A Tale of Two C(...)s: Competence and Complementarity*

Simeon D. Alder[†] Fane N. Groes[‡]

October 9, 2015

Abstract

We build a tractable assignment model to characterize the matching and separation patterns of CEOs and their employers. Managers learn about their own type by observing a sequence of public signals. The sorting is *ex ante* perfect across managers of a given cohort whose most recent assignment is the same, but is not typically so *ex post*. Moreover, if matching is costless, perfect *ex ante* sorting occurs across managers of a given cohort regardless of their assignment history. We calibrate the model to match empirical targets from a large matched employer-employee dataset covering the Danish labor force between 2000 and 2009. We exploit the non-monotonicity of executive compensation in the employer type – the firm's productivity, that is – to parameterize the model. We have a particular interest in the degree of complementarity between the characteristics of the manager and those of the firm in the production function and our results fill a gap in the literature on the aggregate effects of a particular form of misallocation, namely mismatch, which depend critically on this elasticity. What's more, our theory is a natural building block for a dynamic theory of entrepreneurship.

JEL Codes: C78, D33, E23, J31, O47

Keywords: Sorting, complementarity, occupational choice, identification, firm size distribution, aggregate productivity

*For many helpful comments we thank seminar participants at CEMFI, Copenhagen Business School, Einaudi, Frankfurt, Notre Dame, and Oslo as well as participants at the 2014 Search and Matching Conference in Edinburgh, the Development and Midwest Search and Matching Workshops at the Federal Reserve Bank of Chicago, the UCLA Alumni Conference, the SED Annual Meeting, and the EEA Annual Congress. A special thank you goes to Hugo Hopenhayn, our discussant at the HEC Paris Entrepreneurship Workshop who offered valuable insights and comments on numerous other occasions. All remaining errors are, of course, our own.

[†]University of Notre Dame, Department of Economics, email: simeon.alder.1@nd.edu

[‡]Copenhagen Business School, email: fg.eco@cbs.dk

1 Introduction

A sizeable literature addresses the allocation of skills across firms or jobs. Beginning with Lucas (1978) it has been argued that idiosyncratic differences in technologies or organization capital, for instance, are key to understanding the size distribution of firms and establishments. Similarly, individuals are endowed with different innate abilities, which, by way of investment or experience, manifest themselves as variations in skill or human capital (as in Bhattacharya et al., 2013, for instance). While the idea of heterogeneity in productivities and skills is uncontroversial, the extent to which complementarities govern their interaction remains an open question. Abowd et al. (1999) find no empirical evidence for worker-job complementarities. Lopes de Melo (2013), on the other hand, argues that the estimates from fixed effects regressions may be biased due to the non-monotonicity of wages in firm types, a feature that Eeckhout and Kircher (2011) also emphasize. Lise et al. (2013) find evidence for positive complementarity, at least for relatively well educated workers in the United States.

But why should we care about complementarities at all? With complementarities, a worker's marginal product is a function of her own skill as well as the characteristics of the job she is assigned to. Clearly then, the prices that decentralize the optimal assignment of workers to jobs depend on the degree to which skills and productivities complement each other. In such an environment, maximizing aggregate productivity is not just a matter of making sure everyone has *a* job but requires that everybody ends up with the *right* kind of job.

The intuition is quite simple and to fix ideas consider the problem of assigning a manager to a firm. In the absence of complementarities, a CEO's marginal product is only a function of the efficiency units she embodies. Her contribution to the total surplus does not vary across firms and the assignment of managerial talent is of no consequence for efficiency. Only the *aggregate* supply of talent matters. In contrast, if managers and firms are complements, a competent CEO's marginal product is a steeper function of firm productivities compared to the profile of an incompetent manager and positive sorting is optimal. The empirically relevant degree of complementarity, however, is an open question so far.

Consider, for example, a Canadian street performer by the name of Guy Laliberté. He started his career as an accordeon player, fire breather, and stiltwalker with a small group of colorful characters in the small Canadian town of Baie-Saint-Paul. In due time he founded a circus and named it *Cirque du Soleil*, which is now a global enterprise with several thousand employees. While Guy reportedly was creative from an early age, his marginal product and hence his compensation rose sharply over the course of his career. Arguably, it was low while he was entertaining pedestrians on the sidewalks of Baie-Saint-Paul. Soon, however, he quit the sidewalks and moved into the first *Cirque du Soleil* tent seating 800 spectators in 1984. After several upgrades, the circus is now on multiple simultaneous tours in addition to resident shows in Las Vegas and Orlando. His creativity is now paired with a technology that allows him to entertain millions of spectators rather than the few hundred or thousand pedestrians who saw him on a street corner in his early days. Back then, he probably never imagined spending $35 million for an eleven-day trip to the *International Space Station*.

In the absence of talent-technology complementarities, Guy's marginal product and compensation are proportional to his skills and that, in turn, would suggest that his performance talent increased by a few orders of magnitude over the course of three decades. While we acknowledge the importance of (endogenous) skill accumulation, we also believe that technology-skill complementarities are critical features of talent markets in general and the market for entrepreneurs and senior managers more specifically. Once we admit such complementarities, identifying the empirically relevant degree turns into a first-order concern.

If managerial abilities and firm productivities were observable attributes, one could identify the extent to which they complement each other simply by matching the salient empirical price moments to their model-generated counterparts. Alas, we cannot observe them directly and have no choice but to back out the distribution of firm and manager attributes as well as the substitution elasticity from price data. Given the data limitations of the time, earlier assignment models like Gabaix and Landier (2008) and Terviö (2008) sidestepped the challenges associated with identifying the relevant degree of complementarity altogether.

While widely available cross-sectional data generates moments that are necessary to calibrate the model, previous work highlighted the importance of the data's *time series* dimension to definitively pin down the substitution elasticity between firms and managers (Alder, forthcoming). Consequently, we need panel data with enough variation in the time series dimension and a *dynamic* assignment model with endogenous separation and turnover to make further progress. With these two ingredients, data and theory, we can overcome most, if not all, the identification problems and calibrate the parameters of a dynamic assignment model to match population moments in a panel data set. The basic idea is the following. To learn something about Guy Laliberté's creative talent separately from the contribution of the technologies he operates, we need to know the 1980s busker/entrepreneur of Baie-Saint-Paul as well as the successful CEO from the 1990s and 2000s. Not only that, we also need a population of entrepreneurs and managers, some of whom have trajectories similar to like Guy's, while others have fairly unremarkable careers, while yet another group may experience some sort of failure.

Our theory accomplishes this goal in a tractable way and we parameterize our model using matched employer-employee data from three different sources that cover the Danish labor force from 2000 to 2009. The first of the data sources is register data that contains socioeconomic variables, employment information, and a complete employer-employee link for the entire population between 1999 and 2009. The second data source is the *Købmandstadens Oplysningsbureau* (KOB) dataset that contains firm accounting data for all Danish limited liability firms from 1991 to the present. Lastly, the third dataset, *Erhvervs- og Selskabstyrelsen* (ES) data, identifies CEOs and board members of limited liability firms from 2000 to 2010. Unique firm and worker identifiers enable us to merge information from these three sources into a single dataset and to construct employment histories for virtually all individuals of interest, namely CEOs and entrepreneurs. This rich panel structure delivers the additional identification we need to jointly characterize the distribution of managerial abilities and productivities as well as the parameter that governs substitutability in the technology that generates the match surplus.

In our model agents do not know their own type but receive a sequence of public signals (productivity shocks) that enables them to update their beliefs using Bayes' rule (as in Groes et al., 2015). In contrast to Jovanovic (1979) agents learn about their own abilities rather than about an *ex ante* unknown match quality.[1] Conditional on age and the most recent assignment, the equilibrium of our benchmark model features perfect sorting in the agents' *priors*, but not *ex post*. CEOs separate from their current match whenever the gap between their prior and posterior is sufficiently large for given prices (i.e. the manager's outside option and the firms' complete profit profile) and matching costs. In the special case where matching is costless, the sorting of CEOs is *ex ante* perfect conditional on age, but does not depend on the manager's assignment history. Moreover, in this frictionless version of the model, the dynamic program "collapses" to a simple sequence of static equilibria.

Each period, participation in the market for executive talent is endogenous and CEOs pick the employer that offers the most lucrative wage contract in expectation. The model does not distinguish between quits and layoffs. This being said, *ex post* realizations that are associated with low CEO pay have a natural interpretation as layoffs whereas payouts above a certain threshold trigger separations that have the look and feel of quits. As in Eeckhout and Kircher (2011) and Groes et al. (2015) compensation is non-monotonic in the project type: the equilibrium wage offers are such that for each CEO – of a certain age and with some belief about her own abilities – there exists a project that offers a more lucrative contract than any other. Importantly, the ideal type is not typically the most productive project. We exploit this non-monotonicity result to discipline the model's sorting and separation patterns using a panel dataset containing virtually all Danish CEOs and their employers.

Unlike Eeckhout and Kircher (2011) we use wage (i.e. CEO compensation) data as well as data on profits and output. While they argue that firm performance does not respond much to vari-

[1] See also Nagypal (2007) for a similar distinction. Jovanovic (2015) considers the implications of *private* output histories.

ations in the productivity of individual workers inside the firm, Bennedsen et al. (2008, 2012) find that CEOs matter. They have a measurable impact on standard measures of firm performance. Moreover, CEOs are unique in this respect: other senior executives do not significantly affect performance. This empirical evidence supports our model's implicit assumption that, in principle, the CEO's contribution (or lack thereof) can impact firm performance. By using payment data on both sides of the market we can characterize the strength *and* the sign of the equilibrium assignment. What's more, our model qualitatively matches the stylized fact that firm performance is an important determinant of hiring, retention, and separation decisions in the market for chief executives and, possibly, entrepreneurs.[2]

Initially, we abstract from matching and/or separation costs. Our calibration results to date, however, suggest that the assignment market for CEOs is subject to substantial frictions. In particular, we find that instances of CEO turnover in the model are counterfactually clustered around a finite number of critical values in the size distribution of firms. In order to improve the model's predictions in this key dimension, we explore the implications of matching frictions in an alternative model specification. The sorting of CEOs to projects is no longer perfect *ex ante*, not even across members of the same cohort. Rather, perfect *ex ante* sorting only prevails inside endogenous subsets of a given age cohort. Importantly, in the presence of positive matching costs, the dynamic equilibrium is no longer a sequence of very tractable static equilibria and we solve for the recursive equilibrium by backward induction.

