Internal Labor Markets and the Competition for Managerial Talent

Benjamin Friedrich
Northwestern University*

August 1, 2017

Abstract

This paper studies how firms use internal promotions and external hiring to recruit managers. Using matched employer-employee data from Denmark, I document that more productive firms hire more talented trainees, are more likely to promote managers internally, and match with better managers in terms of education and ability. Based on these facts, I develop an assignment model of the market for managers with two-sided heterogeneity. In the model internal labor markets arise from asymmetric learning and firm-specific human capital. Production complementarities between firm productivity and manager talent result in better firms investing in promising workers and in developing talent through firm-specific training and internal promotion. I estimate the model using Danish data. Model simulations indicate that reducing information frictions increases aggregate productivity but leads to higher wage inequality because better signals of talent increase competition for the best managers. This mechanism provides a new market-driven explanation for the increase in upper-tail wage inequality.

*Northwestern University, Kellogg School of Management, 2211 Campus Drive, Evanston, IL 60208, Tel.: 1-203-809-7254, e-mail: benjamin.friedrich@kellogg.northwestern.edu.

Acknowledgment: I am grateful to my advisers Joseph Altonji, Lorenzo Caliendo, Penny Goldberg, and Costas Meghir for their invaluable support. I also thank Noriko Amano, Lisa Kahn, Rebecca McKibbin, Corina Mommaerts, Emily Nix, Marcelo Sant’Anna and seminar participants at Carnegie Mellon, CEMFI, Chicago Booth, IIES, Northwestern Econ, Northwestern Kellogg, Princeton, UPenn and Yale as well as the EMC 2016 and SOLE 2017 for helpful comments and discussions. I thank the Labor Market Dynamics Group (LMDG), Department of Economics and Business, Aarhus University and in particular Henning Bunzel for support and making the data available. LMDG is a Dale T. Mortensen Visiting Niels Bohr professorship project sponsored by the Danish National Research Foundation. My research has been supported financially by the Yale Economics Department. All errors are my own.
1 Introduction

Hiring competent and suitable employees is a crucial pillar of firms’ success. Good managerial practices emphasize human capital management and promotion to develop and retain high performers (Bloom and Van Reenen 2007). Managerial hiring is particularly relevant because managers make strategic decisions that affect the productivity of the entire workforce. Understanding how firms recruit and develop managerial talent is important in understanding firm performance, and in turn aggregate productivity.

Firms use internal labor markets - as well as external hiring - to recruit managers. Firm-specific training and superior information about internal candidates help firms to develop managers internally instead of competing for external managers with high general skills. The extent of market competition for managers is closely linked to wages. Understanding the (changing) role of internal labor markets in the competition for talent is therefore important in understanding the continuing rise in manager compensation (Frydman and Jenter 2010) and in upper-tail wage inequality (Autor et al. 2008).

Despite its importance, firm hiring has received little attention from scholars largely due to data limitations (Oyer and Schaefer 2011). Using a uniquely detailed dataset, this paper studies how firms use internal promotions and external hiring to recruit managers. I establish new stylized facts about managerial hiring that emphasize differences across firms. Motivated by these facts, I model how heterogeneous firms compete for talent in an environment with internal labor markets. In particular, I analyze the role of asymmetric information and firm-specific training for promotion and hiring decisions and I characterize the equilibrium compensation and sorting pattern of talent across firms. I then estimate the model to assess how internal labor markets affect wage inequality and the (mis)allocation of resources.

The paper first uses matched employer-employee data from Denmark 1999-2008 to document differences in managerial hiring across firms. I use occupational switching and job-to-job mobility for managers to show that firms differ widely in the use of internal versus external hiring of managers. These differences are systematically linked to labor productivity and total factor productivity; conditional on firm size, firms in the top decile of productivity are about 20 percentage points more likely to use internal promotion than firms in the bottom decile. These results are robust to taking job characteristics, workforce composition, firm growth, timing of hiring decisions and local

\(^1\)Oyer and Schaefer (2011) state that “The relative weakness of the hiring literature is due to idiosyncrasies and data limitations.” Their paper argues that “What is lacking is (a) documentation of across-firm variation in hiring strategies, (b) linkage of this across-firm variation in strategy to firm-level characteristics, and (c) a tie from these facts back to theory.”
labor market conditions into account.

I complement this evidence on managerial hiring with facts about talent recruitment and sorting between managers and firms. First, managers at more productive firms have higher average education, in particular a higher share of postgraduate degrees. Second, I estimate individual unobserved ability as a time-invariant component of wages over the entire career following Abowd et al. (1999). I find evidence of positive sorting with respect to ability of managers across the firm productivity distribution. Third, the positive sorting pattern begins before promotion; more productive firms recruit candidates with higher education and unobserved ability. Fourth, trainees with higher ability are promoted into management positions internally, while candidates with lower ability become managers by switching firms. Average unobserved ability of externally hired managers is significantly lower than for internal promotions. This suggests that firms have superior knowledge about internal candidates.

Based on these stylized facts, I develop an equilibrium model of the market for managers with two-sided heterogeneity. The model illustrates the tradeoffs that firms face when making hiring decisions in a market environment. I assume that firms consist of two hierarchical layers, one manager and a group of workers. Individuals live for two periods and becoming a manager requires a trainee period for young workers. The market observes a signal of managerial talent for each young individual. Firms use internal labor markets because trainees can acquire firm-specific knowledge and because firms learn superior information about their internal candidates relative to the market. These two channels generate rents from internal promotions because managers are paid their best outside option from taking an alternative manager job. The cost of internal promotion includes training costs and the probability of failure, as well as the risk of investing in a low ability trainee. If an internal candidate is revealed unprofitable for promotion, firms compete for managers externally. Asymmetric employer learning leads to adverse selection among external candidates and affects firms’ optimal hiring and training decisions. Production complementarities between firm productivity and manager talent imply that better firms benefit more from better managers. Consequently, a better firm hires more promising trainees and invests more in firm-specific training to improve their managerial skills; both mechanisms suggest that better firms use more internal promotions.

The model has implications for the effect of internal labor markets on wage inequality and the allocation of resources. First, if firm-specific knowledge is more valuable, firms can use internal training to develop good internal candidates with low general skills; this reduces market competition for managers with high general skills. Lower
competition for high skill managers reduces wage differences across managers and lowers the wage gap between managers and workers. Second, more precise signals about true managerial talent, measured by the correlation between observed skills and total talent, increase market competition for the best managers. As a result, wage dispersion among managers increases and the wage gap between managers and workers widens. There is an equity-efficiency tradeoff because information frictions both compress the wage distribution and prevent perfectly assortative matching based on true talent across firms.

I estimate the model using Danish firms with at least 50 employees to quantify the role of internal labor markets for wage inequality and aggregate productivity. I use the method of simulated moments with a Markov-Chain-Monte-Carlo (MCMC) algorithm following Chernozhukov and Hong (2003) to match the model to the empirical distributions of value added, managerial span of control (number of workers per manager), manager and trainee earnings and the share of external hiring across firms. The model fits the data well and the estimates reveal the relative importance of employer learning and firm-specific human capital across industries. Firm-specific human capital is most important in manufacturing - about twice as valuable as in retail and wholesale - but it cannot fully replace a high ability manager.

I use the estimated model to first quantify misallocation of resources due to information frictions. I find an output elasticity with respect to information frictions about managerial talent of 0.11. The full information economy, where true managerial talent is common knowledge at labor market entry, generates a 22.5% increase in aggregate productivity compared to the asymmetric information benchmark. The productivity gain is realized through reallocation of resources from low to high productivity firms as managers sort across firms based on true managerial talent.

The extent of reallocation depends on the size of information frictions and on the relative importance of firm-specific human capital. I find that the productivity gain is only 12.5% in manufacturing, but 28.1% in wholesale and retail and 24.6% in the business sector. Productivity gains in manufacturing are lower because observable skills are a more precise signal of total managerial talent than in other sectors. As a result, even under asymmetric information, manufacturing firms are making better informed decisions than firms in business and the retail sector. Moreover, firm-specific human capital is relatively more important in manufacturing. Consequently, less productive manufacturing firms are better protected from market competition compared to other sectors.

The counterfactual illustrates the equity-efficiency tradeoff of alleviating informa-
tion frictions and the role of asymmetric information for wage inequality in the market for managers. The productivity gain under perfect information comes at the cost of increased inequality both among managers and with respect to the wage gap between managers and workers. More precise signals of managerial talent intensify market competition for the most talented individuals and lead to a large increase in compensation for top talents. This mechanism - in combination with evidence of reduced information frictions through recruitment websites, social media, and increasing use of executive search firms (headhunters) and machine learning techniques in hiring - provides a new market-driven explanation for the increase in upper-tail wage inequality.

I provide additional counterfactuals to show more generally how a reduction in the attractiveness of internal labor markets can help explain three secular trends in the U.S. and Denmark: (1) an increase in manager compensation, (2) an increase in the wage gap between managers and workers and (3) a decrease in internal promotions. In addition to reductions in information frictions, I consider a decrease in the value of firm-specific human capital motivated by homogenization of management practices and increased transferability of skills. I also study an increase in the productivity of high ability managers that may follow from innovations in monitoring and communication technology. All scenarios lead to a larger average wage gap between managers and workers. Moreover, all counterfactuals decrease the share of internal promotions and lead to more external hiring. Average manager compensation increases in particular for an increase in the value of high ability managers.

I compare these effects to an upward shift in the firm productivity distribution. Consistent with the previous literature, this simulation shows large elasticities of average manager salaries with respect to average firm size (Tervio 2008; Gabaix and Landier 2008). Yet the effects on the external hiring share and wage inequality between managers and workers are small. The simulation suggests that the observed secular trends are consistent with firm growth and a simultaneous reduction in the attractiveness of internal labor markets.

This paper relates to several strands of the literature on internal labor markets and the market for managers. The paper complements the literature on the internal labor market of one particular firm (Lazear 1992; Baker et al. 1994) or a small sample of firms (Lazear and Oyer 2004) by using administrative matched employer-employee data to compare hiring strategies across firms within and across industries. My focus on managerial hiring is related to the literature on the importance of managers for firm performance that documents the impact of CEOs (Bertrand and Schoar 2003; Perez-Gonzalez 2006; Bennedsen et al. 2007) and the effect of supervisors on worker
productivity (Lazear et al. 2015). I complement existing management surveys (Bloom and Van Reenen 2007; Bloom et al. 2014) by using matched panel data to study manager careers and to illustrate how firms recruit and promote promising talents.\(^2\)

The literature on internal labor markets has developed different competing theories about the existence of internal labor markets, in particular emphasizing the role of firm-specific human capital (Becker 1962; Jovanovic 1979; Demougin and Siow 1994) and asymmetric employer learning (Waldman 1984; Greenwald 1986; Gibbons and Katz 1991; Acemoglu and Pischke 1998). Yet these studies typically ignore firm heterogeneity and consider a partial equilibrium setting. My paper is the first to study the effects of asymmetric information and firm-specific human capital on hiring and promotion strategies in an equilibrium model with heterogeneous firms and managers.\(^3\) Moreover, the empirical literature has almost exclusively considered human capital accumulation and employer learning separately.\(^4\) I estimate the model to quantify the relative importance of information frictions and firm-specific human capital for the Danish economy.

Finally, the implications of the model relate to the literature on the market for managers. Murphy and Zabojnik (2004) and Frydman (2005) argue that a reduction in the importance of firm-specific knowledge relative to general manager ability can help explain the increase in manager compensation, see also the extended analysis in Murphy and Zabojnik (2007).\(^5\) Another strand of the literature discusses moral hazard in the market for managers, which motivates large efficiency wages to prevent shirking (Gayle et al., 2015). I complement these studies by emphasizing the role of hidden information for manager compensation and wage inequality. I model how the interaction of information frictions and firm-specific human capital shapes the competition for managerial talent in general equilibrium and derive implications that help reconcile different secular trends in the market for managers.

\(^2\)There is a related literature on hiring practices. See Burks et al. (2015) for recent work on employee referrals, Hoffman et al. (2015) on discretionary versus standardized hiring procedures and Pallais (2014) on inefficiencies in hiring low-skill workers. Pallais and Sands (2016) use randomized experiments to quantify the informational value of referrals for low-skill workers.

\(^3\)I abstract from imperfect monitoring of effort and optimal promotion tournaments (Lazear and Rosen 1981) within the firm to focus on the interaction between internal and external labor markets. Including moral hazard into this framework is a challenge for future research.


\(^5\)The main explanations for the increase in manager compensation in the literature are changes in the size distribution of firms that affect demand for talent (Gabaix and Landier 2008; Tervio 2008) or rent extraction of managers (Bebchuk and Fried 2003).
The rest of the paper is structured as follows. Section 2 introduces the data and provides stylized facts about heterogeneous hiring strategies, the relationship between firm productivity and internal promotions, and about sorting between managers and firms. Section 3 develops an assignment model of the market for managers with two-sided heterogeneity, asymmetric employer learning and firm-specific human capital. I estimate the model in Section 4 and provide counterfactual analysis to quantify the costs of information frictions and the relationship between internal labor markets and wage inequality in Section 5. Section 6 concludes.

2 Stylized Facts on Managerial Hiring

2.1 Data Sources and Main Sample

The main data source is administrative matched employer-employee data (IDA) for the universe of workers and firms in Denmark 1980-2008. The data contain detailed information about the career history of workers, formal education, as well as hourly wages and annual earnings. Starting in 1991, I observe occupations for the primary job of each worker per year as reported by their employer. Occupations are defined based on the international standard classification of occupations (ISCO). Management occupations consist of three different types of manager positions - top executives, production managers and department managers. Occupational switching and employer switching of workers allow me to measure managerial hiring at each firm. External hires enter the firm as managers, whereas internal promotion is defined as a new manager who was employed at the same firm in previous years with a non-managerial occupation. The firm data contain detailed balance sheet and industry information.6

The main sample consists of all private sector firms with at least 50 full-time equivalent workers each year over 1999-2008. The size cutoff follows the convention in the literature (see Bloom and Van Reenen 2007, for example) to selects firms with a meaningful role for internal labor markets. Moreover, I limit my attention to firms with at least ten managerial hires over the sample period. This threshold reduces small sample bias in the subsequent analysis of managerial hiring within firms. The full sample consists of 953 firms; these firms represent two-thirds of value added, 61.1% of employment, and 73.3% of manager hours worked by all firms above the size cutoff over the period 1999-2008. I report descriptive statistics about the sample in Table 1. The number of firms by industry will enable me to compare hiring behavior for firms within narrowly

---

6I provide details on definitions and data cleaning in Appendix A.1.
defined industries. Most firms are in the sample for the full time period and they hire new managers almost every year.

### 2.2 Institutional Background

Danish labor market regulation is flexible in terms of hiring and firing, which means that institutional restrictions have limited influence on internal labor markets. I find that 24.1% of all private sector employees switch jobs per year over the sample period 1999-2008. This transition rate is comparable to job-to-job mobility in the U.S. where Moscarini and Thomsson (2007) report 3.2% mobility per month in the CPS over the period 1994-2006. Moreover, decentralization of wage bargaining over the 1980s and 1990s has increased the share of workers who directly negotiate wages with their employer (see Dahl et al. 2013). This is true in particular for manager occupations that are not covered by industry or occupation-based wage bargaining. In Denmark, there is a specific trade union for managers and executives (Lederne) that advises managers about career development and individual wage negotiations, provides legal council and unemployment insurance.

### 2.3 Stylized Facts

In this section, I first analyze managerial hiring strategies across firms in different industries and I provide new evidence on sorting between managers and firms.
Fact 1: Internal versus external hiring decisions for managers vary across firms.

I first compute the share of external hiring for all new manager positions by firm over 1999-2008. The left panel of Figure 1 shows the distribution of external hiring shares for different sectors. The striking fact from this figure is that there are wide differences across firms within an industry but the differences across industries are small. Notice that the sample only includes firms that fill at least ten new manager positions over the sample period and many of them hire dozens of new managers as illustrated in Table 1. Nevertheless, the share of firms with very low or very high external hiring shares is surprisingly large; the standard deviation of the external hiring share within sectors is 23 percentage points. Technology and institutional differences across industries cannot explain large differences in internal versus external hiring within sectors. As the right panel shows, the best fit for a Binomial distribution does not come close to replicating the wide dispersion within sector. Kolmogorov-Smirnov tests for the full sample and by industry all strongly reject a simple Binomial distribution as the data generating process.

