Internal Labor Markets and the Competition for Managerial Talent

Benjamin Friedrich*

January 6, 2016

[Job Market Paper]

Please download the most recent version here.

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 removing information frictions increases aggregate productivity by 22.5 percent. This gain is accompanied by 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.

*Yale University, 37 Hillhouse Avenue, New Haven, CT 06511, Tel.: 1-203-809-7254, e-mail: benjamin.friedrich@yale.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, Julia Garlick, Lisa Kahn, Rebecca McKibbin, Corina Mommaerts, Emily Nix, Marcelo Sant’Anna and seminar participants at the Yale Labor/Public Economics Workshop 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 therefore important in understanding firm performance, and in turn aggregate productivity.

Firms use internal labor markets - as well as external hiring - to recruit managers. Firmspecific 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, Katz, and Kearney 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. Following the international standard classification of occupations, managers consist of top executives, production managers and department managers. I measure internal promotions as occupational switching of non-managerial workers into manager positions of their incumbent firm, whereas external hires enter the firm at the management level. The data 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 labor market conditions into account.

---

1Oyer 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.”
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 a higher average level of schooling; in particular they are more likely to have postgraduate degrees. Second, I estimate individual unobserved ability as a time-invariant component of wages over the entire career following Abowd, Kramarz, and Margolis (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. Defining the pool of candidates as all individuals who eventually obtain manager promotion, more productive firms recruit candidates with higher unobserved ability. Fourth, trainees with higher unobserved 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. Finally, information frictions prevent perfectly assortative matching based on true talent across firms. Firm-specific training decisions under uncertainty may exacerbate misallocation of resources.

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 measure the distribution of key firm-level outcomes of value added, the span of control of managers (number of workers per manager), manager and trainee earnings and the share of external hiring. I use the method of simulated moments with a Markov-Chain-Monte-Carlo (MCMC) algorithm to match the model to these empirical distributions. The model fits the data well and the estimates reveal the relative importance of employer learning and firm-specific human capital. Training cannot fully replace high ability managers, who are estimated to be about three times more productive than a candidate with low ability and firm-specific human capital.

I find significant differences across industries in the role of firm-specific human capital and the value of talent. Firm-specific human capital is most important in manufacturing - about twice as valuable as in retail and wholesale. The productive value of high ability managers is highest for business activities, which includes finance, consulting and law firms. The firm productivity distribution is highly skewed with a long tail of highly productive firms, whereas the distribution of observable manager characteristics is close to uniform.

I use the estimated model to first quantify misallocation of resources due to information frictions. I consider a scenario with full information, where true managerial talent is common knowledge at labor market entry and the best firms can hire the best trainees from the start. The full information economy 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 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. If the market observes true talent, firms will strongly compete for the best managers in the market. This increases the wages for highly talented managers. In contrast, the average manager with
lower ability no longer benefits from a noisy signal of true manager talent and receives lower real wages. This result illustrates the role of asymmetric information for wage inequality in the market for managers. 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) - 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 supervision. 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, Gibbs, and Holmstrom 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, Shaw, and Stanton 2012). I complement existing management surveys (Bloom and Van Reenen 2010; Bloom, Sadun, and Van Reenen 2015) by using matched panel data to study manager careers and to illustrate how firms recruit and promote promising talents. See Castanias and Helfat (1991) for the business literature on managers as strategic resources of the firm.

There is a related literature on hiring practices. See Burks et al. (2015) for recent work on employee referrals, Hoffman, Kahn, and Li (2015) on discretionary versus standardized hiring procedures and Pallais
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. Moreover, the empirical literature has almost exclusively considered human capital accumulation and employer learning separately. 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 skills can help explain the increase in manager compensation. Yet none of these papers considers the interaction between internal and external labor markets for the recruiting and development of managerial talent in a general equilibrium setting. A recent literature discusses moral hazard in the market for managers, which motivates large efficiency wages to prevent shirking (Gayle, Golan, and Miller, forthcoming). I quantify the implications of internal labor markets, and specifically the role of hidden information, for manager compensation and wage inequality.

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 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.

(2014) on inefficiencies in hiring low-skill workers. Pallais and Sands (forthcoming) use randomized experiments to quantify the informational value of referrals for low-skill workers.

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.


See also the related analysis in Murphy and Zabojnik (2007). The main proposed explanations for the increase in manager compensation 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).
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 1980-2011. 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.

I use the Firm-Integrated Database for Labor Market Research (FIDA) over 1995-2011 to merge information about industry and gross profits 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. 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.

I define the sample for the subsequent analysis as all private sector firms with at least 50 full-time equivalent workers each year over 1999-2008. A major reclassification of occupational codes in 2009 makes comparison with the previous period more difficult. The size cutoff selects firms with an internal structure of workers and managers that makes the operation of an internal labor market in terms of training and promotions possible. Moreover, I limit my attention to firms with at least ten managerial hires over the sample period for the baseline specification. This cutoff level is motivated by the subsequent analysis of managerial hiring within firms, which requires a sufficiently large number of new manager hires to reduce small sample bias. 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 with at least 50 full-time employees in their industries over the period 1999-2008. I report descriptive

---

7 A small share of occupations are measured based on employee reporting (1%) and union membership (7.6%) for the main sample 1999-2008 defined below.
8 I provide more details on measurement and data cleaning in Appendix A.1.
9 The measure of value added in the data differs from gross profits because it takes rental and leasing costs, secondary costs and changes in debt positions into account.
10 The World Management Survey also considers firms above a size threshold of 50 employees (Bloom and Van Reenen 2007).
11 I provide robustness checks using a minimum of five hires per firm.
12 Appendix A.1 reports more details about the sample.
Table 1: Sample Statistics

<table>
<thead>
<tr>
<th>Industry</th>
<th>Firms</th>
<th>Sales</th>
<th>VA</th>
<th>Empl</th>
<th>Span</th>
<th>Manager Hires</th>
<th>Years</th>
<th>Years (Hiring)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>60</td>
<td>214.7</td>
<td>52.0</td>
<td>800.2</td>
<td>41.9</td>
<td>43.1</td>
<td>8.8</td>
<td>7.2</td>
</tr>
<tr>
<td>Textiles</td>
<td>14</td>
<td>49.7</td>
<td>12.6</td>
<td>216.0</td>
<td>27</td>
<td>17.8</td>
<td>8.8</td>
<td>5.5</td>
</tr>
<tr>
<td>Wood Products</td>
<td>59</td>
<td>58.5</td>
<td>22.6</td>
<td>413.8</td>
<td>39</td>
<td>24.8</td>
<td>8.3</td>
<td>6</td>
</tr>
<tr>
<td>Chemicals</td>
<td>55</td>
<td>144.2</td>
<td>55.6</td>
<td>574.0</td>
<td>32.8</td>
<td>33.3</td>
<td>9</td>
<td>6.7</td>
</tr>
<tr>
<td>Mineral Products</td>
<td>27</td>
<td>58.3</td>
<td>24.1</td>
<td>396.9</td>
<td>30.1</td>
<td>27</td>
<td>8.4</td>
<td>6.5</td>
</tr>
<tr>
<td>Metal Products</td>
<td>214</td>
<td>62.1</td>
<td>19.7</td>
<td>347.0</td>
<td>39</td>
<td>24.5</td>
<td>8.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Furniture, NEC</td>
<td>31</td>
<td>69.1</td>
<td>23.2</td>
<td>365.8</td>
<td>36.1</td>
<td>26.6</td>
<td>8.5</td>
<td>6.2</td>
</tr>
<tr>
<td>Construction</td>
<td>58</td>
<td>84.8</td>
<td>25.1</td>
<td>489.6</td>
<td>39.6</td>
<td>47.4</td>
<td>8.3</td>
<td>6.1</td>
</tr>
<tr>
<td>Sale: Motor Vehicles</td>
<td>24</td>
<td>77.6</td>
<td>10.3</td>
<td>200.1</td>
<td>30.4</td>
<td>18.5</td>
<td>9.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Wholesale</td>
<td>153</td>
<td>137.3</td>
<td>20.2</td>
<td>299.0</td>
<td>25.7</td>
<td>32.1</td>
<td>8.5</td>
<td>6.1</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>62</td>
<td>319.2</td>
<td>48.9</td>
<td>1217.0</td>
<td>64.2</td>
<td>113.8</td>
<td>7.7</td>
<td>6.1</td>
</tr>
<tr>
<td>Hotels and Restaurants</td>
<td>13</td>
<td>25.3</td>
<td>11.7</td>
<td>262.2</td>
<td>87</td>
<td>19.7</td>
<td>9.9</td>
<td>6.2</td>
</tr>
<tr>
<td>Transport</td>
<td>35</td>
<td>332.1</td>
<td>65.4</td>
<td>569.6</td>
<td>72.2</td>
<td>17.5</td>
<td>8.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Post and Telecomm</td>
<td>19</td>
<td>324.7</td>
<td>121.4</td>
<td>1155.9</td>
<td>82.8</td>
<td>38.9</td>
<td>7.1</td>
<td>5.7</td>
</tr>
<tr>
<td>Real Estate</td>
<td>14</td>
<td>24.4</td>
<td>12.0</td>
<td>198.7</td>
<td>34.1</td>
<td>17.3</td>
<td>9.1</td>
<td>6.5</td>
</tr>
<tr>
<td>Consulting, Law</td>
<td>115</td>
<td>61.4</td>
<td>30.3</td>
<td>505.9</td>
<td>57.8</td>
<td>32.4</td>
<td>8.1</td>
<td>6.1</td>
</tr>
<tr>
<td>Total</td>
<td>953</td>
<td>120.7</td>
<td>30.9</td>
<td>486.5</td>
<td>43.6</td>
<td>35.3</td>
<td>8.5</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Note: Sales and value added (VA) are reported in millions of USD, deflated using CPI with base year 2000. Employment reports the average full-time equivalent (FTE) employees over the sample period, the span of control is the average number of FTE workers in nonmanagerial occupations per manager. Years (Hiring) are the average number of years in which at least one manager position is filled.

statistics about the sample in Table 1. The number of firms by industry will enable me to compare hiring behavior for firms within narrowly 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 or whose salary is only subject to a lower bound instead of being determined by collective bargaining (see Dahl, Le Maire, and Munch 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

Recent assignment models for matching between managers and firms have emphasized the role of heterogeneity in firm productivity for sorting and compensation in the manager market (Gabaix and Landier 2008; Tervio 2008). In this section, I provide new evidence about sorting between managers and firms and I explicitly analyze differences in internal versus external hiring across firms in different industries and across the productivity distribution.

