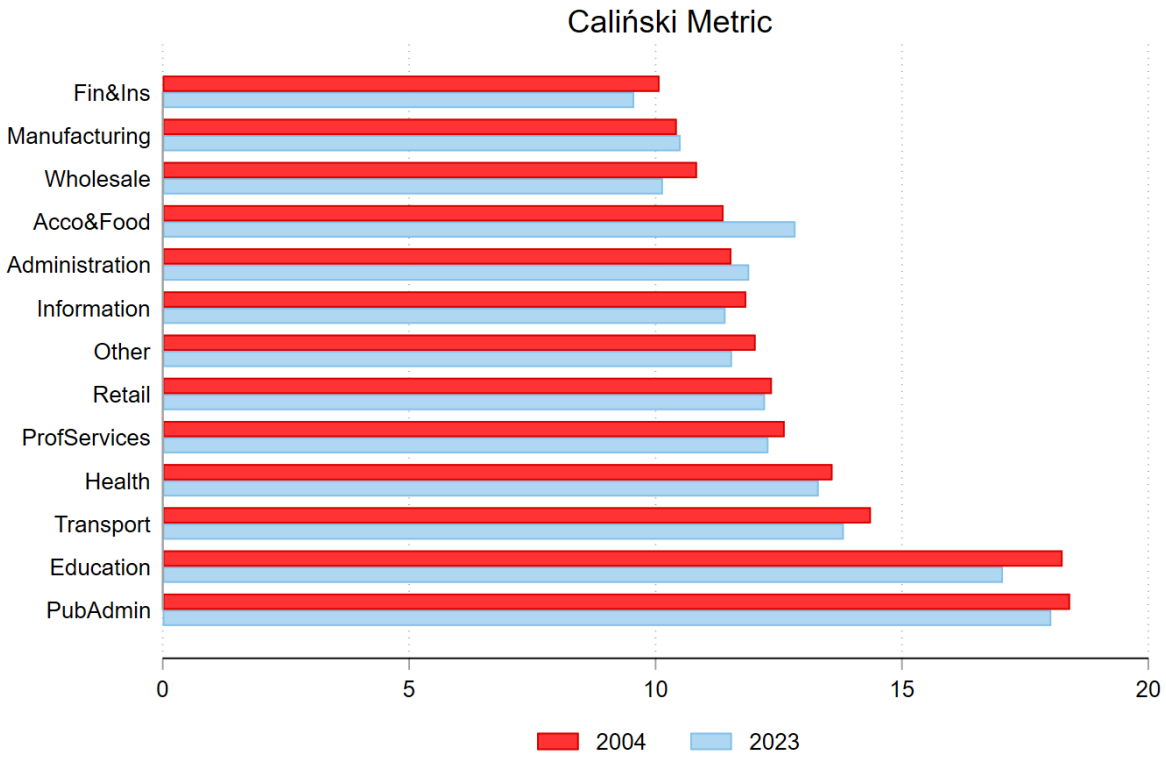
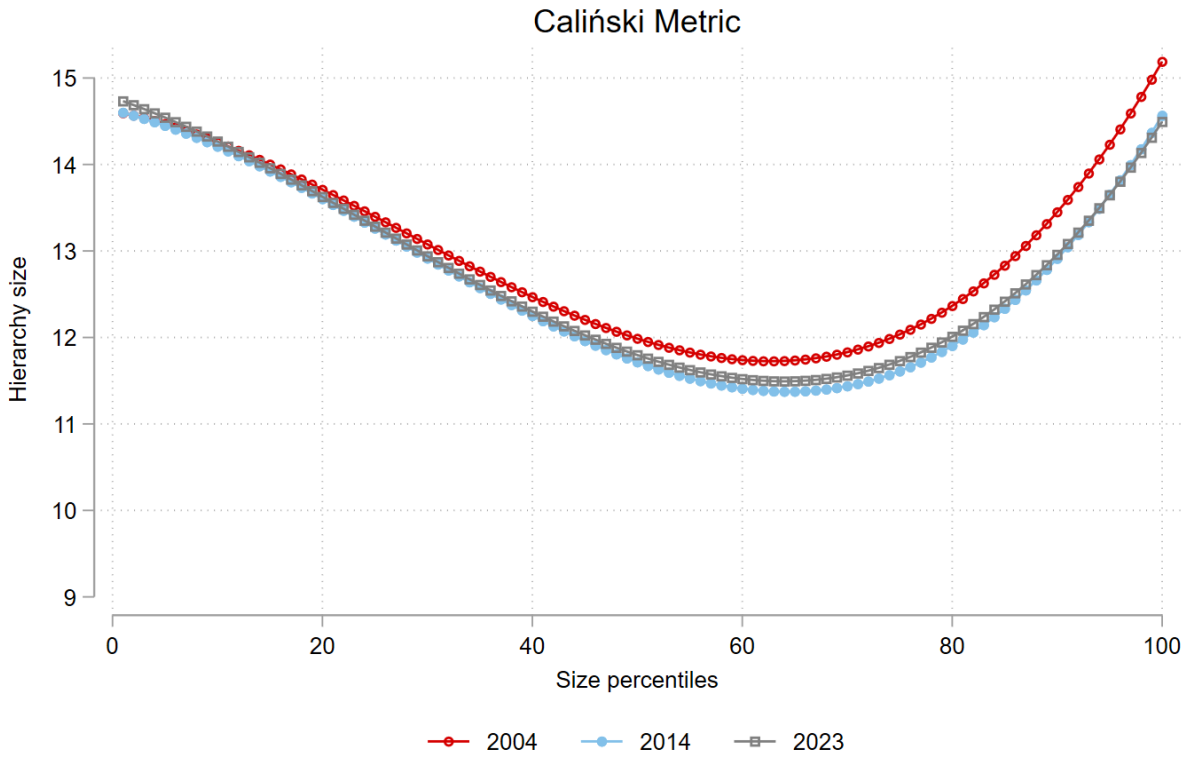