I further emphasize this finding of heterogeneous managerial hiring strategies by

---

7Manufacturing includes food, textiles, wood products, chemicals, mineral products, metal products and furniture. Wholesale and retail includes the sale of motor vehicles, wholesale and other retail trade. Finance and business consists of real estate, consulting and legal activities.

8I nonparametrically test the hypothesis of an underlying Binomial distribution by comparing the empirical distribution to the normal limit distribution $N(np, np(1-p))$ where $n$ is the number of hires by firm and I use the sample average external share as an estimate for $p$. The $p$-value from this test for all samples is less than 0.0005. Standard rules of thumb suggest that the normal approximation is appropriate for the case of $n \geq 10$ and $p \approx 0.5$, but I confirm the robustness by restricting the sample to larger cutoffs as well.
estimating a linear probability model of external managerial hiring,

\[ H_{ijt} = \beta X_{it} + \gamma Z_{jt} + u_j + u_{rt} + u_{st} + \epsilon_{ijt}. \]  

(1)

\( H_{ijt} \) takes the value 1 if firm \( j \) hires manager \( i \) at time \( t \) externally and value 0 for internal promotions. I control for job characteristics \( X_{it} \), firm characteristics \( Z_{jt} \) such as lagged employment, employment growth, turnover rate among employees, lagged sales and assets, as well as industry-time and region-time effects \( u_{rt}, u_{st} \). I include firm-fixed effects \( u_j \) to measure the residual share of external hiring by firm. Using different methods from the literature on teacher value added such as shrinkage estimates and a two-step procedure as in Chetty et al. (2014) to account for sampling error, I find that the residual dispersion in the share of external hiring across firms within industries remains large and economically important, ranging from 16.5 to 19.4 percentage points.\(^9\)

**Fact 2: Conditional on size, more productive firms use more internal promotions**

Next, I analyze which firm characteristics can explain this variation in internal versus external hiring across firms even within industries. Specifically, I regress the estimated firm fixed effects on observable firm characteristics,

\[ \hat{u}_j = \alpha Z_j + \zeta_j. \]  

(2)

I focus on measures of (i) firm productivity, (ii) firm size, (iii) the span of control per manager and (iv) the skill composition of the workforce. Since these characteristics vary over time, I compute firm-level averages over the sample period.\(^{10}\)

Table 2 reports the results. Each row in column (1) reports results for a separate regression using a different firm characteristic. Given the focus of the previous literature on complementarities between managers and firm productivity, I particularly analyze different firm productivity measures, with gross profits per manager as the preferred measure that will have a close counterpart in the subsequent theory and estimation sections.\(^{11}\) The key finding is the strong negative correlation between firm productivity

---

\(^9\)I describe the details of the regression and these different adjustment methods in Online Appendix A.2.2.

\(^{10}\)I compute weighted averages by the number of new positions in each year to account for the fact that the situation at a firm may have been different for managerial hiring decisions in different years. Using lagged firm characteristics in these weighted measures ensures that the timing is not reversed because the situation improved significantly after a particular manager was hired for example.

\(^{11}\)Gross profits are total revenue minus purchases of raw materials, inputs, energy and subcontracting. Value added differs from gross profits because it takes rental and leasing costs, secondary costs
and the share of external hiring. A 1% increase in gross profits per manager is associated with a decrease in the share of external hiring by 0.063 percentage points and differences in this productivity measure explain 7% of residual hiring strategies across firms.\textsuperscript{12} This strong relationship is robust to using alternative measures of productivity such as value added per employee and TFP estimates following Ackerberg et al. (2015). Second, measures of firm size such as sales and employment also show a negative correlation with external hiring shares with estimated coefficients about half as large as for productivity. Third, measures of the pool of internal candidates suggest that firms with more workers per manager, i.e. a larger span of control, are less likely to hire managers externally. Fourth, differences in workforce composition are only weakly related to managerial hiring. Firms with a higher share of workers with vocational training use more internal promotion, but the relationship is not statistically significant.

The results on firm productivity are robust to adding other firm characteristics in terms of size and workforce composition in columns (2)-(4). As a result, this pattern is not explained by more productive firms being larger and mechanically facing a larger internal pool of candidates that increases the probability of internal promotion. Instead, there is a systematic pattern between internal promotions and firm productivity after controlling for position and firm characteristics, in particular firm size.\textsuperscript{13}

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
 & (1) & (2) & (3) & (4) \\
\hline
Profits/Manager & -0.0628*** & 0.068 & -0.0613*** & \\
 & (0.008) & & (0.009) & \\
VA/Employee & -0.0686*** & 0.077 & -0.067*** & \\
 & (0.008) & & (0.009) & \\
Span of Control & -0.0650*** & 0.075 & -0.064*** & \\
 & (0.007) & & (0.008) & \\
TFP & -0.0550*** & 0.099 & -0.056*** & \\
 & (0.022) & & (0.025) & \\
Employment & -0.0290*** & 0.015 & -0.0031 & 0.0256 & -0.0185 \\
 & (0.007) & (0.016) & (0.016) & & (0.018) \\
Sales & -0.0277*** & 0.018 & -0.0027 & -0.0074 & -0.032** & -0.0136 \\
 & (0.006) & (0.015) & (0.014) & (0.014) & (0.016) & \\
Vocational% & -0.0061 & 0.000 & -0.0587 & -0.0475 & -0.0241 & -0.0852 \\
 & (0.050) & (0.051) & (0.050) & (0.050) & & (0.054) \\
\hline
Observations & 953 & 953 & 953 & 953 & \\
R-squared & 0.070 & 0.078 & 0.081 & 0.029 & \\
\hline
\end{tabular}
\caption{Managerial Hiring Strategies and Firm-level Characteristics}
\end{table}

Notes: Standard errors in parentheses. The dependent variable in all regressions is the firm-fixed effect of a regression model (1). Firm characteristics are averages over the sample period, weighted by next period managerial hiring. TFP is estimated by industry following Ackerberg et al. (2015) using capital, labor and energy inputs. All regressions include detailed industry FE as well. Column (1) reports separate regressions for each firm characteristic, columns (2)-(5) include all firm characteristics jointly.

\textsuperscript{12}The R-squared is high despite potential sampling error in the fixed effect estimates. If sampling error accounts for 25% of the variation, then the true explanatory power is one-third higher.

\textsuperscript{13}Probit models at the level of individual hiring decisions provide additional evidence for the role of firm productivity, see the Online Appendix for details.
Figure 2: External Hiring across the Firm Productivity Distribution

Notes: The results are for the full sample in Table 1. Firm productivity is measured as average gross profit per manager over 1999-2008. Gross profits per manager for each firm are computed as a weighted average over the sample period with the share of next period hiring relative to all new manager positions as weights. The figure on the left measures the raw data share of external managerial hiring by firm over the period 1999-2008 and displays a first-order local polynomial regression of these shares on firms’ productivity rank using Epanechnikov kernel with bandwidth 9.03. The figure on the right measures external hiring as the firm FE from regression (1). The fixed effects are normalized to zero for the industry average. The figure displays a first-order local polynomial regression of these FE estimates on firm productivity rank using Epanechnikov kernel with bandwidth 8.55. Confidence intervals are based on 1000 bootstrap replications of the smoothing procedure to yield pointwise standard errors.

Figure 2 provides further nonparametric evidence for the positive relationship between firm productivity and internal promotions. The left panel shows a local polynomial regression of the share of external hiring on firm productivity rank based on gross profits per manager. The figure illustrates that the share of external hiring is significantly lower for firms above the 70th percentile compared to firms that are ranked lower in the productivity distribution. In the right panel of Figure 2, I use the residual hiring shares of firms estimated in equation (1). The plot shows significantly more external hiring compared to the industry average (normalized to zero) among low productivity firms, whereas high productivity firms are about 15 percentage points less likely to use external hiring to fill any managerial position.

Fact 3: Better firms match with better managers, even before the first promotion

This section provides evidence about positive assortative matching between managers and firms. Figure 3 documents that formal education of managers increases along the firm productivity ranking.\textsuperscript{14} The difference in average schooling of managers at low and high productivity firms is about one year. The right panel of Figure 3 shows that more productive firms have significantly more managers with Master degrees, MBAs and PhDs, replacing both college graduates and managers with vocational training or

\textsuperscript{14}All figures use the same firm ranking based on gross profits per manager for consistent comparisons across the empirical analysis. Yet the results are not sensitive to using alternative rankings based on TFP or VA/worker.
Notes: The population consists of all managers who work at firms in the main sample from Table 1. The level of observation is a manager-year to capture the average education level of managers by firm over the sample period. Education and schooling are based on the highest completed education each period and they adjust to additional training over time. Vocational training includes managers whose highest level of education is primary education or a high school degree. The trainee population is defined as all individuals who are employed at firms in the main sample 5 years before their first manager promotion (but promotion can occur at any firm). Unobserved quality is the residual person fixed effect from the AKM regression after controlling for formal education.

The results based on observable formal education are supported by sorting according to characteristics that are unobserved by the econometrician. I run a log wage regression on second-order polynomials in age, experience and tenure, time effects and both individual and firm fixed effects for the universe of workers in Denmark over 1995-2008 following Abowd et al. (1999). I subsequently refer to this estimation as AKM regression. This setting estimates ability as a time-invariant component of wages over the entire career. I subsequently focus on all individuals who ever work as managers.

The AKM regression sample consists of about 3.7 million unique individuals at a total of 304,783 public and private firms. The largest group of firms connected by job movers across firms comprises 291,774 unique firms and 29,373,378 job spells for 3,672,648 unique individuals. Firm fixed effects are identified based on 2,708,338 movers who work at more than one firm in the connected sample; person fixed effects of individuals at these firms are identified up to scale and I choose as a normalization that the sum of all person effects is zero.

The identifying assumption of this decomposition is strict exogeneity for residual errors, which implies random job mobility. I use event studies for workers and managers as in Card et al. (2013).
at a firm in the main sample, corresponding to a total of 52,000 individuals. I separate time-invariant manager quality into an observed component that is due to differences in education and an orthogonal component of residual quality.

The bottom left panel of Figure 3 shows a strongly positive sorting pattern with respect to unobservable quality of managers across the firm distribution. The average manager at a low (high) productivity firm is about 10% less (more) productive than the average manager in the sample. Moreover, I can use residual manager quality to compare the pool of candidates for manager positions across firms. Due to a lack of training data within firms, I define candidates as all individuals who are ever promoted into a management position. The bottom right panel of Figure 3 shows that better firms recruit better manager candidates in terms of unobservable characteristics even five years before the first promotion.

**Fact 4: Firms select good internal candidates for promotion. They know less about external candidates.**

In Figure 3, the slope of the empirical matching function between managers and firms is steeper than for trainees. This pattern suggests more uncertainty about the quality of trainees compared to managers. Figure 4 shows much higher average quality of promoted managers compared to the average quality of the pool of candidates five
years before promotion. Consistent with employer learning, Figure 4 shows that this selection pattern is mainly driven by differences in unobserved quality as opposed to differences in educational attainment.

Moreover, Figure 5 considers first-time manager promotions to illustrate asymmetric information about internal and external managerial candidates. The composition of talent for the two groups is very different. External hires have higher formal education, but internal promotions are superior in terms of their unobserved ability conditional on education. Average unobserved quality of external hires is lower than for internal promotions across the entire firm distribution.\textsuperscript{17} This result suggests that firms substitute observable skills for ability when hiring managers externally.

3 Model

Based on the stylized facts about (1) heterogeneous hiring strategies, (2) the positive relationship between firm productivity and internal promotions, (3) the empirical sorting patterns of trainees and managers across firms, and (4) employer learning and adverse selection on unobservables, I develop a new model of the market for managers. The goal is to characterize the tradeoffs between internal and external hiring in a setting where internal labor markets arise from firm-specific knowledge and asymmetric employer learning. The model will illustrate the role of internal labor markets for market competition and wage inequality, as well as for the allocation of talent across firms.

\textsuperscript{17}These differences are pointwise statistically significant along the firm distribution.
3.1 Basic Setting

Assume a mass of $M$ potential firms in the market of a homogeneous good with normalized price $p = 1$. Firms differ in their permanent productivity $\phi$, drawn from a distribution $\phi \sim \Gamma(\phi)$. Labor is supplied inelastically by overlapping generations of $N$ individuals. Individuals differ by education or other general skills $e \sim F(e)$ that are observable at the beginning of their career and that determine human capital as a worker, $h(e), h'(e) > 0$. Secondly, individuals are characterized by managerial ability $a \sim F_a|e(a)$ that is initially unknown to both firms and individuals. For simplicity, there are two ability types, $a_H > a_L$. Let $p(e) = Pr(a_H|e)$ and assume a positive relationship between skills and managerial ability, $p'(e) > 0$. Skills, managerial ability and firm-specific human capital jointly determine total managerial talent $z$ as specified in detail below. Individuals live for two periods and maximize lifetime wages.

Production

Firms consist of two hierarchy levels, they hire $n$ units of labor and one manager of type $z$, who jointly produce output according to

$$y(\phi, z, n) = (\phi \cdot z)^{1-\alpha} \cdot n^\alpha,$$

where $0 < \alpha < 1$.\(^{18}\) I assume that workers with different levels of human capital $h$ are perfect substitutes in production to abstract from sorting of workers across firms. Each unit of worker human capital conducts a fixed number of tasks that are supervised by the manager with diminishing returns.

The key assumption in the production technology is that managerial talent and firm technology are complements, $\frac{\partial^2}{\partial \phi \partial z} y(\phi, z, n) > 0$. The production technology in (3) models separate roles for firm and manager productivities. This approach is consistent with the notion that management practices can be considered intangible capital of firms which matters for total factor productivity as captured by the firm type $\phi$, but the talent of the manager also plays an explicit role through $z$.\(^{19}\)

\(^{18}\)This functional form builds on Lucas (1978), Rosen (1982) and has been used in a similar context by Tervio (2008). In contrast, most of the literature on worker careers assumes no complementarities between jobs in a firm, see Gibbons and Waldman (1999); Pastorino (2015).

\(^{19}\)Bloom et al. (2013) provide evidence of the former in a RCT setting, Lazear et al. (2015) estimate the value of supervisors on worker productivity in a large service company.
Trainees and Training

Every period, firms hire one trainee for their manager succession. Only old individuals with experience as trainees can become managers; thus, firms have to replace their manager every period in a cycle of promotion and retirement. After the trainee period, firms decide whether they want to promote the internal candidate or hire a new manager externally. There are two advantages of having a trainee. First, the incumbent firm learns about the individual’s true unobserved ability $a$ over time and therefore acquires an informational advantage over competitors in the market. Second, the firm can invest resources to teach the trainee firm-specific knowledge. Yet at the time of the training decision, there is still uncertainty about the candidate’s ability $a$. Training success is stochastic and depends on investment $x$ according to $\kappa(x) \in [0,1]$, $\kappa'(x) > 0$, $\kappa''(x) < 0$. Candidates with successful internal training acquire firm-specific human capital $f$, where $f$ is a constant and thus does not vary with $x$. Note, however, that firm-specific training is not necessary to become a manager in the second period. Successful training only increases a manager’s productivity at the incumbent firm. In particular, I assume that firm-specific knowledge is a perfect substitute to general observed and unobserved skills in the managerial talent function,

$$z(e,a,f) = e + a + f.$$ (4)

This additive specification is a useful benchmark because it does not generate increasing incentives for internal training for workers with higher skills. Instead, this assumption highlights the role of firm heterogeneity in training and hiring decisions. Differing incentives for training will follow from the matching pattern in equilibrium and the complementarity between firm productivity and manager talent.20

The two mechanisms of learning and training are distinct because learning provides the incumbent firm with superior knowledge about general skills whereas training is a firm-specific investment to improve the productivity of an internal candidate at this particular company. As a result, the firm faces a tradeoff in how to use internal labor markets to find a good internal candidate. Firms can either compete for more promising talents early in their career or they can use more resources for firm-specific training to make mediocre candidates more productive.