Fact 1: Internal versus external hiring decisions for managers vary across firms.

I compute the share of external hiring for all new manager positions by firm over 1999-2008. 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 around 23 percentage points. Technology and institutional differences across industries cannot explain large differences in internal versus external hiring within sectors.

Next, I estimate a linear probability model of external managerial hiring controlling for job characteristics, firm characteristics, as well as industry-time and region-time effects. I measure the residual share of external hiring by firm as the firm’s fixed effect. This residual propensity to hire managers externally emphasizes large differences across firms. Using different methods from the literature on teacher value added such as shrinkage estimates and a two-step procedure similar to Chetty, Friedman, and Rockoff (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.

This new stylized fact about firm heterogeneity in managerial hiring raises the question why firms differ in terms of their hiring strategies and which firm characteristics can explain variation in internal versus external hiring even within industries.

---

13 Manufacturing 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.

14 In addition to small differences in the distribution of external hiring shares of firms across sectors, mobility of managers across sectors is quite common, in particular for department managers. I find that about 40% of all external managerial hires in the sample previously worked in a different sector, see Appendix A.2.1 Table 8.

15 I regress a dummy of external hiring $H_{ijt}$ on characteristics of the position $X_{it}$, firm characteristics $Z_{jt}$ and a firm fixed effect $u_j$, $H_{ijt} = \beta X_{it} + \gamma Z_{jt} + u_j + \epsilon_{ijt}$. The threshold of at least 10 managerial hires per firm reduces noise in this procedure. I provide details in Appendix A.2.2.

16 I describe the details of these different adjustment methods in Appendix A.2.2.
Fact 2: Conditional on size, more productive firms use more internal promotions

Given the focus of the previous literature on complementarities between managers and firm productivity, I compare firms according to their productivity measured as gross profits per manager over the sample period. This productivity measure will have a close counterpart in the subsequent theory and estimation sections. I consider alternative measures of productivity such as value added per employee and TFP estimates to show that the results are robust. To capture firm characteristics before the hiring decision, I use lagged gross profits per manager, weighted by the share of new manager positions filled in the next period relative to the total number of new manager jobs at this firm over the sample period. Defining low productivity firms as those below the 25th percentile of the distribution and high productivity firms as those above the 75th percentile, Figure 2 shows that firms with high productivity are more likely to promote managers internally.

I estimate the correlation between external managerial hiring and firm productivity using a simple Probit model in Table 2. I control for job type and time varying firm characteristics such as lagged employment, employment growth, turnover rate among employees, lagged sales and assets. These measures are lagged by one period to describe the status quo of the firm before the hiring decision. All specifications include a full set of industry-time and region-time fixed effects. The key finding is the negative correlation between firm productivity and the probability of hiring externally. This result holds for measuring firm productivity by gross profits per manager, value added per manager or value added per employee. The pattern also holds using TFP estimates following Ackerberg, Caves, and Frazer (forthcoming) for a subsample of manufacturing firms for which detailed accounting data and industry price
I also find that larger firms in terms of employment, assets and sales use more internal promotions in the data. Yet, the role of firm productivity is robust to controlling for firm size 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 even after controlling for job characteristics, firm characteristics and the timing of hiring.

I provide further evidence for the positive relationship between firm productivity and internal promotions from a nonparametric regression of the share of external hiring on firm productivity rank. The left panel in Figure 3 shows a local polynomial regression for the raw data. 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 3, I use the residual hiring shares of firms estimated in the previous section. 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 positions.

---

Notes: The share of external hiring is measured by firm over the period 1999-2008. New manager positions include jobs at all management levels. Low productivity firms are defined as below the 25th percentile of average gross profits per manager, high productivity firms are above the 75th percentile of the full sample in Table 1. 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 kernel density uses Epanechnikov kernel with bandwidth 0.1.

---

17I provide evidence on the relationship between my firm ranking and TFP estimates in Appendix A.2.3. 18The coefficient on TFP in column (5) is more precisely estimated if I only include one measure of firm size, for example sales. Sales, employment and assets are highly correlated and the results for the subset of manufacturing firms become more noisy. 19The results are also robust to including second-order polynomials in sales, employment or assets.
Table 2: Probit Model of Managerial Hiring

Dependent Variable: Dummy for External Hiring of Manager $H_{ijt}$

(1): Separate Regressions (2)-(5): One Regression per Column

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(gross profits/manager)</td>
<td>-0.0274***</td>
<td>-0.0136***</td>
<td>-0.0203***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(VA/employee)</td>
<td>-0.0261***</td>
<td></td>
<td></td>
<td>-0.0278***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>log(TFP)</td>
<td>-0.1064***</td>
<td></td>
<td></td>
<td>-0.0297</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td></td>
<td></td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Time trend</td>
<td></td>
<td>0.0216***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top Manager</td>
<td>0.0712***</td>
<td>0.1002***</td>
<td>0.0976***</td>
<td>0.0906***</td>
<td>0.0921***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Department Manager</td>
<td>0.0281***</td>
<td>0.0489***</td>
<td>0.0490***</td>
<td>0.0497***</td>
<td>0.0501***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Employment Growth</td>
<td>0.1479***</td>
<td>0.1178***</td>
<td>0.1131***</td>
<td>0.1214***</td>
<td>0.3191***</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.025)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>Turnover rate</td>
<td>0.1247***</td>
<td>0.2365***</td>
<td>0.2067***</td>
<td>0.2631***</td>
<td>0.4624***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.051)</td>
<td>(0.055)</td>
<td>(0.051)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>log(sales)</td>
<td>-0.0321***</td>
<td>0.0021</td>
<td>-0.0053</td>
<td>-0.0073</td>
<td>-0.0480***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>log(assets)</td>
<td>-0.0172***</td>
<td>-0.0052*</td>
<td>-0.0015</td>
<td>0.0085***</td>
<td>0.0215***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>log(employment)</td>
<td>-0.0349***</td>
<td>-0.0312***</td>
<td>-0.0226***</td>
<td>-0.0308***</td>
<td>-0.0090</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.018)</td>
</tr>
</tbody>
</table>

Observations 25,917 25,858 29,645 8,540

Notes: The results report marginal effects at the variable means for all specifications. Robust standard errors in parentheses. The first column reports separate regressions for each independent variable, the sample size varies according to data availability. Job positions "top manager" and "department manager" are included in one regression. Columns (2)-(5) report one regression each, including all control variables jointly. 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. TFP is estimated following Ackerberg, Caves, and Frazer (forthcoming) using labor, capital and energy inputs. Column (2) includes a linear time trend, region, industry and year FE. All other regressions include industry-time and region-time fixed effects.
Figure 3: 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 a linear probability model controlling for job type, firm characteristics such as employment growth, turnover, sales, assets, industry-time and region-time FE. 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.

Finally, I use the estimated firm residuals to analyze which observable firm characteristics can explain the large and economically significant differences across firms. 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.

The results are reported in Table 3. First, firms at the top of the productivity distribution are 19 percentage points more likely to promote internally than firms with the lowest productivity rank. A 1% increase in log profits per manager is associated with a decrease in the share of external hiring by 0.055 percentage points. Second, measures of firm size such as sales and employment also show a negative correlation with external hiring shares but the estimated elasticities and the R-squared of the regression are much smaller. Third, measures

---

20 Differences in internal promotions go along with differences in manager turnover as illustrated by Figure 15 in Appendix A.2.4. Turnover for all employees is more similar across firms, suggesting that internal labor markets particularly matter at the management level.

21 I regress estimated firm fixed effects on observable firm characteristics, \( \hat{u}_j = \alpha Z_j + \zeta_j \), see Appendix A.2.2 for details on estimation of \( u_j \).

22 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.
Table 3: Managerial Hiring Strategies and Firm-level Characteristics

<table>
<thead>
<tr>
<th>Firm characteristics</th>
<th>(1): Separate Regressions</th>
<th>(2)-(5): One Regression per Column</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>R-squared</td>
</tr>
<tr>
<td>Productivity Ranking</td>
<td>-0.1895***</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Profits per Manager</td>
<td>-0.0554***</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>VA per Manager</td>
<td>-0.0639***</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Span of Control</td>
<td>-0.0620***</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>TFP</td>
<td>-0.0187*</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td>-0.0267***</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Sales</td>
<td>-0.0224***</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Share vocational</td>
<td>-0.0359</td>
<td>0.001</td>
</tr>
<tr>
<td>training or less</td>
<td>(0.040)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Observations</td>
<td>953</td>
<td>953</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.062</td>
<td>0.075</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. The dependent variable in all regressions is given by the firm-fixed effect estimates of external hiring from specification (5) in Table 9. Firm characteristics are averages over the sample period, weighted by next period managerial hiring. TFP is estimated following Ackerberg, Caves, and Frazer (forthcoming) for 202 manufacturing firms in the main sample. Column (1) reports separate regressions for each firm characteristic, columns (2)-(5) include all firm characteristics jointly.

