CEO Behavior and Firm Performance*

Oriana Bandiera
London School of Economics

Stephen Hansen
University of Oxford

Andrea Prat
Columbia University

Raffaella Sadun
Harvard University

September 10, 2016

****PRELIMINARY AND INCOMPLETE****

Abstract

We measure the behavior of over 1,100 CEOs in six countries (Brazil, France, Germany, India, UK and US) using a new methodology that combines (i) survey data on every activity the CEOs undertake in a random work-week and (ii) a machine learning algorithm that projects these high dimensional data onto one behavior index. A simple firm-CEO assignment model yields the null hypothesis that, in absence of frictions, CEO behavior is uncorrelated with firm performance. Combining the CEO behavior index with firm level accounting data we reject this null: the correlation between CEO behavior and firm performance is large and precisely estimated. The fact that it appears only after the CEO is appointed casts doubt on the relevance of unobservable firm traits. Structural calibration rules out that the same behavior is optimal for all firms but that, nevertheless, 17% of firms get the "wrong" CEO. These mismatches are more frequent in poorer countries (45% vs 10%) and account for 12% of the productivity gap between rich and middle-income countries.

*This project was funded by Columbia Business School, Harvard Business School and the Kauffman Foundation. We are grateful to Morten Bennedsen, Wouter Dessein, Bob Gibbons, Rebecca Henderson, Ben Hermalin, Paul Ingram, Amit Khandelwal, Michael McMahon, Antoinette Schoar, Steve Tadelis and seminar participants at Asian Band Development Institute, Bocconi, Cattolica di Milano, Chicago, Columbia, Copenhagen Business School, Cornell, the Economics of Organization Workshop, Harvard Business School, Insead, LSE, MIT, NBER, Politecnico di Milano, Princeton, Science Po, SIOE, Stanford Management Conference, Tel Aviv, Tokyo, Toronto, Uppsala, Warwick for useful suggestions.

1 Introduction

The impact of CEOs on firm performance is at the core of many economic debates. The conventional wisdom, backed by a growing body of empirical evidence (Bertrand and Schoar 2003, Bennedsen et al. 2007, Kaplan et al. 2012) is that the identity of the CEO matters for firm performance. But what do CEOs actually do? The key role of the firm top executive is to co-ordinate the activities performed by different individuals (Coase 1937). Yet co-ordination can be achieved in different ways, so do different CEOs perform this role differently? And is there a "best practice", or do different circumstances call for different behaviors?

In this paper we develop a new methodology to measure CEO behavior in large samples combining (i) a survey that records each activity the CEOs undertake in a random work-week, and (ii) a machine learning algorithm that projects the several dimensions of CEO behavior onto a low-dimensional behavior index. We use this data to study the correlation between CEO behavior and firm performance within the framework of a firm-CEO assignment model, and show that the data is consistent with the hypothesis that CEOs differentiate horizontally, that is we rule out the "best practice" hypothesis, but assignment is not perfect thus some firms are allocated the wrong CEO. We calibrate the productivity loss due to wrong assignments and show that it is substantial.

Our survey methodology is inspired by the classic study by Mintzberg (1973), who shadowed five CEOs over the course of one week. We scale up this methodology by calling the CEOs or their PAs to record the CEOs' diaries rather than shadowing individuals directly. This approach allows us to collect detailed and comparable data on the behavior of 1,114 CEOs of manufacturing firms in six countries: Brazil, France, Germany, India, UK and the US. To do so, we record the same five features for each activity that involves the CEOs interactions with other people: its type (e.g. meeting, site visits etc.), planning horizon, number of participants involved, number of different functions, and the participants' function (e.g. finance, marketing, clients, suppliers, etc.). Overall, we collected data on 42,233 activities covering an average of 50 working hours per CEO.[1] Each of these activities can take one of the 4,253 combinations of the five features described above.

We use an unsupervised Bayesian machine learning algorithm, Latent Dirichlet Allocation (Blei et al. 2003) to project this high-dimensional feature space onto a lower-dimensional behavior space in a non-subjective fashion.[2] We begin by estimating "pure" prototype

[1]In earlier work (Bandiera et al., 2013) we use the same data to measure the CEOs' labor supply and assess whether and how it correlates with differences in corporate governance (and in particular whether the firm was led by a family CEO).

[2]The typical application of LDA is to natural language, where it is widely cited (for an example in

CEO behaviors as probability vectors over the activities. We then estimate a CEO-specific behavior index as the distribution over the prototype behaviors - namely we allow, but do not force, each CEO to have a different mix of behaviors. In our baseline approach we allow for the lowest possible number of pure behaviors, two, and estimate a uni-dimensional behavioral index that ranges from 0 (for CEOs who follow a pure behavior 0) to 1 (for CEOs who follow a pure behavior 1). The two behaviors differ considerably: the activities that are most frequent in behavior 0 are least frequent in behavior 1, and vice versa. Low values of the CEO behavior index are associated with direct monitoring of production, low levels of planning, and one-on-one meetings. In contrast, high values are associated with high levels of planning, and large meetings that involve the upper levels of management (typically C-suite).

While the diary data reveal that different CEOs behave differently, there is no theoretical reason to expect either type of behavior to lead to better performance, or to be more or less costly for individual CEOs. To the contrary, the fact that different behaviors coexist suggests that they might be best responses to different circumstances faced by the firm. To illustrate this and its implications for firm performance, we develop a firm-CEO assignment model with two types of firms and two types of CEOs. Firm type determines which CEO behavior is most productive given its specific features, while CEO type determines the cost of adopting a certain behavior. The pool of potential CEOs is larger than the pool of firms seeking a CEO, and one type of CEO is relatively more abundant than the other, i.e. its share is higher than the share of firms who seek CEOs of that type. The model thus nests the pure vertical differentiation case, in which all firms seek the same type of CEO.We allow for two types of frictions in the market for CEOs. First, the screening technology is imperfect, so that it cannot always correctly identify the actual CEO type. Second, after hiring the CEO, the firm can offer him incentives to adopt the behavior that is most suitable for the firm, but these are limited due to poor governance or labor laws that make dismissals costly.

We show that, if frictions are small, all firms will hire CEOs of the right type, and that these will adopt the behavior that is optimal for the firm. Therefore, in these circumstances – once we control for firm type – the correlation between CEO behavior and firm performance will be zero. This is the null hypothesis we test. In alternative, if frictions are large enough, in equilibrium some of the abundant type CEOs will match with the wrong type of firms. In this case, we will observe some variation in CEO type within the same broad type of

economics, see Hansen et al. 2014). It is less commonly used for survey data, but in practice it is able to usefully reduce the dimensionality of any dataset of counts.

firms, the mismatched firms will perform worse, and – again controlling for firm type – the correlation between having a CEO of the scarce type and firm performance would be positive.

Guided by the model, we combine our estimated CEO behavior index with firm level accounting data. Using the set of 920 firms (82% of the CEO sample) for which accounting data is available, we find that high values of the CEO behavior index are significantly correlated with higher firm productivity, a key metric of firm performance (Syverson, 2011). A standard deviation increase in the CEO behavior index is associated with a 0.12 log points increase in productivity, which is about 15% of the increase associated with a standard deviation increase in capital. Our identifying assumption, transparent in the model, that firm traits that determine which CEO behavior is optimal are orthogonal to determinants of firm productivity that are not observed in our data. If this assumption fails, unobservable firm traits that lead to low productivity might lead to the appointment of a type of CEO that prefers low index behavior, or make the CEO adopt a low index behavior. To allay this concern, we use a within-firm estimator that exploits accounting data for the period before and after the current CEO was appointed. If indeed firm's unobservables drove the earlier estimates, the fixed effects will account for these and we will find no difference between productivity before and after the appointment of the current CEO. In contrast, we find that the correlation between CEOs behavior and firm performance materializes only after the CEO appointment, and more specifically only four years after the CEO appointment. This approach also allows us to verify that the correlation between CEO behavior and firm productivity are not driven by different trends in performance across firms before the appointment of the current CEO. While these tests do not rule out all factors that can invalidate our identifying assumption, we argue that those that cannot be ruled out require implausible assumptions on the determinants and timing of CEO turnover.

The model makes precise that these results are consistent with two interpretations: either all firms prefer high index CEOs or some prefer low index CEOs but frictions are sufficiently large to create some mismatches between firms and CEOs, and that the (unobserved) CEO type that leads to high values of the CEO behavior index is relatively scarce in the population.

To provide evidence on the interpretation that best fits the data, we estimate hte model structurally to derive the share of firms that demand high index CEOs and the share of firms that get the "wrong" CEO. We can reject that all firms prefer a high index CEO, yet 17% of those that do end up with a low index CEO. In other words, our estimates indicate that while low index CEOs are optimal for some firms, their supply over strips demand and frictions in the market for CEOs must be sufficiently large that some of them end up working in firms that would have been more productive with the other type.

To conclude, we exploit the fact that our sample covers countries at different levels of development to show that these "bad" matches are more frequent in poorer countries (45% vs 10%) where frictions are presumably larger. These differences imply that the misallocation of CEOs to firms accounts for 12% of the labor productivity gap (XX% of the TFP gap) between rich and poor countries.

This study contributes a new method to measure CEO behavior in large samples and evidence on the link between CEO behavior and firm performance. The management literature contains some examples of time use analyses, but typically on much smaller samples and for managers on lower rungs of the hierarchy.[3] In economics, our findings are complementary to the literature that studies the correlation between CEO *traits* and firm performance. Kaplan et al. (2012) and Kaplan and Sorensen (2016) have detailed data on skills and personality traits of several CEOs candidates; they show the CEOs mostly differ along three dimensions: managerial talent, execution skills and interpersonal skills. Of these, only talent and execution skills correlate with firm performance but interpersonal skills increase the likelihood that the candidate is hired.[4] This is consistent with our assumption that screening is imperfect and firms can end up hiring the wrong CEOs. Our methodology is complementary to Mullins and Schoar (2013), who use self-reported survey questions to measure the management style and values of 800 CEOs in emerging economies. Their focus however differs from ours, as they aim to explain variation in style and values rather than the link with performance. This paper is complementary to a growing literature documenting the role of managers and management processes on firm performance (Bloom and Van Reenen (2007) and Bloom et al. (2016)). The relationship between CEO behavior and firm performance that we identify is of the same order of magnitude as the effect of management practices. Furthermore, for a subset of our firms we have both CEO behavior data and management scores (measured at middle managerial levels) and we are able to check that both variables retain independent explanatory power, thus suggesting that these might reflect two distinct channels through which managerial choices influence firm performance. Finally, we share Lippi and Schivardi's (2014) objective to quantify the output reduction caused by distortions in the allocation of

[3] The largest shadowing exercise on top executives known to us – Kotter (1999) – includes 15 general managers, not CEOs. The largest time use study of managerial personnel we are aware of is Luthans (1988), which covers 44 mostly middle managers. Some professional surveys ask large numbers of CEOs general questions about their aggregate time use (e.g. McKinsey 2013), but they do not collect detailed calendar information and do not study the correlation between CEO behavior and firm performance.