Our theory has potential applications beyond the CEO-to-project assignments considered here. In particular, it is a natural building block toward a dynamic theory of entrepreneurship. In our model, the returns to entrepreneurship are uncertain while the outside option is associated with a constant flow payment.[3] Occupational choice theories like Johnson (1978), Jovanovic (1979), or Miller (1984) suggest that individuals select risky occupations when they are young. Evans and Jovanovic (1989), in contrast, argue that liquidity constraints are tighter for younger individuals since they have, on average, accumulated less wealth. This constraint acts as a countervailing force to the rookies' appetite for risk. Evans and Leighton (1989) find that the hazard rate into self-employment is constant in age and it is natural to conclude that liquidity constraints are binding in the data. Our theory offers alternative support for the Knightian view according to which risk-bearing is a defining trait of entrepreneurship (Knight, 1921). Rather than liquidity constraints, our theory's countervailing force is fueled by the fact that inexperienced entrepreneurs have less precise beliefs about their own abilities. With diminishing marginal returns to talent the market is more selective for young entrepreneurs compared to older ones. Since, in addition, the cross-sectional distribution of

[2] See, for instance, the discussion in Eisfeldt and Kuhnen (2013).

[3] One may think of the outside option as the wage associated with supplying a raw unit of labor, as in Lucas' canonical span-of-control model (Lucas, 1978).

beliefs varies by age cohort, the model delivers a rich pattern of entrepreneurial dynamics and selection that we can calibrate to corresponding moments in our matched employer-employee dataset.

We emphasize that our model is a building block rather than a full-fledged theory of entrepreneurship. Most importantly, we abstract from entry, exit, and other project dynamics that are distinct from those of the entrepreneurs themselves. We believe, however, that this is a promising direction for future research.[4]

The remainder of the paper is organized as follows. Section 2 details the model environment and defines the equilibrium. Section 3 characterizes the patterns of CEO mobility. We describe the data in section 4 and the specifics of our calibration strategy in section 5. The counterfactuals in section 6 quantify the effects of mismatch (to be completed). Section 7 concludes (to be completed).

2 Model

2.1 Population and Endowment

The economy is populated by a unit measure of individuals who are endowed (at birth) with skill a. This ability to manage a firm is drawn from a known normal distribution with mean zero and precision ϕ. The draws from this distribution are i.i.d. across agents.

Individuals live for R periods. Each period, a new cohort of size $\frac{1}{R}$ enters the economy, while the oldest cohort (of the same size) retires. New entrants do not know their type. Instead, they observe a *public* signal $a_0 = a + \alpha_0$. The innovation α_0 is drawn from a normal distribution with mean zero and precision ψ. At time $t \in \{1, 2, \ldots\}$, a manager's contemporaneous ability is given by $a_t = a + \alpha_t$. The uncertainty associated with the CEOs' type, α_t, is drawn from the same distribution as α_0.[5] The innovations are independent cross-sectionally and over time. In addition, individuals have a periodic unit endowment of time and each period they decide what fraction to spend on managerial tasks and on an alternative occupation or activity that is compensated at the exogenous rate w per unit of time.

The economy is also endowed with a unit measure of long-lived projects with idiosyncratic attributes denoted by q. The qualities or productivities of the trees are drawn from a c.d.f. $F(\cdot)$ with corresponding density $f(\cdot)$. Whenever q follows a discrete distribution with K distinct values, the corresponding p.m.f. is denoted by $f(q_k) = \gamma_k$, for $k \in \{1, \ldots, K\}$ and for simplicity we adopt the notation $F(q_k) = \sum_{\ell=1}^{k} \gamma_\ell = \Gamma_k$. Each individual owns one such project for

[4] See, for instance, Alder et al. (2014).

[5] Relaxing this assumption generates the same qualitative results but requires heavier notation.

the duration of her lifetime. When generation g dies after R periods, the "orphaned" trees are bequested to randomly chosen individuals of the cohort with birth date $g + R$.

2.2 Preferences and Technology

Individuals have linear preferences over their lifetime consumption stream:

$$U\big(\{c_t\}\big) = \sum_{r=t}^{t+R} \beta^{r-t} c_r$$

Since there is no consumption-saving tradeoff, c_r is simply capped by the sum of the outside option w, if taken, managerial income, and the period flow return π from ownership of a tree. A CEO-firm pair produces a contemporaneous surplus x_t, which is a function of their respective contributions a_t and q:

$$x_t = x(a + \alpha_t, q) \tag{1}$$

The technology is continuous and increasing in both the manager's effective contribution a_t and the tree's quality q. Moreover, $x(\cdot,\cdot)$ satisfies the (weakly) increasing differences property.

2.3 Beliefs

The distribution of true, but unobserved managerial talent is stationary. Each cohort of size $\frac{1}{R}$ draws types from the same normal distribution with mean zero and variance ϕ^{-1}. Clearly then, the aggregate distribution of talent follows that same distribution. In contrast, individual realizations of a are not observable. Instead, at time t each agent receives a public signal $a_t = a + \alpha_t$ and forms a belief about her own type. Since a_t is public information, everybody has the same belief about any one individual's type. Recall that α_t is drawn from a mean-zero normal distribution with precision ψ:

$$\alpha_t \sim \mathrm{N}\big(0, \psi^{-1}\big), \text{ for all } t \tag{2}$$

Each signal enables individuals to update their prior using a Kalman filter. Since the innovation is normal, the posterior $\tilde{a}'$ of someone who is r years old at time t is normal with mean

$$\hat{a}' \equiv \mathbb{E}\big(\tilde{a}'\big) = \tfrac{\phi_i + r\psi}{\phi_i + (r+1)\psi}\hat{a} + \tfrac{\psi}{\phi_i + (r+1)\psi} a_t, \tag{3}$$

where $\hat{a}$ denotes the mean of the agent's prior $\tilde{a}$. The posterior's variance is independent of the realization of a_t and given by

$$\text{Var}(\tilde{a}') = \tfrac{\psi}{(\phi+r\psi)(\phi+(r+1)\psi)}. \tag{4}$$

Note that the variance of the posterior is independent of the individuals life-time history of shocks and decreasing in age: older workers have a more precise idea about their own abilities. Since there is no private information, this implies that everybody in this economy has more precise beliefs about the abilities of older workers compared to younger ones. Since $\frac{\phi+r\psi}{\phi+(r+1)\psi} + \frac{\psi}{\phi+(r+1)\psi} = 1$ and the second term is decreasing in r, agents put more and more weight on their prior $\hat{a}_r$ at the expense of the innovation a_t as they approach retirement at age R.

To build some intuition, we first describe the contracts in an environment where matching and separation are free as in Groes et al. (2015), among others. The dynamics of this economy are particularly simple and tractable. Since learning does not depend on a CEO's assignment, the dynamic equilibrium is a sequence of (stationary) static equilibria where only the cross-section of beliefs matters. Individual assignment histories, on the other hand, are of no consequence. In section 2.5 we consider a more general version of the model where a CEO pays a fixed matching cost μ. In this case, there are no shortcuts and we must solve the full dynamic program where a CEO's most recent assignment – including the outside option, if she happens to have taken it – is an endogenous idiosyncratic state.[6] Just like before, a manager's belief and age are the stochastic and deterministic exogenous states, respectively. Costly matching will turn out to be a salient feature of the data and we use the second version of the model in the quantitative parameterization in section 5 (currently work in progress).

2.4 Contracts when Matching is Free ($\mu = 0$)

As in Groes et al. (2015), we consider the set of contracts that is bounded by two polar types, where either the owner or the manager assumes all risks associated with a particular match:

1. The wage offers are output-contingent contracts. The firm holds on to a reservation profit and the CEO is the residual claimant (and hence assumes all the risk).
2. The wage offers are *ex ante*-type-contingent contracts. The firm offers a non-contingent payment to the CEO and claims the residual surplus.

[6] To keep the dimensionality of the model in check, we solve this model only for discrete distributions of productivities.

In the former case, the owners of a project seize a guaranteed profit π and pay out the residual surplus to the CEO. Note that for particularly bad realizations of the signal (or productivity shock), the CEO's compensation can be negative.[7] In equilibrium, projects with the same characteristics earn identical profits and we denote them by $\pi(q)$.[8] The cross-sectional distribution of beliefs – which is stationary and common knowledge in this economy – is sufficient to characterize the profit profile that clears the market and firm owners hire anyone who applies for the job. The assignment of CEOs is self-selective. In contrast, wage contracts that do not depend on the *ex post* surplus require that project owners know as much about the applicants as the CEOs themselves do. If they did not, CEOs would have incentives to misrepresent their type in order to extract a bigger share of the match value. Both contracts are efficient in that they decentralize the planner's solution that features *ex ante* assortative matching.

The realized split of the surplus between managerial compensation and payments to shareholders depends on the type of contract. In particular, the joint pattern of pay volatility and switching (turnover) differs between the two types of contracts. When pay is output-contingent then deviations from expected compensation are associated with an increased probability of separation. In contrast, turnover is orthogonal to contemporaneous pay, but correlated with firm profits when the terms of the contract are type-contingent. The data we have suggests that, in practice, the wage contract is somewhere between these types. We begin by spelling out the wage offers in the polar self-selection and full insurance cases.