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 significant. Note that except for span of control, other firm characteristics have much lower explanatory power for firm-level hiring strategies than measures of firm productivity. The results on firm productivity are robust to adding other firm characteristics in terms of size and workforce composition in columns (2)-(4). The overall R-squared only increases slightly, emphasizing the large explanatory power of firm productivity alone.

23The R-squared values are 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.
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 4 documents that managers at more productive firms have higher formal education than managers at less productive firms. The difference in average schooling of managers at low and high productivity firms is about one year. The right panel of Figure 4 illustrates that more productive firms have significantly more managers with secondary degrees such as Master degrees, MBAs and PhDs, replacing both college graduates and managers with vocational training or less.

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, Kramarz, and Margolis (1999). 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

---

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. These managers account for less than one-third of the group with vocational training.

---

24 The 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.

25 As discussed by Abowd, Kramarz, and Margolis (1999) the identifying assumption of this decomposition is strict exogeneity for residual errors, which implies random job mobility. The recent literature on sorting with search frictions (Eeckhout and Kircher 2011; Melo 2013; Lise, Meghir, and Robin 2013) proposes alternative ways to identify sorting of workers across firms, but the firm-individual decomposition remains an important benchmark that has been used to estimate unobserved manager characteristics by Bertrand and Schoar (2003).
Notes: The manager 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. 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). Total quality is measured as individual fixed effects from a Mincer regression on the universe of firms and workers including age, experience and tenure profiles, year and firm fixed effects. The scale is normalized to zero for the average manager and the average trainee in the sample. Unobserved quality is the residual fixed effect after controlling for formal education.
Figure 6: Internal Candidates and Promotions

Notes: The population of all candidates is defined as all individuals who are employed at firms in the main sample from Table 1, 5 years before their first manager promotion (but promotion can occur at any firm). The population of internal promotion includes those candidates that become managers at their incumbent firm. Manager quality is measured as individual fixed effects from a Mincer regression on the universe of firms and workers including age, experience and tenure profiles, year and firm fixed effects. Observed quality is the component of formal education in the person fixed effect. The scale is normalized to zero for the average candidate in the sample. Unobserved quality is the residual fixed effect after controlling for formal education.

managers at a firm in the main sample, corresponding to a total of 52,000 individuals. The first finding from this exercise in the top left of Figure 5 is that the average manager at a low productivity firm is about 10% less productive in terms of time-invariant characteristics than the average manager in the sample, whereas the average manager at a highly productive firm is 15% more productive. In the bottom left panel of Figure 5, I control for level differences in individual fixed-effects due to education differences; each manager’s residual quality is measured as a deviation from the average quality manager, conditional on schooling and highest degree. The sorting pattern with respect to unobservable ability is strongly positive across the firm distribution.

Moreover, I can use unobserved idiosyncratic productivity to compare the pool of candidates for manager positions across firms. Due to a lack of training data within firms, I consider candidates as all individuals who are ever promoted into a management position. The right hand side of Figure 5 shows that better firms recruit better manager candidates both in terms of observable and 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 5, the slope of the matching function between managers and firms is steeper than for trainees. This pattern suggests more uncertainty about the quality of trainees compared and Graham, Li, and Qiu (2012) for example.
Figure 7: First-Time Managers: Internal versus External Hiring

Notes: The population of the figures are first-time managers at firms in the main sample from Table 1. Each figure distinguishes between first-time managers who are promoted internally and managers who are hired externally. Manager quality is measured as individual fixed effects from a Mincer regression on the universe of firms and workers including age, experience and tenure profiles, year and firm fixed effects. Observed quality is the component of formal education in the person fixed effect. The scale is normalized to zero for the average first-time manager in the sample. Unobserved quality is the residual fixed effect after controlling for formal education.

to managers. Figure 6 compares the average quality of the pool of candidates five years before promotion with the average quality of promoted managers. I find a strong selection procedure at the time of promotion, which leads to a much higher average quality of internally promoted managers compared to all candidates. Consistent with employer learning, Figure 6 shows that this selection pattern is mainly driven by differences in unobserved quality.

Moreover, Figure 7 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. As a result, firms substitute observable skills from education and experience 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 unobserved...
servables, 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.

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 according to their permanent productivity $\phi$, drawn from a distribution $\phi \sim \Gamma(\phi)$. Labor supply is provided by overlapping generations of $N$ individuals with heterogeneous skills. 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 according to $h(e), h'(e) > 0$. Secondly, individuals are characterized by managerial ability $a \sim F_a(e)$ that is initially unknown to both firms and individuals. For simplicity, there are only two ability types,

$$a_H > a_L$$

and I assume a positive relationship between skills and managerial ability, $p'(e) > 0$ where $p(e) = Pr(a_H|e)$. Skills and managerial ability determine general managerial talent according to

$$z_{gen} = e + a.$$

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 $0 < \alpha < 1$.\(^{27}\) I assume that workers with different levels of human capital $h$ are perfect substitutes in production such that a firm simply chooses the total units of worker human capital $n$ that are supervised by the manager.\(^{28}\)

The key assumption in the production technology is that managerial talent and firm

\(^{27}\)This functional form builds on Lucas (1978), Rosen (1982) and has been used in a similar context by Tervio (2008) as well. In contrast, a large strand of the literature on worker careers assumes no complementarities between jobs in a firm, see Gibbons and Waldman 1999a; Gibbons and Waldman 1999b; Pastorino 2015.

\(^{28}\)This simplifying assumption keeps the focus of the analysis on manager careers and abstracts from sorting of workers across firms.
technology are complements,
\[ \frac{\partial^2}{\partial \phi \partial z} y(\phi, z, n) > 0. \]

The production technology in (1) 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 \( z \).\(^{29}\)

**Workers, Trainees and Managers**

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 \( \kappa(x) \in [0, 1], \kappa'(x) > 0, \kappa''(x) < 0 \).\(^{30}\) 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,

\[ z(e, a, f) = e + a + f. \] (2)

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.\(^{31}\)

\(^{29}\)Bloom et al. (2013) provide evidence of these separate roles in a randomized experiment where a change in management practices with the same managers as before the intervention leads to significant improvements in firm performance.

\(^{30}\)If training success depends on ability \( a \) as well, there is an additional incentive for training of high skill trainees because training success will be most likely for a very promising candidate.

\(^{31}\)In contrast, complementarity between unobserved ability and training is a key prerequisite for the results in Acemoglu and Pischke (1998). Their paper shows that if there is superior knowledge of the incumbent firm about ability \( a \), and if talent and training are complementary, then homogeneous firms will invest in
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.

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 do not poach managers from other firms but only 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 $\delta$. This is a common assumption in the literature (Greenwald 1986; Acemoglu and Pischke 1998), which leads to a mix of types in the market. 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 $e$, $\tilde{p}_{\text{ext}} (e)$,

$$\tilde{p}_{\text{ext}} (e) = \frac{H (e)}{H (e) + L (e)}.$$  

(3)

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

---

training even if training is general and completely transferable across firms. The same mechanism would work in my model, but I focus on training decisions for firm-specific knowledge in a market with heterogeneous firms instead.

32If firms observe promotion, but not the true ability of a manager, competitors face the risk of poaching a low ability type who was promoted because of successful training. If the value of firm-specific human capital is high enough and if the share of low ability types in the population is sufficiently large, outside firms will optimally choose not to poach internal candidates.

33Assuming 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 $\tilde{p}_{\text{ext}} (e)$ instead of a distribution function over continuous types.
**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 denoted by the market wage \( w_n \). I assume that firms have full bargaining power to capture all rents from internal labor markets. Managers are paid their best outside option.\(^{34}\)

Observed skills \( e \) are a signal for total managerial talent and they determine the market wage functions for trainees, \( w_\tau(e) \), and for managers, \( w_m(e) \). A participation constraint ensures that individuals who become trainees prefer this career path to being a worker,

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

(4)

and individuals who completed the trainee stage prefer to work as managers instead of going back to being a worker,

\[
w_m(e) \geq h(e) w_n.
\]  

(5)

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

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

(6)

Moreover, this constraint implies that for the lowest skill trainee, \( e^\tau \), who in the second period becomes the lowest skill manager \( e^m \),

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

(7)

This wage setting is consistent with trainee wages increasing by less than the worker salary as a function of observed skills, if these wage cuts are overcompensated 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 8.

I solve the firm’s problem backwards to illustrate the profit maximization problem in

---

\(^{34}\)It is straightforward to allow for bargaining between firm and manager over the surplus from internal promotion using fixed bargaining shares. This channel will further strengthen wage differences for managers across firms but I abstract from this type of bargaining in the presentation of the model.
Beginning of period t:
Young generation enters, skills e observed.
Each firm hires one trainee and chooses training investment x.

End of period t:
Incumbent firm learns trainee ability a.
Training success \{0,1\} is revealed.

Beginning of period t+1:
Trainee is promoted internally or meets another firm in the external manager market.
Manager ability is revealed after external hiring.
Firms hire optimal workforce for their manager.