[4] Malmendier and Tate (2005) and Malmendier and Tate (2009) focus on overconfidence; they find that this is correlated with higher investment–cash flow sensitivity and mergers that destroy value.

managerial talent.

The paper is organized as follows. Section 2 describes the data and the machine learning algorithm that yields the CEO behavior index. Section 3 presents the assignment model, which is then used to inform the empirical analysis in section 4. Section 5 quantifies the share of mismatches and their consequences for productivity differentials across high and middle income regions. Section 6 concludes.

2 Measuring CEO Behavior

2.1 Sample

The survey covers CEOs in six of the world's ten largest economies: Brazil, France, Germany, India, the United Kingdom and the United States. For comparability, we chose to focus on established market economies and opted for a balance between high and middle-low income countries. While titles may differ across countries (e.g. Managing Director in the UK) we always interview the highest-ranking authority in charge of the organization who has executive powers and reports to the board of directors. For brevity, we refer to them as CEOs in what follows.

Our sampling frame was drawn from ORBIS, restricted to manufacturing to maintain comparability, and contained 6,527 eligible firms in 32 two-digit SIC industries.[5] Of these, 1114 (17%) participated in the survey. [6] Table A1 shows that sample firms have on average slightly lower log sales (coefficient 0.071, standard error 0.011) but we do not find any significant selection effect on performance variables, such as labor productivity (sales over employees) and return on capital employed (ROCE) (see Appendix X for details).

Table 2.1 shows descriptive statistics on the sample CEOs and their firms. Sample CEOs are 52 years old on average, nearly all (96%) are male and have a college degree (92%). About half of them have an MBA and a similar share has studied abroad. The average

[5]Among firms in this sector we selected those with available sales and employment data, yielding 11,500 potential sample firms. We could find CEOs contact details for 7,744 firms and, of these, 1,217 later resulted not to be eligible.The reasons for non eligibility included recent bankruptcy or the company's not being in manufacturing. 310 of the 1217 could not be contacted to verify eligibility before the project ended.

[6]This figure is at the higher end of response rates for CEO surveys, which range between 9% and 16% (Graham et al 2011). Our final sample thus comprises of 1,114 CEOs, of which 282 are in Brazil, 115 in France, 125 in Germany, 356 in India, 87 in the UK and 149 in the US. 1,131 CEOs agreed to participate but 17 dropped out before the end of the data collection week for personal reasons.

Table 1: Summary Statistics

Variable	Mean	Median	Standard Deviation	Observations
A. CEOs Traits				
CEO age	50.93	52.00	8.45	1107
CEO gender	0.96	1.00	0.19	1114
CEO has college degree	0.92	1.00	0.27	1114
CEO has MBA	0.55	1.00	0.50	1114
CEO has studied abroad	0.48	0.00	0.50	1114
CEO tenure in post	10.29	7.00	9.55	1110
CEO tenure in firm	17.10	16.00	11.58	1108
CEO belongs to the owning family	0.41	0.00	0.49	1114
B. Firms Traits				
Employment	1275.5	300.0	6497.7	1114
Sales ('000 $)	222033.9	35340.5	1526261.0	920
Capital ('000 $)	79436.7	10029.0	488953.6	618
Materials ('000 $)	157287.1	25560.0	1396475.0	448
Profits per employee ('000 $)	8.6	2.5	14.9	386
C. Regional Traits				
Log Regional Income per Capita	9.36	9.48	1.08	1111

Notes: Variables in Panel A and B are drawn from our survey and ORBIS, respectively. In Panel C, Log regional income per capita in current purchasing-power-parity (PPP) dollars is drawn from Gennaioli et al (2013).

tenure is 10 years, with a standard deviation of 9.6; the heterogeneity is mostly due to the distinction between family and professional CEOs as the former have much longer tenures.[7]

2.2 The Executive Time Use Survey

2.2.1 Data collection

The data were collected by a team of enumerators through daily phone calls with the personal assistant (PA) of the CEO, or with the CEO himself (43% of the cases), over a week randomly chosen by us.[8] On day one of this week (typically a Monday), the enumerator called in the

[7]In our sample 57% of the firms are owned by a family, 23% by disperse shareholders, 9% by private individuals, and 7% by private equity. Ownership data is collected in interviews with the CEOs and independently checked using several Internet sources, information provided on the company website and supplemental phone interviews. We define a firm to be owned by an entity if this controls at least 25.01% of the shares; if no single entity owns at least 25.01% of the share the firm is labeled as "Dispersed shareholder".

[8]The data collection methodology discussed in this section is an evolution of the approach followed in Bandiera et al. (2012) to collect data on the agenda of 100 Italian CEOs. While the data collection of the

morning and gathered detailed information on all the activities planned in the CEO diary for the day. The enumerator then called again in the evening, to gather information on the actual activities undertaken during the day (including those that were not originally planned), and the activities planned for the following day. On subsequent days, the enumerator called in the evening, again to collect data on the actual activities undertaken during the day, and the activities planned for the next day.[9]

Figure A.2 shows a screen-shot of the survey tool.[10] The survey collects information on all the activities lasting longer than 15 minutes in the order they occurred during the day. To avoid under (over) weighting long (short) activities we reshape the data so that the unit of analysis is a 15-minute time block.

Overall we collect data on 42,233 activities of different duration, equivalent to 225,721 15-minute blocks, 90% of which cover work activities. Of these, 127,660 (or 62.6% of total time), involve at least one other person and can be classified as the "co-ordination activities" that are the focus of this study.[11] The average CEO has 202 15-minute time blocks, adding up to 50 hours per week.

2.2.2 Feature description and combinations

For each activity we collect information on the following features: (1) type (e.g. meeting, public event, etc.); (2) duration (15m, 30m, etc.); (3) whether planned or unplanned; (4) number of participants; (5) functions of participants, divided between employees of the firms or "insiders" (finance, marketing, etc.) and "outsiders" (clients, banks, etc.).

Table 2, panel A shows summary statistics on these features. Within the "type" feature, the most frequent entry is "meeting", which accounts for 74.1% of time. Furthermore, 64% of time is spent in activities lasting more than 1 hour, and 75% of time in activities that are planned in advance, while 62% of activities include more than one other participant beyond

Italian data was outsourced to a private firm, the data collection described in this paper was internally managed from beginning to end. Due to this basic methodological difference and other changes introduced after the Italian data was collected (e.g. the vector of features used to characterize every activity) we decided not to combine the two samples.

[9]For 70% of the CEOs in our sample, the work week consisted of 5 days. The remaining 30% of the CEOs also reported to work during the weekend (21% for 6 days and 9% for 7 days). Analysts were instructed to call the CEO after the weekend to retrieve data on Saturdays and Sundays. On the last day of the data collection, the analysts also interviewed the CEO to validate the activity data (if collected through his PA) and to collect information on the characteristics of the CEO and of the firm.

[10]The survey tool can also be found online on `www.executivetimeuse.org`.

[11]The non-work activities cover personal and family time during business hours. The working time not spent with others is divided between working alone (20%), travelling (9.3%), and sending emails (5.3%); for robustness, we show that the our results are not sensitive to including these activities.

the CEO. Table 2, panel B shows the time the average CEO spends with different functions. Perhaps unsurprisingly, given that we are working with a sample of manufacturing firms, the average CEO is most likely to spend time with employees involved in production. CEOs also spend more time with inside than outside functions. Functions are not mutually exclusive, and CEOs can spend time with more than one function in a single activity; in 39.5% of activities there is more than one function present.

While Table 2 shows average behavior, the data features substantial heterogeneity across CEOs. For example, while the average CEO spends 75% of time in planned activities, the 25th and 75th percentiles are 64% and 91%, respectively. The corresponding percentiles for time spent with production are 19% and 51%.

In order to fully describe each 15-minute block of CEO time, we combine all the features into a single overall variable, which from now on we refer to as an *activity*. More specifically, we define each activity according to the five distinct features described above (type of activity, duration, planning horizon, number of participants, type of functions involved). Using this approach, we obtain 4,253 unique combinations in the data.[12] For example, the most frequent combination is a meeting that lasts longer than 1 hour, is planned in advance and has two participants, both from the production division. There are also many rare activities; we drop those that appear in fewer than 30 CEOs' time use since they do not appear often enough to shed light on general management behaviors. This leaves 654 combinations and 98,347 time blocks in the baseline analysis. Table A.1 in the appendix shows average CEO time shares across features on this subsample, which are very similar to those of the whole sample reported in table 2.

2.3 Projecting high dimensional data onto a CEO behavior index

Our challenge is to reduce the dimensionality of the data to incorporate into standard econometric analysis. In the absence of theories about CEO behavior, we adopt a data-drive approach to identify the key dimensions along which CEO differ.

The algorithm we use for dimensionality reduction is latent Dirichlet allocation (LDA) (Blei et al., 2003), an unsupervised machine-learing algorithm for discrete data.[13] In our

[12]In all cases, the value of the first four features is unique, while the value of the last feature—the functions present in the activity—is a set that contains one or more elements.

[13]An alternative approach would be to use a supervised learning algorithm that used variation in time use to directly predict firm performance. This would "force" the data to explain performance. Instead, we adopt a two-step approach in which we first identify the primary dimensions along which CEOs differ in their time use, and then examine whether variation in these dimensions explains differences in firm performance.