An individual's posterior belief (denoted by $\tilde{a}'$ is characterized by two arguments: (1) the mean of her prior $\hat{a}$, which is a discrete-time Martingale, and (2) the variance associated with a CEO of a certain age, say r. We know from (4) that the variance is decreasing deterministically in r. In the special case of our model where $x(a_t, \cdot)$ is linear in a_t we can ignore the variance altogether and CEOs are sufficiently characterized by $\hat{a}$ for sorting purposes. When the attributes of CEOs and firms are compolements, however, we need to keep track of a CEO's expectation *and* variance.[9] Since the innovation α is normal, the posterior $\tilde{a}$ is also normal.[10]

$$\tilde{a}' \sim \mathrm{N}\Big(\hat{a}, \tfrac{\psi}{(\phi+r\psi)(\phi+(r+1)\psi)}\Big) \tag{5}$$

[7] Since agents are risk neutral, these outcomes are not problematic in any way.

[8] When the distribution is discrete, we index the profits of each type of project by $k \in \{1, \ldots, K\}$.

[9] Curvature requires us to project the expectation and variance of a manager's type onto a one-dimensional space in order to characterize an *ex ante* assortative match.

[10] Put differently, when the technology is super-modular we need to keep track of the entire distribution of $\tilde{a}(\hat{a}, r)$. Since it is normal, mean and variance describe the distribution completely. When the technology is linear, the expectation $\hat{a}$ is a sufficient statistic.

The expected output of a manager with prior $\hat{a}$ and age r who is paired with project q is:

$$\mathbb{E}_{\hat{a},r}\left[x(\tilde{a},q)\right] \equiv \int_{-\infty}^{\infty} x(\tilde{a}',q)\mathrm{d}F_{\hat{a},r}(\tilde{a}'), \tag{6}$$

where $F_{\hat{a},r}$ is the C.D.F. of $\tilde{a}'(\hat{a},r)$, i.e. the distribution of posterior beliefs of CEO whose prior has mean $\hat{a}$ and who has r years of experience.

2.4.1 Output-Contingent Offers

When the CEO's pay is contingent on output, she is a residual claimant and bears all the risk. Firms keep a reservation profit $\pi(q)$ for themselves and offer $\omega(x,q) = x(a+\alpha_t,q) - \pi(q)$ to prospective CEOs. Keep in mind that for particularly unfavorable realizations of $a+\alpha_t$, the CEO may earn $\omega(x,q) < 0$. When the firm retains $\pi(q)$, the sorting of CEOs relies on self-selection and prospective employers do not care about the applicant's type. While project owners want to maximize their share of the surplus, they must offer prospective CEOs terms such that the vacancy is filled. That is, the profile $\pi(\cdot)$ must be such that the market for CEOs clears.

Note: The remainder of section 2.4.1 is written for the special case where projects are drawn from a discrete distribution. The next version of the paper will characterize the decentralization for the general case.

To ensure market clearing we need to verify that CEOs self-select exactly one employer, except for a finite number – $K \times T$ of them, to be precise – with measure zero. We call them "critical" types. To make further progress we need to describe the cross-sectional distribution of beliefs in the economy. To build some intuition, consider a cohort of age r. According to equation (5), the variance of the CEOs' beliefs *within* this cohort is constant and equal to $\frac{\psi}{(\phi+r\psi)(\phi+(r+1)\psi)}$. Since agents will differ in their expected ability $\hat{a}$, we need to characterize the cohort-specific distribution of these mean beliefs. One can show that mean beliefs of CEOs of age r are cross-sectionally normal:

$$\hat{a}_r \sim \mathrm{N}(0, \tfrac{r\psi}{\phi(\phi+r\psi)}) \tag{7}$$

We denote the corresponding C.D.F. by F_r. Recall that the true (but unobserved) distribution of abilities is $N(0,\phi^{-1})$. The cross-section of beliefs must be unbiased and therefore the distribution "inherits" the zero mean property. To understand the variance, let us compare the case of a CEO who has not yet observed a signal ($r = 0$) with one who has seen many of them ($r \to \infty$). A group of CEOs who have not received any signals about their type must have the same belief: zero. The distribution collapses to a mass point and the variance is indeed $\frac{0\psi}{\phi(\phi+0\psi)} = 0$. In contrast, the Bayesian update in equations (3) and (4) implies that the distribution of expected abilities converges to the distribution of true types a with a variance around each point estimate that decays to zero as $r \to \infty$. In the limit, CEOs know their own type for sure. The cross-sectional distribution of expected abilities approaches the variance of true types in the (very) long run:

$$\frac{r\psi}{\phi(\phi+r\psi)} \xrightarrow{r\to\infty} \frac{1}{\phi}$$

Since the sorting is *ex ante* assortative in our model, we find it helpful to rank CEOs by their expected ability $\hat{a}$ within their cohort. Moreover, it turns out to be useful to define the inverse C.D.F. of expected abilities in cohort r:

$$\hat{a}_r[i] = F_r^{-1}(i) \tag{8}$$

The inverse C.D.F. maps a CEO's rank i and age r into an expected ability $\hat{a}$. When she is paired with a firm of type q_k, the match generates a value $\mathbb{E}\big[x\big(a_t(r,i), q_k\big)\big]$, where the expectation is over the normal realizations of a_t with mean $\hat{a}$ and variance $\frac{\psi}{(\phi+r\psi)(\phi+(r+1)\psi)}$. For a given K-tuple of reservation profits $\{\pi_k\}_{k=1}^K$ and age r, there are K critical ranks $\{i_{k,r}\}_{k=1}^K$ and $i_0(r) \equiv 0$ (for all r) that satisfy:

$$\mathbb{E}\big[x\big(a_t(r, i_{k,r}), q_{k-1}\big)\big] - \pi_{k-1} = \mathbb{E}\big[x\big(a_t(r, i_{k,r}), q_k\big)\big] - \pi_k \tag{9}$$

The "critical" CEO $i_{k,r}$ is indifferent between the contracts offered by type q_k and q_{k-1} firms and $\frac{i_{k,r}-i_{k-1,r}}{S}$ denotes the measure of CEOs who select the contract offered by q_k. Since $x(\cdot,\cdot)$ exhibits increasing differences and given the beliefs about their own type, alternative employers offer less lucrative contracts in expectation. Summing the difference of "adjacent" critical ranks over all cohorts $s \in \{1,\ldots,S\}$ yields the total measure of CEOs who select the contract offered by q_k:

$$j_k - j_{k-1} = \frac{1}{R}\sum_{r=1}^{R}\Big(j_{k,r} - j_{k-1,r}\Big) \tag{10}$$

In addition to the indifference condition, the contracts must satisfy a participation constraint and market clearing conditions. CEOs have an outside option that delivers w for sure.[11] Clearly, then, the marginal CEOs indexed by $j_{1,r}$ and the marginal project q_1 satisfy:

$$\mathbb{E}\big[x\big(a_t(r, i_{1,r}), q_1\big)\big] = w + \pi_1 \tag{11}$$

[11]The equilibrium is partial. It is straightforward to characterize the general equilibrium in a version of the theory with a Lucas (1978) span-of-control production function $x(\cdot,\cdot)^{1-\gamma}\ell^{\gamma}$ that requires managers as well as workers to produce the final good. In such an economy workers earn a competitive wage w, which is determined in general equilibrium.

Knowing the distribution of beliefs across all cohorts, firms choose the reservation profits $\{\pi_\ell\}_{\ell=1}^K$ that satisfy (11) and clear the market for CEOs:

$$j_{\ell+1} - j_\ell = \gamma_\ell, \text{ for } \ell = 1, \dots, K-1 \text{ and } j_{K+1} = 1 \tag{12}$$

We assume that there is no use for projects outside the assignment market, which in turn implies that $\pi_1 = 0$ in equilibrium.

Definition 1 (Equilibrium with Output-Contingency) *An equilibrium is a set of reservation profits $\{\pi_k\}_{k=1}^K$ that satisfy the participation constraint in* (11) *and the market clearing conditions in equation*(12) *by way of the indifference conditions in* (9) *and the cross-cohort aggregation in* (10).

Since agents are risk-neutral, they select projects based on expected ability and compensation. Figures 1 and 2 plot the expected output-contingent compensation against the expected contribution against the CEO's belief about his own type for a simple version of the model with $K = 3$, $T = 5$ and outside option $w = 0.986$. This competitive wage implies that $i_1 = 0.95$ and only the right tail of the distribution is sufficiently competent to manage a project. Importantly, after controlling for expected ability, older CEOs receive higher compensation than young ones. Since they have more precise beliefs about their own type, they can command higher compensation in expectation when the production technology exhibits diminishing marginal returns to ability. As $\rho \to \infty$, precision ceases to matter and the expected compensation is determined by the expected ability alone.

Figures 3 and 4 highlight the *ex ante* selection: CEOs never accept a job that lies inside the envelope covering the K different contracts. The lower panel shows that older cohorts have more mass in the tail of the belief distribution and the occupational cutoff for the youngest CEOs is higher (1.77) than for the oldest (1.42). Once the uncertainty is realized, of course, the output-contingent compensation need not lie on the envelope. The plot of realized ability $a + \alpha_t$ against actual compensation lies *on* the wage offer curve, but not necessarily on the *envelope* of all offers. Clearly, CEOs bear all the risk. Alternatively, firms may offer a type-contingent contract.