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 (1), yields the optimality condition

$$n^* = \phi z \cdot \left[ \frac{\alpha}{w_n} \right]^{\frac{1}{1-\alpha}}. \quad (8)$$

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. I rewrite the production technology after substituting in the optimal workforce from (8) as

$$y^* (\phi, z) = \phi \cdot z \cdot \Psi (\alpha, w_n)$$

where $\Psi (\alpha, w_n) = (1 - \alpha) \left( \frac{\alpha}{w_n} \right)^{\frac{\alpha}{1-\alpha}}$ is a function of the market wage and technology. The Cobb-Douglas production technology yields output as a linear function of manager skills conditional on optimal labor demand for workers. In combination with the assumption of perfect substitutability of skill components in (2), this technology will provide a tractable analytical solution to the model.

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 $\tilde{p}_{ext}(e)$ and the wage schedule for managers $w_m(e)$ as given. Denote expected profits from external hiring as

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

where $E_{a|e,ext}[\cdot]$ 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,

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

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 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.

$$I_0 = \{ (a, e) \ s.t. \pi^*(\phi, z(a, e, 0)) \geq E_{a|e,ext}[\pi^*(\phi, e(\phi), a)] \} \quad (10)$$

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

Based on the set $I_0$, I define the default profit for firm $\phi$ after hiring trainee $e$ with revealed ability $a$ and no internal training as

$$A(\phi, e, a) = 1 \{ (a, e) \in I_0 \} \cdot \pi^*(\phi, z(a, e, 0)) + 1 \{ (a, e) \notin I_0 \} \cdot E_{a|e,ext}[\pi^*(\phi, e^*_m(\phi), a)].$$

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.

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,

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

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.
where rents from training, \( R(\phi, e, a) \), are defined by

\[
R(\phi, e, a) = \max \{0, \pi^*(\phi, z(a, e, f)) - A(\phi, e, a)\}.
\]

Training generates rents in addition to the default profit \( A(\phi, e, a) \). 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. Rents are only realized with probability \( 1 - \delta \) because some trainees leave the firm for exogenous reasons.\(^{36}\)

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 same skill level, in the hope of finding a high ability type. The full first order condition of training takes these cases into account.\(^{37}\)

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

where \( D(\phi, e, a, f) \) defines the difference in profits between any trainee \( \{e, a, f\} \) and the best external hire,

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

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

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

\(^{36}\)Expected rents from training may further decrease if firms can poach managers from competitors. Intuitively, this possibility may reduce the attractiveness of training more for low productivity firms that cannot counter the offer of highly productive firms.

\(^{37}\)Using Inada conditions on the probability of training success as well as a sufficiently large value for firm specific training, I can ensure that there is an interior solution for training investment at all firms and skill levels.
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, $e_m^*(\phi)$.

Firm $\phi$ chooses a trainee $e^*_\tau(\phi)$ to maximize expected profits next period,

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

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 always 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\textsuperscript{38}

\begin{equation}
\begin{align*}
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) \cdot \left[ p'(e_{\tau}) - \tilde{p}'_{ext}(e_{\tau}) \right] \right. \\
& \left. \equiv T_1: \text{marginal rents with successful training} \right. \\
& + (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}'_{ext}(e_{\tau}) \right] \\
& \equiv T_2: \text{marginal rents without successful training} \right. \\
\end{align*}
\end{equation}

The marginal change in trainee wages reflects the probability-weighted sum of marginal rents

\textsuperscript{38}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 as well. More generally, one has to distinguish different cases analogous to (12).
from a better candidate with and without successful training.\textsuperscript{39} 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 $p(e)$.

### 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 $\{w_r(e), w_m(e)\}$, assignment functions for trainees $\phi_r(e)$ and managers $\phi_m(e)$, a set of active firms $\phi \in [\phi^l, \phi^\text{max}]$, a set of trainees $e \in [e^l, e^\text{max}]$, 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

\[
N \int_0^{e^l} h(e) \, dF(e) + N \int_0^{e^m} h(e) \, dF(e) = M \int_\phi^{\phi^\text{max}} s(\phi, e^*_e(\phi)) \int n(\phi, e^*_m(\phi), a', 0) \, dF_{e|m|\text{ext}}(a') \, d\Gamma(\phi) \\
+ M \int_\phi^{\phi^\text{max}} (1 - s(\phi, e^*_e(\phi))) \int n(\phi', e^*_e(\phi), a', f') \, dF_{e|m|\text{int}}(a') \, dF_{f|\text{int}}(f') \, d\Gamma(\phi)
\]

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.\textsuperscript{40} The ability type

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

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.\textsuperscript{40}

More formally, the probability of separation from any trainee type $e$ for firm $\phi$ is given by

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

\[
N \int_{e^r}^{e^m} dF(e) = M \int_{e^r}^{\phi_{\text{max}}} d\Gamma(\phi) \\
N \int_{e^r}^{e^m} dF(e) = M \int_{\phi}^{\phi_{\text{max}}} d\Gamma(\phi).
\]

Lastly, free entry determines the lowest productivity firm in the market, denoted by \( \hat{\phi} \).

New entrants make zero expected profits by hiring a trainee today and maximizing future profits,

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

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

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

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

\( ^{41} \)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.
\begin{equation}
\frac{1+\beta}{\beta} \cdot \frac{E[z|\text{ext}, e]}{E[z|f, e]} \cdot \eta \frac{z|\text{ext}, e}{z|f, e} (e) > \eta_h (e) \forall e \text{ where } E[z|\text{ext}, e] = e + a_L + \frac{\delta_p(e)}{s + (1-s)(1-p(e))} [a_H - a_L]
\end{equation}

is the minimum expected talent of skill type \( e \) under external hiring.

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 (11). 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 (4) 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 highest possible selection scenario, the elasticity of talent with respect to skill is sufficiently large to overcompensate improvements in worker human capital.

As a result, condition (5) 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, \( w'_t(e) > 0 \).

I provide a constructive proof for the equilibrium with PAM. With PAM, market clearing for managers and trainees pins down the assignment function conditional on the exogenous distributions of types,

\begin{equation}
M (1 - \Gamma (\phi(e))) = N (1 - F(e))
\end{equation}

\begin{equation}
\phi(e) = \Gamma^{-1} \left[ 1 - \frac{N}{M} (1 - F(e)) \right].
\end{equation}
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 (12), external managerial hiring (9) and the optimal trainee hiring decision (13). 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 (14) and (15) 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 (14) strictly decreases in both unknowns, this condition is represented by a downward-sloping line in $\{\phi, w_n\}$-space. In contrast, profits in equation (15) are increasing in the cutoff but decreasing in the wage rate for workers, which corresponds to an upward sloping line in $\{\phi, w_n\}$-space to satisfy this constraint. Hence, the two conditions cross at most 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 of Proposition 1 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, $\tilde{p}_{\text{ext}}(e)$, increases in observed skills $e$.

I characterize the interaction of firm specific human capital and information asymmetries in equilibrium in Proposition 2, 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.\(^{42}\)

\(^{42}\)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.
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

\[(A8) \quad -\kappa''(x) \leq \kappa'(x) \forall x \geq 0.\]

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 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 \(\tilde{\phi}\).

I provide intuition for the results in the main text and report the formal proof in Appendix B.3. Consider first the effect of an increase in the value of firm-specific human capital \(f\) on the market wage. 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 counter-acting 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.
Proposition 4. 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 \( e \in [e, e(\tilde{\phi})] \). For skill levels \( e \geq e(\tilde{\phi}) \), 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

\[
w_m(e) = h(e) w_n + \int_e^{e'} w_m(e') \, de
\]

using equation (9) to yield\(^{43}\)

\[
\frac{d\left(\frac{w_m(e)}{w_n}\right)}{df} = \frac{dw_m}{df} + \frac{e^m}{\text{scale}<0} + \frac{e_m}{\text{entry}<0} + \frac{e_m}{\text{selection}>0} \frac{dx}{df}.
\]  

\(17\)

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. Second, firm exit increases the cutoff level for trainees and managers, \( de > 0 \) and decreases inequality because the competitive equilibrium rewards managers and trainees for being better than all lower ranked candidates. 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. 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. If internal candidates become 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 by limiting production scale.\(^{44}\) Finally, for all firms above some threshold \( \tilde{\phi} \), increased training and the reduction in adverse selection counteract the wage compression for

\(^{43}\)Details of the derivation are provided in Appendix B.4.

\(^{44}\)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 the market. Similarly, if the share of high ability types is much larger at the top end of the skill distribution, then the best firms benefit much more from the change in managerial productivity than all other firms and this scenario would again suggest firm exit.
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. The best managers are rewarded the most for this improvement in expected productivity from external hiring and wage dispersion among managers increases.

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 (13),

$$\frac{d\left(\frac{w_{x}(e)}{w_n}\right)}{df} = \epsilon_{w}^{\text{scale}} \frac{dw_{n}}{df} + \epsilon_{e}^{\text{entry}} \frac{de}{df} + \epsilon_{x}^{\text{selection}} \frac{dx}{df}. \quad (18)$$

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 productive, competition for trainees is reduced for two reasons. First, higher market wages reduce production scale, $\epsilon_{w}^{\text{scale}} < 0$, 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, $\epsilon_{x}^{\text{selection}} < 0$. 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** In addition to the insights about competition for managerial talent and its implications for wage inequality, the model also implies misallocation 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 $e + a$. This misallocation may yield considerable productivity losses if highly talented individuals are matched with low productivity firms because of a weak signal $e$. 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 with $e_1 = e_2 + \xi$, and $0 < \xi < a_H - a_L$. Suppose trainee 1 has low unobserved ability whereas trainee 2 has high ability. PAM according to $e$ implies that $\phi(e_1) > \phi(e_2)$, but the supermodularity in the production technology suggests gains from resorting,

$$y(\phi(e_1), e_1 + a_L) + y(\phi(e_2), e_2 + a_H) < y(\phi(e_1), e_2 + a_H) + y(\phi(e_2), e_1 + a_L).$$
The magnitude of this productivity gain will depend on the full parameterization of the model, in particular the production technology, the ability difference between high and low types and the productivity distribution of firms. Yet in a counterfactual world with full information, there will be an efficiency-equity tradeoff because low types can no longer benefit from a noisy signal that suggests they might be high ability types. Instead, high types are revealed and they will be paid according to their true marginal product. This suggests larger wage dispersion among managers in the market.