20Additivity in (4) implies that only the firm-specific component generates rents. This is in contrast to the complementarity between unobserved ability and training in Acemoglu and Pischke (1998), where homogeneous firms invest in general training if they have superior knowledge about ability $a$. 
The External Market for Managers

The second fundamental tradeoff that firms face is between hiring managers externally or using internal labor markets to develop and promote talent. The simplest version of the model considers a scenario where firms meet external candidates in a secondhand market for managers. Managers search for a job in the external manager market for two reasons. First, firms separate from candidates endogenously if their revealed type lies outside the internal promotion range. Second, trainees separate from their incumbent firm for exogenous reasons at rate $\delta$ (Greenwald 1986). One interpretation is that individuals receive personal preference shocks to switch jobs. I assume that firms only observe the skill level $e$ of external managers but not the reason of separation at the previous job. Otherwise unsuccessful candidates who anticipate endogenous separation would preemptively leave their firm to avoid a bad signal. I characterize the quality of the pool of external candidates by the share of high ability types for each skill type $e$, $\bar{p}_{ext}(e)$,

$$\bar{p}_{ext}(e) = \frac{H(e)}{H(e) + L(e)}.$$  \hspace{1cm} (5)

$H(e)$ and $L(e)$ denote the mass of high and low ability types with skill level $e$ in the external market respectively. The share of high-ability managers $\bar{p}_{ext}(e)$ is an equilibrium object that I characterize below. The distribution of $\bar{p}_{ext}(e)$ differs from the population share $p(e)$ because there is adverse selection in the pool of external candidates who were not internally promoted.

Wage Setting

Wages are determined according to perfect competition for workers, trainees and managers. A worker with skill level $e$ supplies $h(e)$ units of labor. The price of one efficiency unit of human capital is the market wage $w_n$. I assume that firms have full bargaining power to capture rents from internal labor markets. Managers are paid their best outside option. Observed skills $e$ are a signal for total managerial talent and they determine the market wage functions for trainees, $w_t(e)$, and for managers, $w_m(e)$. A participation constraint ensures that at discount rate $\beta$, individuals who become trainees prefer this career path to being a worker, and individuals who completed the trainee stage

\footnote{If firm knowledge is sufficiently valuable or high ability types are sufficiently rare at less productive firms, the winner’s curse is exacerbated and there will be no poaching in equilibrium, see section 3.5.}

\footnote{Assuming only two unobserved ability types provides a key simplification of the problem because the firm’s belief about the quality of external hires of skill $e$ can be summarized by a scalar valued function $\bar{p}_{ext}(e)$ instead of a general distribution function over continuous types $F_{n|e,ext}$.}
prefer to work as managers instead of going back to being a worker,

\[ w_\tau(e) + \beta w_m(e) \geq (1 + \beta) h(e) w_n \]  
\[ w_m(e) \geq h(e) w_n. \]  

As a result, there is an indifference condition for the lowest skill manager \( e^m \),

\[ w_m(e^m) = h(e^m) w_n. \]  

If the lowest skill trainee, \( e^\tau \) becomes the lowest skill manager \( e^m \) in the next period, the boundary condition for trainees simply yields

\[ w_\tau(e^\tau) = h(e^\tau) w_n. \]  

This wage setting is consistent with trainee wages increasing by less than the worker salary as a function of observed skills. These wage cuts can be (over)compensated by high manager wages in the second period.

### 3.2 The Firm’s Problem

In period \( t \), a firm maximizes expected profits in \( t + 1 \) by hiring a trainee and optimally investing in training. This decision anticipates the optimal choice between internal and external hiring in \( t + 1 \) after the trainee’s ability type and training outcome have been observed. After the firm chooses between internal and external hiring, the true type of the new manager is revealed and the firm hires the optimal workforce corresponding to the manager’s talent. The timing of the model is summarized in Figure 6.

I solve the firm’s problem backwards to illustrate the profit maximization problem
in detail. In period \( t + 1 \) after the managerial hiring decision has been made, the firm optimally adjusts the workforce conditional on manager type \( z \). Under the assumption of efficiency units of labor, the firm is indifferent about the exact composition of the workforce because workers with different levels of human capital are perfect substitutes. The firm chooses labor demand for the total units of human capital \( n \) according to

\[
y^*(\phi, z) = \max_n \{y(\phi, z, n) - n \cdot w_n\}
\]

which, using (3), yields the optimality condition

\[
n^* = \phi z \cdot \left[ \frac{\alpha}{w_n} \right]^{\frac{1}{1-\alpha}}.
\]

One can see that \( n^* \) is proportional to \( \phi \) and \( z \), so better firms hire a larger workforce and better managers supervise a larger group of workers. The span of control of a manager is limited by diminishing returns to supervision (smaller \( \alpha \)) and by the market wage rate \( w_n \) that reduces the scalability of manager talent.\(^{23}\)

Firms anticipate the optimal choice of the workforce when deciding between external hiring and internal promotion. Firm \( \phi \) determines the optimal external candidate without firm-specific knowledge, denoted by \( e^*_m(\phi) \). This decision takes the share of high ability types in the external market \( \bar{p}_{ext}(e) \) and the wage schedule for managers \( w_m(e) \) as given. Denote expected profits from external hiring as

\[
E_\text{a|e,ext}[\pi^*(\phi, e^*_m(\phi), \alpha)] = \max_e \{\bar{p}_{ext}(e) \cdot y^*(\phi, a_H, e, 0) + (1 - \bar{p}_{ext}(e)) \cdot y^*(\phi, a_L, e, 0) - w_m(e)\},
\]

where \( E_\text{a|e,ext}[\cdot] \) is the expectation over the distribution of ability \( a \) given observed skill \( e \) and the fact that the firm hires externally. The first order condition of the external hiring choice characterizes the slope of the managerial wage function,

\[
w'_m(e) = \phi^*_m(e) \Psi(\alpha, w_n) \cdot \left[ 1 + (a_H - a_L) \frac{\partial}{\partial e} \bar{p}_{ext}(e) \right]
\]

where \( \Psi(\alpha, w_n) = (1 - \alpha) \left( \frac{\alpha}{w_n} \right)^{\frac{\alpha}{\alpha-1}} \) is a function of the market wage and technology. The slope of the wage profile for managers equals the expected marginal product of a manager type \( e \) at the equilibrium firm assignment \( \phi^*_m(e) \). This expectation takes into account the firms' anticipation of the optimal choice of the workforce when deciding to hire externally.

---

23This result shows the analytical convenience of the Cobb-Douglas production technology because it yields output as a linear function of firm type and manager talent conditional on optimal labor demand for workers, \( y^*(\phi, z) = \phi \cdot z \cdot (1 - \alpha) \left( \frac{\alpha}{w_n} \right)^{\frac{\alpha}{\alpha-1}} \).
account that hiring better observed skill types $e$ from the external market also changes the probability of finding a high ability manager.

The firm compares profits for each potential trainee outcome $z(e,a,f)$ with the optimal external manager to determine sets $I_f$ and $I_0$ of profitable internal promotions with and without successful training.$^{24}$

$$
I_0 = \{(a,e) \ s.t. \ \pi^*(\phi,z(a,e,0)) \geq E_{a|e,ext}[\pi^*(\phi,e(\phi),a)]\} \\
I_f = \{(a,e) \ s.t. \ \pi^*(\phi,z(a,e,f)) \geq E_{a|e,ext}[\pi^*(\phi,e(\phi),a)]\}.
$$ 

Firms make training investments under uncertainty because the true type $a \in \{a_L,a_H\}$ is only revealed at the end of the trainee period. Firm $\phi$ chooses optimal investment $x^*(\phi,e)$ for any trainee type $e$ to maximize expected profits from firm-specific training,

$$
\max_x \left\{ -x + \beta (1 - \delta) \kappa(x) E_{a|e}[R(\phi,e,a)] \right\}.
$$

$R(\phi,e,a) = \max \{0, \pi^*(\phi,z(a,e,f)) - A(\phi,e,a)\}$ denotes rents from training relative to a default profit for firm $\phi$ when hiring trainee $e$ with revealed ability $a$ and no internal training, $A((\phi,e,a))$. $^{25}$ The expectation over rents from training is with respect to the population distribution of ability types $p(e)$ because there is no adverse selection as young workers enter the labor market. Future rents are discounted at rate $\beta$ and only realized with probability $1 - \delta$ because some trainees leave the firm for exogenous reasons.

Firms may benefit from training in two cases. First, successful training improves rents for high ability trainees that would have been promoted anyway (intensive margin). For any trainee of skill $e$, intensive margin gains only occur if $\{e,a_H\} \in I_0$. Second, successful training may turn some trainee types into profitable promotions (extensive margin). These extensive margin gains occur for high ability trainees if $\{e,a_H\} \notin I_0$ and $\{e,a_H\} \in I_f$ and for low ability trainees if $\{e,a_L\} \in I_f$. Trainees with low ability and without successful training can always be replaced by an external manager with the

$^{24}$ Note that $I_0 \subset I_f$ because firm-specific knowledge creates an additional rent for the firm that makes internal promotion weakly more profitable for any skill level.

$^{25}$ If $\{e,a\} \in I_0$, then the firm prefers to promote the internal candidate even without firm specific knowledge; otherwise the firm prefers to hire the best external candidate instead. Formally,

$$
A(\phi,e,a) = 1\{(a,e) \in I_0\} \cdot \pi^*(\phi,z(a,e,0)) + 1\{(a,e) \notin I_0\} \cdot E_{a|e,ext}[\pi^*(\phi,e_m(\phi),a)].
$$
same skill level, in the hope of finding a high ability type. The full first order condition of training takes these cases into account, 26

\[
1 \geq \beta (1 - \delta) \kappa'(x) \left\{ p(e) \cdot \mathbb{1}\{(e, a_H) \in I_0}\cdot \phi\Psi(\alpha, w_n)f \right. \\
+ p(e) \cdot \mathbb{1}\{(e, a_H) \notin I_0, (e, a_H) \in I_f\}\cdot D(\phi, e, a_H, f) + (1 - p(e)) \cdot \mathbb{1}\{(e, a_L) \in I_f\}\cdot D(\phi, e, a_L, f) \left. \right\}
\]

where \( D(\phi, e, a, f) \) defines the difference in profits compared to the best external hire,

\[
D(\phi, e, a, f) = \pi^*(\phi, z(a, e, f)) - E_{\alpha'|z,ext}[\pi^*(\phi, e^*_m(\phi), f)].
\]

Note that under positive assortative matching (PAM), high ability trainees will always satisfy \( \{e, a_H\} \in I_0 \). If firm-specific human capital is sufficiently valuable (see assumption (A1) below), low ability types with training are promoted and the optimality condition for training simplifies to

\[
1 = \beta (1 - \delta) \phi\Psi(\alpha, w_n) \kappa'(x) \left[ f - (1 - p(e)) \tilde{p}_{ext}(e(\phi), x)(a_H - a_L) \right].
\]

The right-hand side captures the marginal gains from an increased probability of training success. A high ability type would be promoted even without training; thus, the marginal productivity gain of the manager is \( f \) times the increase in production scale \( \phi\Psi(\alpha, w_n) \), discounted at rate \( \beta (1 - \delta) \). With probability \( 1 - p(e) \), the internal candidate is a low ability type and therefore the alternative would be to hire a manager externally. This alternative implies a chance of finding a high type according to the external market share \( \tilde{p}_{ext}(e) \). The risk of hiring and training a low ability type internally discourages training because of the option of hiring a high type through the external market.

Firms anticipate optimal training investment, the decision between internal promotion and external hiring, and the optimal choice of the workforce when hiring a trainee. For simplicity, I assume that trainees do not participate in production; they only spend time in managerial training. Their value is determined by potential rents from promoting them internally next period instead of hiring the best external candidate, \( e^*_m(\phi) \).

26 Using Inada conditions on the probability of training success as well as a sufficiently large value for firm specific training, there is an interior solution for training investment at all firms and skill levels.
Firm $\phi$ chooses trainee $e^*_\tau(\phi)$ to maximize expected profits next period,

$$\max_{e^*_{\tau}} \left\{ -w_{\tau}(e^*_{\tau}) - x^*(\phi, e_{\tau}) + \beta (1 - \delta) E_{\alpha, e_{\tau}} \left[ \kappa(x^*(\phi, e_{\tau})) R(\phi, e_{\tau}, a) + A(\phi, e_{\tau}, a) \right] \right\}.$$

The trainee will remain at the firm with probability $1 - \delta$, future profits are discounted by $\beta$. For any realized ability level $a$, the firm will earn the default profit $A(\phi, e_{\tau}, a)$ without training, either through promotion or through external hiring. Hiring a trainee with higher observed skills means better chances to find a high ability candidate that generates rents even without training. Moreover, successful training implies additional rents $R(\phi, e_{\tau}, a)$ that will be more or less likely to be realized depending on the optimal training intensity. The first order condition characterizes the equilibrium wage schedule for trainees according to

$$w'_{\tau}(e_{\tau}) = \beta (1 - \delta) \left\{ \kappa(x^*(\phi, e_{\tau})) \cdot \phi^*_{\tau}(e_{\tau}) \Psi(\alpha, w_n)(a_H - a_L) \left[ p'(e_{\tau}) - \tilde{p}'_{\text{ext}}(e_{\tau}) \right] \right\} \equiv T_1: \text{marginal rents with successful training}$$

$$+ (1 - \kappa(x^*(\phi, e_{\tau}))) \cdot \left\{ p'(e_{\tau}) D(\phi, e_{\tau}, a_H, 0) - \phi^*_{\tau}(e_{\tau}) \Psi(\alpha, w_n)(a_H - a_L) p(e_{\tau}) \tilde{p}'_{\text{ext}}(e_{\tau}) \right\} \equiv T_2: \text{marginal rents without successful training}$$

The marginal change in trainee wages reflects the probability-weighted sum of marginal rents from a better candidate with and without successful training.\(^{28}\) Hiring a better trainee is more valuable if there is a large difference in productivity between unobserved high and low ability types, $a_H - a_L$. The wage schedule for trainees is closely related to how much the probability of finding a high ability trainee increases in observable skill $e$, $p'(e)$. Moreover, the larger the improvement in the population share of high types when hiring a trainee compared to the share of high ability managers that end up in the external market, the stronger the competition for the most promising young candidates at the trainee stage. Finally, note that the trainee wage only depends on observed skills $e$ through the population share of high ability types $p(e)$.

\(^{27}\)I focus on the scenario where in equilibrium $\{e^*_\tau(\phi), a_H\} \in I_0(\phi)$. I show below that this is the relevant scenario in an equilibrium with PAM. More generally, one has to distinguish different cases analogous to (14).

\(^{28}\)Note that under PAM, marginal rents without successful training in equation (15) simplify to

$$T_2 = \phi^*_{\tau}(e_{\tau}) \Psi(\alpha, w_n)(a_H - a_L) \left[ p'(e_{\tau}) - \tilde{p}'_{\text{ext}}(e_{\tau}) p'(e_{\tau}) - p(e_{\tau}) \tilde{p}'_{\text{ext}}(e_{\tau}) \right]$$

which reflects the tradeoff between increasing the probability of a high ability candidate today and increasing the chances of finding a high ability manager in the external market.
3.3 General equilibrium

Definition. A competitive equilibrium in this model is defined by the worker wage rate \( w_n \), wage schedules for trainees and managers \( \{ w_\tau(e), w_m(e) \} \), assignment functions for trainees \( \phi_\tau(e) \) and managers \( \phi_m(e) \), a set of active firms \( \phi \in \Phi, \phi_{\text{max}} \), a set of trainees \( e \in [e^\tau,e^{\text{max}}] \), a set of managers \( e \in [e^m,e^{\text{max}}] \), training investment \( x(\phi) \) and internal promotion ranges \( \{ I_0(\phi), I_f(\phi) \} \) for each firm type such that firms maximize profits, individuals maximize lifetime wages and labor markets for managers, trainees and workers clear.