Table 2: Average Time Shares for all CEOs

(a) Distribution of time within features

Type		**Duration**		**Planned**		**Participants**	
value	share	value	share	value	share	value	share
meeting	0.741	1hr+	0.642	planned	0.754	size2+	0.62
business_meal	0.07	1hr	0.198	unplanned	0.244	size1	0.362
phone_call	0.06	30m	0.138	missing	0.002	missing	0.018
site_visit	0.059	15m	0.022				
conference_call	0.033						
public_event	0.02						
workrelated_leisure	0.011						
video_conference	0.005						
other	0.0						

(b) Distribution of time across functions

Inside Functions		**Outside Functions**	
function	share	function	share
production	0.354	clients	0.108
mkting	0.224	suppliers	0.069
finance	0.173	others	0.059
hr	0.082	associations	0.036
groupcom	0.081	consultants	0.035
bunits	0.055	govoff	0.023
other	0.049	compts	0.02
board	0.043	banks	0.018
admin	0.042	lawyers	0.015
cao	0.036	pemployee	0.015
coo	0.03	investors	0.014
strategy	0.022		
legal	0.018		

Notes: The top table shows the amount of time the average CEO spends on different options within features for the 127,660 interactive 15-minute unit of time in the data. The bottom table shows the amount of time the average CEO spends with different functions. Since there are typically multiple functions in a single activity, these shares sum to more than one.

specific application, observations are activities with specific features conducted by different CEOs. LDA posits that the actual behavior of each CEO is a mixture of a small number of "pure" CEO behaviors, and that the creation of each activity is attributable to one of these pure behaviors.

To be more concrete, suppose all CEOs have A possible ways of organizing each unit of their time, which we define for short activities, and let x_a be a particular activity. Let $X \equiv \{x_1, \ldots, x_A\}$ be the set of activities. A *pure behavior* k is a probability distribution $\boldsymbol{\beta}^k$ over X that is common to all CEOs. That is, every CEO who adopts behavior k draws elements from X according to the same distribution $\boldsymbol{\beta}^k$. The ath element of the vector $\boldsymbol{\beta}^k$, namely β_a^k, gives the probability of generating x_a when adopting behavior k. All behaviors are potentially associated with all elements of X: it may be that, for the same activity a, both $\beta_a^{k'} > 0$ and $\beta_a^{k''} > 0$ for two distinct behaviors k' and k''. Importantly, the model allows for arbitrary covariance patterns among features of different activities. For example, one behavior may be characterized by large meetings whenever the finance function is involved but small meetings whenever marketing is involved.

LDA can also be understood in terms of a two-step generative process for observed activities, which is illustrated in figure 1 for the case of two pure behaviors. The *behavior* of CEO i is given by a mixture of the two pure behaviors according to weight $\theta_i \in [0,1]$. So, the probability that CEO i assigns to activity x_a is $\chi_a^i \equiv (1-\theta_i)\beta_a^0 + \theta_i \beta_a^1$. We refer to the weight θ_i as the *behavior index* of CEO i.[14] For each activity of CEO i, one of the two pure behaviors is drawn independently from θ_i. Then, given the pure behavior, an activity is drawn according to its associated distribution (either $\boldsymbol{\beta}^0$ or $\boldsymbol{\beta}^1$). Each CEO is represented on a low-dimensional latent space of pure behaviors, which are themselves linked to the high-dimensional observed space of activities through the $\boldsymbol{\beta}$ distributions.

In our baseline specification, we focus on the simplest possible case in which there exist only two possible pure behaviors: $\boldsymbol{\beta}^0$ and $\boldsymbol{\beta}^1$. This choice—while most likely an underestimate of the actual number of different pure behaviors CEOs can choose from—allows us to have the most parsimonious description of heterogeneity in CEO behavior. Note that, however, this does not imply that CEOs can only take one of two types, rather, the probability that CEO i generates activity a can lie anywhere between β_a^0 and β_a^1. Even if two pure behaviors are very distinct, some CEOs can have very similar, somewhat different, or very different probability distributions over activities. In contrast, in a traditional clustering

[14] Since we are working with only two pure behaviors, this is a one-dimensional index. This approach can be extended to K rather than two pure behaviors, in which case the behavioral index becomes a point on a $K-1$-dimensional simplex.

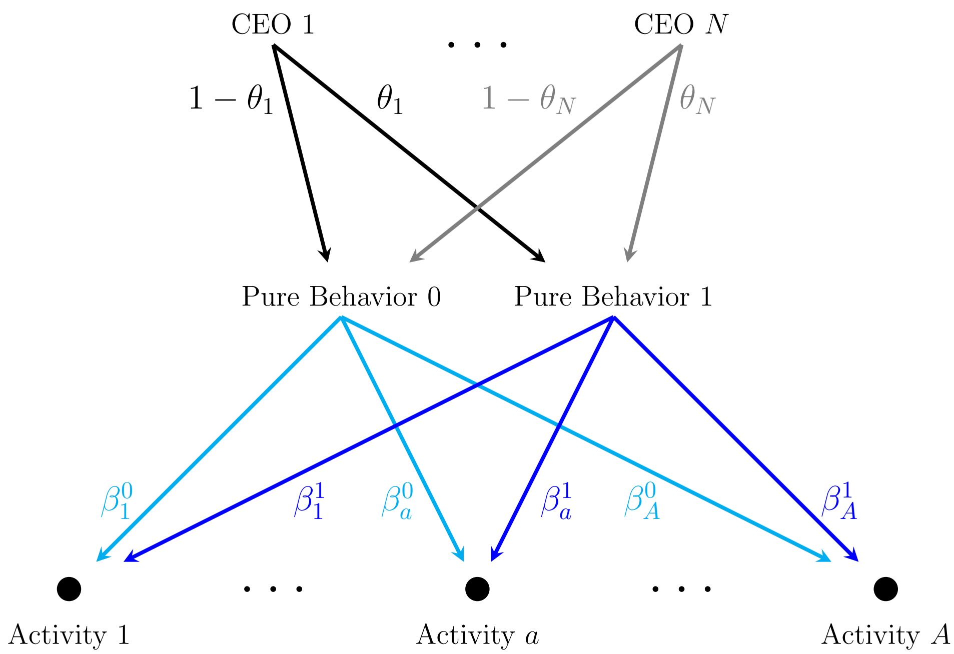


Figure 1: Data-Generating Process for Activities with Two Pure Behaviors

model, each CEO would be associated with one of the two pure behaviors, which corresponds to restricting $\theta_i \in \{0, 1\}$.

An alternative approach is to choose the number of pure behaviors according to a statistical criterion. Accordingly, we have also used a cross-validation approach in which we estimate LDA on a subset of the data, and then use the estimated parameters to predict the held-out data. A model with ten pure behaviors is best at prediction, so we also present results with this model in the appendix. We choose the model with two pure behaviors as the baseline because its output is easier to interpret.[15]

If we let $n_{i,a}$ be the number of times activity a appears in the time use of CEO i, then by independence the likelihood function for the model is simply $\prod_i \prod_a \chi_i^{n_{i,a}}$. The independence assumption of time blocks within a CEO may appear strong, since one might imagine that CEO behavior is persistent across a day or week. However, our goal in this initial application of machine learning methods in economics is to understand overall patterns CEO behavior rather than issues such as the evolution of behavior over time, or other more complex dependencies. These are of course interesting, but outside the scope of the paper.

Finally, while we have opted for LDA other simpler techniques like principal components

[15]The tradeoff between interpretability (which favors a small number of pure behaviors) and goodness-of-fit (which favors a greater number) is well known in the unsupervised learning literature. See, for example, Chang et al. (2009).

analysis (PCA) or k-means clustering are also possible. The advantage of LDA is that it is a generative model that provides a complete probabilistic description of time-use patterns linked explicitly to statistical parameters. In this sense, LDA is akin to structural estimation in econometrics, and allows for a transparent interpretation of the estimated parameters. In contrast, neither PCA or clustering estimates the parameters of a statistical model,[16] which can make interpreting their output difficult.

2.4 Estimation Results

While in principle one can attempt to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ via direct maximum likelihood or the EM algorithm, in practice the model is intractable due to the large number of parameters that need to be estimated (and which grow linearly in the number of observations). LDA overcomes this challenge by adopting a Bayesian approach, and placing Dirichlet priors on the $\boldsymbol{\beta}$ and θ terms. For posteriors we follow the Markov Chain Monte Carlo (MCMC) approach of Griffiths and Steyvers (2004). Details of these procedures are presented in the Appendix. Here we discuss the estimated object of interest, which are the two estimated pure behaviors $\widehat{\boldsymbol{\beta}}^0$ and $\widehat{\boldsymbol{\beta}}^1$, as well as the estimated behavioral indices $\widehat{\boldsymbol{\theta}}_i$ for every CEO $i = 1, \ldots, N$.

2.4.1 Pure behaviors

To illustrate differences in estimated pure behaviors, in figure2 we order the elements of X according to their estimated probability in $\widehat{\boldsymbol{\beta}}^0$ and then plot the estimated probabilities of each element of X in both behaviors. The figure shows that the combinations that are most likely in pure behavior 0 have low probability in pure behavior 1 and vice versa. To construct a formal test we simulate data by drawing an activity for each time block in the data from a probability vector that matches the raw empirical frequency of activities. We then use this simulated data to estimate the LDA model with two pure behaviors from our baseline analysis. We perform 1,000 simulations, and store the estimated pure behaviors and behavior indices for each one. A standard distance metric for probability distributions is Hellinger distance, which lies in $[0, 1]$. The Hellinger distance between the distributions in figure 2 is 0.776. The maximum Hellinger distance in the simulated data is 0.412, while the mean is 0.327 and the minimum is 0.176. Hence we conclude that the observed differences

[16] PCA performs an eigenvalue decomposition of the variance-covariance matrix, while k-means solves for centroids with the smallest squared distance from the observations.