2.4.2 Type-Contingent Offers

A CEO *type* is characterized by the pair (r, i), that is, an age r (associated with belief precision $\phi + r\psi$) and a rank i (corresponding to expected ability $\hat{a}_r[i]$). When the offers are type-contingent, the firms are residual claimants and assume all the risk. Instead of picking the reservation profits $\{\pi_k\}_{k=1}^K$, firms offer a set of wage functions $\{\omega_k(r,i)\}_{k=1}^K$. Assuming that

Figure 1: Youngest CEOs

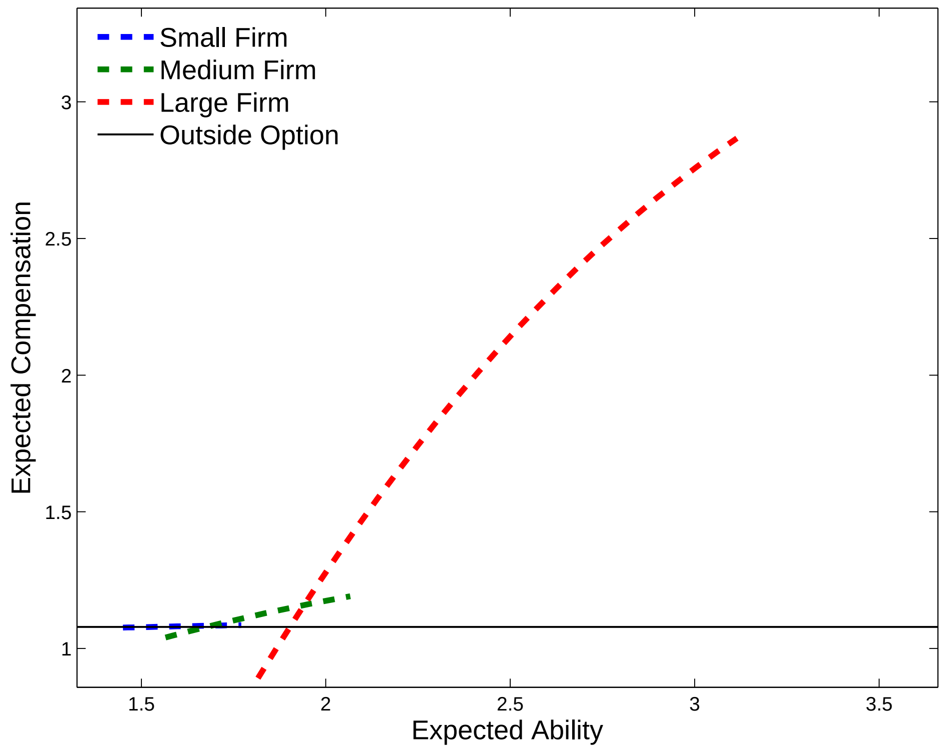


Figure 2: Oldest CEOs

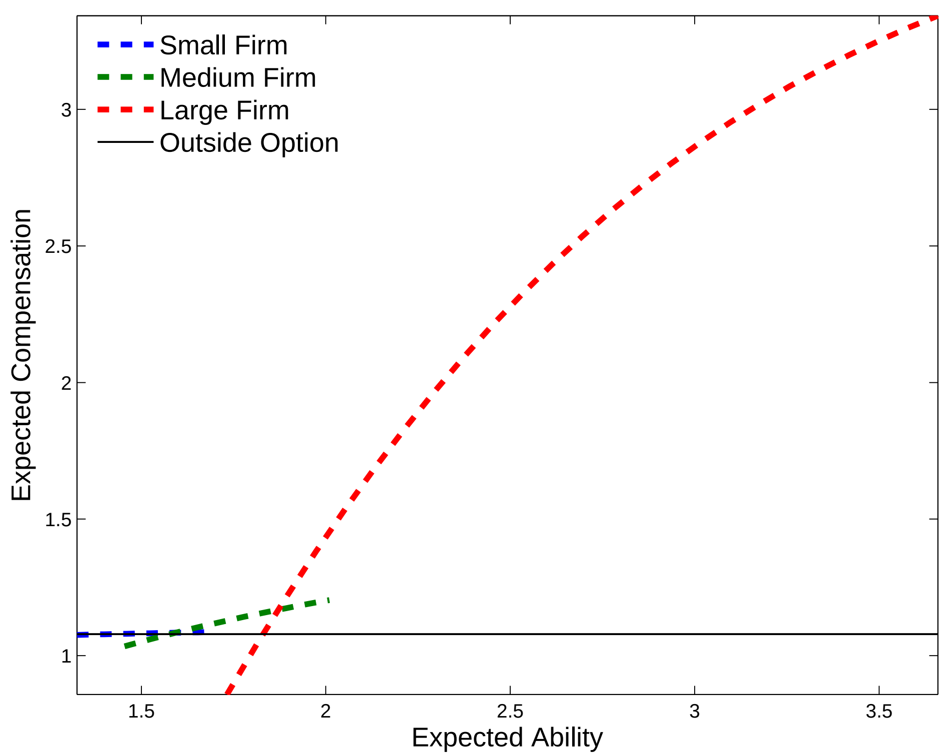


Figure 3: Youngest CEOs

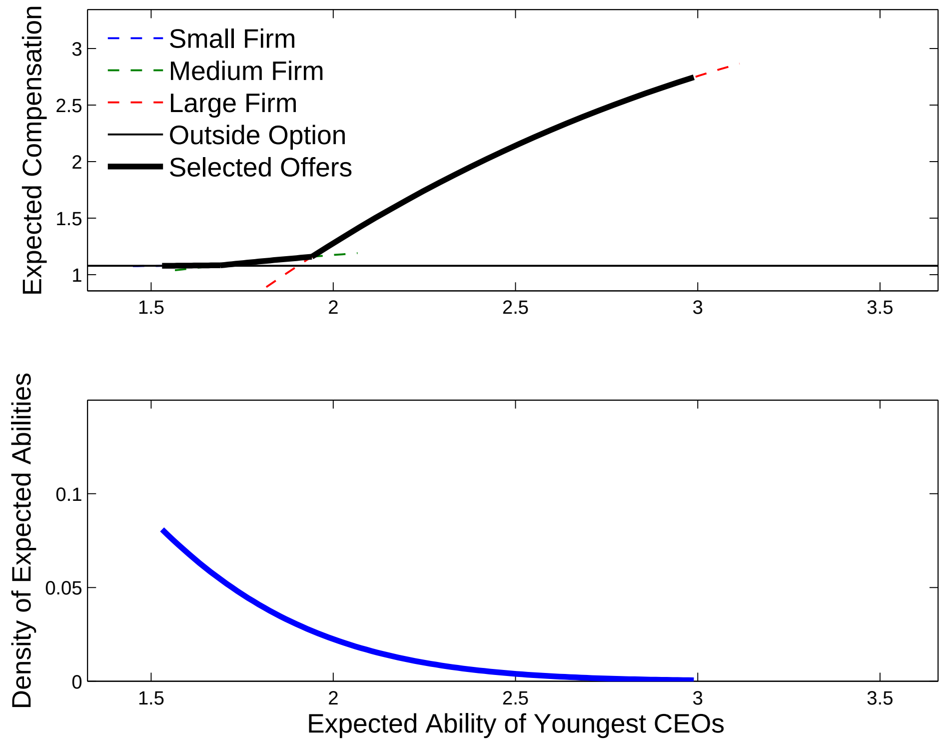
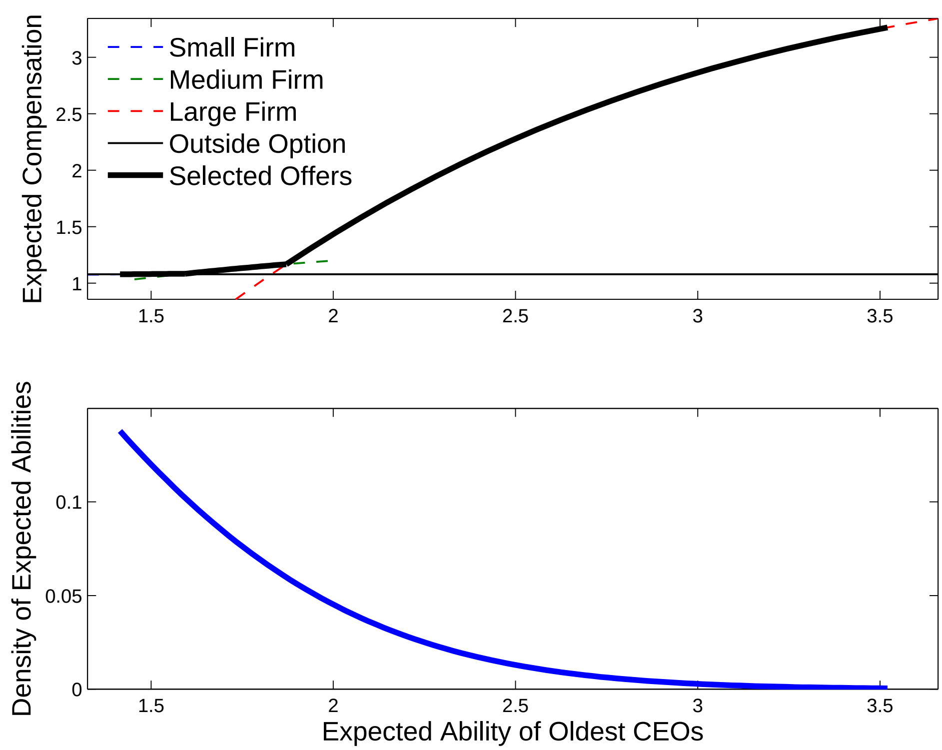


Figure 4: Oldest CEOs



firm k knows as much about prospective CEOs as the candidates themselves do, the offers are such that it is indifferent between all of them *ex ante*:

$$\mathbb{E}\big[x\big(a_t(r,i),q_k\big)\big]-\omega_k(r,i)=\mathbb{E}\big[x\big(a_t(r',i'),q_k\big)\big]-\omega_k(r',i') \tag{13}$$

for $(i,i')\in[0,1]$, $(r,s')\in\{1,\dots,S\}$, and $k\in\{1,\dots,K\}$.