Labor market clearing for workers requires

\[
\begin{align*}
N \int_0^{e^\tau} h(e) dF(e) + N \int_{e^\tau}^{e^m} h(e) dF(e) &= M \int_{\Phi} s(\phi,e^*_\tau(\phi)) \int n(\phi,e^*_m(\phi),a',0) dF_{a|e,\text{ext}}(a') d\Gamma(\phi) \\
&+ M \int_{\Phi} (1 - s(\phi,e^*_\tau(\phi))) \int n(\phi',e^*_\tau(\phi),a',f') dF_{a|e}(a') dF_{f|\phi'}(f') d\Gamma(\phi)
\end{align*}
\]

where the supply of workers is given by two active generations in the market who have not been chosen as trainees and managers and labor demand at each firm depends on the realized trainee type or external manager type along the equilibrium assignment. The share of separations by firm type, denoted by \( s(\phi) \), depends on exogenous shocks \( \delta \) and endogenous layoff probabilities based on realized ability types and training success.\(^{29}\) The ability type distribution for internal candidates depends on the population distribution \( F_{a|e}(a') \), whereas the distribution for external hires \( F_{a|e,\text{ext}}(a') \) takes adverse selection in the external market according to (5) into account. Under the assumption that each firm hires one trainee and one manager each period, market clearing for trainees and managers requires that the mass of trainees and the mass of

\(^{29}\) More formally, the probability of separation from any trainee type \( e \) for firm \( \phi \) is given by

\[
s(\phi,e) = \delta + (1 - \delta) p(e) \left[ (1 - \kappa(e)) \mathbb{I} \left\{ e + a_H \notin I_0(\phi) \right\} + \kappa(e) \mathbb{I} \left\{ e + a_H + f \notin I_f(\phi) \right\} \right] \\
+ (1 - \delta) (1 - p(e)) \left[ (1 - \kappa(e)) + \kappa(e) \mathbb{I} \left\{ e + a_H + f \notin I_f(\phi) \right\} \right]
\]
managers are equal to the mass of active firms respectively,

\[ N \int_{e^T}^{e^{max}} dF(e) = M \int_{\phi}^{\phi^{max}} d\Gamma(\phi) \]
\[ N \int_{e^T}^{e^{max}} dF(e) = M \int_{\phi}^{\phi^{max}} d\Gamma(\phi). \] (17)

Lastly, free entry determines the lowest productivity firm in the market, denoted by \( \phi \). New entrants make zero expected profits by hiring a trainee today and maximizing future profits,

\[ 0 = -x^*(\hat{\phi}, e^T) - w^*_r(e^T) \]
\[ + \beta(1 - \delta) E_{a|e}[\kappa(x^*(\hat{\phi}, e^T)) R(\hat{\phi}, e^T, a) + A(\hat{\phi}, e^T, a)] + \beta \delta E_{a|e,ext}[A(\hat{\phi}, e^T, a)]. \] (18)

Note that since the problem is sequential and I focus on a steady state equilibrium, this firm will also make zero profits on all trainee hiring decisions in future periods of their infinite horizon problem. I therefore simplify the zero profit condition by focusing only on the current decision to invest in a trainee.

**Characterization**

In order to characterize the equilibrium in a simple way, I make the following assumptions that are sufficient for Proposition 1 below. Throughout, I denote the elasticity of a function \( g(y) \) with respect to \( y \) as \( \eta_g(y) \).

**A1** \( f > \max_e \{p(e)\} \cdot (a_H - a_L) \),

**A2** \( \max_e \{p(e)\} < 0.5 \)

**A3** \( \frac{E[\phi]}{\phi_{min}} > \left[ \frac{\beta(1-\alpha)}{\alpha(1+\beta)} \right]^{1-\alpha} \cdot \frac{2(N-M)}{M} \)

**A4** \( \eta_{m(e)}(e) > \eta_h(e) \ \forall e \) where \( m(e) = \Gamma^{-1} \left[ 1 - \frac{N}{M} (1 - F(e)) \right] \)

**A5** \( \eta_{E[z|f,e]}(e) > \eta_h(e) \ \forall e \) where \( E[z|f,e] = e + a_L + p(e) [a_H - a_L] + (1 - \delta) f \) is the expected talent of skill type \( e \) under maximum training,

**A6** \( \frac{1+\beta}{\beta} \cdot \frac{E[z|ext,e]}{E[z|f,e]} \cdot \eta_{E[z|ext,e]}(e) > \eta_h(e) \ \forall e \) where the minimum expected talent of skill type \( e \) under external hiring is \( E[z|ext,e] = e + a_L + \frac{\delta p(e)}{(1-\delta)(1-p(e))} [a_H - a_L] \).

\(^{30}\)A new firm spends one period setting up production to start producing in the next period. New firms cannot hire a manager because the mass of available managers is limited by the mass of trainees in the previous period.
Assumption (A1) implies that firms will always prefer trainees with successful training over external hiring for a particular skill level $e$. This assumption will help to simplify the taxonomy of cases in equation (13). Assumption (A2) implies that high ability types are sufficiently scarce to make internal training attractive. Assumption (A3) states that the least productive firm must be sufficiently worse than the average productivity type in the market. This condition will ensure that an equilibrium with a cutoff $\phi > \phi_{\text{min}}$ exists because the worst firm requires a very low market wage to break even and at this wage rate there will be excess labor demand. Higher discounting (lower $\beta$) and higher returns from supervision (higher $\alpha$) will alleviate this condition because they lower the profits of the cutoff firm and further increase labor demand in the market respectively. Assumption (A4) states that the match quality of skill type $e$, $m(e)$, has to increase sufficiently quickly to overcompensate the elasticity of worker human capital with respect to skill. Intuitively, this condition will be satisfied if the distribution of firm types is sufficiently skewed. Assumption (A5) requires that the elasticity of expected talent with respect to skill level $e$ is larger than for worker human capital. Under this condition, the incentive constraint (6) will be satisfied. The larger firm-specific human capital, the less firms are willing to bid up wages for trainees and managers, and so the improvement in talent has to be larger to satisfy (A5). On the other hand, higher separation rates increase the competition for external managers and make training less valuable. This reduces the restrictiveness of this condition. Assumption (A6) states that even when hiring a manager externally under the worst adverse selection scenario, the elasticity of talent with respect to skill is sufficiently large to overcompensate improvements in worker human capital. As a result, condition (7) holds and managers will not want to become workers in the second period instead.

**Proposition 1.** Suppose assumptions (A1)-(A6) hold. (1) There exists a unique equilibrium with positive assortative matching (PAM) according to observed skills $e$ for both trainees and managers. (2) The equilibrium is characterized by upward sloping wage schedules for managers, $w'_{m}(e) > 0$ and trainees, $w'_{t}(e) > 0$.

The proof is constructive. With PAM, market clearing for managers and trainees pins down the assignment function conditional on the exogenous distributions of types,

$$M \left(1 - \Gamma (\phi(e))\right) = N \left(1 - F(e)\right)$$

$$\phi(e) = \Gamma^{-1} \left[1 - \frac{N}{M} (1 - F(e))\right]. \quad (19)$$

Note that the subscripts for managers and trainees in the assignment function can be dropped because in equilibrium both assignments follow the identical PAM pattern.
Moreover, conditional on the assignment function and the market wage for workers \( w_n \), I can determine training, adverse selection, managerial wages and trainee wages from the firm’s optimality conditions for training (14), external managerial hiring (11) and the optimal trainee hiring decision (15). The remaining endogenous variables are the entry cutoff level for firm productivity \( \phi \) and the market wage for workers \( w_n \).

Under PAM, the firm cutoff will immediately pin down cutoff levels for trainees \( e^t \) and managers \( e^m \) as well.

Equilibrium conditions (16) and (18) are two equations in two unknowns, the worker wage \( w_n \) and the firm entry cutoff \( \phi \) that determine the equilibrium. Since excess labor demand from equation (16) strictly decreases in both unknowns, this condition is represented by a downward-sloping curve in \( \{\phi, w_n\} \)-space. In contrast, profits in equation (18) are increasing in the cutoff but decreasing in the wage rate for workers, which corresponds to an upward sloping curve in \( \{\phi, w_n\} \)-space to satisfy this constraint.

Under (A1)-(A6), the two conditions cross exactly once and there is a unique equilibrium with PAM. This equilibrium will be supported by increasing wage functions for both trainees and managers as I illustrate in the remaining proof in Appendix B.1.

I now make an additional assumption to further characterize the equilibrium analytically.

\[(A7) \quad (1 - \delta) \eta_{m(e)}(e) > \eta_p(e).\]

Assumption (A7) restricts the elasticity of the share of high ability managers compared to the assignment function of managers to firms. As a result, the external market of managers does not quickly become more attractive because of large differences in the population shares of high types for managers with similar observed skills.

**Proposition 2.** Suppose assumptions (A1)-(A7) hold. (1) In equilibrium, training investment increases in trainee skill level \( e \). (2) In equilibrium, the share of high ability managers with skills \( e \) in the external market, \( \tilde{p}_{ext}(e) \), increases in observed skills \( e \).

Proposition 2 characterizes the interaction of firm specific human capital and information asymmetries in equilibrium, the proof is in Appendix B.2. More productive firms benefit more from firm-specific human capital because of complementarities between manager and firm productivity in the production technology. As a result, better firms have a higher incentive to invest in firm-specific training to make their internal candidate more productive. As a counteracting force, better firms have a better chance to find a high ability manager in the external market. Assumption (A7) provides a sufficient condition such that this mechanism is overcompensated by higher gains from
training as firm productivity improves.\textsuperscript{31} There is a positive relationship between the success rate of training and the share of low ability managers promoted internally. More training reduces the amount of low ability managers in the external market and thereby improves the mix of external candidates. Since better trainees receive higher training investment at more productive firms, the share of high ability managers in the external markets is higher for candidates with higher observed skills.

### 3.4 Comparative Statics

As a next step, I further analyze the role of firm-specific human capital and adverse selection with respect to wage inequality and competition for managerial talent. The model shows that firm-specific human capital reduces market competition for managers, but asymmetric information increases competition for managerial talent of trainees. In order to provide sharp predictions, I assume

$$\text{(A8)} \quad -\kappa''(x) \leq \kappa'(x) \quad \forall x \geq 0.$$

Assumption (A8) states that the probability of training success $\kappa(x)$ is not too concave. As a result, additional training compared to any arbitrary baseline investment level $x$ sufficiently improves the chances of training success.

**Proposition 3.** Consider the effect of an increase in the productivity of firm knowledge $f$ on wage inequality under assumptions (A1)-(A8). (1) The resulting increase in the market wage strictly decreases the wage gap between managers or trainees at all skill levels $e \geq e$ compared to workers. (2) The resulting increase in adverse selection reduces the wage gap between managers and workers but increases the wage gap between trainees and workers for all skill levels $e \in [e, e(\hat{\phi})]$. For skill levels $e \geq e(\hat{\phi})$, the resulting decrease in adverse selection increases the wage gap between managers and workers but reduces the wage gap between trainees and workers. (3) Resulting firm entry (exit) increases (decreases) the wage gap between managers and workers at all skill levels $e \geq e$.

I determine the change in the wage gap between managers and workers in response to an increase in $f$ from totally differentiating the equilibrium wage function for managers \textsuperscript{31}This result illustrates that the model provides a mechanism that allows for lower training at high productivity firms despite the complementarity between training and firm productivity. Better firms have an advantage in competing for the best talent externally and this opportunity may deter training investment.
\[ w_m(e) = h(e) \, w_n + \int_e^e w'_m(e) \, de \] using equation (11) to yield\(^{32}\)

\[
\frac{d}{df} \left( \frac{w_m(e)}{w_n} \right) = \frac{\epsilon'_w}{\text{scale}<0} \frac{dw_n}{df} + \frac{\epsilon'_e}{\text{entry}<0} \frac{de}{df} + \frac{\epsilon'_x}{\text{selection}>0} \frac{dx}{df}. \tag{20}
\]

First, stronger internal labor markets reduce the wage gap between any manager and the average worker through higher equilibrium wages for workers. Better managers are less valuable than before because they now supervise smaller teams and cannot leverage their high skills as much as in the baseline scenario (scale effect). Second, firm exit increases the cutoff level for trainees and managers, \(d_x > 0\) and decreases inequality because the competitive equilibrium rewards managers and trainees for being better than all lower ranked candidates.\(^{33}\) Yet without additional assumptions about the type distributions of firms and managers, it is ambiguous whether an increase in the value of firm knowledge will lead to firm entry or exit. Firm profits increase through higher rents from firm-specific training, but higher market wages limit production scale.\(^{34}\) Finally, for all firms above some threshold \(\tilde{\phi}\), increased training and the reduction in adverse selection counteract the wage compression for managers. Training reduces adverse selection because a higher share of low ability candidates receive successful training and remain at their previous firm instead of entering the external market. As a result, managers are rewarded for a better pool of external candidates. Since the increase in training is largest for the best firms, adverse selection is reduced more at the top of the manager distribution and the best managers receive the largest wage gains.

The competition for talent is reflected by the wage function for trainees. The change in the wage gap between trainees and workers in response to a change in \(f\) is given by totally differentiating the trainee wage function using equation (15),

\[
\frac{d}{df} \left( \frac{w_T(e)}{w_n} \right) = \frac{\epsilon'_w}{\text{scale}<0} \frac{dw_n}{df} + \frac{\epsilon'_e}{\text{entry}<0} \frac{de}{df} + \frac{\epsilon'_x}{\text{selection}>0} \frac{dx}{df}. \tag{21}
\]

In general, internal labor markets weaken the competition for trainees in the market because firms have means of training candidates with lower general skills internally. If internal labor markets become more attractive because firm-specific knowledge is more

---

\(^{32}\)Details of the derivation are provided in Appendix B.4.

\(^{33}\)Assumption (A5) ensures that the outside option of being a worker improves too slowly to counteract the lower market competition for trainees by higher opportunity costs.

\(^{34}\)For example, if the firm distribution is highly dispersed, high productivity firms benefit more than proportionately and drive up equilibrium wages such that some low productivity firms prefer to exit.
productive, competition for trainees is reduced for two reasons. First, higher market wages reduce production scale, $\epsilon^w < 0$, which means recruiting a trainee with higher expected productivity is less valuable. Second, lower adverse selection in the market for managers reduces the value of better trainees because firms can more easily find a high ability manager in the external market, $\epsilon^x < 0$. The only potential counteracting force is firm entry that might increase competition for trainees, but firm entry or exit depends on additional assumptions about the type distributions.

**Asymmetric information and Sorting** The model also has implications about the allocation of resources in the market for managers. Under asymmetric information, sorting between trainees or managers and firms is based on a noisy signal of talent $e$ instead of true talent $e + a$. If the signal $e$ is weak, some highly talented individuals are matched with low productivity firms. The production complementarity between firm and manager type implies suboptimally low output compared to sorting according to true talent $e + a$. As an example, consider two trainees, $z_1 = e_1 + a_L$, $z_2 = e_2 + a_H$ with $e_1 = e_2 + \xi$, and $0 < \xi < a_H - a_L$. PAM in $e$ implies that $\phi(e_1) > \phi(e_2)$, but the supermodularity in the production technology yields gains from resorting,

$$y(\phi(e_1), z_1) + y(\phi(e_2), z_2) < y(\phi(e_1), z_2) + y(\phi(e_2), z_1).$$

The magnitude of this productivity gain will depend on the production technology, the ability difference between high and low types and the productivity distribution of firms. Yet reducing information frictions imply an efficiency-equity tradeoff because low types can no longer benefit as much from a noisy signal, while high types are more strongly rewarded. This suggests larger wage dispersion among managers in the market.

### 3.5 Extensions

The model is based on several stylized assumptions that emphasize core tradeoffs related to asymmetric information and firm-specific human capital that firms face when recruiting managers. This section discusses the intuition for three extensions to the baseline model: general training, top management promotions observable by the market, and endogenous choice of the number of trainees. The conclusion proposes additional directions for future research.

**General Training** First, if trainees acquire some general skills through training, there will be knowledge transfer from external hiring. The external market will re-
ward managers from highly productive firms because higher training intensity at these firms implies a higher probability of successful training and general human capital accumulation. Note that even if training can be decomposed into a firm-specific and a general component, the additive structure of manager talent in (4) implies that only the firm-specific component generates rents.

**Market Signals and Poaching** Second, if the market receives additional signals about the quality of a manager over time, for example by observing promotions into the top management level, competition for the best managers will be intensified (Frydman 2005). Yet the fixed organizational structure prevents firms from choosing inefficiently low promotion rates to hide talent as in Waldman (1984). In general, firms face a winner’s curse that will impede direct poaching from competitors. Firms can only hope to make a profit by poaching managers from less productive firms. Yet hiring a manager from a lower ranked firm is risky because the promotion is more likely due to firm-specific knowledge. If firm knowledge is sufficiently valuable or if high ability types are sufficiently rare at less productive firms, there will be no poaching in equilibrium. This is consistent with a low share of manager transitions from low to high productivity firms in the data, see Appendix Figure A.2.