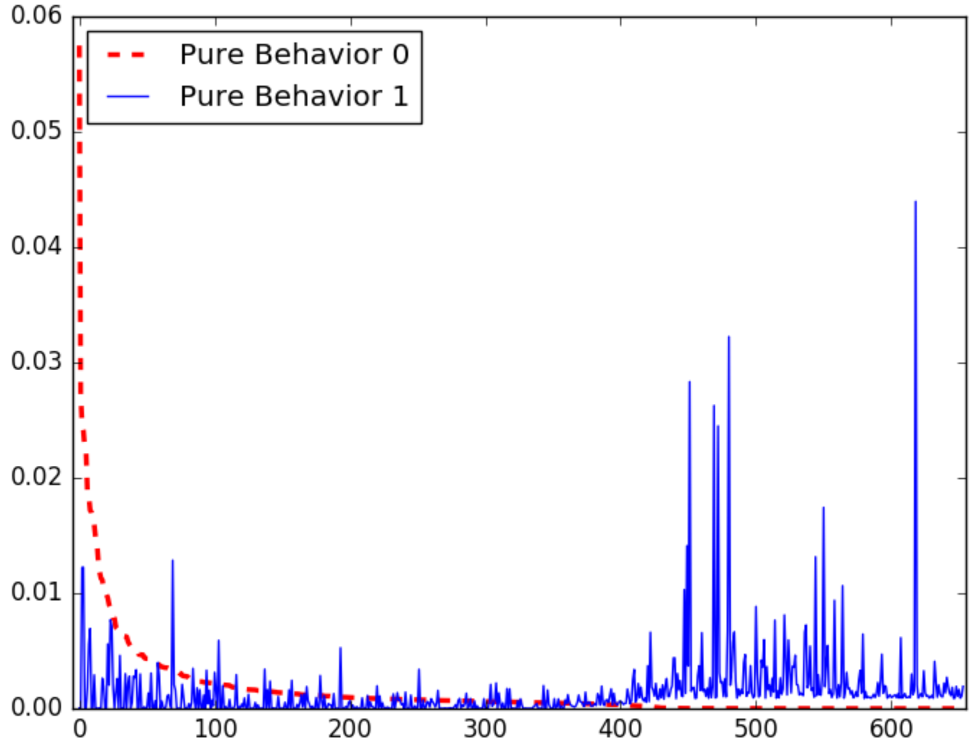


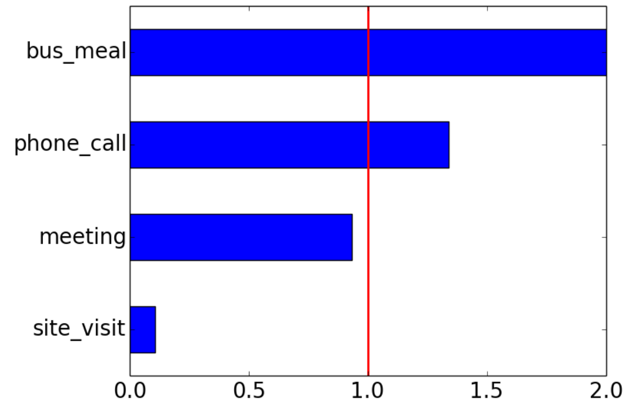
Figure 2: Probabilities of Activities in Estimated Pure Behaviors
Notes: The red dotted line plots the estimated probabilities of different activities in pure behavior 0, the blue solid line plots the estimated probabilities of different activities in pure behavior 1. The 654 different activities are ordered left to right in descending order of their estimated probability in pure behavior 0.

in pure behaviors are not consistent with their being no actual underlying heterogeneity in CEO time use.

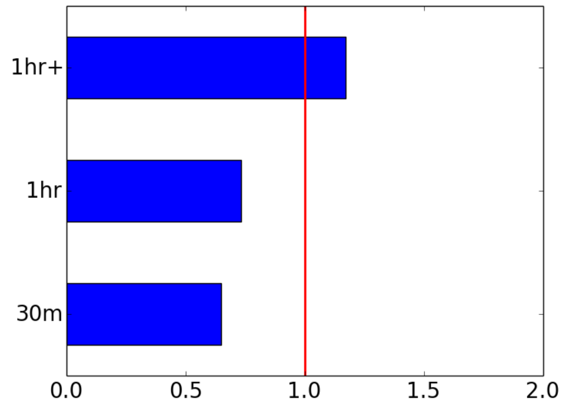
To aid the interpretation of pure behaviors we compute marginal distributions over individual features. Figure 3 displays the ratios of all the marginal distributions that we compute. We have constructed simulated standard errors for the differences in probabilities of each feature reported in the figure based on draws from the Markov chains used to estimated the reported means. All differences are highly significant, except for time spent with insiders alone.

A value of 1 for the ratio indicates that both pure behaviors place the same probability on the feature category; a value greater than (less than) 1 indicates a higher (lower) probability for behavior 1. Finally, where bars extend to the edges of the figure, we have truncated the ratio for visual coherence.

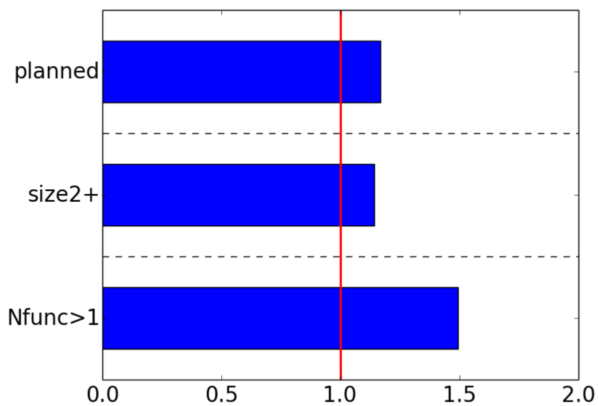
For activity types the most prominent distinction is site visits—i.e. the CEO physically



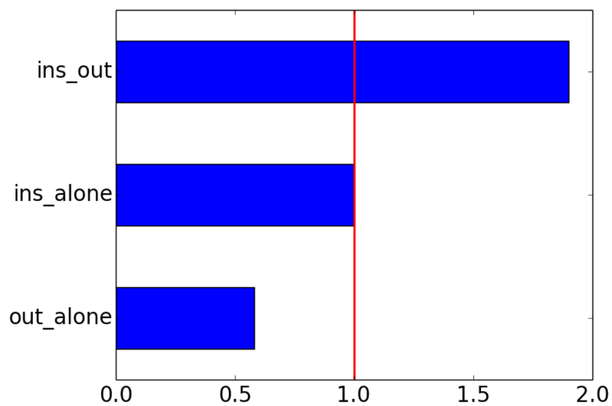
(a) Activity Type



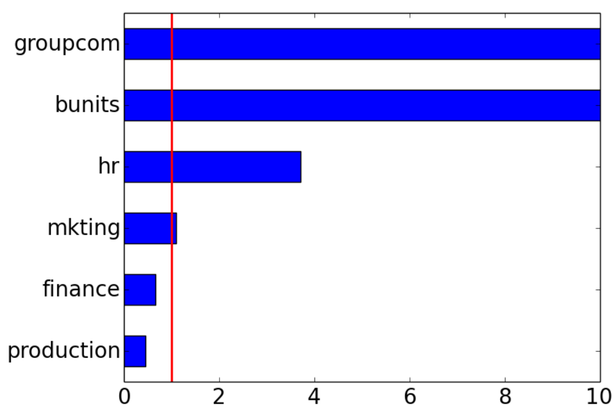
(b) Duration



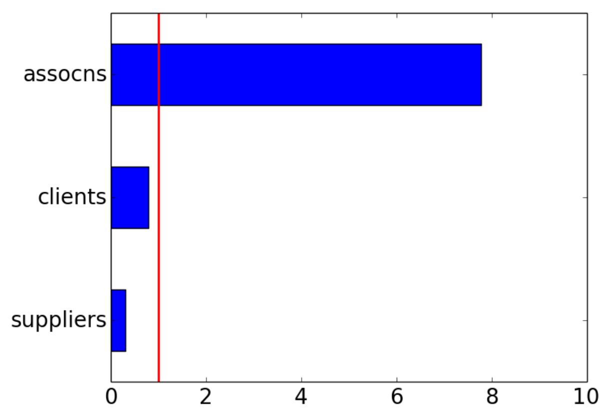
(c) Planning; Size; Number of Functions



(d) Inside vs Outside Time



(e) Inside Functions



(f) Outside Functions

Figure 3: Ratios of Marginal Distributions

Notes: We generate these figures in two steps. First, we create marginal distributions for each behavior along several dimensions. Then, for each category that has more than 5 per cent probability in either behavior, we report the probability of the category in behavior 1 over the probability in behavior 0. The third panel represents three separate marginal distributions. Each has two categories, so we report the ratio for only one of them.

visiting the manufacturing shop floor—which is ten times more likely in pure behavior 0. Another notable difference is for business meals, which pure behavior 1 is over twice as likely to generate. Less prominent differences exist for phone calls, which are 34% more likely in pure behavior 1 and meetings, which are 7% more likely in pure behavior 0.

For meeting duration, pure behavior 0 is clearly more associated with shorter engagements, with 30-minute durations 54% more likely and 1-hr durations 36% more likely. In contrast, pure behavior 1 is 17% more likely to generate activities that last more than one hour. Pure behavior 1 is also more likely to engage in planned activities (17% more likely); activities with two or more participants (14% more likely); and especially activities with two or more functions (50% more likely).

While both pure behaviors spend time in activities with only inside functions in equal amounts, pure behavior 1 is twice as likely to spend time with both inside and outside functions together, and pure behavior 0 is twice as likely to spend time with only outside functions. Very stark differences emerge in time spent with specific inside functions. Pure behavior 1 is over ten times as likely to spend time in activities with commercial-group and business-unit functions, and nearly four times as likely to spend time with the human-resource function. On the other hand, pure behavior 0 is over twice as likely to engage in activities with production. Smaller differences exist for finance (50% more likely in pure behavior 0) and marketing (10% more likely in pure behavior 1) functions. In terms of outside functions, pure behavior 0 is over three times as likely to spend time with suppliers and 25% more likely to spend time with clients, while pure behavior 1 is almost eight times more likely to attend trade associations.