Figure C3: Hierarchy Size vs Firm Size: Maximum of Caliński Metric



Notes: This figure presents relationship of hierarchy size to percentiles of firm size. Firm hierarchy size was regressed on a cubic polynomial in firm size percentiles with NAICS 3-digit fixed effects.

Figure C4: Hierarchy Size vs Firm Size: Maximum of Caliński Metric

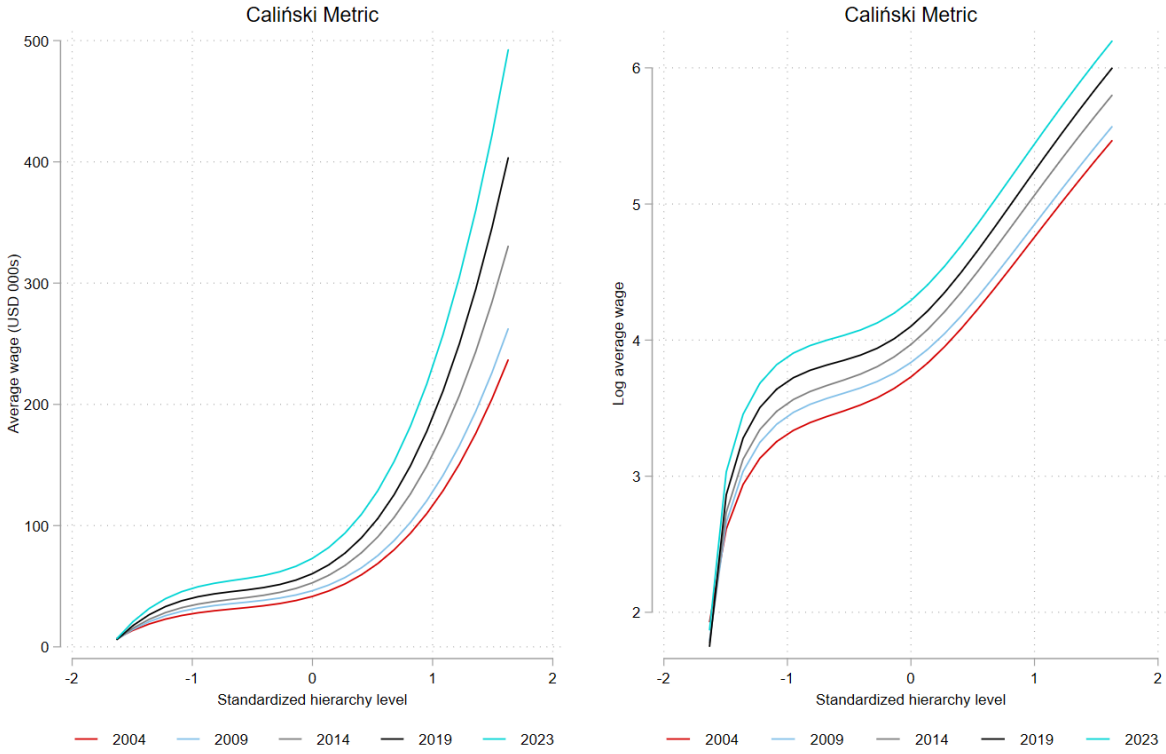


Table C1: Correlation of Hierarchy Size with Future Firm Performance

	Five Year Survival Dummy		Five Year Employment Growth		Five Year Payroll Growth	
	(1)	(2)	(3)	(4)	(5)	(6)
(Number of Clusters by Jump Metric)/100	-0.04123*		-0.2236***		-0.1801**	
	(0.01839)		(0.06204)		(0.05898)	
(Number of Clusters by Caliński (Max D2) metric)/100		-0.03508*		-0.1587***		-0.1281**
		(0.01401)		(0.04327)		(0.0411)
Adjusted R-square	0.071	0.071	0.091	0.091	0.09	0.09
Number of Observations (Rounded)	243000	243000	243000	243000	243000	243000
Fixed Effects	State-NAICS3-Year-Initial Size Quintile-MultiUnit		State-NAICS3-Year-Initial Size Quintile-MultiUnit		State-NAICS3-Year-Initial Size Quintile-MultiUnit	
Mean of DV	0.9193	0.9193	9.728	9.728	9.852	9.852
SD of DV	0.2724	0.2724	0.7215	0.7215	0.7621	0.7621
Mean of Number of Clusters/100	0.1281	0.1175	0.1281	0.1175	0.1281	0.1175
SD of Number of Clusters/100	0.06996	0.06931	0.06996	0.06931	0.06996	0.06931
Effect Size (1SD RHS as proportion of DV mean)	-0.31%	-0.26%	-0.16%	-0.11%	-0.13%	-0.09%
Effect Size (1SD RHS as proportion of DV SD)	-1.06%	-0.89%	-2.17%	-1.52%	-1.65%	-1.17%