4 Estimation

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

4.1 Measuring firm productivity

The first step for this estimation procedure 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. 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 3 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.

45 In fact, this ranking will be preserved if there are proportional coordination costs associated with adding manager teams, see Appendix C.1.

46 Another common perspective on firms’ management level is that there are different managerial positions in middle and top management. As a result, the managerial hierarchy and the quality of managers relative to the workforce are also crucial determinants of firm productivity. Modeling the endogenous choice of hierarchical structure is beyond the scope of this paper, but it provides an interesting direction to link this approach to the literature on knowledge-based and incentive-based hierarchies (Calvo and Wellisz 1978; Garicano 2000; Caliendo and Rossi-Hansberg 2012).
4.2 Identification

The estimation matches five key 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. Output per manager is closely related to the distribution of firm and manager types and the technology parameter $\alpha$ in equation (1). Firm types are separately identified from manager types through the wage function for managers in equation (9), which only depends on the distribution of firm productivity. The value of high ability and the share of high ability managers across the observed skill distribution are closely related to the slope of the manager wage function and the trainee wage function in equations (9) and (13) which differentially depend on these market characteristics. If there is a large difference in the productivity between high and low ability types and the chances of finding a high ability candidate strongly increase in observed skills, then firms will bid up wages for these most promising candidates early in their career. The value of firm-specific human capital relates to the span of control in (8) and the internal promotion share of firms based on optimal training decisions in the optimal assignment, (12), while differences in external versus internal hiring across the firm distribution are closely related to training success and the distribution of high ability types.

4.3 Estimation Procedure

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. 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. 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. Finally, I use the results for the firm-level

\footnote{I use approximately the mode annual hours for this time period, 1670 hours per year.}

\footnote{The average candidate spends 5-8 years at a firm before management promotion, see Appendix A.2.5 Table 11. 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.}

\footnote{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.}
share of external hiring from Figure 3 above.

I estimate the distribution of each firm-level outcome across the productivity distribution using first-order local polynomial regression weighted by an Epanechnikov kernel. Dashed lines in Figure (10) summarize the empirical distributions of these moments including 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.

4.4 Parameterization

I make several parametric assumptions in order to fit the model to the data in a parsimonious way. They are listed in Table 4. In particular, I assume bounded Pareto distributions for both firm productivity and observable manager skills. Each of these type distributions is characterized by their lower and upper bound and a Pareto shape parameter \( \theta \). I normalize the low ability type to zero, which means that the estimated distribution of observed skills measures \( e + a_L \). Furthermore, I choose a flexible cubic function for the share of high ability types at each observed skill level \( p(e) \). This function provides sufficient flexibility to match the curvature in the wage profiles of trainees and managers that I observe empirically. I specify a concave success rate of learning between zero and one according to \( \kappa(x) = \frac{\sqrt{x}}{\sqrt{\lambda + \sqrt{x}}} \) where larger \( \lambda \) will imply a slower improvement in training success as investment increases.\(^{51}\)

Finally, I fix the discount rate at 0.9.

I normalize efficiency units of human capital in terms of the lowest skill trainee who is indifferent between being a worker or entering a management career. I use earnings of the lowest skill trainee as a normalization for \( w_n \) in the estimation.\(^{52}\) As a result of this normalization, I measure the span of control of managers by the wage bill of their supervised workers relative to earnings of the lowest skill trainee. Note that in the data, starting salaries for trainees at the lowest productivity firms are around $38,000 whereas starting salaries for managers are around $55,000. One explanation for the level differences in earnings between managers and trainees is experience gains over time. Altonji and Shakotko (1987) estimate experience gains of 40 log points over 30 years of labor market experience and Dustmann and Meghir (2005) 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). I assume that worker human capital increases by 20% between the trainee and manager stage to pin down the outside option of the

\(^{50}\)This assumption is flexible because it allows either the productivity distribution or the skill distribution or both to drive the Pareto shape of profits and span of control.

\(^{51}\)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 \( y = \lambda x^2 \).

\(^{52}\)Since span of control is a key data moment that the model matches, the simulated economy satisfies market clearing at this wage rate. Alternatively, I could take the size of the market as given and use the worker wage as an additional moment in the estimation.
Table 4: 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_φ(x) = 1 - \left( \frac{φ_{\text{min}}}{φ_{\text{max}}} \right)^{θ_φ} )</td>
<td>( φ_{\text{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>( φ_{\text{max}} )</td>
<td></td>
<td></td>
<td>( p_0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( θ_φ )</td>
<td></td>
<td></td>
<td>( p_1 )</td>
</tr>
<tr>
<td>distr of obs skills</td>
<td>( F_e(x) = 1 - \left( \frac{e_{\text{min}}}{e_{\text{max}}} \right)^{θ_e} )</td>
<td>( e_{\text{min}} )</td>
<td>technology</td>
<td>( y = (φ_2)^{1-α} n^α )</td>
<td>( α )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( e_{\text{max}} )</td>
<td>exog. separations</td>
<td></td>
<td>( δ )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( θ_e )</td>
<td>training success</td>
<td></td>
<td>( \kappa(x) = \frac{\sqrt{7}}{\sqrt{x} + \sqrt{7}} )</td>
</tr>
<tr>
<td>distr of ability</td>
<td>( a_L = 0 )</td>
<td></td>
<td>manager premium</td>
<td>( w_{\text{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></td>
<td>( β )</td>
</tr>
</tbody>
</table>

lowest skill manager in the market, \( e \). I interpret the remaining salary gap between trainees and managers as a premium \( w_p \) that compensates effort from working longer hours or higher responsibility for managers.\(^{53}\)

### 4.5 Results

I estimate the model using simulated method of moments based on the MCMC algorithm suggested by Chernozhukov and Hong (2003).\(^{54}\) Table 5 reports the parameter estimates. As illustrated by Figure 9, 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. According to the estimates for the function \( p(e) \), the probability of finding a high ability manager at the bottom of the observed skill distribution is around 2\%, whereas it reaches almost 30\% at the top of the skill distribution. This result suggests that initial sorting is based on a noisy signal of true managerial talent. Moreover, I estimate that firms face a risk of 38\% that candidates will leave after their trainee stage. This finding is related to a high share of external hiring in the data despite high rents from training.

Figure 10 evaluates the goodness of fit of the estimation procedure. The simulated mo-

\(^{53}\)I consider these effort costs under perfect and costless monitoring to abstract from moral hazard problems.

\(^{54}\)MCMC is a derivative-free method that can deal with non-smooth objective functions and a high-dimensional parameter vector. MCMC circumvents the curse of dimensionality because I only need to evaluate the objective function at many different points. Details about the MCMC algorithm are in Appendix C.2.
ments are within the confidence interval of the smoothed data functions and remarkably close to the data moments across all figures, emphasizing the ability of the model to capture the key features of the data. The parametric assumption of a 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 implied by the Pareto shape, whereas there is a faster improvement in firm productivity at the upper tail in the data as indicated by the increase in profits and manager salaries than reflected by the Pareto distribution.

Since 75% of managers are either internally promoted or hired externally within the same sector, I also estimate the model separately for three main sectors: (i) manufacturing, (ii) wholesale and retail and (iii) consulting and other business activities. The results in columns

---

55 I illustrate the deviation between model and data, weighted by the inverse variance of the data moments, in the bottom right panel in Figure 10.