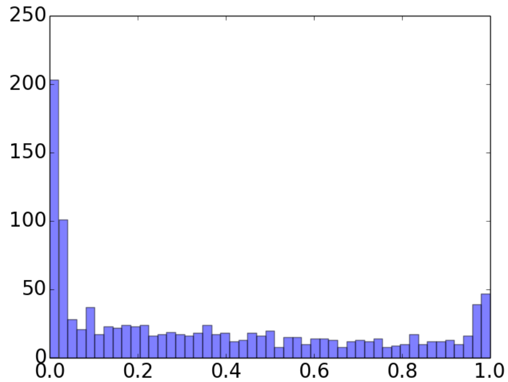
In summary, an overall pattern arises in which pure behavior 0 generates short, small, production-oriented activities and pure behavior 1 generates long, planned activities that combine numerous functions, especially high-level insiders.

2.4.2 The CEO Behavior Index

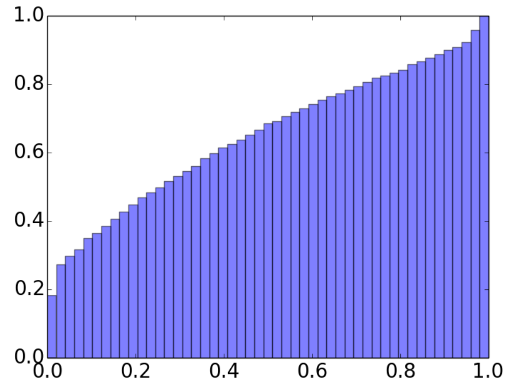
The two pure behaviors we estimate represent extremes. As discussed above, individual CEOs generate time use according to the behavioral index θ_i that gives the probability that any specific time block's feature combination is drawn from pure behavior 1. Figure 4 plots both the frequency and cumulative distributions of the $\widehat{\theta}_i$ estimates across CEOs.

Many CEOs are estimated to be mainly associated with one pure behavior: 316 have a behavioral index less than 0.05 and 94 have an index greater than 0.95. As Figure 4 shows, though, away from these extremes the distribution of the index is essentially uniform, and the

Figure 4: CEO behavior and Index Distribution



(a) Frequency Distribution



(b) Cumulative Distribution

Notes: The left-hand side plot displays the number of CEOs with behavioral indices in each of 50 bins that divide the space [0,1] evenly. The right-hand side plot displays the cumulative percentage of CEOs with behavioral indices lying in these bins.

bulk of CEOs draw their time use from both pure behaviors. This again highlights the value of using a model that allows CEOs to be associated with both estimated behaviors. Finally, we calculate the estimated time shares the average CEO spends within different categories for each feature displayed in Table 2 from the marginal distributions computed in Figure 3 and the estimated behavioral indices displayed in Figure 4. Table A.2 in appendix contains the results, which track very closely the actual time shares computed on the subsample used in estimation contained in Table 2. This provides assurance that the differences between pure behaviors that LDA uncovers are consistent with the raw time-use data.

Finally, we analyze whether the CEO behavior index is correlated with firm characteristics (see Appendix Table A.4). The index is significantly correlated with firm size, in line with the intuition that larger firms have greater demand for structured and multilateral interactions.[17] This correlation is of interest, since it suggests matching between CEOs and firms along at least one important firm characteristic. What is still unclear, however, is the extent to which the match between CEOs and firms is optimal. The next section presents a simple

[17]Beyond firm size, the CEO behavior index is also correlated with industry characteristics, and in particular with variables proxying for skill intensity and complexity of tasks in production. However, since the analysis is always conducted controlling for industry dummies, this cross industry variation does not contribute to the estimates presented in Section 4. We also find that individual CEO characteristics—and in particular whether the CEO has an advanced degree—is significantly correlated with the CEO behavioral index. Conditioning on CEO characteristics does not, however, affect the magnitudes and the significance of the coefficient on firm size.

assignment model to guide the empirical answer to this question.

3 The assignment of CEOs to firms.

In this section we develop a model to guide the empirical analysis of the correlation between CEO behavior and firm performance. The starting point of the model is the observation that to interpret the correlation between behavior and performance in observational data one needs to understand the source of variation in behavior, and this requires modeling how CEOs are assigned to firms. This section develops a simple CEO-firm assignment model that yields the null hypothesis of zero correlation as the equilibrium in which CEOs behave optimally for the firm they are assigned to. The model also illustrates a simple mechanism through which mismatch between CEOs and firms may appear in equilibrium, and how a cross-sectional correlation between CEO behavior and firm performance can be symptomatic of a mismatch.

This minimalistic CEO-firm assignment model is based on two assumptions. First, as in Dessein and Santos (2016), both CEOs and firms have "types". The type of a firm determines which CEO behavior makes it most productive and the type of the CEO determines how willing or able he is to adopt a certain behavior. The model nests the case of pure vertical differentiation, where one CEO type always dominates the other for all firms. One type of CEO may be relatively more abundant than the other type, in the sense that its share is higher than the share of firms who seek CEOs of that type. Second, there are frictions in the market for CEOs. On the hiring side, the firm's screening technology is imperfect and it cannot always correctly identify the CEO's type. On the dismissal side, firing a CEO may be a lengthy process. This second assumption distinguishes our approach from existing theories of manager-firm assignment, where the matching process is frictionless, and the resulting allocation of managerial talent achieves productive efficiency (Gabaix and Landier (2008), Tervio (2008), Bandiera et al. (2015)).

3.1 Model

There are two possible behaviors that CEO i can adopt: $x_i = 0$ and $x_i = 1$. Once a CEO is hired, he decides how he is going to manage the firm that hired him. CEO i has a type $\tau_i \in \{0, 1\}$. Type 0 prefers behavior 0 to behavior 1. Namely, he incurs a cost of 0 if he selects behavior 0 and cost of $c > 0$ if he selects behavior 1. Type 1 is the converse: he incurs a cost of 0 if he selects behavior 1 and cost of c if he selects behavior 0. The cost of choosing a certain behavior can be interpreted as coming from the preferences of the CEO (i.e. he may find one behavior more enjoyable than the other), or his skill set (i.e. he may find one behavior less costly to implement than the other).

Firms also have types. The type of firm f is $\tau_f \in \{0, 1\}$. The output of firm f assigned to CEO i is

$$y_{fi} = \lambda_f + \mathbb{1}(\tau_f = x_i)\Delta$$

for some $\Delta > 0$. Hence firm f's productivity depends on two components. The first is a firm-specific component that we denote λ_f. In principle, this can depend on observable firm characteristics, unobservable firm characteristics, and the firm's type. These dependencies are crucial for identification, and we discuss them further in the empirical section below. The second component is specific to the behavior of the CEO. Namely, if the CEO's behavior matches the firm's type, then productivity increases by a positive amount Δ. The captures the fact that different firms require different behaviors: there not necessarily a "best" behavior in all circumstances. We assume that $c < \Delta$ so that it is efficient for the CEO to always adopt a behavior that corresponds to the firm's type.

Firms offer a linear compensation scheme that rewards CEOs for generating good performance. The wage that CEO i receives from employment in firm $f$is

$$w\left(y_{fi}\right) = \bar{w} + B(y_{fi} - \lambda_f) = \overline{w} + B\mathbb{1}(\tau_f = x_i)\Delta,$$

where $\bar{w}$ is a fixed part, and $B \geq 0$ is a parameter that can be interpreted directly as the performance-related part of CEO compensation, or indirectly as how likely it is that a CEO is retained as a function of his performance (in this interpretation the CEO receives a fixed per-period wage but he is more likely to be terminated early if firm performance is low).

The total utility of the CEO is equal to compensation less behavior cost, i.e. $w(y_{fi}) - \mathbb{1}(\tau_i \neq x_i)c$. After a CEO is hired, he chooses his behavior. If the CEO is hired by a firm with the same type, he will obviously choose the behavior that is preferred by both parties. The interesting case is when the CEO type and the firm type differ. If $B > \frac{c}{\Delta}$, the CEO will adapt to the firm's desired behavior, produce an output of $\lambda_f + \Delta$, and receive a total payoff of $\bar{w} + B\Delta - c$. If instead $B < \frac{c}{\Delta}$, the CEO will choose $\tau_i = x_i$, produce output λ_f and receive a payoff $\bar{w}$. We think of B as a measure of governance. A higher B aligns CEO behavior with the firm's interests.

3.2 Assigning firms and CEOs

Now that we know what happens once a CEO begins working for a firm, let us turn our attention to the assignment process. There is a mass 1 of firms. A proportion ϕ of them are of type 1, the remainder are of type 0. The pool of potential CEOs is larger than the pool of firms seeking a CEO. There is a mass $m >> 1$ of potential CEOs. Without loss of generality, assume that a proportion $\gamma \leq \phi$ of CEOs are of type 1. The remainder are of type 0. From now on, we refer to type 1 as the *scarce* CEO type and type 0 as the *abundant* CEO type. We emphasize that scarcity is relative to the share of firm types. So, it may be the case that the share of type-1 CEOs is actually more numerous than the share of type-0 firms.

The market for CEOs works as follows. In the beginning, every prospective CEO sends his application to a centralized CEO job market. The applicant indicates whether he wishes to work for a type-0 or type-1 firm. All the applications are in a large pool. Each firm begins by downloading an application meant for its type. Each download costs k to the firm. After receiving an application, firms receive a signal about the underlying type of the CEO that submitted it. If the type of the applicant corresponds to the type of the firm, the signal has value 1. If the type is different, the signal is equal to zero with probability $\rho \in [0, 1]$ and to one with probability $1 - \rho$. Thus, $\rho = 1$ denotes perfect screening and $\rho = 0$ represents no screening.[18]

Potential CEOs maximize their expected payoff, which is equal to the probability they are hired times the payoff if they are hired. Firms maximize their profit less the screening cost (given by the number of downloaded application multiplied by k). Clearly, if k is low enough, firms download applications until they receive one whose associated signal indicates the CEO type matches the firm type, which we assume holds in equilibrium.

Let us begin the analysis with an efficient benchmark. If there is no scarce CEO type ($\gamma = \phi$), a CEO has no reason to apply to a firm of a different type. If screening is perfect ($\rho = 1$), a CEO who applies to a firm of the other time is always caught (and hence he won't do it). If governance is good ($B < \frac{c}{\Delta},$), a CEO who is hired by a firm of the other type will always behave in the firm's ideal way (and hence there will either be no detectable effect on firm performance or CEOs will only apply to firms of their type). Thus, under any of these cases, we will have perfect assignment of CEOs to firms, which we can summarize as follows:

[18] The implicit assumption is that CEOs have private information about their types, while firms' types are common knowledge. However, we could also allow firms to have privately observed types; in equilibrium, they will report them truthfully. Moreover, if CEOs have limited or no knowledge of their own type, it is easy to see that our mismatch result would hold a fortiori.