CEOs, on the other hand, are not indifferent between prospective employers except, again, for $K\times S$ of them with measure zero. These critical CEOs, indexed by $i_{k,r}$ as in section 2.4.1, are indifferent between the contracts offered by firm k and $k-1$:

$$\omega_k\big(r,i_{k,r}\big)=\omega_{k-1}\big(r,i_{k,r}\big) \tag{14}$$

By combining (13) and (14) we can characterize the discrete increase in expected profits between firm $k-1$ and k by the difference in the expected match value when q_{k-1} and q_k are paired with the critical CEO indexed by $i_{k,r}$:

$$\begin{aligned}\mathbb{E}\Big[x\Big(a_t\big(r,i_{k,r}\big),q_k\Big)\Big]&-\mathbb{E}\Big[x\Big(a_t\big(r,i_{k,r}\big),q_{k-1}\Big)\Big]\\&=\mathbb{E}\Big[\pi_k\Big(a_t\big(r,i_{k,r}\big)\Big)\Big]-\mathbb{E}\Big[\pi_{k-1}\Big(a_t\big(r,i_{k,r}\big)\Big)\Big]\end{aligned} \tag{15}$$

The cross-cohort aggregation continues to follow (10) and the participation constraint is:

$$\mathbb{E}\big[x\big(a_t(r,i_{1,r}),q_1\big)\big]=w+\mathbb{E}\Big[\pi_1\Big(a_t\big(r,i_{1,r}\big)\Big)\Big] \tag{16}$$

Definition 2 (Equilibrium with Type-Contingency) *An equilibrium is a set of wage contracts $\{\omega_k(r,i)\}_{k=1}^{K}$ that satisfy the participation constraint (16) and the market clearing conditions in equation (12) by way of the indifference conditions in (14) and the cross-cohort aggregation in (10).*

Just as with output-contingent offers, CEOs select an offer that compensates them for their expected ability on the envelope. In contrast, since the contract payout does not depend on the realizations of the stochastic processes, the realized compensation is always a point on the envelope of offered contracts (see the plot of "Selected Offers" in figures 3 and 4 for young and old CEOs in our simple example).

Since everyone is risk-neutral, the type of contract on offer does not affect the optimal assignment of CEOs to firms.

Proposition 1 *The ex ante sorting of CEOs to firms is invariant to the identity of the residual claimant.*

Proof See appendix (forthcoming). Q.E.D.

2.5 Contracts when Matching is Costly

When matching entails no costs, the mean of a CEO's prior and her age are sufficient to characterize the equilibrium assignment. We now introduce assignment frictions by way of a fixed matching cost μ. In addition to a CEO's belief and experience, we must also keep track of her assignment history in order to solve the dynamic program. More formally, a CEO with prior $\hat{a}$, r years of experience, and who was most recently matched with a project of productivity q, solves the following Bellman equation:

$$V(\hat{a}, q, s) = \max_{q' \geq 0} \Big\{ \mathbb{E}_{\hat{a},s}\Big[x(a_t, q')\Big] - \pi(q') - \mu \times \mathbb{1}_{q \neq q' > 0} + \beta \mathbb{E}_{\hat{a},s}\Big[V\big(\hat{a}'(a_t), q', s+1\big)\Big] \Big\}, \tag{17}$$

where $x(a_t, 0) = w$ for all a_t, $s \in \{1, \dots, S\}$ and $V(\cdot, \cdot, S+1) \equiv 0$.

The effect of costly matching is to discourage CEOs from leaving their current employer when their posterior $\hat{a}'$ is "close" to the prior $\hat{a}$. This friction, together with the precision of the period innovation α_t, allows us to calibrate the frequency of separations as well as the extent to which individual CEOs are assigned to projects with distinct productivities over the course of their careers. Note that we allow CEOs to select the outside option, denoted by $q = 0$, at no cost. The reverse, however, is not true. If she selects any $q \neq 0$ after a spell in $q = 0$, she incurs the fixed cost μ.

TO BE COMPLETED.

3 Patterns of Mobility

The mobility patterns of CEOs in this model are governed by the evolution of the age-specific cross-sectional distribution of mean beliefs, the precision (i.e. the inverse of the variance) of the individuals' beliefs, the size of the matching cost μ, and the amount of curvature in the production technology $x(\cdot, \cdot)$. While matching is assortative in beliefs across members of a given age cohort who have their most recent assignment (including the outside option) in common, the matching cost generates considerable assignment overlap when we consider the various decisions of an entire cohort. CEOs with identical beliefs and of the same age may endogenously choose different assignments and this complicates the CEOs' participation, separation, and exit decisions. In fact, we have to rely on numerical solutions to describe these lifecycle patterns.

When the fixed matching cost is positive, the sorting of agents across productivities *within* age cohorts is not perfect and it is a bit of a challenge to highlight the role of the remaining parameters transparently. To build some intuition for the various forces that shape sorting patterns in the economy, let us consider a very stylized model endowed with a unit measure of potential managers, costless matching ($\mu = 0$), five different productivities (each with measure $\frac{1}{5}$), and three age cohorts.

We start this exercise with the following set of parameter values: $\psi \to \infty$, $\rho \to \infty$, and $w = 0$. Put differently, the agents' beliefs coincide with the distribution of true types in the economy (and hence do not evolve with the agents' age), the surplus technology is (almost) linear, and all agents participate.[12] As Figure 6a illustrates, the assignment of managers to projects is identical across the three age cohorts.

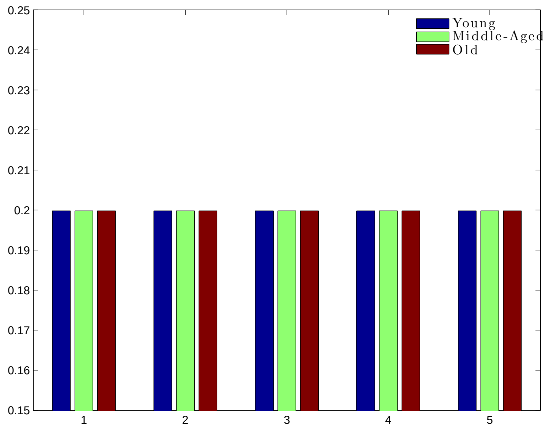
In Figure 6b, managers have imprecise beliefs and sorting with $\psi < \infty$ is affected by two distinct forces: First, the mean beliefs of young agents are more tightly clustered around the mean of the distribution and that generates the inverted U-shape of the blue bars. Second, since the managers' contributions are effectively log-normal, the assignment of young CEOs is skewed to the right of the size distribution of firms. Since all agents of all age cohorts participate in this assignment market, market clearing implies that assignment of older CEOs is skewed to the left of the size distribution.

When we tighten the degree of complementarity as in 5c, the curvature partially offsets the skewness of the managers' effective contributions to the match surplus. Young CEOs are more tightly clustered around the median productivity whereas older managers tend to be assigned more frequently to extreme project types in either tail of the size distribution.

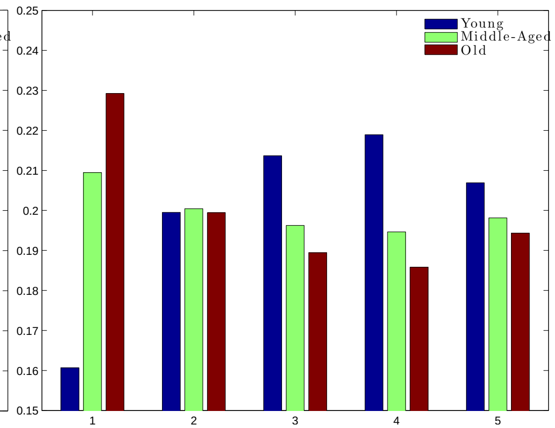
Next, let's consider the case where the participation constraint is binding, that is, $w > 0$. In the example illustrated in Figure 5d, we set w such that 10 percent of all potential agents participate in the market. The remainder prefer to take up their outside option. Given that the distribution of mean beliefs varies across cohorts, their participation rates are no longer equalized. Older agents are more likely to be managers than their more junior colleagues. This is, in fact, a pattern that we also observe in the data.

Lastly, if participation is not universal, the precision of the productivity shocks (ψ) affects the hazard rate into managerial work or entrepreneurship across different age groups. In Figure 5e, for instance, as the innovations become less precise, young managers are less likely

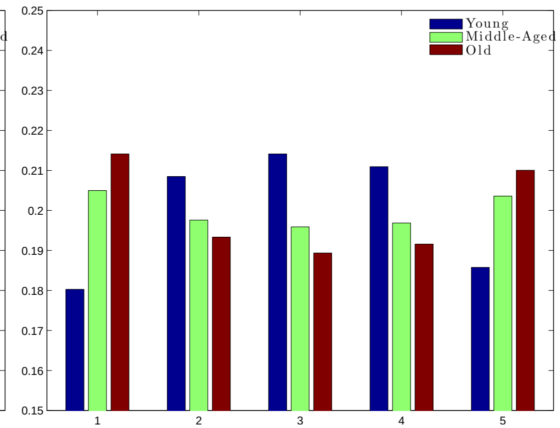
[12] In this stylized thought experiment, the substitution elasticity is high, but finite. When the technology is linear, the assignment problem is uninteresting since optimality does not require perfect sorting. Any assignment maximizes the total surplus. A tiny bit of curvature is enough, however, to restore sorting while marginal products remain largely functions of the manager's (or firm's) own type and this is exactly the point from where we want to build the intuition for the effects of alternative parameter values.