**Trainee Pool** If firms can hire multiple trainees, the production complementarity implies that more productive firms will hire weakly more trainees. Less productive firms will benefit from access to higher-skill unsuccessful candidates in the external pool of managers. The identity of the training firm will affect the extent of the negative signal of searching in the external market because trainees at more productive firms have to compete with more high quality candidates internally. The model assumption of one trainee per manager yields a tractable solution to the assignment problem and captures the advantage for highly productive firms through the underlying distributions of manager types and firm types and the training technology. In the data, a large share of firms only hire one trainee per manager position, see Appendix Figure A.2.\(^{35}\) The increase in the number of trainees at the top of the firm productivity distribution is consistent with the model mechanism.

\(^{35}\)Note that this fact represents a lower bound because trainees who never become a manager at any firm are unobserved.
Table 3: Parameters for Estimation

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameterization</th>
<th>Parameters</th>
<th>Description</th>
<th>Parameterization</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>distr of productivity</td>
<td>$F_{\phi}(x) = \frac{1 - (\frac{\phi_{\min}}{\phi_{\max}})^{\phi_{\max}}}{1 - (\frac{\phi_{\min}}{\phi_{\max}})^{\phi_{\min}}}$</td>
<td>$\phi_{\min}$</td>
<td>share of high types</td>
<td>$p(e) = p_0 + p_1 (e - p_2)^3$</td>
<td>$p_2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\phi_{\max}$</td>
<td></td>
<td>$p_0$</td>
<td>$p_0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\theta_{\phi}$</td>
<td></td>
<td>$p_1$</td>
<td>$p_1$</td>
</tr>
<tr>
<td>distr of obs skills</td>
<td>$F_{e}(x) = \frac{1 - (\frac{e_{\min}}{e_{\max}})^{e_{\max}}}{1 - (\frac{e_{\min}}{e_{\max}})^{e_{\min}}}$</td>
<td>$e_{\min}$</td>
<td>technology</td>
<td>$y = (\phi z)^{1-\alpha} \kappa^{\alpha}$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$e_{\max}$</td>
<td>exog. separations</td>
<td>$\delta$</td>
<td>$\delta$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\theta_{e}$</td>
<td>training success</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distr of ability</td>
<td>$a_L = 0$</td>
<td>$a_H$</td>
<td>manager premium</td>
<td>$w_{in}^{data}(e) = w_p + 1.2 w_n(e)$</td>
<td>$w_p$</td>
</tr>
<tr>
<td>firm knowledge</td>
<td>$f$</td>
<td></td>
<td>discount factor</td>
<td>fixed</td>
<td>$\beta$</td>
</tr>
</tbody>
</table>

4 Estimation

4.1 Estimation Procedure

In order to quantify misallocation of talent and to analyze wage inequality between managers and workers, I estimate the model on the Danish firms described in Table 1.

Parameterization I estimate a fully parametric version of the model as listed in Table 3. In particular, I assume bounded Pareto distributions for both firm productivity and observable manager skills to allow both to drive the long upper tail of profits and span of control. I normalize low ability to zero and I choose a flexible cubic function for the share of high ability types at each observed skill level $p(e)$ to match the curvature in the empirical wage profiles of trainees and managers. I specify a concave success rate of learning between zero and one where a larger $\lambda$ implies more costly improvement in training success.\(^{36}\) I normalize efficiency units of human capital in terms of the lowest skill trainee who is indifferent between worker and manager career. Thus, earnings of the lowest skill trainee determine $w_n$.\(^{37}\) I further assume that worker human capital increases by 20% between the trainee and manager stage to set the outside option of the lowest skill manager in the market, $e$.\(^{38}\) I interpret the remaining salary gap between trainees and managers as a premium $w_p$ that compensates for working longer hours or

\(^{36}\)This function can be derived from a simple probability of success $\kappa(y) = y/(1+y)$ where $y$ are training units that become increasingly expensive according to the quadratic function $y = \lambda x^2$.

\(^{37}\)Since span of control is a key data moment that the model matches, the simulated economy satisfies market clearing at this wage rate.

\(^{38}\)This is consistent with Dustmann and Meghir (2005) who find large annual experience gains for skilled workers early in their career (7.1% in the first year, then decreasing to 1.2% after 4 years).
higher responsibility of managers under perfect and costless monitoring. Finally, I fix
the discount rate at 0.9 to reflect a multi-year time horizon of the trainee period.

**Estimation Method** I use simulated method of moments based on the MCMC al-
gorithm suggested by Chernozhukov and Hong (2003). This derivative-free procedure
is computationally attractive because it can deal with non-smooth objective functions
and a high-dimensional parameter vector.\(^{39}\)

The estimation matches five outcomes for each firm that are measurable in the
data and that identify the key parameters in the model. In particular, I use output
per manager, manager salaries, trainee salaries, span of control per manager, and the
share of internal promotions in managerial hiring. Since firms typically have several
managers in the data, I assume a constant returns technology to accumulate output
from multiple manager-worker teams within a firm. I estimate the distribution of each
firm-level outcome across the productivity distribution using first-order local polynomial
regression weighted by an Epanechnikov kernel. Solid lines in Figure 8 summarize the
empirical distributions of these moments including dotted 95% bootstrap confidence
intervals. I evaluate each empirical distribution at 2% increments over the entire firm
distribution, which yields a total of 255 (five times 51) moments for estimation. I
use the inverse bootstrap variances at each 2% increment as weights in the estimation
procedure.

**Identification** I provide an intuitive argument for identification. Output per man-
ger per manager at each firm percentile is closely related to the distribution of firm and manager
types and the technology parameter \(\alpha\) in equation (3). Firm types are separately
identified from manager types through the slope of the wage function for managers in
equation (11), which only depends on the distribution of firm productivity. Moreover,
the bounds of the firm type distribution are identified by the empirical curvature of span
of control and output per manager because increasing the lower (upper) bound affects
these outcomes at a decreasing (increasing) rate throughout the firm distribution.

The distinction between the value of high ability and the value of firm-specific
human capital is mainly driven by the shape of the wage functions for trainees and
managers across firms compared to the share of external versus internal hiring across
the firm distribution. A higher value of unobserved general ability generates steeper
wage schedules for trainees because firms will bid up wages for the most promising
candidates early in their career. At the same time, the manager wage function also

\(^{39}\)Details about the MCMC algorithm are in Appendix C.2.
becomes steeper because of a higher general ability premium. In contrast, a higher productivity of firm-specific human capital is associated with flatter trainee wages as internal training becomes a better substitute for competing for promising young talent. As a result, the share of internal promotions increases. Manager wages will experience a level shift through GE wage adjustments.

The parameters for the share of high ability managers across the observed skill distribution are identified based on the shape of the external hiring share across firms in the data and the curvature in the empirical wage profiles of trainees and managers.

The share of exogenous separations is mainly related to the level of external hiring across firms in the data and to the shape of the wage function for managers. More exogenous separations yield higher external hiring shares and steeper wages for managers because of lower adverse selection, while at the same time reducing training success and thereby lowering span of control, output of managers and trainee salaries.

Finally, the cost of internal training affects training investment across firms holding the productive value generated by successful training constant. This parameter is determined by the zero profit condition for the lowest productivity firm in the market.

4.2 Results

Table 4 reports the parameter estimates. As illustrated by Figure 7, the distribution of firm productivity shows a long Pareto tail with a small share of high productivity firms as opposed to a large mass of firms with much lower productivity. In contrast, the shape parameter for the observed skill distribution is close to zero, which implies that the distribution of observable skills is close to uniform. The results show large productivity gains from both firm-specific knowledge and high ability. The value of a manager with firm knowledge at the median firm is USD 530,000 per year, while the value of a high ability manager is USD 2.2 million. Initial sorting is based on a noisy signal of true managerial talent because Figure 7 shows a probability of finding a high ability manager at the bottom of the observed skill distribution of around 2%, whereas it reaches almost 30% at the top of the skill distribution. Moreover, I estimate that firms face a risk of 38% that candidates will leave after their trainee stage.

Figure 8 emphasizes the ability of the model to closely match the key features of the data.\textsuperscript{40} The Pareto distribution for firm productivity prevents an even closer fit in the tails of the distribution for profits and span of control as well as in the shape of the external hiring distribution. The lower tail of firms in the data is less productive than

\textsuperscript{40}I illustrate the deviation between model and data, weighted by the inverse variance of the data moments, in the bottom right panel in Figure 8.
Table 4: Estimation Results

<table>
<thead>
<tr>
<th>Param.</th>
<th>All firms</th>
<th>Manufacturing</th>
<th>Retail</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{min}$</td>
<td>91.0 [49.2, 246.1]</td>
<td>26.4 [18.5, 76.2]</td>
<td>6.1 [4.5, 11.1]</td>
<td>6125 [4290, 51587]</td>
</tr>
<tr>
<td>$\phi_{max}$</td>
<td>855.8 [410.8, 2086.6]</td>
<td>162.5 [120.4, 424.2]</td>
<td>53.2 [32.3, 101.8]</td>
<td>75933 [36166, 509559]</td>
</tr>
<tr>
<td>$\theta_\phi$</td>
<td>1.660 [1.448, 2.167]</td>
<td>2.245 [1.60, 3.48]</td>
<td>1.182 [0.816, 1.657]</td>
<td>1.020 [0.192, 1.785]</td>
</tr>
<tr>
<td>$e_{min}$</td>
<td>32.0 [31.3, 37.1]</td>
<td>80.0 [41.3, 107.9]</td>
<td>46.7 [40.4, 51.4]</td>
<td>29.5 [1.6, 81.7]</td>
</tr>
<tr>
<td>$e_{max}$</td>
<td>46.6 [43.6, 59.3]</td>
<td>102.0 [62.8, 131.5]</td>
<td>54.0 [49.1, 58.5]</td>
<td>43.0 [17.8, 132.1]</td>
</tr>
<tr>
<td>$\theta_e$</td>
<td>0.000 [0.000, 0.000]</td>
<td>0.000 [0.000, 0.464]</td>
<td>1.548 [0.02, 11.7]</td>
<td>1.799 [0.035, 58.2]</td>
</tr>
<tr>
<td>$a_H$</td>
<td>3053 [2933, 4488]</td>
<td>4532.0 [2621, 5834]</td>
<td>1884.7 [1243, 3939]</td>
<td>5084.1 [2670, 8129]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>51.0 [43.9, 76.4]</td>
<td>126.0 [84.1, 170.8]</td>
<td>53.5 [48.6, 59.8]</td>
<td>43.3 [24.7, 127.9]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.299 [0.222, 0.446]</td>
<td>0.500 [0.406, 0.500]</td>
<td>0.345 [0.232, 0.494]</td>
<td>0.247 [0.11, 0.498]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.040 [0.005, 0.07]</td>
<td>0.005 [0.001, 0.01]</td>
<td>0.651 [0.086, 2.44]</td>
<td>0.044 [0.000, 0.367]</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.380 [0.343, 0.422]</td>
<td>0.375 [0.289, 0.470]</td>
<td>0.327 [0.264, 0.372]</td>
<td>0.425 [0.278, 0.601]</td>
</tr>
</tbody>
</table>

Value of general versus firm-specific skills

$a_H / f$ | 3.19 [2.34, 5.59] | 2.19 [1.52, 4.35] | 5.50 [2.44, 8.67] | 3.28 [0.95, 8.69] |

Notes: Firm types are measured in millions of USD, the scaling factor of $p(e)$ is multiplied by 1000 to improve readability. $w_p$ denotes the wage premium for managers. 95% Bootstrap confidence intervals in brackets.

Figure 7: Illustration of Results

implied by the Pareto shape, whereas there is a faster improvement in firm productivity at the upper tail in the data.

Since 75% of managers are either internally promoted or hired externally within the same sector, I also estimate the model separately for three sufficiently large sectors, (i) manufacturing, (ii) wholesale and retail, and (iii) business activities.\(^\text{41}\) The results in columns (2)-(4) of Table 4 show that firm-specific human capital is most valuable in manufacturing. A high ability manager in manufacturing is only about twice as productive as a low ability manager with firm knowledge, while the advantage of high ability managers in retail is by a factor of 5.5. These differences in the relative value of firm-specific human capital across sectors are consistent with firm-specific production

\(^{41}\)See Appendix Table A.1 for details on mobility across sectors conditional on external hiring. I report the sector-specific data moments and the goodness of fit by sector in Appendix C.3.
Figure 8: Data Moments and Goodness of Fit

- **Output**
  - Value added per manager (in millions of USD)
  - Percentile of Firm Productivity Distribution

- **Manager Salaries**
  - Manager Salaries (in thousands of USD)
  - Percentile of Firm Productivity Distribution

- **Trainee Salaries**
  - Trainee Salaries (in thousands of USD)
  - Percentile of Firm Productivity Distribution

- **Span of Control**
  - Workers per manager
  - Percentile of Firm Productivity Distribution

- **External Hiring and Internal Promotion**
  - External hiring share
  - Percentile of Firm Productivity Distribution

- **Goodness of Fit**
  - Weighted Deviation between Data and Model
  - Percentile of Firm Productivity Distribution
processes across manufacturing firms but more standardized procedures in the wholesale and retail sector which make managerial skills more transferable across firms.\footnote{\textit{Bloom et al. (2015)} argue that more innovative industries in terms of R&D and patents focus more on people management and developing talent. This is consistent with the estimated training profiles. Based on the estimate for $\lambda$ in Table 4 and the simulated training success rate, training investment is in the range of USD 45,000-140,000 in manufacturing and USD 10,000-25,000 in retail.}

The results also illustrate that hard skills matter most in manufacturing. Observable skills are more valuable and they provide a better signal of ability than in other sectors where it is more difficult to distinguish types ex ante. Figure 7 shows that the share of high ability managers strongly increases with observed skills in manufacturing, whereas the distribution of high types is less related to observed skills in the retail sector and the overall economy.

5 Counterfactuals

5.1 Signal Precision and Full Information

Based on the estimated model, I quantify the effect of alleviating information frictions on the allocation of resources and wage inequality. I first consider a marginal improvement in the precision of the initial signal $e$. Specifically, I rotate the function $p(e)$ to increase the coefficient of variation for the distribution $p(e)$ by 1% while holding the total share of high ability managers constant. This corresponds to stronger sorting of high types along the observed skill distribution as illustrated by the dotted line in the right panel of Figure 9. I complement this exercise with a “full information” benchmark where managerial talent is fully observed at labor market entry.

Panel A of Table 5 reports the results for a marginal reduction in information frictions. The first column shows that for a 1% increase in the signal precision, the economy generates a 0.11% increase in total revenue. This gain is realized by reallocating resources from low to high productivity firms, see the left panel of Figure 9.

Importantly, the overall reallocation is mitigated by heterogeneous changes in training intensity. Total training investment in the economy decreases by 0.18%, but Figure 9 shows that the least productive firms in fact increase training investment, whereas the most productive firms reduce training substantially. Intuitively, stronger sorting changes the option value of external hiring along the firm distribution. Highly productive firms now face a better chance of meeting a high ability type through the external market, which makes internal training less valuable. In contrast, less productive firms increasingly rely on internal training because their chances to hire a high ability type...
from the external market have decreased.\textsuperscript{43}

Note that the distribution of the change in revenue depends on the exact distribution of the reduction in information frictions. The right panel of Figure 9 shows that the largest reduction in high ability types occurs at the bottom of the observed skill distribution and there is a substantial increase in the top quintile. Figure 9 illustrates that in addition to the increase in allocative efficiency in the market, this change in the signal precision comes at the cost of increased wage dispersion. First, a level increase in worker wages improves the outside option of trainees and managers. But this worker

\textsuperscript{43}In addition, competition for workers increases as talent sorting improves and more productive firms expand. Higher equilibrium wages make firm-specific human capital less attractive because the optimal span of control decreases. Under the model assumptions, this level effect of higher wages is strongest for the most productive firms.
wage increase is the lower bound of the gains for managers. Managerial wages increase more the stronger the improvement in the signal. Competition for talent also intensifies for trainees, more so for candidates who have become relatively more promising. In sum, Table 5 shows a 0.5% increase in the average manager-worker wage gap.

Based on the estimation results by sector, I compare outcomes across industries to highlight the role that firm-specific human capital plays in the equity-efficiency tradeoff. Both the aggregate revenue gain and the increase in the manager-worker wage gap are smallest for manufacturing and largest in the retail sector. This is consistent with a larger role for firm-specific human capital in manufacturing that better protects less productive firms from market competition compared to other sectors. As a result, firm knowledge particularly benefits less skilled managers employed at low productivity firms and reduces overall wage dispersion.