Proposition 1 *If neither CEO type is sufficiently scarce, if screening is sufficienctly effective, or if governance is sufficiently good, then in equilibrium all CEOs are correctly assigned and there is no residual productivity differential between firms run by CEOs who choose behavior 1 and CEOs who choose behavior 0.*

The main result concerns the case where all three conditions in Proposition 1 fail. Then, as the following proposition details, we should expect misallocation:

Proposition 2 *If the screening process is sufficiently unreliable, governance is sufficiently poor, and one CEO type is sufficiently abundant,*[19] *then in equilibrium:*

- *All scarce-type CEOs are correctly matched;*
- *Some abundant-type CEOs are mismatched;*
- *Both the residual productivity and profit (revenues net of CEO compensation) of firms run by abundant-type CEOs are lower than those of firms run by scarce-type CEOs.*

Proof. See Appendix. ■

The intuition for this result is as follows. If all abundant-type CEOs applied to their firm type, they would have a low probability of being hired and they would prefer to apply to the other firm type and try to pass as a scarce-type CEO. In order for this to be true, it must be that the share of abundant types is sufficiently larger than the share of scarce types, and that the risk that they are screened out is not too large. If this is the case, then in equilibrium some abundant-type CEOs will apply to the wrong firm type up to the point where the chance of getting a job is equalized under the two strategies.

Under Proposition 2, the economy under consideration does not achieve productive efficiency. As the overall pool of scarce-type CEOs is assumed to be sufficient to cover all firms that prefer that CEO type ($m >> 1$), it would be possible to give all firms their preferred type and thus increase overall production.[20]

[19] Formally, this is given by the conditions: $B < \frac{c}{\Delta}$, and

$$\rho < \frac{\phi - \gamma}{\phi - \gamma\phi}.$$

[20] If side transfers were feasible, this would also be a Pareto improvement as a type-1 CEO assigned to type-0 firm generates a higher bilateral surplus than a type-0 CEO matched with a type-1 firm, and the new firm-CEO pair could therefore compensate the now unemployed type-0 CEO for her job loss.

Proposition 2 is the alternative hypothesis to the null hypothesis in Proposition 1. If the conditions of Proposition 1 are satisfied, every firm has a CEO that behaves as the firm wants her to behave. Proposition 1 yields the null hypothesis that the cross-sectional correlation between CEO behavior and firm performance is zero. The alternative is given by Proposition 2, where all scarce-type CEOs are matched to the right firm and will have high performance, while some abundant-type CEOs are matched to the wrong firm, and will thus have low performance. Therefore, when frictions in the assignment process exist, the average performance of firms led by abundant-type CEOs will be lower.

The model makes precise the assumption needed for the converse to hold: if the average performance of firms led by abundant-type CEOs is lower, then there must be frictions in assignment. Both propositions hold conditional on firm's type specific baseline performance level λ_f. In other words, they both assume that it is possible to subtract from observed firm performance y_{fi} the part that is independent of CEO behavior, but due to differences in firm type. Only in this case rejecting the null is symptomatic of matching frictions. This is the identifying assumption that underpins the empirical analysis, which clearly fails if there are unobservable factors that drive firm performance and, at the same time, CEO behavior. In this case, the cross-sectional correlation between CEO behavior and firm performance would reflect these firm unobservables, rather than the existence of matching frictions. We return to this identification problem and we provide supporting evidence using changes in performance through time in Section 4 below.

Finally, suppose we are within the conditions of Proposition 2, namely there is some mismatch between CEOs and firms. Some firms who want the scarce-type CEO end up with an abundant-type CEO (type 0). Within this case, there is an extreme case where no firm actually wants a type-0 CEO. This happens when $\phi = 1$. Type 0 is simply an inferior type that reduces productivity in all firms. This case is nested in Proposition 2 and yields this additiuonal prediction:

Corollary 1 (Pure Vertical Differentiation) *If no firm wants the abundant type CEO ($\phi = 1$), then in equilibrium all employed abundant-type CEOs underperform.*

Under Corollary 1, all type-0 CEOs are unwanted and they all reduce the productivity of the firm they work for.

Note that the null hypothesis (Proposition 1) can be interpreted as a pure-horizontal differentiation case, where different firms have CEOs with different behaviors, but all assignments are correct. Proposition 2 also covers intermediate cases, where there is an element of horizontal differentiation (there is heterogeneity in firms' prefererences) and vertical differentiation (some firms end up hiring an inferior CEO type).

In the empirical part of the paper, we will first attempt to distinguish between the alternative hypothesis in Proposition 2 and the null in Proposition 1 (Section 4). We will then move to the distinction between the hypothesis in Proposition 2 and the nested null in Corollary 1 (Section 5).

Finally, one can tweak the model by assuming that some CEOs have observable attributes that make them more or less likely to be one type of CEO. For instance, assume that the share of CEOs with an MBA degree is μ_0 in the abundant type and μ_1 in the scarce type, with $\mu_1 > \mu_0$. However, it can be easily seen that, in equilibrium, the performance of a CEO who works for a type-1 firm cannot be predicted by whether the CEO has an MBA. If all type-1 firms used the presence of an MBA degree to screen applicants, then only abundant-type CEOs with an MBA would apply. However, that cannot be an equilibrium outcome, because having an MBA would be a "negative" signal. In equilibrium, it must be that the abundant-type CEOs who are hired by type-1 firms have the same share of MBA degrees as scarce-type CEOs.

Corollary 2 *If CEOs have stable observable attributes that are correlated to their type, in equilibrium such characteristics will not predict firm performance given behavior.*

Note that the corollary does not imply that the presence of visible CEO attributes is inconsequential. Type signals may make it harder for abundant-type applicants to pretend to be scarce-type applicants, which in turn reduces CEO type mismatch and improves firm performance.[21]

4 CEO Behavior and Firm Performance

4.1 A test of pure horizontal differentiation.

Guided by the model, we now test the null hypothesis of pure horizontal differentiation, namely that CEOs' behavior differ because different firms need different behaviors. The model makes clear that if this hypothesis holds the correlation between CEO behavior and

[21] If the number of MBAs increases so much that it eliminates the incentive for abundant-type CEOs to apply to type-1 firms, then the Corollary is no longer applicable.

firm performance is zero because each firm gets the "right" CEO and hence, controlling for firm traits, firms led by different CEOs perform equally well.[22]

We estimate a production function of the form:

$$y_{ifts} = \alpha\widehat{\theta}_i + \delta^E e_{ft} + \delta^K k_{ft} + \delta^M m_{ft} + \mathbf{Z}_i\gamma + \zeta_t + \eta_s + \varepsilon_{ifts} \tag{1}$$

where y_{ifts} is the log sales (in constant 2010 USD) of firm f, led by CEO i, in period t and sector s. To smooth out short run fluctuations and reduce measurement error in performance, y_{ifts} is average sales computed over up to the three most recent years pre-dating the survey, conditional on the CEO being in office.[23] $\widehat{\theta}_i$ is the estimated behavior index of CEO i, e_{ft}, k_{ft}, and m_{ft} denote, respectively, the natural logarithm of the number of firm employees and, when available, capital and materials. $\mathbf{Z}_i$ is a vector of CEO characteristics (MBA dummy and CEO tenure), ζ_t and η_s are period and SIC2 sector fixed effects, respectively.[24] We include country by year dummies throughout, as well as a set of noise controls.[25]

In Column 1, Table 3 shows the estimates of Equation (1) controlling for firm size, country by year and industry fixed effects, and noise controls. The estimate of α is positive and we can reject the null of zero correlation between firm productivity and CEO behavior at the 1% level (coefficient 0.372, standard error 0.089). To estimate the correlation between TFP and CEO behavior, Column (2) adds capital and Column (3) materials. In these smaller samples, the effect of both inputs has the expected magnitude and is precisely estimated but the magnitude and the precision of coefficient of the CEO behavior index remains unchanged.

[22]Data on firm performance was extracted from ORBIS. We were able to gather at least one year of sales and employment data in the period in which the sampled CEO was in office for 920 of the 1,114 firm with time use data. Of these: 29 did not report sales information at all; 128 were dropped in cleaning, 37 had data that referred only to years in which the CEO was not in office, or for years $< t-3$. The accounting data used in the regressions covers the time period 2003-2014 (this is the maximum number of years of data which can be retrieved from Orbis). See the data Appendix for more details.

[23]In practice we have three years for 58% of the sample, two years for 24% and one year for the remainder of firms.

[24]Since the data is averaged over three years, year dummies are set as the rounded average year for which the performance data is available. The results discussed in this section are robust to using multiple years and clustering the standard errors at the firm level instead of using averages.

[25]The noise controls included throughout the analysis are: a dummy to denote whether the data was collected through the PA (rather than the CEO himself), a reliability score attributed by the analyst at the end of the week of data collection, a set of dummies to denote the specific week in which the data was collected and a dummy to denote whether the CEO formally reported to another manager (this was the case in 6% of the sample). Furthermore, since the behavior index is meant to represent "typical" CEO behavior, regardless of the specific week in which the data was collected, all regressions in this table and throughout the analysis are weighted by a score (ranging between 1 and 10) attributed by the CEO to the survey week to denote its level of representativeness. Finally, we cluster the standard errors at the industry level throughout the table.

Column (4) restricts the sample to firms that, in addition to having data on capital and materials, are listed on stock market and hence have higher quality data. The coefficient of the CEO behavior index is larger in magnitude (0.595) and significant at the 5% level (standard error 0.253) in this sample.

In column (5) we test Corollary 2, namely whether CEO traits predict performance once we control for behavior. If firms are using an observable CEO trait to select among candidates, then in equilibrium that trait cannot predict performance; if it did, it would mean the firm has not used that information optimally. We control for whether the CEO has an MBA degree, and for his tenure in the post. In line with the model, including these variables hardly changes the magnitude of the CEO behavior index (coefficient 0.352, standard error 0.09), and the variables themselves are not significant at standard levels. As Corollary 2 predicts, observable CEO characteristics should not be correlated with performance. As firms are able to screen CEOs on the basis of observables, we should not observe mismatch along those dimensions.