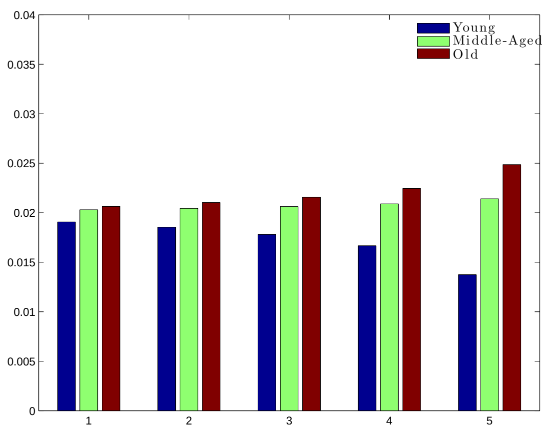
(a) $\psi \to \infty,\ \rho$ high, $w = 0$



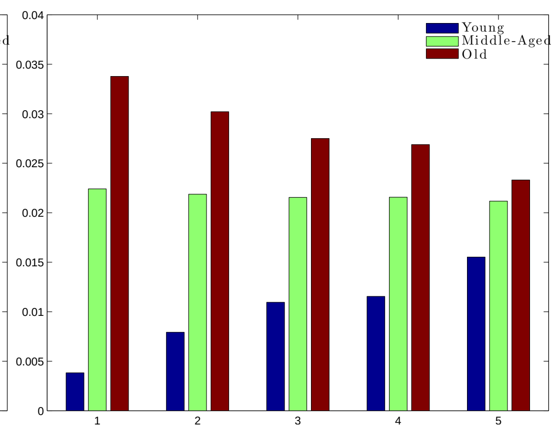
(b) ψ low, ρ high, $w = 0$



(c) ψ low, ρ low, $w = 0$



(d) ψ low, ρ low, $w > 0$ (10% participation)



(e) ψ even lower, ρ low, $w > 0$

Figure 5: Stylized Sorting Patterns by Age for $K = 5,\ S = 3$, and $\mu = 0$

to participate in the market, but there is a differential effect across the size distribution of firms. High productivity are less averse to “hiring” a young CEO compared to those with low productivity.

In sum, the three key parameters generate a fairly rich sorting pattern across the managerial talent and productivity distributions. Quantitatively, the challenge is to find a set of moments in the data with corresponding counterparts in the model in order to discipline the parameterization. This is, in fact, what we are turning our attention to in the next two section.

4 Data

We use information on CEOs and firm accounting data from three different data sources during the period 2000 to 2009. The first is the Danish administrative register data covering 100% of the population of individuals and firms in the years 2000 to 2009. The second is from *Købmandstadens Oplysningsbureau* (KOB), which contains accounting data from the firm population dating back to 1991. The third is from the Danish Commerce and Companies Agency (*Erhvervs- og Selskabstyrelsen*, or ES) at the Ministry of Economic and Business Affairs and has information on managers of firms from 2000 to 2010. Personal information on managers can be linked from the ES data to the register data through a personal id number and the information on firm accounting data can be linked from the KOB data to the register data through a firm identifier. The register data further contains a link between the firm and the manager identifiers.

Using all three datasets to we can identify the CEOs of all Danish publicly and privately held stock companies and link this to information on CEO compensations well as accounting data for the firms the CEOs work in.

Our first dataset is administrative register data, from Statistics Denmark, covering 100% of the population in the years 2000 to 2009. The Statistics Denmark data is from the Integrated Database for Labor Market Research (IDA), which contains annual information on socioeconomic variables (e.g., age, gender, education, etc.), characteristics of employment (e.g., wages, earnings, occupations, industries, etc.), and a employer-employee link for the population. The employer-employee link only exist for those individuals who held a job during the last week of November in any given year.

We use a measure of yearly earnings that contains yearly salaries including perks, tax-free salaries, jubilee- and severance payments, and the value of stock options and futures. These values are recorded when they are taxed, which happens at the year excercise (when they get sold or ceded/waived). The measure of earnings further contains payment for work on any board of directories. Payments for consultancy work and other work related to giving presentation etc. are not included in the earnings variable.

The Statistics Denmark individual register data can be linked, through a firm identifier, to our second dataset, KOB, that contains firm accounting data, covering 100% of the publicly and privately held stock companies in the population since 1991. The KOB dataset is collected and digitalized from scanned documents at the Ministry of Economic and Business Affairs by a private data collector called Experian. According to Danish law, all stock companies must provide the Ministry of Economic and Business Affairs with information about firms' asset and measures of profitability such as operating cost and net income. Other types of firm information (eg. number of employees, total wages, and sales) is provided voluntarily for a large majority of the firms. The firms' reported accounting data are subject to random auditing by external accountant which makes the reliability of the accounting data high.

Our third dataset, ES, is provided by the Danish Ministry of Economic and Business Affairs and contains information on all CEOs and board members in the publicly and privately held stock companies in Denmark as well as information about the founders of the firms. The dataset contains information about the starting and finishing date of the spell for each CEO, board member, and founder and it includes an individual id number such that it is possible to link the the managers and founders, to their background characteristics in the Statistics Denmark dataset. The dataset also contains information on the type of firm the managers work in and the starting year of the firm.

We deflate the labor earnings, assets, and value added to the 2000 level using Statistics Denmark's consumer price index.

4.1 Sample Selection

We use privately and publicly held stock market firms in Denmark containing employees who work either part time or full time during the period 2000-2009. We then select those firms where the CEO has wage work in the firm as his or her main occupation during the last week of November of a given year.

While we can match a CEO to 94% of all stock market firms in Denmark, only 74% of the firms have a CEO who has his or her main employment in the firm. Since our focus is on the match between CEOs and firms, we exclude all firms without a CEO with a main employment in the firm. These excluded firms are primarily smaller than average since the included firms contains 80% of all the workers. Of the matched firms, 8% has more than one registered CEO and for these firms, we identify the top CEO as the CEO with the highest occupational code (eg. "Directors and chief executives" is a higher occupational code than "Productions and operations manager")[13]. If a firm has two or more registered CEOs who have the same highest occupation code, we identify the CEO as the one who has the highest earnings. For 0.2% of the firms there are two or more CEOs with the same highest earnings and we also exclude these firms.

For our estimation we need accounting information from the firm and we therefore match the Statistics Denmark data to the accounting data from KOB. We match 95.5% of workers and 93.5% of the firms where the unmatched firms are smaller than average or firms that terminate during our selected sample period. We exclude the firms with unmatched KOB accounting data and we also exclude firms in the years they have missing information on value added, total assets, or fixed assets. There are less than 1% of firm-years with missing data on total assets or fixed assets and there are 4% of firm-years with missing value added. The firms with missing information are also primarily firms that are in their termination year or firms that are new in the sample. In total this gives a base sample of 149,272 stock companies that all have non-missing accounting data and an identified CEO whose main occupation is as CEO of the company.

We think of a CEO as one who manages people and we therefore select CEOs from firms with five or more employees. We also exclude stock companies where the CEO is is the founder of the company because of problems separating company profit from CEO earnings when the CEO is also the founder. This reduces the sample of stock companies to 85,791. In the remaining sample of CEOs 7% are women and we exclude the women to get a more homogenous sample of CEO. We also exclude CEOs without a valid education, which leaves a sample of 79,995 male CEOs. In order to avoid complication with CEO retirement we further exclude all companies with CEOs of age 60 and above in 2009.[14] We also exclude all CEO with ages below 25 in year 2000, which is less than 1 percent of the sample. This gives a final sample of 57,279 CEOs in stock companies employing 3.985 million workers. Descriptive statistics of the CEO sample and the firm sample used in the analysis are provided in Table 1 and Table 2.

In order to match moments from the data, we divide the firms into bins by using deciles created either by the number of employees or value added. The bins are computed using the firms in our sample and we use these bins when we calculate the moments relating to CEO transitions.

[13]The occupation code are from the Danish version of the ISCO codes.

[14]selecting CEOs of age 59 or below in year 2009, means that the oldest cohort of CEOs we include in the sample are the ones who are 50 in year 2000 (and therefore 59 years old in year 2009).

	Mean	Median	25 percentile	75 percentile
Number of CEOs	57,279			
CEO to CEO switchers across firms	0.0326			
CEO to CEO switchers (10 employment bins)	0.0242			
CEO to CEO switchers (10 value added bins)	0.0253			
Males	1			
Age	44.2	44	39	49
Potential Experience	24.3	24	19	30
CEO tenure in firm	5.88	6	3	9
Hourly nominal wage in 2000 DKK	415.8	340	234	500
Total nominal earnings in 2000 DKK	709,811	571,824	390,309	848,562
Total nominal income in 2000 DKK	950,403	696,544	485,778	1,031,895

Table 1: Summary Statistics for CEOs

	Mean	Median	25 percentile	75 percentile
Number of firms	57,279			
Number of employees	60	19	10	41
Total assets in 1,000 DKK	615,777	16,926	7,558	47,330
Intangible assets in 1,000 DKK	536,348	10,877	4,785	30,760
Value added in 1,000 DKK	44,262	9,838	5,037	22,837

Table 2: Summary Statistics for Firms

4.2 Empirical Moments

We use six empirical moment in our calibration, which are as follows. Roberts' Law, which is the elasticity of CEO pay with respect to firm size, is calculated by using either employment or value added as firm size. We find that the elasticities are around 0.27 and are robust to conditioning on 1 to 4-digit industry fixed effects. The tail index of the Generalized Pareto distribution of firm sizes is similarly calculated using either employment or value added as our measure of firm size. In the former case, the parameter ξ is estimated at 0.63 while in the latter the index is one half.

The average CEO share of total output is 0.33 and is calculated by finding the average of CEO earnings over the sum of CEO earning and the flow value of intangible assets. [15]

[15] If we use the value added as total output the CEO's compensation share is 0.131.

The annual wage growth of CEO compensation over the lifecycle is found by calculating the average growth rate in earning by age for each cohort of CEO's. To estimate this semi-elasticity we follow Beaudry and Green (2000) and regress log earnings on a cubic polynomial in age, a quadratic polynomial in cohort, and an linear interaction term between age and cohort while controlling for business cycle conditions.[16] As in Beaudry and Green (2000) we perform this regression for the log of cohort-year averages in earnings, which after taking averages has 90 observations. The marginal effect of age on log earnings equals the annual growth rate for CEO compensation, which is 1.93.