56 See Appendix Table 8 for details on mobility across sectors conditional on external hiring. I consider sectors with a sufficiently large number of firms. I report the sector-specific data moments and the goodness of fit by sector in Appendix C.3.
Figure 10: Data Moments and Goodness of Fit
Table 5: Estimation Results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>All firms</th>
<th>StDev</th>
<th>Manufacturing</th>
<th>StDev</th>
<th>Retail</th>
<th>StDev</th>
<th>Business</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{\text{min}}$</td>
<td>91.0</td>
<td>3.8</td>
<td>26.4</td>
<td>3.9</td>
<td>6.1</td>
<td>0.6</td>
<td>6124.6</td>
<td>2601.0</td>
</tr>
<tr>
<td>$\phi_{\text{max}}$</td>
<td>855.8</td>
<td>34.8</td>
<td>162.5</td>
<td>23.7</td>
<td>53.2</td>
<td>5.5</td>
<td>75933.1</td>
<td>33766.5</td>
</tr>
<tr>
<td>$\theta_{\phi}$</td>
<td>1.660</td>
<td>0.016</td>
<td>2.245</td>
<td>0.045</td>
<td>1.182</td>
<td>0.042</td>
<td>1.020</td>
<td>0.069</td>
</tr>
<tr>
<td>$e_{\text{min}}$</td>
<td>32.0</td>
<td>1.0</td>
<td>80.0</td>
<td>5.9</td>
<td>46.7</td>
<td>3.2</td>
<td>29.5</td>
<td>6.5</td>
</tr>
<tr>
<td>$e_{\text{max}}$</td>
<td>46.6</td>
<td>1.0</td>
<td>102.0</td>
<td>6.0</td>
<td>54.0</td>
<td>3.2</td>
<td>43.0</td>
<td>8.3</td>
</tr>
<tr>
<td>$\theta_{e}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.548</td>
<td>1.033</td>
<td>1.799</td>
<td>1.269</td>
</tr>
<tr>
<td>$a_{H}$</td>
<td>11659.8</td>
<td>186.0</td>
<td>9924.3</td>
<td>335.4</td>
<td>10365.6</td>
<td>919.1</td>
<td>16685.9</td>
<td>2424.4</td>
</tr>
<tr>
<td>$f$</td>
<td>3652.5</td>
<td>133.0</td>
<td>4532.0</td>
<td>159.0</td>
<td>1884.7</td>
<td>193.2</td>
<td>5084.1</td>
<td>1018.0</td>
</tr>
<tr>
<td>$p_2$</td>
<td>51.0</td>
<td>1.0</td>
<td>126.0</td>
<td>6.3</td>
<td>53.5</td>
<td>3.2</td>
<td>43.3</td>
<td>7.6</td>
</tr>
<tr>
<td>$p_0$</td>
<td>0.299</td>
<td>0.006</td>
<td>0.500</td>
<td>0.000</td>
<td>0.345</td>
<td>0.023</td>
<td>0.247</td>
<td>0.050</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.040</td>
<td>0.001</td>
<td>0.005</td>
<td>0.000</td>
<td>0.651</td>
<td>0.022</td>
<td>0.044</td>
<td>0.010</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.532</td>
<td>0.001</td>
<td>0.503</td>
<td>0.003</td>
<td>0.482</td>
<td>0.002</td>
<td>0.606</td>
<td>0.004</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.380</td>
<td>0.001</td>
<td>0.375</td>
<td>0.001</td>
<td>0.327</td>
<td>0.005</td>
<td>0.425</td>
<td>0.008</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>15164.9</td>
<td>502.4</td>
<td>24461.2</td>
<td>1036.7</td>
<td>12177.7</td>
<td>1105.1</td>
<td>7968.6</td>
<td>1465.1</td>
</tr>
<tr>
<td>$w_{p}$</td>
<td>9543.5</td>
<td>97.7</td>
<td>9878.2</td>
<td>124.1</td>
<td>11692.9</td>
<td>224.4</td>
<td>13180.5</td>
<td>471.6</td>
</tr>
<tr>
<td>$a_{H}/f$</td>
<td>3.19</td>
<td>(0.13)</td>
<td>2.19</td>
<td>(0.11)</td>
<td>5.50</td>
<td>(0.73)</td>
<td>3.28</td>
<td>(0.89)</td>
</tr>
</tbody>
</table>

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. I report the standard deviation of the converged MCMC chain using a diagonal weight matrix with inverse variances of the data moments.

(2)-(4) of Table 5 differ from the full sample in column (1) in interesting and intuitive ways. Firm-specific human capital is most valuable in manufacturing and least valuable in retail and wholesale. A high ability manager in manufacturing is only about twice as productive as a low ability manager with firm knowledge. High ability managers in retail and business are more productive by a factor of 5.5 and 3.3 respectively. These differences in the relative value of firm-specific human capital are statistically significant. This result is consistent with firm-specific production processes across manufacturing firms but relatively standardized procedures in the wholesale and retail sector which make managerial skills more transferable across firms.57

Differences in profits and the level of manager salaries across sectors suggest that retail firms are less productive than manufacturing and business firms. Finally, observable skills are more valuable in manufacturing than in retail or business. Hard skills matter more for the productive value of a manager in manufacturing. These observable skills are a better signal of ability than in other sectors where it is more difficult to distinguish types ex ante. Figure 11 shows that the share of high ability managers strongly increases with observed skills in manufacturing, whereas the distribution of high types is only weakly related to observed skills.

57Bloom, Sadun, and Van Reenen (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 5 and the simulated training success rate, training investment is in the range of USD 45,000-140,000 in manufacturing, USD 10,000-25,000 in retail and USD 20,000-100,000 in business.
in the other sectors.

These differences across sectors follow from the prevalence of internal promotion and the gradients of the wage functions for trainees and managers. A high share of internal promotions indicates that manufacturing firms value internal training. Yet the increase in manager salaries with firm productivity is larger in manufacturing than in retail and business. This stronger competition for the best managers in manufacturing implies that expected talent increases quickly across the manager distribution. A low span of control per manager suggests a lower value of high ability compared to other sectors. Hence, the share of high ability managers must quickly increase in observed skills to explain the increase in managerial wages across the skill distribution.

5 Internal Labor Markets, Aggregate Productivity and Wage Inequality

5.1 Full Information

Based on the estimated model, I first quantify the productivity loss due to imperfect information. The question is to what extent information frictions nourish internal labor markets and lead to misallocation of resources. I compare the benchmark situation where only incumbent firms learn about the true ability of managers with the first best allocation where managerial talent is fully observed at labor market entry. In the first best, trainees can be ranked according to true talent and 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. This scenario captures full information about general skills, but there will still be uncertainty about training success within the firm.
I report the results in the first column of Table 6. The full information economy generates a 22.5% increase in total revenue compared to the asymmetric information benchmark. This gain is realized through reallocation of resources from low productivity to high productivity firms as summarized by the revenue plot in Figure 12. Less productive firms use a larger share of resources in the benchmark economy with asymmetric information because they are able to promote high ability trainees with a positive probability. With full information, firms up to the 80th percentile of the productivity distribution lose resources because all high ability managers are now employed by the most productive firms. The parameter estimates yield $e^{max} + a_L < e^{min} + a_H$, which explains the discontinuity in the allocation of resources under full information. The log scale understates the dramatic change in relative size between high and low productivity firms in this counterfactual.

The overall reallocation also depends on changes in training investment. Training investment in the economy increases by 24%, but Figure 12 shows that the average change hides large firm heterogeneity. The least productive firm decreases training investment by 15.8%, whereas the most productive firm increases training by 47%. Intuitively, competition for workers has become stronger as high productivity firms hire the best managers and try to expand. Higher equilibrium wages make firm-specific human capital less attractive for low productivity firms because they cannot expand as much as in the benchmark scenario. However, this effect of higher wages is overcompensated for a large group of firms by the fact that there is no longer uncertainty about the trainee type at the time of training investment. Highly productive firms no longer face the risk of investing in a low ability type instead of the chance of meeting a high ability type through the external market. Less productive firms no longer face the option value of trying to hire a high ability type from the external market. As a result, many firms that only hire low ability managers under full information will increase their training investment compared to the asymmetric information benchmark to improve the quality of their manager.

The managerial wage profile in Figure 12 illustrates that the increase in allocative efficiency in the market comes at the cost of increased dispersion in manager compensation. Low ability managers with observable skills $e$ can no longer benefit from their skill being a noisy signal of ability. True talent is revealed and as a consequence, manager compensation among low ability types is very flat. In contrast, high ability managers are compensated for the fact that high productivity firms around the 80th percentile of the firm distribution would have liked to hire them as well, and hence the best firms have to pay their top managers according to this outside option. The strong competition for talent leads the best firms to pay their managers millions of dollars per year, whereas observable low manager types receive only a small fraction of that amount.

Finally, reallocation of resources, adjustments in training investment and changes in the wage distribution affect firm profits differentially in different parts of the distribution. For
low productivity firms that no longer receive any high ability managers, profits strictly fall. Within this group of firms, more productive firms lose more in absolute terms because they used to recruit trainees with a higher chance of high ability types. Moreover, the group of high productivity firms that always hires high ability managers under full information does not uniformly gain in terms of profits. The least productive firms within this group, between the 80th and 90th percentile of the firm distribution, would prefer to stay in a world with asymmetric information. As Figure 12 illustrates, the strong competition for high ability managers and high manager wages for high ability types reduce their profits. Only firms in the top decile of the productivity distribution increase profits under full information because gains from production complementarities overcompensate higher salaries for the best managers.

Based on the estimation results by sector, I compare the effects of full information for different subsets of firms in Table 6. Compared to the economy-wide productivity gain of 22.5%, I find 12.5% gain in manufacturing, 28.1% in wholesale and retail, and 24.6% in the business sector. Intuitively, 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 compared to other sectors. In sum, Figure 13 illustrates that full information yields less reallocation from low to high productivity firms in manufacturing.

Productivity gains from full information depend on adjustments in training investment and firm exit. I find the largest training effect in the retail and wholesale sector, where the share of managers with firm-specific knowledge increases across the entire firm distribution if there is no uncertainty about the true trainee type. Depending on the sector-specific firm productivity distribution and the share of high ability managers in the market, some firms at the bottom of the productivity distribution will make losses in a full information scenario as illustrated by Figure 13. As a result, Table 6 provides a lower bound for the productivity gains in retail and business if the distribution of firms is held fixed. Firm exit will yield further reallocation of resources to high productivity firms.