The full information scenario illustrates the total potential for productivity gains across sectors and further emphasizes the equity-efficiency tradeoff. If trainees can be ranked according to true talent, the best firms match with the best trainees without uncertainty about their true type. The only remaining value of internal labor markets is to accumulate firm-specific human capital. Low ability managers can no longer benefit from their skills \( e \) being a noisy signal of ability. As a consequence, manager compensation among low ability types is very flat. In contrast, high ability managers benefit from strong competition among highly productive firms to hire them.\(^4\)

Panel B of Table 5 estimates an economy-wide productivity gain of 22.5%, which compares to a 12.5% gain in manufacturing and 28.1% in wholesale and retail. Intuitively, potential productivity gains in manufacturing are lower because observable skills are a more precise signal of general managerial talent than in other sectors. As a result, even under asymmetric information, manufacturing firms are making better informed decisions than firms in business and the retail sector. Moreover, firm-specific human capital is relatively more important in manufacturing and better protects firms from market competition even under full information about general ability.

With full information, firms up to the 80th percentile of the productivity distribution lose resources and profits because all high ability managers are now employed by the most productive firms. Firms no longer receive rents from asymmetric information and total profits decrease substantially. Even high productivity firms that always hire high ability managers under full information do not uniformly gain in terms of profits. The least productive firms within this group, between the 80th and 90th percentile of the

\(^4\)The best firm hiring a low ability manager bids up the managerial wage rate at the worst firm still hiring a high type. Hence, the best firms pay their managers millions of dollars per year, whereas observable low manager types receive only a small fraction of that amount.
firm distribution, would prefer to stay in a world with asymmetric information to avoid the intense competition for high ability managers.

Finally, productivity gains from full information are magnified by adjustments in training investment and firm exit. I find large increases in training across all sectors if there is no uncertainty about the true trainee type and no option value of external hiring. Moreover, some firms at the bottom of the productivity distribution in retail and business will make losses under full information. As a result, Table 5 provides a lower bound for the productivity gains if the distribution of firms is held fixed. Firm exit will yield further reallocation of resources to high productivity firms.

5.2 Secular Trends in the Market for Managers

The second part of the counterfactual analysis contributes to the ongoing debate about different secular trends that have been documented in the U.S. and that I find in Denmark as well. First, there is a steep increase in executive compensation over time. Frydman and Jenter (2010) report an increase in average CEO compensation at S&P500 firms by 310% over 1992-2001. Looking at top managers more broadly, I find a 40% real increase in compensation over 1991-2008 in Denmark.\(^{45}\) Second, there is a continuing increase in the wage gap between managers and workers. Autor et al. (2008) report that the 90/50 wage gap grows by 1 log point per year in the CPS 1963-2005 and Murphy (1999) shows the increasing gap between CEOs and production workers. For Denmark 1991-2008, I find that the wage gap between top managers and workers grows by 1.4 log points per year. Third, Murphy and Zabojnik (2007) document that external hiring of CEOs at Forbes 800 companies increased from 15% in the 1970s to 26% in the 1990s. For Denmark, I estimate an increase in external hiring by 1.7% per year over 1999-2008.\(^{46}\)

The literature has argued that firm growth plays an important role in understanding the increase in manager compensation (Gabaix and Landier, 2008). I first simulate the effects of a 1% increase in average firm productivity, modeled as a proportional increase in the boundaries of the firm distribution. The first column in Table 6 shows that this counterfactual indeed yields a 0.4% increase in average manager compensation but it cannot capture the simultaneous increase of inequality and external hiring in the data.

I then analyze three different scenarios that affect the attractiveness of internal labor markets. The results suggest an important role for internal labor markets to reconcile

---

\(^{45}\)I provide details about the secular trends in Denmark in Appendix A.2.6.

\(^{46}\)I estimate a Probit model of external hiring with a linear time trend, controlling for region and industry FE, as well as the full set of job and firm characteristics as described in section 2.
the different secular trends. Consider first a 1% increase in the value of high ability managers in the second column of Table 6. This scenario is motivated by improvements in monitoring and information technology that increase the leverage of a highly talented manager more than for less able candidates. The model simulation suggests that this situation will increase average manager compensation, but at the same time lead to higher inequality and more external hiring. Intuitively, general talent becomes relatively more valuable and market competition for the best managers increases. However, the elasticity of manager compensation is an order of magnitude larger than the change in manager-worker wage gap and external hiring share.

I compare this scenario to a pure improvement in information as in the previous subsection, for example motivated by increasing use of career platforms, headhunter services and machine learning techniques to improve hiring. The results of this exercise show a strong positive response in salaries and inequality, while the increase in hiring is an order of magnitude smaller. Finally, a 1% decrease in the value of firm-specific human capital is consistent with both the increase in inequality and higher external hiring. This scenario is motivated by an increase in transferability of skills because of MBA programs and standardized management practices.

In sum, these simulations illustrate that a relative decrease in the value of internal labor markets can help explain secular trends in hiring and wage inequality that productivity growth cannot fully capture. These counterfactuals illustrate the value of future research to empirically estimate the causal effect of changes in information frictions, improvements in monitoring and in the transferability of skills on wage inequality and hiring behavior.

6 Conclusion

This paper studies how firms compete for managerial talent. New stylized facts from matched employer-employee data show large heterogeneity in internal promotions versus
external hiring across firms. More productive firms hire more talented trainees, are more likely to promote managers internally, and match with better managers in terms of education and ability. Firms select the best internal candidates for promotion, but know less about the ability of external hires.

Based on these facts, I develop and estimate an assignment model of the market for managers with two-sided heterogeneity and with an explicit role for internal labor markets. The model illustrates the main tradeoffs that a firm faces about internal versus external hiring of managers in an environment with asymmetric employer learning and firm-specific human capital. Firms can compete for the most promising talent before promotion to acquire private information about her true talent. Firms can invest in firm-specific training to make their internal candidate more productive, but they face uncertainty about training success and the true ability of their candidate. Alternatively, firms can hire their manager externally, but they face an informational disadvantage and an adversely selected pool of candidates. Production complementarities between firm productivity and manager talent result in better firms investing more in promising workers and in developing talent through firm-specific training and internal promotion.

The counterfactual analysis emphasizes an equity-efficiency tradeoff from reducing information frictions. More generally, I show that a decrease in the attractiveness of internal labor markets - because of a decrease in the value of firm-specific knowledge or lower information frictions - helps explain the secular increase in external hiring and in upper-tail wage inequality that previous models of the market for managers based on firm heterogeneity cannot fully capture.

The model is based on several stylized assumptions that provide a stepping stone for future work on internal labor markets. First, a distinction between top and middle management would allow for a more nuanced information structure, in particular by treating top management promotions as public signals. Second, firms endogenously choose their hierarchical structure and therefore promotion opportunities. This is challenging for the theoretical analysis because the hierarchy becomes an endogenous determinant of firm productivity (Caliendo and Rossi-Hansberg 2012). Yet incorporating training and promotion decisions into models of organizational choice may help understand how firms respond to exogenous shocks by adjusting their internal structure and how restructuring affects career paths (Friedrich 2015). Third, the framework can enrich studies of moral hazard problems at the management level. The relative value of outside options compared to the internal career ladder determines to what extent managers have to be incentivized through performance pay to address monitoring problems. Finally, the model can be used to understand hiring and promotion strategies for other
occupations as well. Collecting additional survey data on talent development within firms that can be linked to administrative data (Bloom et al., 2013a) could complement these theoretical advances with much needed empirical evidence across firms.

References


Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen (2014). *IT and Management in America*. CEP Discussion Papers dp1258. Centre for Economic Performance, LSE.


42


Appendix

A Data

A.1 Data Definitions, Data Cleaning and Sample Selection

I define internal promotions and external hiring based on occupational switching and job switching of workers. The three-digit occupations from the international standard classification of occupations (ISCO) that define managers are 121 - Top executives, senior management; 122 - Management of the main activity, production manager; 123 - management of special areas of large and medium-sized enterprises such as finance, personnel, sales, marketing, IT, purchasing and R&D. Occupational data is available from 1991 and I use the occupational history of individuals since 1991 to determine first-time manager promotion.

I define job movers based on plant and firm identifiers. Firm identifiers are merged using the Firm-Integrated Database for Labor Market Research (FIDA) over 1995-2008. The establishment identifier in the administrative matched data is constant over time unless there is a simultaneous change in owner and address. I track the physical identity of the plant over time using the establishment panel data (IDAS) to avoid spurious job mobility. Movers are defined as individuals who move to a different establishment that does not belong to the previous firm. Movers within the same firm who become managers at a different plant are considered internal promotions.

For the main sample as defined in Table 1, 82% of occupations are directly reported by employers, 8% are imputed based on union membership, 1% are self-reported from the population register and 9% of occupations are unknown. I clean the occupational data in the following way: I impute the occupation for the first period in a new job by the occupation in the second period in two cases: First, a new employee’s occupation is sometimes initially unknown, but the firm reports their occupation in the second period. Second, a new employee’s occupation in the first year on the job may still refer to the previous job because of timing issues in occupational reporting, but the occupation is updated in the subsequent period. Next, I exclude cases where a worker switches from unknown to managerial occupation after more than one period at the incumbent firm, if there are no managers employed at the firm in the previous period. I consider this case measurement error in internal promotions. If a worker has the same occupation over time except for one gap year in between, I impute the stable occupation for this period to avoid spurious occupational switching. Finally, I do not count managerial
hiring in the birth year of a firm because there is no tradeoff between internal and external hiring.

I use FIDA to merge information about industry and value added from General Firm Statistics (GF). Gross profits are reported starting in 1999 and defined as total revenue minus purchases of raw materials, inputs, energy and subcontracting. Value added differs from gross profits because it takes rental and leasing costs, secondary costs and changes in debt positions into account. I use gross profits as the main measure of firm performance throughout the analysis because it is not subject to fluctuations in debt positions that may be unrelated to production.

To construct the main sample, I drop firms from agriculture, mining, electricity, finance and insurance and public services because of missing data and for confidentiality reasons in industries with a small number of firms. The sample in Table 1 accounts for 30.7% (29.0%) of total economic activity in Denmark over the period 1999-2008 in terms of profits and value added respectively.

A.2 Stylized Facts

A.2.1 Manager Transitions Across Sectors

Table A.1 shows the share of manager mobility within sector. Overall, about 40% of all external managerial hires in the sample previously worked in a different sector. This number hides considerable variation across sectors and particularly across different types of manager positions as illustrated in Table A.1. For example, external hiring from other sectors is less common in the retail sector across all managerial positions than in finance and business. Moreover, department managers such as sales and marketing managers move across sectors frequently, suggesting that they possess skills that are better transferable across sectors. In contrast, production managers are more likely to move within a sector, consistent with their product-specific knowledge. As a result, more than 70% of external hires for the latter group move within sectors while only 51.4% of externally hired department managers come from the same sector. In sum, these transition rates do suggest some role for technological restrictions in different industries and for different manager positions, but they also emphasis considerable manager mobility across sectors.
Table A.1: Share of Transitions Within Industry

<table>
<thead>
<tr>
<th>Sector</th>
<th>All External</th>
<th>Top Managers</th>
<th>Production</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td>0.556</td>
<td>0.582</td>
<td>0.661</td>
<td>0.505</td>
</tr>
<tr>
<td>Construction</td>
<td>0.589</td>
<td>0.543</td>
<td>0.652</td>
<td>0.485</td>
</tr>
<tr>
<td>Retail and Wholesale</td>
<td>0.664</td>
<td>0.606</td>
<td>0.778</td>
<td>0.547</td>
</tr>
<tr>
<td>Transport and Telecom</td>
<td>0.457</td>
<td>0.615</td>
<td>0.500</td>
<td>0.425</td>
</tr>
<tr>
<td>Finance and Business</td>
<td>0.519</td>
<td>0.574</td>
<td>0.495</td>
<td>0.518</td>
</tr>
<tr>
<td>Total</td>
<td>0.591</td>
<td>0.587</td>
<td>0.705</td>
<td>0.514</td>
</tr>
</tbody>
</table>

A.2.2 Residual Hiring Strategies

I estimate unobserved heterogeneity in firm-level hiring behavior based on a linear probability model of managerial hiring,

\[ H_{ijt} = \beta X_{it} + \gamma_1 Z_{jt} + \gamma_2 Z_{s(j)t} + u_j + \epsilon_{ijt} \]  

(22)

where the dependent variable \( H_{ijt} \) is a binary variable that takes a value of one if managerial position \( i \) at firm \( j \) in time \( t \) is filled by an external hire and zero if a manager is promoted internally. The main object of interest is firm-level propensity for external hiring, \( u_j \). \( X_{it} \) are characteristics of the particular job \( i \), for example middle management versus top executive level positions, \( Z_{jt} \) are time-varying characteristics of firm \( j \), for example the growth rate of employment, the turnover rate across all employees, revenue and the age of the firm, and \( Z_{s(j)t} \) are characteristics of firm \( j \)’s environment \( s(j) \) that are time-varying. In practice I include industry-time fixed effects and region-time fixed effects to capture local labor market conditions and technological change at the industry level.

I use a fixed effects model to estimate equation (22); this model allows for correlation between firm hiring behavior \( u_j \) and other covariates, in particular firm characteristics \( Z_{jt} \). Firm effects are defined relative to the industry average hiring behavior for managers and illustrate the propensity of a firm towards external hiring as opposed to internal promotion. One main object of interest is the variance of firm-fixed effects, \( \sigma_u \) which characterizes the dispersion in hiring strategies within an industry. This is analogous to the literature on teacher effects that measures the dispersion in value added across teachers and the effect of different quality teachers on test scores of students. Here, industries replace the role of schools in the teacher literature because they provide the environment in which the teacher or the firm operates. For example the quality of students might differ across schools, making it easier or harder for a teacher to influence outcomes. Analogously, industries differ in their technology and the role
Table A.2: First Stage: Firm FE of Managerial Hiring

<table>
<thead>
<tr>
<th>Dependent Variable: Dummy for External Hiring of Manager ( H_{i,j,t} )</th>
<th>No FE</th>
<th>FE</th>
<th>FE</th>
<th>FE</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Top Manager</td>
<td>0.0946***</td>
<td>0.0442***</td>
<td>0.0573***</td>
<td>0.0528***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>Department Manager</td>
<td>0.0468***</td>
<td>0.0125*</td>
<td>0.0261***</td>
<td>0.0238***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>Employment Growth</td>
<td>0.0959***</td>
<td>0.1146***</td>
<td>0.1147***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.018)</td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnover rate</td>
<td>0.2828***</td>
<td>-0.0208</td>
<td>-0.0203</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.073)</td>
<td>(0.074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(sales)</td>
<td>-0.0081</td>
<td>0.0257*</td>
<td>0.0262*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.014)</td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(assets)</td>
<td>0.0055**</td>
<td>-0.0192***</td>
<td>-0.0202***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(employment)</td>
<td>-0.0253***</td>
<td>0.0052</td>
<td>0.0037</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>30,207</td>
<td>33,248</td>
<td>33,248</td>
<td>30,207</td>
<td>30,207</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.066</td>
<td>0.189</td>
<td>0.189</td>
<td>0.196</td>
<td>0.205</td>
</tr>
<tr>
<td>F-test</td>
<td>5.568</td>
<td>5.512</td>
<td>5.018</td>
<td>4.916</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. The omitted job category is production managers; sales, assets and employment are lagged by one period. The turnover rate measures the share of exits between the previous and current period relative to last period’s total employment. All regressions include industry-time fixed effects, model (5) adds region-time fixed effects. The null of the F-test is that all firm fixed effects are jointly zero, which is strongly rejected in all specifications.

One concern with the empirical model in \( (22) \) is that the dependent variable is binary. Yet, one can think of the linear probability model as a linear approximation of a Probit or Logit model around \( p = 0.5 \). Since firms use a mixture of internal and external hiring such that the frequency of external hiring is between 30 and 70%, this approximation seems reasonable for my application. Since the main variable of interest is the firm fixed effect, nonlinear estimation would require explicitly including the full set of dummy variables for firms. This is computationally challenging and leads to the incidental parameters problem because the number of managerial hires per firm is limited.

I report results including different sets of control variables and specifications in Table
A.2. The F-test is highly significant in all specifications and rejects the hypothesis that the firm-fixed effects are zero. The R-squared shows a reasonable fit of the model, with more than 20% of the variation explained by industry, time and firm fixed effects as well as job and firm characteristics.