We calculate the CEO-to-CEO transition probability by keeping track of switches across firm deciles, *conditional* on switching firm. For our purposes a switch is a CEO-to-CEO transition that happens between year t and $t+1$ or between t to $t+2$, separated by a year of not being a CEO in the sample. The reason for allowing a transition year is a peculiarity with respect to occupational classifications in our sample. CEOs are labeled as such if their job in November is their *main* employment during the year. By allowing a transition year we can account for CEO-to-CEO transitions that take place toward the of the year, but before the end of November. In this case, their November employment is not their main employment for that year. In the CEO-to-CEO transitions we therefore condition the CEO in year t on also being a CEO in year $t+1$ or $t+2$. We calculate the overall CEO-to-CEO transition probability across firm as 2.6% and CEO-to-CEO transition probability across firms *and* decile bins as 1.87%.[17]

Table 3: Empirical Moments.

Moment	Value
Elasticity between CEO pay and firm size (value added)	0.27
Generalized Pareto tail index of firm sizes	0.50
Average probability of CEO-to-CEO transition	1.9%
Cross-sectional separation (and exit) patterns	0.37
CEO's share of surplus	0.33
Annual growth of CEO compensation	1.9%

[16]The business cycle conditions is also created as it is in Beaudry and Green (2000) and is the residuals from a linear regression of the unemployment rate of the population of men, age 25-60, on a quadratic time trend.

[17]When looking at CEO-to-CEO transitions, we condition of the CEO being in the CEO sample in the year $t+X$. This reduces the sample to 43,220 observations.

5 Calibration

5.1 Parameters and Empirical Moments

In our calibration strategy we set six parameters in order to match six empirical moments and their model counterparts. The first two parameters are the substitution elasticity ρ and the share parameter λ in the CES function that characterizes the surplus produced by a manager with ability a when she is assigned to a project with productivity q:

$$x(a + \alpha_t, q) = \left(\lambda \exp(a + \alpha_t)^{\frac{\rho-1}{\rho}} + (1-\lambda) q^{\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}}.$$

Recall that her true type is unknown. Instead, her belief follows a normal distribution with mean $\hat{a}$ and variance $\frac{\psi}{(\phi+r\psi)(\phi+(r+1)\psi)}$. While her belief at age r is a function of the *entire* sequence $\{\alpha_t\}_{t=0}^{r}$, we can conveniently characterize the evolution of $\hat{a}$ recursively as in (3). The variance, in contrast, is simply a function of r, the precision ϕ of the zero-mean managerial ability distribution, and the precision ψ of the normal innovations. We use these precisions to match two additional targets. The fixed matching cost, which managers must pay when they switch their assignment to a project with a different productivity, is denoted by μ. Note that exits from the assignment market do not entail a fixed cost, whereas entry does. We find this assumption to be quite natural. Finally, we assume that the productivities are drawn from a discretized unit-scale Generalized Pareto distribution with tail index $\xi > 0$. There are $L \in \mathbb{Z}^{++}$ distinct productivities. The corresponding CDF and PMF are given by:

$$\begin{aligned} \Gamma_\ell &= G(q_\ell; \xi) = 1 - (1 + \xi q_\ell)^{-\frac{1}{\xi}}, \text{ where } q_0 = 0 \\ \gamma_\ell &= \Gamma_\ell - \Gamma_{\ell-1}, \text{ for } \ell \in \{1, \ldots, L\} \end{aligned}$$

We assume that q_L is large but finite in the numerical exercise and we approximate the CDF with $G(q_L; \xi) = 1$ in our calibration.

The model is exactly identified and the six moments we aim to match are:

1. *Roberts' Law*, i.e. the elasticity of CEO pay with respect to firm size (employment or value added),

2. the tail index of the Generalized Pareto distribution of firm sizes (employment or value added),

3. the average CEO share of total output (match surplus),

4. the probability of CEO-to-CEO transitions,

5. the average lifecycle growth rate of CEO compensation, and
6. the propensity to participate in the assignment market of young CEOs (aged 31-35) relative to older ones (aged 46-50).

To compute individual spells we construct employment histories where we keep track of the assignments to deciles in the firm size distribution rather than firm assignments themselves. In the model, CEOs have no special attachment to a firm. What matters are the productivities and the assignment frictions apply to transitions between firms with *different* productivities. Firms are simply labels and classifying them in terms of their rank in the distribution of firm *sizes* is the natural empirical counterpart. We abstract from potential dynamics associated with individual productivities in the model. Unless ownership and management interact in ways that affect separation and matching, this simplification has no impact on the structural estimates.[18] In the data, spells are terminated when CEOs retire, when they switch from an employer in one decile to another employer in a different decile, or when their current employer moves to a higher or lower decile in the distribution of firm sizes without separation.[19] Through the lens of our model, the spell of a CEO is not terminated when she moves from one firm to another in the *same* decile and we classify turnovers in the data accordingly.

In our benchmark estimation, we use intangible capital to proxy for firm size. We target an elasticity of 27% between the CEOs' compensation and the empirical counterpart to the firms' payments to intangibles, which we compute as the perpetuity value of the firms' capitalization net of the (book) value of tangible assets. The corresponding estimates in Gabaix and Landier (2008) and Alder (forthcoming) are around 25% for firms included in the ExecuComp database of large US corporations. The point estimate for Denmark hardly varies when we use employment, another widely used proxy for firm size, instead.[20] However, since the model does not generate a distribution of non-managerial employment across firm productivities,

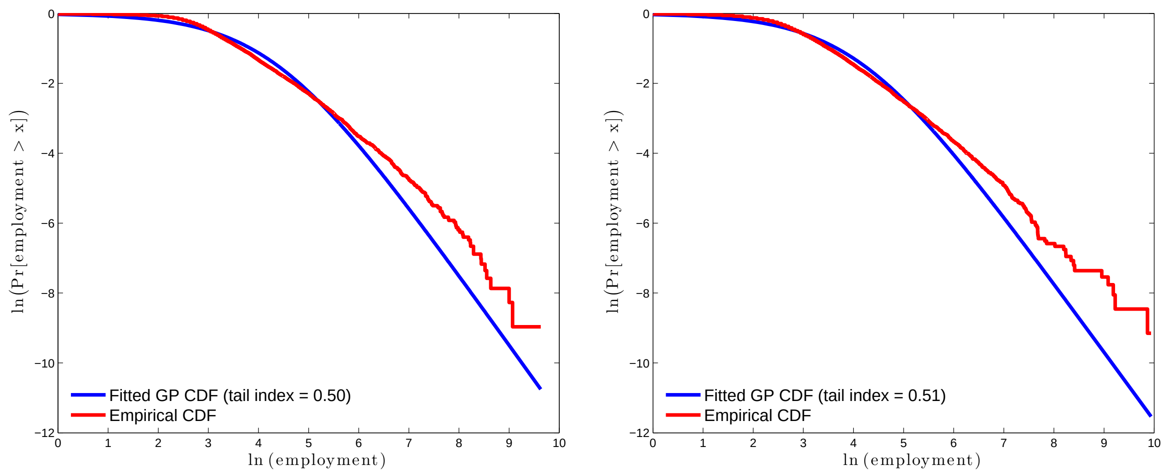
Roberts' Law is a description of the CEO's marginal contributions across the size distribution of firms. Next, we need to "anchor" the level of the CEOs' importance by computing her *relative* rather than marginal contribution to the match surplus.[21] In the data, the average

[18]This is an area for future research. A dynamic theory of entrepreneurship, for instance, requires that we think about the dynamics of productivities more carefully. This is part of an ongoing research agenda in collaboration with Moritz Meyer-ter-Vehn and Lee Ohanian (Alder et al., 2014).

[19]According to the model, the CEO incurs the fixed matching cost in this final instance, which is admittedly at odds with common sense. This is an issue that we need to address in our next round of revisions.

[20]In Gabaix and Landier (2008) and Alder (forthcoming), the elasticity with respect to employment is roughly 31% and thus slightly higher than the estimate for intangibles. We have not computed the elasticity for a comparable subsample of *large* Danish firms.

[21]Since the model features constant returns to scale in the aggregate our calibration strategy does not depend on the choice of numeraire. For this reason, there is no need to characterize the CEO's contribution in absolute



(a) Distribution in 2000

(b) Distribution in 2008

Figure 6: Danish Firm Sizes

share of the CEOs' compensation relative to value added is 13%. This number may seem exceedingly high, but it is important to keep in mind that our sample contains numerous small firms where the entrepreneur's or CEO's human capital contributes a sizeable fraction of total value added. If instead, we use the firms' intangible assets as a proxy for q and compute the CEO's share relative to the annuity value of intangibles, the estimate more than doubles to 33%. Since our model abstracts from non-managerial labor and capital inputs, this second moment is our preferred calibration target.

We compute the maximum likelihood estimate of the Generalized Pareto (GP) tail index using employment rather than value added. We drop all those firms for which we cannot identify a CEO and our point estimates are 0.50 in 2000 and 0.51 in 2004 and 2008. We target a value of .50 in our benchmark calibration. Figure 6 shows that the GP distribution is a good approximation of the size distribution of Danish firms with the usual caveat that it tends to underestimate the probability mass in the far right tail.[22] Compared to the US size distribution of firms or establishments, which is approximately unity, the Danish tail index is quite low (see, for instance Luttmer, 2004; Rossi-Hansberg and Wright, 2007). This discrepancy is an area in need of further investigation.

We follow Beaudry and Green (2000) to estimate the average annual growth rate of the CEOs' compensation in the sample. The idea is to isolate the forces behind growth in executive compensation that are distinct from wage growth in the economy more broadly. In our benchmark

terms. This feature has been labeled "scale invariance" in Alder (forthcoming).