5.2 Manager Compensation, Wage Inequality and External Hiring

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. Second, there is a continuing increase in the wage gap between managers and workers. Autor, Katz, and Kearney (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

---

Table 6: Welfare Effects: Full Information

<table>
<thead>
<tr>
<th>Outcome</th>
<th>All firms</th>
<th>Manufacturing</th>
<th>Retail</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>Total</td>
<td>22.54%</td>
<td>12.48%</td>
<td>28.09%</td>
</tr>
<tr>
<td>Worker Wage Rate</td>
<td>Total</td>
<td>22.53%</td>
<td>12.48%</td>
<td>28.08%</td>
</tr>
<tr>
<td>Manager-Worker Wage Gap</td>
<td>Average</td>
<td>667.74%</td>
<td>711.02%</td>
<td>709.95%</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>-17.51%</td>
<td>-12.50%</td>
<td>-25.72%</td>
</tr>
<tr>
<td>Training Investment</td>
<td>Total</td>
<td>23.98%</td>
<td>22.35%</td>
<td>146.04%</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>46.97%</td>
<td>41.46%</td>
<td>232.77%</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>-15.77%</td>
<td>-9.92%</td>
<td>22.72%</td>
</tr>
<tr>
<td>Total Profits</td>
<td>Total</td>
<td>-37.47%</td>
<td>-44.18%</td>
<td>-39.86%</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>63.76%</td>
<td>30.00%</td>
<td>55.11%</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>-97.71%</td>
<td>-78.58%</td>
<td>-130.79%</td>
</tr>
</tbody>
</table>

---

58I provide details about the secular trends in Denmark in Appendix A.2.6.
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 has increased from 15% in the 1970s to 26% in the 1990s. For Denmark, the Probit model in Section 2 shows that external hiring over the period 1999-2008 increases by 1.7% per year.

The literature has argued that firm growth plays an important role in understanding the increase in manager compensation (Gabaix and Landier, 2008). My counterfactual analysis simulates the effects of firm productivity growth on the three margins of manager compensation, manager-worker wage gap, and external hiring. I consider a 1% increase in average firm productivity, modeled as a proportional increase in the boundaries of the firm distribution. The first column in Table 7 shows that this simulation yields a 0.4% increase in average manager salaries but this scenario cannot capture the simultaneous increase of inequality and external hiring in the data.

I then consider three different scenarios that affect the attractiveness of internal labor markets. The results suggest an important role for internal labor markets to reconcile the different secular trends. Consider first a 1% increase in the value of high ability managers in
Table 7: Simulation: Trends in Compensation, Inequality and Hiring

<table>
<thead>
<tr>
<th>Outcome (Average across Firms)</th>
<th>1 percent increase in...</th>
<th>1 percent decrease in...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>average firm productivity</td>
<td>high ability $a_H$</td>
</tr>
<tr>
<td>Manager Salary</td>
<td>0.4183</td>
<td>0.3251</td>
</tr>
<tr>
<td>Manager-Worker Gap</td>
<td>-0.0589</td>
<td>0.0443</td>
</tr>
<tr>
<td>External Hiring Share</td>
<td>-0.0403</td>
<td>0.0731</td>
</tr>
</tbody>
</table>

the second column of Table 7. 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 response of the manager-worker wage gap and the external hiring share.

I compare this scenario to a pure improvement in information, for example motivated by increasing use of career platforms and headhunter services. The value of high ability managers is fixed but the observed signal becomes more precise with respect to true talent. A 1% increase in precision corresponds to a steepening in the profile of $p(e)$ such that the area between the new and old curve corresponds to 1% of the total share of high types. 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 recruit managers in a market environment with information frictions and specific human capital. 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 firms 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.

I estimate the model to examine the extent of misallocation of resources caused by information frictions and to illustrate the role of internal labor markets for wage inequality. I find that removing information frictions increases aggregate productivity by 23%. Full information about managerial talent allows for perfectly assortative matching between managers and firms and leads to reallocation of resources from low to high productivity firms. The productivity gain is accompanied by higher wage inequality because better signals of talent increase competition for the best managers. Better information provides a new mechanism that increases the competition for talent. 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 increase in external hiring and in upper-tail wage inequality over time.

This paper provides a theoretical framework to analyze the interaction between promotion and training decisions within a particular firm and market competition across heterogeneous firms. I briefly discuss several directions to extend the model to take additional features of the market for managers into account. First, if trainees acquire some general skills through training, there will be additional strategic considerations of knowledge transfer from external hiring. 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. Third, internal labor markets 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, firms endogenously choose

---

59 Frydman (2005) presents a simple framework for CEOs that builds on this idea of promotion as signaling from Waldman (1984) to study the increase in manager compensation over time.
their hierarchical structure; incorporating training and promotion decisions into models of organizational choice is a promising avenue to understand how firms respond to exogenous shocks by adjusting their internal structure (Friedrich, 2015). More generally, the model can be used to understand firm-level hiring and promotion strategies for other occupations as well.

Empirically, this paper emphasizes the value, but also the limitations, of matched employer-employee data to study hiring decisions of firms. Recent management surveys have started to collect additional information about talent development within firms that can be linked to administrative data (Bloom et al., 2013b). My analysis suggests the benefit of expanding these surveys to learn about how firms select and invest in trainees or how they search for external candidates. These insights will help to empirically estimate the effect of changes in information technology and transferability of skills on talent development and managerial hiring in future research.

References


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. These definitions have been used to define hierarchical layers of managers in current work by Friedrich (2015) building on related work by Caliendo, Monte, and Rossi-Hansberg (2015). Occupational data is available from 1991 and I start my analysis in 1999. I use the occupational history of individuals since 1991 to determine first-time manager promotion.

I first define a variable for job movers in the administrative matched employer-employee data (IDA) 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 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. In particular, this includes movers to the firm headquarters.

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.

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 size cutoff of 50 full-time equivalent workers allows for entry and exit but avoids issues of very small firms growing large over time. I focus on firms that have a significant role for management for their entire observed period. 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. I add 653 firms with at least 50 employees and 5-9 managerial hires in robustness checks. These firms account for an additional 13.6% of management employment and 6.1% of total value added, while the remaining firms with less than 5 managerial hires account for 13.1% of manager hours worked and 8.8% of total value added in the relevant industries. There are many small firms in Denmark that I ignore for the analysis of the market for managers.

A.2 Stylized Facts

A.2.1 Manager Transitions Across Sectors

In addition to small differences in the distribution of external hiring shares of firms across sectors, mobility of managers across sectors is quite common, in particular for department managers. 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 8. For example, external hiring from other sectors is much less common in the retail sector across all managerial positions than in finance and business. Moreover, department managers such as sales and marketing managers possess skills that are better transferable across sectors, whereas production managers from within the sector are often preferred because of 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 a high degree of manager mobility across sectors.

A.2.2 Residual Hiring Strategies

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

\[ H_{ijt} = \beta X_{it} + \gamma_1 Z_{jt} + \gamma_2 Z_{s(j)t} + u_j + \epsilon_{ijt} \]  (19)
Table 8: 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 Telecomm</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>

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 (19); 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 (see Rockoff 2004 for a brief overview). 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 of firm-specific human capital for example, thereby placing a natural constraint on the extent of external hiring for important management positions.

One concern with the empirical model in (19) 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
Table 9: First Stage: Firm FE of Managerial Hiring

<table>
<thead>
<tr>
<th></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>
</tbody>
</table>

Observations: 30,207 33,248 33,248 30,207 30,207
R-squared: 0.066 0.189 0.189 0.196 0.205
F-test: 5.568 5.512 5.018 4.916

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. The number of managerial hires per firm is limited. Monte Carlo studies suggest that inconsistencies from using a fixed-effects model are small even in short panels (see Heckman 1981). To limit the amount of bias, I only include firms with at least 10 new managers over the sample period and as illustrated in Table 1, I observe on average 35 managerial hiring decisions per firm.

I report results including different sets of control variables and specifications in Table 9. 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 9. I use different strategies from the literature on teacher value
added to account for sampling error. First, analogous to Aaronson, Barrow, and Sander (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 \( \sigma_u^2 \).

\[
\sigma_u^2 = \sigma_{\hat{u}}^2 - \frac{1}{n} \sum_{j=1}^{n} \hat{se}(\hat{u}_j)^2. \tag{20}
\]

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 (20) 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, Friedman, and Rockoff (2014) by first consistently estimating the coefficients on the control variables from a fixed effects regression according to equation (19) to get residuals

\[
e_{ijt} = u_j + e_{ijt}.
\]

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 in the sample.

The results are reported in Table 10. The key finding across different methodologies and model specifications is that there is a large variation in firm-level hiring strategies within

<table>
<thead>
<tr>
<th>Fixed Effects</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>Adjusted standard deviation</td>
<td>0.192</td>
<td>0.191</td>
<td>0.194</td>
<td>0.194</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>Bayes</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>
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 section illustrates the external validity 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. I use industry-specific price deflators for output, intermediates and capital from the OECD database. I also apply the Ackerberg, Caves, and Frazer (forthcoming) 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. Data availability restricts this robustness check to the subsample of manufacturing firms. In order to increase precision, I use all available firms in each industry to estimate the industry production function, which explains that average productivity across firms with at least 50 employees is larger than the industry benchmark across the distribution. The results in Figure 14 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 150-250% of the industry average, whereas average TFP for the lowest ranked firms is slightly below the industry average.

A.2.4 Managerial Hiring and Turnover

Long-term employment is another dimension of internal labor markets besides internal promotion ladders. I define the share of long-term employment among managers as the share of managers who are still employed at the incumbent firm five years after internal promotion or external hiring. Figure (15) shows that managers at firms with external hiring shares below industry average are also much more likely to stay at the firm over five years. In contrast, firms with high external hiring shares experience much higher turnover among managers. The same pattern holds for all employees, but turnover rates for the entire workforce are more similar across all firms. In other words, some firms engage in long-term relationships and internal promotions for a small subset of their workforce, for whom internal labor markets play an important role. At the same time there is considerable turnover for the average employee across all firms.
Figure 14: Firm Ranking and TFP Estimates

The sample consists of 423 manufacturing firms from the main sample for which data on value added, capital, employment and energy is available over the sample period. TFP estimates report the log difference to the industry average TFP in 2000.