In the next step, I quantify the dispersion in firm-level hiring strategies based on the fixed-effect estimates in Table A.2. In addition to the unweighted standard deviation, I compute the weighted dispersion based on the number of hires. I also use different strategies from the literature on teacher value added to account for sampling error. First, analogous to Aaronson et al. (2007), I adjust the variance of $\hat{u}_j$ by assuming that the estimated firm fixed effects consist of the true firm-specific propensity to hire externally, $u_j$ and an additive error term $\xi_j$ that is uncorrelated with the true firm fixed effect,

$$\hat{u}_j = u_j + \xi_j.$$  

I estimate adjusted dispersion in hiring strategies $\sigma_u^2$ by subtracting the variance of $\xi$, measured by the average over the squared standard error estimates for $\hat{u}$, from the observed variance in firm-fixed effects $\sigma^2 u$,

$$\sigma^2 u = \hat{\sigma}^2 u - \frac{1}{n} \sum_{j=1}^{n} \hat{se}(\hat{u}_j)^2. \quad (23)$$

Second, another strand of the literature computes empirical Bayes estimates by multiplying the firm-fixed effect $\hat{u}_j$ by a shrinkage factor that reflects the signal to noise ratio (see for example Kane and Staiger 2008). In my application, the estimate of firm hiring behavior for each firm is scaled by the ratio between the true variance in firm fixed effects from equation (23) and total variation for this particular firm given by the sum of $\sigma_u^2$ and the standard error for this firm’s fixed effect estimate. Then the true dispersion in hiring strategies across firms follows from the variance over the empirical Bayes estimates. Third, I follow the two-step procedure proposed by Chetty et al. (2014) by first consistently estimating the coefficients on the control variables from a fixed effects regression according to equation (22) to get residuals

$$e_{ijt} = u_j + \epsilon_{ijt}.$$  

I then estimate firm-fixed effects as best linear predictors of the residual terms in the second step to minimize the mean squared error of the forecasts. Finally, I compare these measures to the weighted standard deviation of firm-fixed effects where the weight for each firm is given by the number of managerial hires relative to total observed hires.
Table A.3: Dispersion of hiring strategies relative to industry average

<table>
<thead>
<tr>
<th>FE Specification</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.22</td>
<td>0.22</td>
<td>0.224</td>
<td>0.225</td>
</tr>
<tr>
<td>Weighted standard deviation</td>
<td>0.188</td>
<td>0.187</td>
<td>0.196</td>
<td>0.195</td>
</tr>
<tr>
<td>Adjusted standard deviation</td>
<td>0.192</td>
<td>0.191</td>
<td>0.194</td>
<td>0.194</td>
</tr>
<tr>
<td>Empirical Bayes Method</td>
<td>0.166</td>
<td>0.165</td>
<td>0.167</td>
<td>0.166</td>
</tr>
<tr>
<td>Chetty Method</td>
<td>0.198</td>
<td>0.197</td>
<td>0.188</td>
<td>0.232</td>
</tr>
</tbody>
</table>

in the sample.

The results are reported in Table A.3. The key finding across different methodologies and model specifications is that there is a large variation in firm-level hiring strategies within industries. Using the different adjustment methods, the dispersion in hiring relative to the industry average is between 16.5 and 23.2 percentage points.

A.2.3 Firm Ranking

This subsection illustrates the robustness of measuring firm productivity by gross profits per manager. Alternatively, I use accounting data for capital stock, employment, investment, energy expenses in order to estimate TFP based on standard models for value added. I use alternative identification assumptions based on investment decisions (Olley and Pakes, 1996) or energy consumption (Levinsohn and Petrin, 2003) to estimate firm-level TFP from industry-specific models of production. The main results apply the Ackerberg et al. (2015) method for weaker identifying assumptions. I normalize the estimates by the average productivity in the baseline year 2000 in order to make the results more easily interpretable across industries and time. The results in Figure A.1 show that TFP estimates from standard models of industrial organization correspond closely to the firm ranking based on gross profits per manager. TFP at firms with the highest productivity rank is estimated to be 30 log points above the industry average in 2000, whereas average TFP for the lowest ranked firms is 10 log points below the industry average. The dispersion using the Olley-Pakes method yields similar productivity dispersion, while the results using Levinsohn-Petrin suggest slightly larger differences at the top of the distribution. In sum, all estimates emphasize that profits per managers and other measures of firm productivity are highly correlated.
The sample consists of 927 firms from the main sample for which data on value added, capital, employment and investment is available over the sample period. TFP estimates report the log difference to the industry average TFP in 2000.

### TFP estimates: Ackerberg-Caves-Frazer

<table>
<thead>
<tr>
<th>Productivity Rank: Profits per Manager</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFP estimates: Levinsohn-Petrin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Productivity Rank: Profits per Manager</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFP estimates: Olley-Pakes</td>
</tr>
</tbody>
</table>

### Table A.4: Internal Candidates versus External Hires

<table>
<thead>
<tr>
<th>Manager Type</th>
<th>Obs</th>
<th>Male</th>
<th>Age</th>
<th>Schooling</th>
<th>Experience</th>
<th>Tenure</th>
<th>Wages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>943</td>
<td>0.933</td>
<td>46.27</td>
<td>14.58</td>
<td>22.23</td>
<td>8.11</td>
<td>70.66</td>
</tr>
<tr>
<td>External</td>
<td>1474</td>
<td>0.929</td>
<td>44.60</td>
<td>14.64</td>
<td>19.70</td>
<td>1.90</td>
<td>74.53</td>
</tr>
<tr>
<td>Diff</td>
<td>-0.004</td>
<td>-1.666***</td>
<td>0.055</td>
<td>-2.531***</td>
<td>-6.213***</td>
<td>3.872</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.361)</td>
<td>(0.089)</td>
<td>(0.347)</td>
<td>(0.220)</td>
<td>(2.742)</td>
<td></td>
</tr>
<tr>
<td>Production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>5791</td>
<td>0.792</td>
<td>38.82</td>
<td>13.35</td>
<td>17.17</td>
<td>6.23</td>
<td>30.16</td>
</tr>
<tr>
<td>External</td>
<td>6818</td>
<td>0.808</td>
<td>36.95</td>
<td>13.54</td>
<td>15.07</td>
<td>1.53</td>
<td>31.45</td>
</tr>
<tr>
<td>Diff</td>
<td>0.016**</td>
<td>-1.871***</td>
<td>0.194***</td>
<td>-2.099***</td>
<td>-4.706***</td>
<td>1.287***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.182)</td>
<td>(0.036)</td>
<td>(0.155)</td>
<td>(0.076)</td>
<td>(0.331)</td>
<td></td>
</tr>
<tr>
<td>Department</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>8028</td>
<td>0.75</td>
<td>41.08</td>
<td>13.94</td>
<td>18.44</td>
<td>6.62</td>
<td>35.93</td>
</tr>
<tr>
<td>External</td>
<td>10194</td>
<td>0.764</td>
<td>39.00</td>
<td>14.24</td>
<td>15.81</td>
<td>1.38</td>
<td>38.19</td>
</tr>
<tr>
<td>Diff</td>
<td>0.014**</td>
<td>-2.077***</td>
<td>0.295***</td>
<td>-2.639***</td>
<td>-5.244***</td>
<td>2.263***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.131)</td>
<td>(0.034)</td>
<td>(0.124)</td>
<td>(0.063)</td>
<td>(0.338)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The sample includes all managerial hires in the main sample by manager position. Wages are hourly wages in USD deflated by CPI with base year 2000. Standard errors in parentheses.
A.2.4 Differences between Internal and External Hires

Table A.4 compares age, education, experience and wages for internal promotion versus external candidates by manager category. Across both top and middle management positions, internal candidates are significantly older on average, they have lower formal education but higher labor market experience and 6-8 years of firm tenure until promotion (including the year of promotion) on average. Moreover, external candidates receive higher hourly wages than internal candidates at all management levels. But the advantage for external hires is only statistically significant for production and department managers who receive one or two dollars more per hour worked.

A.2.5 Trainee-Manager Transitions and Earnings

The top left of Figure A.2 shows the number of trainees per manager within the sample across the firm productivity distribution. The most productive firms hire about two trainees per manager position filled, while the least productive firms only hire about half as many trainees as managers. Note that by definition, trainees are employees who become manager within the sample period. Second, the top right figure illustrates that across the manager type distribution, external hires are more likely to move to lower ranked firms. This pattern is more pronounced for managers with lower unobserved ability. In the second row of Figure A.2, I illustrate the origin of external hires across the firm productivity distribution. Since the share of hires from lower or higher productivity firms mechanically depends on a firm’s position in the ranking itself, it is best to interpret these figures compared to a benchmark of random hiring. Firms with similar productivity are defined as within 10 percentage points from a firm’s ranking, so random external hiring would suggest a share of 20% from these firms. The results show that the most productive firms hire more than proportionately from other high productivity firms, whereas less productive firms often hire external managers from higher ranked firms. The most productive firms also hire a substantial share from less productive firms, but the share is lower than predicted by random hiring.

A.2.6 Trends in Manager Compensation and the Manager-Worker Gap for Denmark

This subsection summarizes trends in the market for managers in Denmark based on the manager panel in IDA. The left panel in Figure A.3 shows a 40% increase in average real earnings for top managers (ISCO code 121) over 1991-2008. Earnings for middle managers, defined as production and department managers (ISCO codes 122, 123),
Figure A.2: Model Extensions: Stylized Facts

Notes: For trainees per manager, trainees are counted once, 5 years before their first manager promotion, and managers are counted once at the time of their first promotion, to reflect trainee positions per manager position. Transitions across firms use all manager moves within the firm sample. Manager quality is estimated as the manager FE from the AKM regression in section 2. Firm productivity rank is determined on a scale from 0 to 1 based on gross profits per manager. For the bottom row figures, similarly ranked firms are firms up to 10 percentage points higher or lower than each firm’s own rank.

increased by 24%, while professionals (ISCO groups 2 and 3) gained 18.5% and real earnings for blue-collar and white-collar workers (ISCO-88 groups 4-9) only increased by 4%. The right panel of Figure A.3 shows that the log difference in average hourly wages between top managers and workers increased continuously over time, by a total of 30 log points over an 18 year period. The wage gap between middle managers and workers increased quickly over the 1990s but remained stable during the 2000s. Finally, I illustrate the increase in external hiring for first-time manager promotions over time in the bottom panel. The figure shows cyclicality in internal versus external hiring, but also a clear trend towards more external hiring over time. Using a Probit model with the full set of job and firm characteristics, region and industry fixed effects as controls, I estimate a highly significant coefficient on the time trend for external managerial hiring of 1.7 percentage points per year.
Notes: The results are based on all private sector employees by occupation groups over 1991-2008. All wages are deflated by CPI in base year 2000. Top managers are defined by ISCO=121, middle managers consist of production managers (ISCO 122) and department managers (ISCO 123). Professionals are defined as main ISCO-88 categories 2 and 3, while blue-collar and white-collar workers consist of occupation categories 4-9 in ISCO-88. The left panel uses full-time, full-year equivalent earnings based on average hours and hourly wages reported in the data. The right panel computes the log difference in hourly wages. The bottom panel considers first-time manager promotions.

B Model

B.1 Proof: Proposition 1

What remains to be shown for part (1) is that under the PAM assignment, the remaining equilibrium objects \(w_n\) and \(\phi\) can be uniquely determined from the market clearing condition for workers and from the zero profit entry condition.

I first show that labor demand strictly decreases in both the equilibrium wage rate and the firm entry cutoff. An increase in the entry cutoff means a corresponding increase in the manager cutoff \(c\) such that labor supply of workers increases. At the same time, with the wage rate in the market fixed, all remaining firms will keep the same hiring and training strategy, so total labor demand falls by the lower mass of active firms in
the market. An increase in the market wage rate lowers the optimal span of control for any manager in the market according to (10). Moreover, a higher wage rate will ceteris paribus reduce training investment at all firms according to equation (14). In particular,

$$\frac{dx}{dw_n} = \frac{\beta (1-\delta) \phi \kappa' (x) [f - (1-p(e)) \bar{p}(e) (a_H - a_L)] (\partial \Psi / \partial w_n)}{\beta (1-\delta) \kappa' (x) (1-p(e)) \phi \Psi (e) (a_H - a_L) - \beta (1-\delta) \phi \Psi (f - (1-p(e)) \bar{p}(e) (a_H - a_L))} < 0$$

because the term in brackets in the numerator is positive by assumption (A1) and

$$\frac{\partial \Psi}{\partial w_n} = -\alpha \frac{1}{1-a} w_n - \alpha a - 1 < 0.$$

Both of these mechanisms will lead to a decrease in excess labor demand as the wage rate increases. Hence, labor market clearing for workers is represented by a downward-sloping line in \{\phi, w_n\}-space.

Secondly, consider the free entry condition. An increase in the cutoff firm type will increase profits under assumption (A4) which guarantees that profits from a better trainee overcompensate the increase in their outside option as a worker. However, an increase in the market wage rate for workers will reduce profits of the cutoff firm because it reduces training and the optimal span of control for each manager type. Hence, the free entry condition can be illustrated as an upward sloping line in \{\phi, w_n\}-space.

Both the labor market clearing condition and the free entry condition are continuous functions in \(w_n\) and \(\phi\). The equilibrium exists by the intermediate value theorem if the wage rate that satisfies labor market clearing for \(\phi = \phi_{min}, w_n^{LC} (\phi_{min})\), is larger than the wage rate that solves the free entry condition, \(w_n^{FE} (\phi_{min})\). Assumption (A3) provides a sufficient condition that this is the case. Moreover, the relationship has to be reversed at \(\phi = \phi_{max}, w_n^{LC} (\phi_{min}) < w_n^{FE} (\phi_{min})\). This will always be true because \(w_n^{FE} (\phi_{max}) > 0\) and \(\lim_{\epsilon \to 0} w_n^{LC} (\phi_{max} - \epsilon) = 0\).

For part (2), I first characterize the equilibrium share of high ability types in the external market with positive assortative matching for both trainees and managers. In this case, firms compare the same observable skill type for the internal and external candidate. The expected difference in profits between a low type with training and a high type without training under PAM is

$$\phi \Psi \cdot [f - \bar{p}(e) [a_H - a_L]].$$

Assumption (A1) implies that trainees with successful training will never be laid off. Moreover, firms always prefer to promote their internal candidate even without firm-
specific knowledge if the person has high ability. In sum, the share of high ability managers in the external market from (5) simplifies to

$$\tilde{\phi}(e) = \frac{\delta p(e)}{\delta + (1 - \delta)(1 - \kappa(e))(1 - p(e))}$$

(24)

because high types only separate for exogenous reasons, whereas low types are laid off if their training is unsuccessful.

The result in part (2) then follows directly from the optimality conditions (11) and (15). Using (24), the partial derivative is given by

$$\frac{\partial}{\partial e} \tilde{\phi}(e) = \frac{\delta [\delta + (1 - \kappa(x)) \cdot (1 - \delta)]}{[\delta + (1 - \delta)(1 - \kappa(x))(1 - p(e))]^2} \cdot p'(e).$$

(25)

Since the partial effect $\frac{\partial}{\partial e} \tilde{\phi}(e) > 0$ if $p'(e) > 0$, the manager wage function is upward sloping, $w'_m(e) > 0$.

Similarly, I can show that under assumption (A2) and using the results in (24) and (25), $p'(e) - \tilde{\phi}'(e) > 0$ and $[p'(e)(1 - \tilde{\phi}(e)) - \tilde{\phi}'(e)p(e)] > 0$. As a result, $w'_r(e) > 0$.

The exact shape of the wage profiles for trainees and managers depends on additional assumptions about the function $p(e)$.