[22] This is a standard observation in the literature on the distribution of firm and establishment sizes. If we include firms without known CEOs the estimate of the tail index rises to the 0.61-0.63 range.

calibration, we attribute this differential growth to the managers' incentive to leave an existing assignment (including exit, which entails taking up the exogenous outside option) or to seek out a new employer (including entry). While we do not claim that the occurrence and resolution of mismatch is the only source of wage growth we do believe that it is an important conduit. In our robustness check we explore an alternative mechanism whereby managers accumulate human capital through a learning-by-doing process in our robustness checks (to be completed). We will return to this later.

Our final parameter is the fixed matching cost μ. The data contain no direct evidence for the magnitude of this cost and we calibrate the parameter indirectly. It is precisely to the details of this strategy that we turn our attention to in the next section.

5.2 Strategy

How exactly do the moments generated by the parameterized model map into their empirical counterparts? While we review them individually in the order that they are introduced in section 5.1 and highlight the one-to-one mapping from parameters to moments as much as possible in order to provide some intuition for our strategy, it is important to realize that they are identified jointly in the numerical calibration. Unfortunately, the computational procedure is somewhat cumbersome and we are, for that reason, only able to provide some generic results at this stage.

We previously mentioned that Roberts' Law is a characterization of the manager's marginal appropriation of the match surplus in the *cross-section* of the size distribution of firms. To build some intuition for this "law", consider a version of the model with a single cohort (which implies that the model is essentially static), a continuous distribution of productivities (that is, $L \to \infty$), and a linear surplus technology with $\rho \to \infty$. While this linear case is uninteresting with respect to optimal sorting it has the advantage that the marginal products of managers and firms only depend on their own respective types. Clearly then, the average elasticity is simply a function of the relative magnitudes of ϕ and ξ. For a given precision ϕ, the elasticity of pay with respect to firm size is decreasing in ξ. Similarly, for a given ξ, the elasticity is decreasing in ϕ, the inverse of the variance of the zero-mean distribution of managerial abilities.[23] Once we allow for finite values of ρ the increasing differences property of $x(\cdot,\cdot)$ implies that the slope of the CEOs' cross-sectional pay profile is not just increasing in the variance of the ability distribution but also in the shape index ξ of the productivity

[23] Alder (forthcoming) calibrates a model where a and q follow Generalized Pareto distributions and $\psi \to \infty$ (that is, managers know their own abilities with certainty). In the version of that model where $\rho \to \infty$, it can be shown that the elasticity of managerial compensation with respect to firm size is a monotone increasing function of the ratio of the tail indices.

distribution. By the same token, the slope of the profit profile is increasing in both ϕ^{-1} and ξ. The tight relationship between the elasticity of managerial pay and the relative magnitudes of the two parameters governing the shape of the ability and productivity distributions is further perturbed when the economy is populated by agents belonging to multiple age cohorts ($S > 1$) with imprecise beliefs ($\psi < \infty$) and a discrete distribution of productivities ($L < \infty$). The basic intuition and the monotonicity of the relationship, however, still prevail in these more general cases and for given values of all other parameters we select the ϕ such that the elasticity in the model coincides with its empirical counterpart.

The distribution of firm sizes is a function of both the distribution of abilities and productivities. This relationship is particularly easy to grasp and we simply set ξ to generate a tail index of match surpluses equaling one half. These first two moments identify to a large extent – though not exclusively – the empirically relevant values of ξ and ϕ.

The parameter $\lambda \in [0, 1]$ pins down the CEOs' average share of the surplus mostly by governing the rate at which it declines from unity for the marginal firm-manager pair, for which the match surplus equals the manager's outside option regardless of λ in equilibrium, across the size distribution of firms. In the special case where $\lambda = 1$ and $\rho \to \infty$ one can easily show that all CEOs appropriate the entire match surplus. Conversely, CEOs earn the outside option w regardless of the productivity q they are assigned to when $\lambda = 0$ and that maximizes the rate of decline of her surplus share across the size distribution.

The precision of normal productivity shocks (or signal) is denoted by ψ. It determines, via equation (3), the relative importance of priors and innovations in the agents' Bayesian updates of their posterior mean beliefs. All else equal, a low ψ generates more volatile beliefs and hence triggers more frequent separations, entry and exit. Moreover, it governs the rate at which the distribution of mean beliefs converges toward the true mean-zero normal with variance ϕ^{-1}. But that it is not all. Since we exponentiate $a + \alpha_t$ in the surplus technology, the distribution of the CEO's contribution is effectively log-normal and hence skewed.[24] The distribution of mean beliefs interacts with the skewness of effective contributions and with the curvature parameter ρ to generate a rich pattern of participation and assignment across productivities. In our calibration strategy, we only target the 1.9 percent average probability of CEO-to-CEO transitions and use the additional moments to check for robustness.

Due to Bayesian learning, agents form more accurate and precise (i.e. lower variance) beliefs about their true ability as they age. Since agents are more likely to be mismatched early in their careers, the curvature on the technology that determines the match surplus generates "secular" growth in compensation that is distinct from any productivity process that my be

[24] Any convex transformation of normally distributed beliefs generate such skewness. See Bolton et al. (2013) for a more detailed discussion.

underlying wage growth outside the managerial assignment market (which would be crudely captured by exogenous growth of the outside option). The extent to which mismatch generates compensation growth depends on the degree of complementarity: all else equal, lower values of ρ generate higher growth rates simply because the cost of being mismatched is more severe and that is far more likely at the beginning of a manager's career compared to the end. In our benchmark parameterization, we match a 1.9 percent growth rate. In our robustness checks we allow for learning-by-doing to soak up some of the compensation growth, but we can rule out that human capital accumulation of this sort is the sole driver of compensation growth. In particular, it would have the counterfactual prediction that CEOs move across the size distribution only from left (small) to right (large). The data suggest that managers move both up and down over the course of their careers.

Finally, μ inhibits reassignments and hence can discourage participation and separation in the assignment market. The effect of this fixed cost, however is not uniform across the talent distribution. Relatively low ability CEOs find the cost more onerous compared to high-ability ones and since the distribution of beliefs is more compressed (around the mean of zero) in young cohorts compared to older ones, it tends to discourage participation by the former. It is worth noting that ψ and ρ generate fairly strong age-specific participation patterns but since we already use these two parameters to match other targets, we need one additional degree of freedom to simultaneously generate the average CEO-to-CEO transition probability, the compensation growth rate, and the hazard rate into managerial work (or entrepreneurship).

6 Counterfactual Experiments

TO BE COMPLETED.

7 Conclusion

TO BE COMPLETED.

References

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

ALDER, S. D. (forthcoming): "In the Wrong Hands: Complementarities, Resource Allocation, and Aggregate TFP," *American Economic Journal: Macroeconomics*.

ALDER, S. D., M. MEYER-TER VEHN, AND L. E. OHANIAN (2014): "Dynamic Sorting," Working Paper.

BEAUDRY, P. AND D. A. GREEN (2000): "Cohort patterns in Canadian earnings: assessing the role of skill premia in inequality trends," *Canadian Journal of Economics*, 33, 907–936.

BENNEDSEN, M., F. PÉREZ-GONZÁLEZ, AND D. WOLFENZON (2008): "Do CEOs Matter?" Working Paper.

BENNEDSEN, M., F. PEREZ-GONZALEZ, AND D. WOLFENZON (2012): "Evaluating the Impact of The Boss: Evidence from CEO Hospitalization Events," Working Paper.

BHATTACHARYA, D., N. GUNER, AND G. VENTURA (2013): "Distortions, Endogenous Managerial Skills and Productivity Differences," *Review of Economic Dynamics*, 16, 11–25.

BOLTON, P., M. K. BRUNNERMEIER, AND L. VELDKAMP (2013): "Leadership, Coordination and Corporate Culture," *Review of Economic Studies*.

EECKHOUT, J. AND P. KIRCHER (2011): "Identifying Sorting – In Theory," *Review of Economic Studies*, 78, 872–906.

EISFELDT, A. L. AND C. M. KUHNEN (2013): "CEO Turnover in a Competitive Assignment Framework," *Journal of Financial Economics*, 109, 351–372.

EVANS, D. S. AND B. JOVANOVIC (1989): "An estimated model of entrepreneurial choice under liquidity constraints," *Journal of Political Economy*, 97, 808–827.

EVANS, D. S. AND L. S. LEIGHTON (1989): "Some empirical aspects of entrepreneurship," *American Economic Review*, 79, 519–535.

GABAIX, X. AND A. LANDIER (2008): "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123, 49–100.

GROES, F., P. KIRCHER, AND I. MANOVSKII (2015): "The U-shapes of Occupational Mobility," *Review of Economic Studies*, 82, 659–692.

JOHNSON, W. R. (1978): "A theory of job shopping," *Quarterly Journal of Economics*, 92, 261–278.

JOVANOVIC, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87, 972–990.

——— (2015): "Learning and Recombination," *Working Paper*, 1–34.

KNIGHT, F. H. (1921): *Risk, Uncertainty, and Profit*, New York: Houghton Mifflin.

LISE, J., C. MEGHIR, AND J.-M. ROBIN (2013): "Mismatch, Sorting and Wage Dynamics," *Working Paper*.

LOPES DE MELO, R. (2013): "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," *Working Paper*, 1–39.

LUCAS, R. E. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523.

LUTTMER, E. G. J. (2004): "The Size Distribution of Firms in an Economy with Fixed and Entry Costs," *Working Paper*, 1–36.

MILLER, R. A. (1984): "Job matching and occupational choice," *Journal of Political Economy*, 92, 1086–1120.

NAGYPAL, E. (2007): "Learning by doing vs. learning about match quality: Can we tell them apart?" *Review of Economic Studies*, 74, 537–566.

ROSSI-HANSBERG, E. AND M. L. WRIGHT (2007): "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*.

TERVIÖ, M. (2008): "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, 98, 642–68.