Figure 15: Managerial Hiring and Turnover

Notes: The share of external hiring for managers is given by the estimates for $u_j$ from equation (19) for all firms in the main sample of Table 1. Long-term employment is defined as the share of individuals that are still employed at the firm after five years. The left-hand side considers the retention rate for managers from the time of promotion or external hiring. The right-hand side considers long-term employment for all employees from the time of entry into the firm.
Table 11: 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.2.5 Differences between Internal and External Hires

Table (11) 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.2.6 Trends in Manager Compensation and the Manager-Worker Gap for Denmark

The left panel in Figure 16 shows the 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), increase by 24%, while professionals (ISCO groups 2 and 3) gain 18.5% and real earnings for blue-collar and white-collar workers (ISCO-88 groups 4-9) only increase by 4%. I report the log difference in average hourly wages between managers and blue-collar and white-collar workers over time in the right panel of Figure 16. The wage gap between top managers and workers increases continuously over time, by a total of 30 log points over roughly a 20 year period. The wage gap between middle managers and workers increases quickly over the 1990s but remains stable during the 2000s. Finally, I illustrate the increase in external hiring for first-time manager promotions over time in the
Figure 16: Manager Earnings and the Manager-Worker Wage Gap

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.

The figure shows cyclicality in internal versus external hiring, but also a clear trend towards more external hiring over time.

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 $g$ 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 (8). Moreover, a higher wage rate will ceteris paribus reduce training investment at all firms according to equation (12). In particular,

$$dx = \beta (1 - \delta) \phi \nu(x) \frac{[f - (1 - p(e)) \tilde{\rho}(e)(a_H - a_L)]}{\beta (1 - \delta) \kappa(x) \phi \Psi [f - (1 - p(e)) \tilde{\rho}(e)(a_H - a_L)]} < 0$$

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

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

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

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

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 (3) simplifies to

\[ \tilde{p}(e) = \frac{\delta p(e)}{\delta + (1 - \delta)(1 - \kappa(e))(1 - p(e))} \]  

(21)

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 (9) and (13). Using (21), the partial derivative is given by

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

(22)

Since the partial effect \( \frac{\partial}{\partial \phi} \tilde{p}(e) > 0 \) 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 (21) and (22), \( p'(e) - \tilde{p}'(e) > 0 \) and \( \left[p'(e)(1 - \tilde{p}(e)) - \tilde{p}'(e)p(e)\right] > 0 \). 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 (4) is satisfied in equilibrium. For any skill type \( e \geq \bar{e} \), 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 (5) 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 (12),

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

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 (21) with respect to training is
\[
\frac{\partial}{\partial x} \tilde{p}(e, x) = \delta p(e) \kappa'(x) (1 - \delta) (1 - p(e)) \left[ \delta + (1 - \delta) (1 - \kappa(x)) (1 - p(e)) \right]^2 > 0,
\]
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 (22) 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

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
\[
\begin{align*}
A_\phi \frac{d\phi}{de} + A_e \frac{de}{de} + A_x \frac{dx}{de} + A_w \frac{dw_n}{de} + A_f \frac{df}{de} = 0.
\end{align*}
\]
I provide more detailed derivations of these and the subsequent results in the Online Appendix. Next, totally differentiate the free entry condition for firms, (15), using the boundary conditions (6) and (7) to yield
\[
\begin{align*}
B_\phi \frac{d\phi}{df} + B_e \frac{de}{df} + B_x \frac{dx}{df} + B_w \frac{dw_n}{df} + B_f \frac{df}{df} = 0.
\end{align*}
\]
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,

\[
\frac{C_x}{<0} dx + \frac{C_w}{<0} dw_n + \frac{C_f}{>0} df = 0. \tag{25}
\]

Finally, note that the equilibrium PAM assignment in (16) implies

\[
d_e = M\gamma(\phi) \frac{N_f(e)}{D_e \phi} \equiv D_{e\phi}. \tag{26}
\]

Now I jointly solve equations (23), (24), (25) and (26) for the change in the market wage rate in response to an increase in \( f \). I define functions

\[
\Omega_A = [\begin{bmatrix} -A \phi - A_e D_{e\phi} \end{bmatrix}]^{-1} > 0 \\
\Omega_B = [\begin{bmatrix} -B \phi - B_e D_{e\phi} \end{bmatrix}]^{-1} < 0.
\]

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 (23) and (24) as

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

Substituting equation (25) into these conditions, and solving for \( \frac{dw_n}{df} \) yields

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

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 (25), 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
\[ \kappa' (x (\phi)) \phi > \frac{\alpha}{1 - \alpha} \frac{1}{w_n \Psi \beta (1 - \delta)} \cdot \frac{dw_n}{df} \]
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
\[ \frac{d\phi}{df} = \Omega_A \left\{ [A_w - A_x C_x^{-1} C_w] \frac{dw_n}{df} + \left[ A_f - A_x C_x^{-1} C_f \right] \right\} \]
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
\[ \frac{w_m (e)}{w_n} = h (\xi) + \int_\xi^e \frac{\Psi}{w_n} \phi (\xi) (1 + (a_H - a_L) \tilde{p}' (\xi)) d\xi. \]

Totally differentiating this function implies
\[ d \left( \frac{w_m (e)}{w_n} \right) = \epsilon_w^m dw_n + \epsilon_x^m de + \epsilon_x^m dx \]
where
\[ \epsilon_w^m = - \int_\xi^e \alpha \frac{\alpha}{1 - \alpha} \frac{1 - \alpha}{w_n} \phi (\xi) (1 + (a_H - a_L) \tilde{p}' (\xi)) d\xi < 0 \]
\[ \epsilon_x^m = h' (\xi) - \frac{\Psi}{w_n} \phi (\xi) (1 + (a_H - a_L) \tilde{p}' (\xi)) < 0 \]
\[ \epsilon_x^m = \int_\xi^e \Psi \frac{\partial^2 \tilde{p} (\xi, x)}{\partial e \partial x} d\xi > 0. \]

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,

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

I find

$$d \left( \frac{w_T(e)}{w_n} \right) = \epsilon_w \, dw_n + \epsilon_x \, de + \epsilon_{\phi} \, dx$$

where

$$\epsilon_w = -\frac{\alpha}{1 - \alpha} \left[ \frac{w_T(e)}{w_n} - h(e) \right] < 0$$

$$\epsilon_x = h'(e) - \frac{w_T'(e)}{w_n}$$

$$\epsilon_{\phi} = \beta (1 - \delta) \Psi (a_H - a_L) \left\{ \int_{e}^{\phi} \kappa (x(e)) \phi(e) \left[ -\tilde{p}'(e) (1 - p(e)) + p'(e) \tilde{p}(e) \right] de - \int_{e}^{\phi} \phi(e) \left\{ \kappa (x(e)) \frac{\partial^2 \tilde{p}(e,x)}{\partial e \partial x} + (1 - \kappa (x(e))) \left[ p'(e) \frac{\partial \tilde{p}(e,x)}{\partial x} + p(e) \frac{\partial^2 \tilde{p}(e,x)}{\partial e \partial x} \right] \right\} \right\} < 0.$$

The first expression is negative because the term in brackets is strictly positive for any skill level above $e$. 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 Manager-Worker Teams

Consider an extended model where firms consist of $N_m$ manager-worker teams. Formally, suppose total production of each firm is given by

$$Y = \gamma^{N_m-1} \sum_{m=1}^{N_m} \left[ (\phi z_m)^{1-\alpha} n_m^\alpha \right]$$

where $\gamma \leq 1$ is the cost of coordination across teams. The estimation setting in the main text corresponds to $\gamma = 1$ which means there are no coordination frictions between different teams. Yet the productivity ranking by value added per manager will also hold under weaker
conditions. Consider the case of increasing coordination costs, $\gamma < 1$. Despite adding value in terms of output, each additional team also reduces revenue of all existing teams through higher overall communication and coordination requirements. As a result, adding more and more teams becomes prohibitively expensive. The optimal number of manager teams depends on market conditions, but more productive firms will optimally hire (weakly) more managers in order to maximize overall profits. Since the number of manager teams is discrete, there will be a range of firms with different productivities and identical number of managers under these assumptions. Yet note that conditional on the same number of manager teams, more productive firms will have higher revenue per manager. Secondly, adding a manager team is more attractive for more productive firms because they generate higher revenue and profits for each team. In the model with one manager per firm, the most productive firm makes the highest profits. The cost of additional teams in terms of wages increases linearly for each firm, but revenue decreases at a geometric rate $n\gamma^n$. As a result, more productive firms will have both more managers and higher revenue per manager, such that the productivity ranking is preserved.

C.2 Estimation Algorithm

This section provides details about the MCMC algorithm following Chernozhukov and Hong (2003). The GMM objective function is given by

$$L_n(\theta) = -\frac{n}{2} (g_n(\theta))^\top W_n(g_n(\theta))$$

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 for each function.

I use the Metropolis-Hastings algorithm to simulate a chain of parameters that converges to a probability distribution over the parameter vector,

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

68
In practice, 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.

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

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

Using the correct weight matrix, standard errors are the standard deviation of the \( \{\theta^{(k)}\} \) chain. Chernozhukov and Hong (2003) provide large-sample properties for this class of estimators.

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 (17)-(19). 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 17: Goodness of Fit: Manufacturing
Figure 18: Goodness of Fit: Wholesale and Retail
Figure 19: Goodness of Fit: Business Activities