Finally, notice that the incentive compatibility constraint for trainees and managers from equation (6) is satisfied in equilibrium. For any skill type $e \geq e$, it must be that

$$w'_r(e) + \beta w'_m(e) > h'(e) w_n \forall e,$$

which will hold under assumption (A5) above. Moreover, equation (7) holds because assumption (A6) guarantees that

$$w'_m(e) > h'(e) w_n \forall e.$$

### B.2 Proof: Proposition 2

The result on optimal training follows from the implicit function theorem using optimality condition (14),

$$\frac{dx}{d\phi} = -\frac{\beta (1 - \delta) \kappa'(x) [f(1 - p(e)) \tilde{\phi}(e)(a_H - a_L)] + \theta' \phi (a_H - a_L)[p'(e) \tilde{\phi}(e) - (1 - p(e)) \tilde{\phi}'(e)]}{\beta (1 - \delta) \kappa'(x)(1 - p(e)) \tilde{\phi} \phi \tilde{\phi}(e)(a_H - a_L) - \beta (1 - \delta) \kappa''(x) \phi \Psi [f(1 - p(e)) \tilde{\phi}(e)(a_H - a_L)]} > 0.$$

The denominator is strictly positive if $\kappa'(x) > 0$, $\kappa''(x) \leq 0$ and under assumption (A1). Assumption (A4) is sufficient for the numerator to be positive. The second term is negative, reducing the incentive to invest in training for better firms because they
have a better chance to find a high ability type externally. Under assumption (A4), observed characteristics of managers improve sufficiently slowly and hence better firms find it more attractive to invest in internal training. Hence under PAM,

\[
\frac{dx}{de} = \frac{dx}{d\phi} \cdot \frac{d\phi}{de} > 0.
\]

Next, consider the level of adverse selection in equilibrium. The partial derivative of (24) with respect to training is

\[
\frac{\partial}{\partial x} \tilde{p}(e,x) = \delta p(e) \kappa'(x)(1-\delta)(1-p(e)) \left[ \delta + (1-\delta)(1-\kappa(x))(1-p(e)) \right] > 0,
\]

which illustrates the alleviating effect of training on adverse selection. More training makes internal promotion for low ability types more likely and therefore improves the average quality of the external pool of candidates. Using the partial derivative in (25) as well as the previous results yields

\[
\frac{d\tilde{p}(e,x)}{de} = \frac{\partial}{\partial x} \tilde{p}(e,x) \frac{dx}{de} + \frac{\partial}{\partial e} \tilde{p}(e,x) > 0.
\]

The unobserved quality of external managers strictly increases in their skill level \( e \) in equilibrium. This is due to both a positive correlation between ability and skill level in the population and to a stronger mitigating effect of training on adverse selection for better skilled trainees.

**B.3 Proof: Proposition 3**

**Lemma.** Consider an increase in the productivity of firm knowledge \( f \) under assumptions (A1)-(A7). (1) The equilibrium wage rate \( w_n \) strictly increases. (2) Further assume that assumption (A8) holds. Then training investment increases for all firms above a threshold \( \phi \geq \tilde{\phi} \).

If the value of firm-specific human capital \( f \) increases, the market wage rate \( w_n \) increases as a result of higher labor demand from two sources; directly because any manager with successful training is more productive and can supervise more workers and indirectly because higher training investment increases the frequency of training success. This result takes potential firm entry or exit into account. Next, the effect of an increase in the value of firm knowledge on training follows from two counteracting forces. On the one hand, higher worker wages make firm-specific training investment less attractive because the equilibrium wage limits the scale of production. If workers
are more expensive, firms optimally increase the span of control less in response to successful training and therefore the gains from training are smaller. On the other hand, the complementarity between firm productivity and manager talent implies that the increase in $f$ benefits high productivity firms more than proportionately. As a result, the most productive firms will strictly increase their training investment and the frequency of internal promotions. For less productive firms below the threshold $\tilde{\phi}$, the first mechanism dominates and they will reduce their training investment.

Consider an increase in the productivity of firm knowledge $f$. The equilibrium assignment is unchanged for firms that remain in the market in these comparative statics since the observed ranking of manager types is unaffected. The proof is based on totally differentiating the labor market clearing, free entry and optimal training conditions and to show the total effect of a change in $f$ on wages, training and the firm entry cutoff.

The total differential of the market clearing condition for workers yields

$$A_\phi \, d\phi + A_e \, de + A_x \, dx + A_w \, dw_n + A_f \, df = 0.$$  \hfill (26)

I provide more detailed derivations of these and the subsequent results in the Online Appendix. Next, totally differentiate the free entry condition for firms, (18), using the boundary conditions (8) and (9) to yield

$$B_\phi \, d\phi + B_e \, de + B_x \, dx + B_w \, dw_n + B_f \, df = 0.$$  \hfill (27)

Third, I take into account optimal adjustment in training according to the total differential of the first order condition for training investment at any firm productivity level,

$$C_x \, dx + C_w \, dw_n + C_f \, df = 0.$$  \hfill (28)

Finally, note that the equilibrium PAM assignment in (19) implies

$$de = \frac{M_\gamma (\phi)}{N_f (e)} \, d\phi.$$  \hfill (29)

Now I jointly solve equations (26), (27), (28) and (29) for the change in the market
wage rate in response to an increase in \( f \). I define functions

\[
\Omega_A = \left[ -A_\phi - A_e D e \phi \right]^{-1} > 0
\]

\[
\Omega_B = \left[ -B_\phi - B_e D e \phi \right]^{-1} < 0.
\]

The last result follows from assumption (A4) which is sufficient to sign the total effect of firm entry on profits of the cutoff firm, taking into account that the quality and salary of the manager that this firm hires will also change. I use these expressions to rewrite equations (26) and (27) as

\[
\frac{d\phi}{dx} = \Omega_A A_x dx + \Omega_A A_w dw_n + \Omega_A A_f df
\]

\[
\frac{d\phi}{dx} = \Omega_B B_x dx + \Omega_B B_w dw_n + \Omega_B B_f df.
\]

Substituting equation (28) into these conditions, and solving for \( \frac{dw_n}{df} \) yields

\[
\frac{dw_n}{df} = \frac{\Omega_B \left( B_f - B_x C_x^{-1} C_f \right) + \Omega_A \left( A_x C_x^{-1} C_f - A_f \right)}{\Omega_B (B_x C_x^{-1} C_w - B_w) + \Omega_A (A_w - A_x C_x^{-1} C_w)} > 0
\]

because both numerator and denominator are strictly negative. As a result, the market wage strictly increases if the productivity of firm-specific human capital increases, \( df > 0 \).

Second, consider the effect on training. From equation (28), there is a positive direct effect of the value of firm-specific human capital on training investment, but there is also a counteracting force through the increase in the market wage:

\[
\frac{dx}{df} = -C_x^{-1} C_w \frac{dw_n}{df} - C_x^{-1} C_f.
\]

This condition is positive if

\[
\kappa' (x (\phi)) \phi > \frac{\alpha}{1 - \alpha w_n \Psi (1 - \delta)} \cdot \frac{dw_n}{df}
\]

where only the left-hand side depends on firm type \( \phi \). Under Assumptions (A7) and (A8), \( \kappa' (x (\phi)) \phi \) is increasing in firm type and so there exists a threshold type \( \tilde{\phi} \) such that \( \frac{dx}{df} > 0 \) for all \( \phi > \tilde{\phi} \). For firms below the threshold, the wage effect dominates and training is reduced.
Third, the change in the firm cutoff is given by

\[
\frac{d\phi}{df} = \Omega_A \left\{ \left[ A_w - A_x C_{x}^{-1} C_w \right] \frac{dw_n}{df} + \left[ A_f - A_x C_{x}^{-1} C_f \right] \right\}
\]

which again shows a tradeoff through the mechanisms of the model. If internal candidates become more productive, there is a direct effect of firm entry, lowering the entry cutoff. More productive training makes it easier for firms to break even because they can achieve higher rents from internal promotion. However, the increase in the market wage for workers counteracts this force because especially low productivity firms suffer from a higher price for workers. The overall outcome is ambiguous and will depend on further assumptions about the underlying type distributions, the probability of high types and the training success rates.

**B.4 Proof: Proposition 4**

Rewrite the managerial wage function as

\[
\frac{w_m(e)}{w_n} = h(e) + \int_e^c \frac{\Psi}{w_n} \phi(e) \left( 1 + (a_H - a_L) \tilde{p}'(e) \right) \) de.
\]

Totally differentiating this function implies

\[
d \left( \frac{w_m(e)}{w_n} \right) = \epsilon_w dw_n + \epsilon_e de + \epsilon_x dx
\]

where

\[
\epsilon_w = - \int_e^c \frac{2-\alpha}{\alpha} \frac{\phi(e)}{w_n} \left( 1 + (a_H - a_L) \tilde{p}'(e) \right) de < 0
\]

\[
\epsilon_e = h'(e) - \frac{\Psi}{w_n} \phi(e) \left( 1 + (a_H - a_L) \tilde{p}'(e) \right) < 0
\]

\[
\epsilon_x = \int_e^c \frac{\Psi}{w_n} \phi(e) (a_H - a_L) \frac{\partial^2 \tilde{p}(e,x)}{\partial e \partial x} de > 0.
\]

The second result holds by the incentive compatibility constraint for managers under assumption (A6). The managerial wage has to increase more quickly than the outside option of being a worker. Otherwise better managers want to switch back to being workers. The last result holds under assumption (A2) because that implies \( \frac{\partial^2 \tilde{p}(e,x)}{\partial e \partial x} > 0. \)
Analogously for the wage function for trainees,

\[ w_{\tau}(e) = h(\epsilon) w_n + \beta (1 - \delta) \int_{\epsilon}^{e} \kappa(x^*(\phi,e)) \phi^*_e(e) \Psi(a_H - a_L) \left[ p'(e) - \tilde{p}'(e) \right] de \]
\[ + \beta (1 - \delta) \int_{\epsilon}^{e} \left( 1 - \kappa(x^*(\phi,e)) \right) \phi^*_e(e) \Psi(a_H - a_L) \left[ p'(e) (1 - \tilde{p}(e)) - \tilde{p}'(e) p(e) \right] de , \]

I find

\[ d \left( \frac{w_{\tau}(e)}{w_n} \right) = \epsilon^r_w dw_n + \epsilon^r_e de + \epsilon^r_x dx \]

where

\[ \epsilon^r_w = -\frac{\alpha}{1 - \alpha} \left[ \frac{w_{\tau}(e)}{w_n} - h(\epsilon) \right] < 0 \]
\[ \epsilon^r_e = h'(\epsilon) - \frac{w_{\tau}'(\epsilon)}{w_n} \]
\[ \epsilon^r_x = \beta (1 - \delta) \frac{\Psi}{w_n} (a_H - a_L) \left\{ \int_{\epsilon}^{e} \kappa'(x(e)) \phi(e) \left[ -\tilde{p}'(e) (1 - p(e)) + p'(e) \tilde{p}(e) \right] de \right. \]
\[ \left. - \int_{\epsilon}^{e} \phi(e) \left\{ \kappa(x(e)) \frac{\partial^2 \tilde{p}(e,x)}{\partial e \partial x} + (1 - \kappa(x(e))) \left[ p'(e) \frac{\partial \tilde{p}(e,x)}{\partial x} + p(e) \frac{\partial^2 \tilde{p}(e,x)}{\partial e \partial x} \right] \right\} de \right\} < 0. \]

The first expression is negative because the term in brackets is strictly positive for any skill level above \( \epsilon \). Assumption (A2) ensures that the third condition holds. The second condition may be positive because trainees have to prefer a manager career to a worker career. Yet, if manager wages increase steeply in the second period, trainee wages might be lower than the outside option of being a worker.

C Estimation

C.1 Data Moments

The first step to compute the data moments is to establish a ranking of firm productivity in the data. I use gross profits per manager as a proxy for firm productivity to measure the relevant firm-level outcomes along the productivity distribution. Since firms typically have more than one manager, I assume that firms in the data use a constant returns technology to accumulate output from different manager-worker teams. In other words, there are no complementarities between different managers within a firm. This ranking will be preserved if there are proportional coordination costs associated with adding manager teams, for example. Since firms have many managers in the data,
I ignore randomness in the composition of each firm’s management level to construct the firm ranking and I consider the observed gross profit per manager as equivalent to the expected gross profit by firm type. Moreover, note that the model suggests several measures that have perfect rank correlation with productivity. Consistent with the theory, Table 2 documents that other proxies of firm productivity such as firm size, span of control or value added per employee are also negatively related to external hiring. I choose gross profits per manager as the preferred proxy of firm productivity because it directly captures factors such as pricing markups, brand value and technology that are part of the firm type.

For each firm-level outcome, I compute average values for each firm over the sample period. I measure average earnings for full-time employees by using hourly wages and full-time equivalent hours, specifically the mode annual hours for this time period, 1670 hours per year. The use of full-time employees and hours adjustments are important to prevent differences in average earnings for managers and trainees from being affected by different levels of job mobility and part-time work. One empirical challenge is that I do not directly observe training or trainee hiring. In order to measure trainee wages, I define trainees as future managers five years before their first promotion into the management level. This definition captures the fact that the trainee period in the data spans several years. The average candidate spends 5-8 years at a firm before management promotion, see Table A.4. The results are robust to the time definition for trainees, the shape of the trainee wage function under alternative distance to promotion is very similar. I measure trainee wages at their current firm five years before promotion, although trainees may move to a different firm to become a manager. For the wage schedule of trainees, the productivity of their current firm is the relevant criterion. Because of the five year lag before promotion, trainee earnings are measured in earlier years than manager earnings on average. In order to account for productivity growth over time, I control for year fixed effects from a standard Mincer regression with age, experience and education before computing average earnings across firm types. As a result of normalizing efficiency units of human capital by the lowest trainee type, I measure the span of control of managers as the wage bill of their supervised workers relative to earnings of the lowest skill trainee. Finally, I use the results for the firm-level share of external hiring from Figure 2 above.
C.2 Estimation Algorithm

The GMM objective function is given by

\[ L_n(\theta) = -\frac{n}{2} (g_n(\theta))' W_n(g_n(\theta)) \]

where

\[ g_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m_i(\theta) \]

and \( m_i(\theta) \) is a vector of differences between simulated moments \( m^S(\theta) \) and data moments \( m^D \) such that

\[ E[m_i(\theta_0)] = E[m^D - m^S(\theta_0)] = 0. \]

In practice, these moments are average firm-level outcomes computed from the data that are compared to simulated outcomes from the model. The weight matrix \( W_n \) is a diagonal matrix using the inverse variance of each moment based on 1000 bootstrap repetitions of the kernel smoothing procedure in the data.

Chernozhukov and Hong (2003) propose a Markov-Chain Monte-Carlo estimation procedure for this problem. MCMC is a derivative-free method that circumvents the curse of dimensionality because it only requires evaluating the objective function at many different points to simulate a chain of parameters that converges to a probability distribution over the parameter vector,

\[ p(\theta) = \frac{e^{L_n(\theta)} \pi(\theta)}{\int e^{L_n(\theta)} \pi(\theta) d\theta}. \]

The estimator is the average over the \( K \) elements of the converged chain

\[ \hat{\theta} = \frac{1}{K} \sum_{k=1}^{K} \theta^{(k)}. \]

In practice, I use the Metropolis-Hastings algorithm. One starts from a parameter guess \( \theta^{(k)} \) and generates an alternative draw \( \theta' \) from a proposal density \( q(\theta' | \theta^{(k)}) \) which I assume to be a random walk with multivariate normal distribution. I update the parameter guess according to

\[ \theta^{(k+1)} = \begin{cases} \theta' & \text{with probability } \rho(\theta^{(k)}, \theta') \\ \theta^{(k)} & \text{with probability } 1 - \rho(\theta^{(k)}, \theta') \end{cases} \]
where

\[ \rho(x, y) = \min \left( e^{L_n(y) - L_n(x)}, 1 \right) \]

under the assumption of uniform priors and the proposal density a random walk. For each model, I simulate 40 chains of 40,000 elements. Based on the first 20,000 iterations, I adjust the variance matrix of the proposal density and I use the last 5,000 elements of each chain to compute the pooled parameter estimate.

### C.3 Estimation Details by Sector

I report the data moments for profits, manager and trainee salaries, the span of control and external hiring shares in figures (A.4)-(A.6). The span of control normalizes efficiency units of human capital in terms of the lowest skill trainee in each sector respectively. The figures show that the model does a good job matching the key features of the data, although there is clearly more noise in the data moments with a lower number of observations by industry than when jointly analyzing the full sample.
Figure A.4: Goodness of Fit: Manufacturing
Figure A.5: Goodness of Fit: Wholesale and Retail

Output

Manager Salaries

Trainee Salaries

Span of Control

External Hiring and Internal Promotion

Goodness of Fit
Figure A.6: Goodness of Fit: Business Activities

Value Added per manager in millions of USD vs Percentile of Firm Productivity Distribution

Manager Salaries vs Percentile of Firm Productivity Distribution

Trainee Salaries vs Percentile of Firm Productivity Distribution

Span of Control vs Percentile of Firm Productivity Distribution

External Hiring and Internal Promotion vs Percentile of Firm Productivity Distribution

Goodness of Fit vs Percentile of Firm Productivity Distribution