To benchmark the magnitude of the coefficient of the CEO behavior index, the estimate in column 2 implies that a one standard deviation increase in the CEO behavior index is associated with a 0.12 log points higher log sales. This magnitude is similar to the effect of a one standard deviation change in management practices on firm performance (0.15, estimated in Bloom et al, 2016) and about 15% of the effect of a one standard deviation increase in capital (taking the coefficient of 0.396 times the in sample standard deviation of log capital of 1.88). Table A4 shows that the main productivity results are robust to alternative specifications and measurements of the CEO behavior index.

Column (6) analyzes the correlation between CEO behavior and profits per employee. This allows us to assess whether CEOs capture all the extra rent they generate, or whether firms profit from being matched with the scarce type CEO. The results are consistent with the latter interpretation: the correlation between the CEO index and profits per employee is positive and precisely estimated. The magnitudes are also large: a one standard deviation increase in the CEO behavior index is associated with an increase of $3,400 in profits per employee.[26]

[26] Another way to look at this issue is to compare the magnitude of the relationship between the CEO behavior index and profits to the magnitude of the relationship between the CEO behavior index and CEO pay. We are able to make this comparison for a subsample of 196 firms with publicly available compensation data. Over this subsample, we find that a standard deviation change in the CEO behavior index is associated with an increase in profits per employee of $4,900 (which using the median number of employees in the subsample would correspond to $2,686,000 increase in total profit) and an increase in annual CEO compensation of $33,960. According to the point estimates above, the CEO keeps less than 2% of the

Table 3: CEO behavior and firm performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			Log(sales)			Profits/Emp
CEO behavior index	0.374***	0.373***	0.286**	0.595**	0.271**	9.836**
	(0.088)	(0.113)	(0.130)	(0.253)	(0.135)	(4.463)
log(employment)	0.886***	0.555***	0.353***	0.356***	0.355***	0.089
	(0.035)	(0.053)	(0.080)	(0.118)	(0.079)	(0.078)
log(capital)		0.398***	0.211***	0.185**	0.209***	
		(0.031)	(0.051)	(0.087)	(0.052)	
log(materials)			0.428***	0.443***	0.423***	
			(0.063)	(0.100)	(0.062)	
COO Dummy					0.100	
					(0.083)	
log(CEO tenure)					-0.010	
					(0.069)	
CEO has an MBA					-0.051	
					(0.036)	
Adjusted R-squared	0.775	0.839	0.906	0.889	0.906	0.179
Number of observations (firms)	920	618	448	243	448	386
Observations used to compute means	2202	1415	975	565	975	1028
Sample	all	with k	with k & m	with k & m, listed	with k & m	all

Notes: *** (**) (*) denotes significance at the 1%, 5% and 10% level, respectively. We include at most 5 years of data for each firm and build a simple average across output and all inputs over this period. The number of observations used to compute these means are reported at the foot of the table. The sample in Column 1 includes all firms with at least one year with both sales and employment data. Columns 2, 3 and 4 restrict the sample to firms with additional data on capital (column 2) and capital and materials (columns 3 and 4). The sample in columns 4 and 6 is restricted to listed firms. "Firm size" is the log of total employment in the firm, "Log CEO tenure" is the log of 1+number of years CEO is in office, "CEO has an MBA"is a dummy taking value one is the CEO has attained an MBA degree or equivalent postgraduate qualification. All columns include a full set of country by year dummies, industry dummies and noise controls. Noise controls are a full set of dummies to denote the week in the year in which the data was collected, a reliability score assigned by the interviewer at the end of the survey week and a dummy taking value one if the data was collected through the PA of the CEO, rather than the CEO himself. Industry controls are 2 digit SIC dummies. All columns weighted by the week representativeness score assigned by the CEO at the end of the interview week. Errors clustered at the 2 digit SIC level.

In light of the model, the results in Table 3 allow us to reject the null of perfect horizontal differentiation, namely that CEOs' behavior is perfectly tailored to the needs of the firms they lead. The model makes precise two possible alternatives: horizontal differentiation with imperfect matching, whereby only some of the firms require high index CEOs but some of them end up with low index CEOs, and pure vertical differentiation, whereby all firms require high index CEOs. Both interpretations rely on the identifying assumption that, conditional on observable firm characteristics such as factor inputs and sector, the unobserved productivity of firms that need low index CEOs and those that need high index CEOs is the same. If this assumption fails, the fact that a firm hires a low index CEO might just reflect unobservable firm traits that lead to low productivity. We analyze the empirical validity of this assumption in the next section.

4.2 A test of the identifying assumptions.

The theoretical framework shows that the correlation between CEO behavior and firm performance is indicative of mismatching under the identifying assumption that there are no unobservable firm traits that determine both performance and CEO behavior. Proposition 2 specifies that the mismatch between CEO behavior and firm type affects *residual* firm performance, that is, netting out the firm-type specific baseline productivity level λ_f. However, firm type is not directly observable. Our firm-level variables (size, industry, etc) control for some but not all the components of λ_f. The unobservable component of firm type might affect both observed firm performance and observed CEO behavior (or the type of CEO that is hired, which in turn affects behavior), thus creating a correlation between the two even in the absence of matching frictions.

In our static model, unobservables are necessarily time invariant (ν_f). These might determine both productivity and the CEOs firms hire. For instance, high-productivity firms hire high index CEOs while low-productivity firms hire low index CEOs.

In practice, however, unobservables might vary with time (ν_{ft}) and affect CEO behavior with a lag or with a lead. In the first case there is a causal link between between ν_{ft} and θ_{is} with $s \geq t$. One example is that there are "growth firms" and "decline firms". Growth firms are optimally run by high index CEOs while decline firms are optimally run by low index CEOs. Sometimes a growth firm turns into a decline firm – or the other way round.

marginal value he creates through her behavior. This broadly confirms the finding that the increase in firm performance associated with higher values of the CEO behavior index is not fully appropriated by the CEO in the form of rents.

When that happens, the board fires the current high index CEO and replaces him with a low index CEO. Alternatively when the firm type changes, so does performance, and hence the behavior of its CEO. In the second case, there is a causal link between between ν_{ft} and θ_{fs} with $s \leq t$. Namely, what affects CEO behavior is not the change itself but *the expectation of the change.* For instance, a growth firm can turn into a decline firm (and viceversa), and its board predicts this event s year before it actually occurs. To prepare for the change, the board immediately fires the current high index CEO and hires a low index CEO. Thus, future performance causes present CEO behavior.

To provide evidence on the relevance of these concerns, we estimate productivity *within the same firm* before and after the sample CEO is appointed, allowing for the correlation between the CEO behavioral index and productivity to vary by year as follows. First, let τ_{ift} denote the number of years from appointment that CEO i assigned to firm f is at time t. For example, $\tau_{ift} = 3$ indicates that CEO i working at firm f in time t was appointed at time $t - 3$. We adopt the regression specification

$$y_{ift} = \lambda_f + \gamma_t + \sum_{-7 \leq k \leq 7, k \neq 0} \beta_k \mathbb{1}(\tau_{ift} = k) + \sum_{-7 \leq k \leq 7, k \neq 0} \gamma_k \mathbb{1}(\tau_{ift} = k)\tilde{\theta}_i + \delta^E e_{ft} + \varepsilon_{ift} \qquad (2)$$

where (λ_f, γ_t) are firm and year fixed effects. As our sample CEOs have different tenures, and hence have been appointed in different calendar years t, we can estimate the model with both year dummies γ_t and event time dummies. To illustrate the results graphically we use a discretized version of the CEO behavior index so that $\tilde{\theta}_i = 1$ if $\widehat{\theta}_i \geq 0.5$, although the main results hold when we use the continuous CEO behavior index. The coefficients β_k measure the average productivity of CEOs with $\tilde{\theta}_i = 0$, k years after (or before if $k < 0$) they have been appointed. The coefficients γ_k measure the difference in average productivity of CEOs with $\tilde{\theta}_i = 1$ relative to CEOs with $\tilde{\theta}_i = 0$, k years after (or before if $k < 0$) they have been been appointed

If the cross-sectional estimates in Table 3 were due to firm time-invariant unobservables' driving both higher productivity and a tendency to hire CEOs with $\tilde{\theta}_i = 1$, then the firm fixed effects λ_f would capture these and γ_k would be zero for every k, namely if it is a permanent firm trait to drive both productivity and CEO behavior, controlling for all permanent traits will make the effect of behavior disappear. Alternatively, if the cross-sectional estimates were capturing time-varying unobservables that act with a lag at least some γ_k coefficients would be different from zero both before ($k < 0$) and after ($k > 0$) the appointment because, by definition, the firm unobservable trait leads to the appointment of the CEO and hence precedes it. In contrast, if it is the CEO who affects productivity, the γ_k coefficients will be zero for $k < 0$ and positive for $k > 0$.

The results of this analysis are shown in Figure 5 with the underlying coefficients (β_k and $\beta_k + \gamma_k$) reported in Table 4. The sample is restricted to 585 firms with available labor productivity data at least one year after the CEO in our sample has been appointed.[27] This sample includes 3,433 observations, of which 762 in years pre-dating the CEO appointment, and the rest in years in which the CEO was in office.[28] Figure 5, Panel A illustrates the point estimates on the event time dummies for the two types of firms (β_k and $\beta_k + \gamma_k$), while Panel B shows the difference γ_k, as well as their confidence interval. Figure 5 shows that there is no difference between firms that will eventually appoint a low index CEO and those that will eventually appoint a high index CEO before the appointment date, namely γ_k=0 for all $k < 0$. This rules out firm time invariant unobservables because these should always affect productivity, i.e. regardless of whether the current CEO is in place.

The difference appears 4 years after the current CEO is appointed, when the relative performance of firms that hire a low index CEO drops, and that of firms that hire a high index CEO raises. The difference in the point estimates of the tenure dummies after the CEO appointment between the two types of firms, that is the γ coefficients, are significant after year 3.

Table 5 groups the event time dummies into three broader sub periods within which effects are similar: all years before the CEO appointment, between 0 and 3 years after the CEO appointment and 4 years after the CEO appointment. Throughout, the period before the CEO appointment is the omitted category. Column (1) shows that the difference between firms with $\tilde{\theta}_i = 1$ and $\tilde{\theta}_i = 0$ is significant only after year 3, and that $\tilde{\theta}_i = 1$ firms are significantly more productive relative to themselves 3 years after the appointment of the current CEO (as shown by the test on the equality of the interaction terms at the bottom of the table). Column (2) thus repeats the specification only including the interaction term of the CEO behavior index with the dummy denoting the second sub period of the

[27]For comparison with the earlier cross sectional results, note that the coefficient on the CEO behavior index on this subsample using the specification of Table **??**, column (1) and just using the years in which the CEO is in office is 0.427 (standard error .103).

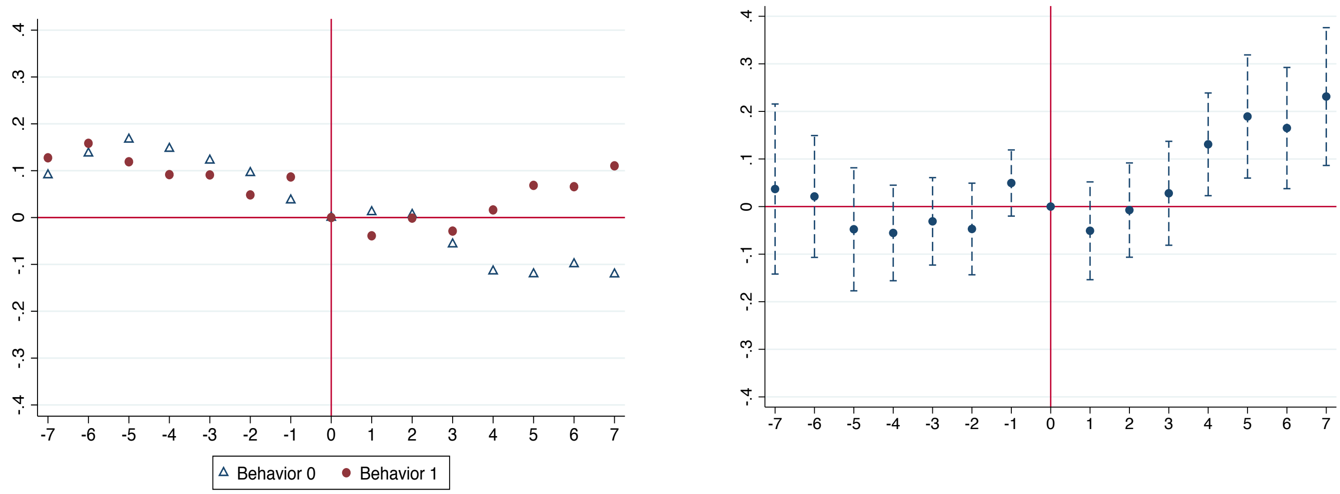
[28]Note that since we have few observations outside the $\tau \in (-7, +7)$ in this subsample which contribute to the estimation of the coefficients in the fixed effects specification, for ease of presentation we group observations with $\tau < -7$ and $\tau > 7$ at the extremes of this interval. This affects 48 observations with $\tau < -7$ and 287 observations with $\tau > 7$. Table AX in Appendix shows the full distribution of firms per year underlying the figure.

Table 4: CEO behavior and productivity by year since and to appointment

Time before/after Ceo appointment	Behavior 0	Behavior 1	Difference Behavior 1 - Behavior 0
-7 and earlier	0.09	0.13*	0.04
	(0.11)	(0.08)	(0.11)
-6	0.14**	0.16**	0.02
	(0.07)	(0.06)	(0.08)
-5	0.17**	0.12**	-0.05
	(0.07)	(0.06)	(0.08)
-4	0.15***	0.09*	-0.06
	(0.05)	(0.05)	(0.07)
-3	0.12***	0.09*	-0.03
	(0.04)	(0.05)	(0.06)
-2	0.10**	0.05	-0.05
	(0.05)	(0.04)	(0.06)
-1	0.04	0.09**	0.05
	(0.04)	(0.03)	(0.05)
0	0	0	0
1	0.01	-0.04	-0.05
	(0.03)	(0.06)	(0.07)
2	0.01	-0.01	-0.01
	(0.04)	(0.05)	(0.07)
3	-0.06	-0.03	0.03
	(0.05)	(0.06)	(0.07)
4	-0.11**	0.02	0.13**
	(0.05)	(0.06)	(0.07)
5	-0.12*	0.07	0.19**
	(0.07)	(0.06)	(0.08)
6	-0.10*	0.07	0.17**
	(0.07)	(0.07)	(0.08)
7 and over	-0.12*	0.11	0.23***
	(0.07)	(0.08)	(0.09)

Notes: *** (**) (*) denotes significance at the 1%, 5% and 10% level, respectively. Columns (1) and (2) report coefficients from the same regression: the natural log of input is regressed on the natural log of employment, firms fixed effects, country by year dummies and noise controls, and the reported dummy variables for whether the firm has a $\theta=0$ or $\theta=1$ CEO in each year relative to the year of CEO appointment. We include all available years with information on sales, employment and capital for all firms with observable data at least one year after the CEO was appointed. Noise controls are a full set of dummies to denote the week in the year in which the data was collected, a reliability score assigned by the interviewer at the end of the survey week and a dummy taking value one if the data was collected through the PA of the CEO, rather than the CEO himself. Industry controls are 2 digit SIC dummies. All columns weighted by the week representativeness score assigned by the CEO at the end of the interview week. Errors clustered by firm and before/after CEO appointment period.

Figure 5: CEO behavior and productivity by year since and to appointment



Notes: These figures accompany Table 4. Panel A plots the point estimates of the coefficients reported in Table 4, columns (1) and (2). Panel B plots the difference between the two points in each year and its confidence interval.

CEO tenure (4 years and + after appointment). Column (3) shows that the results are robust to controlling for differential time trends throughout the pre and post appointment periods. Furthermore, as it can also be inferred by Figure (5), the trend interaction is insignificant. Another possible explanation for the finding that the correlation between behavior and performance emerges only after three years of tenure is that CEOs might take time to learn what the firm needs, and require exactly three years to adjust behavior. Since we measure behavior at different times from the appointment date, we can check whether this drives the tenure result including a variable measuring the distance between the year in which the productivity data is measured and the survey year, interacted with the CEO behavior index in column (4) - this leaves the estimates unchanged, although it reduces power. Finally column (5) shows that the main result is robust to including additional controls for capital and material inputs.[29]

Taken together, the findings in Figure and Table 5 cast doubt on the relevance of time varying unobservable that affects CEO behavior and/or her identity contemporaneously or with a lag. We find no evidence that performance levels or performance trends differed before the years that led to the appointment of the current CEO. In addition, Appendix table XX shows that the behavior of the current CEO is not correlated with the growth of firm sales

[29]To keep firms*year observations in the sample in which we have information on the additional inputs (170 observations for capital and 1,345 for materials), we set these variables to -99 when missing and include a dummy in the set of regressors to keep track of this change.

in years pre-dating the survey, casting doubts on the empirical relevance of the concern that the same CEO changes behavior according to firm performance.

What we cannot rule out are time varying firm traits ν_{ft} that affect the behavior of the CEO with a lead $s \leq -3$. The only type of time-varying firm unobservable that would violate the identifying assumptions requires firms to appoint CEOs in anticipation of events that will take place three years after the appointment date. While our data cannot rule out this possibility, two observations should make us question its plausibility. First, the board strategy must be such that they dismiss CEOs who are currently performing adequately because they might perform poorly four years ahead, which requires exceptional abilities in both forecasting firm's needs and to remove inadequate CEOs – something that is at odds with the general consensus on the difficulty of removing corporate leaders. Second, this strategy implies that boards are willing to bear losses due to the wrong match or CEOs are willing to behave sub-optimally for three years, that is 33% of the average CEO tenure in our sample, rather than appoint a new CEO when he is actually needed.

Overall, the patterns just described are consistent with the interpretation that a CEO with the right behavior improves firm performance, and that this becomes apparent after three years. The effect is not immediate because the actions of the new CEO take time to affect the production process. The existence of significant organizational inertia within firms has been a central theme in the management literature (Cyert and March, 1963), and is central to a recent strand of the organizational economics literature.[30] In fact, the gradual emergence of the CEO behavior effect is consistent with a simple dynamic extension of the model developed in Section 3. In this extension, a firm that hires a behavior 1 CEO experiences on average a gradual increase in productivity as in Figure A.1. This is the combination of two possible cases: (i) The firm already had a behavior 1 CEO, in which case productivity stays constant, and (ii) The firm had a mismatched CEO, in which case productivity increases gradually. This last observation implies that the fixed-effect estimate we obtain in Table 5 represent lower bounds to the effect of replacing a mismatched CEO with the right CEO.

[30]For example, in the model of Halac and Prat (2014), it takes time for a corporate leader to change the existing management practice and to affect the company's culture. Empirically, Bloom et al. (2016) estimate adjustment costs in managerial capital of similar magnitude to the ones estimated for physical capital.

Table 5: CEO behavior and productivity by tenure period

	(1)	(2)	(3)	(4)
Dependent Variable		Log(sales)		
1-3 years after CEO appointment	-0.052			
	(0.035)			
4+ years after CEO appointment	-0.152***	-0.096**	-0.084*	-0.064
	(0.055)	(0.048)	(0.044)	(0.043)
CEO behavior index*1-3 years after CEO appointment	-0.020			
	(0.054)			
CEO behavior index*4+ years after CEO appointment	0.190**	0.203**	0.175**	0.132*
	(0.089)	(0.090)	(0.083)	(0.080)
Trend				0.023
				(0.021)
Trend*CEO behavior index				0.010
				(0.009)
log(employment)	0.557***	0.559***	0.521***	0.519***
	(0.068)	(0.068)	(0.068)	(0.064)
log(capital)			0.052*	0.052**
			(0.027)	(0.026)
log(materials)			0.173***	0.174***
			(0.047)	(0.052)
Adjusted R-squared	0.973	0.973	0.976	0.976
Observations	3433	3433	3433	3433
Number of firms	585	585	585	585
Test CEO behavior index*1-3 years after CEO appointment=CEO behavior index*4+ years after CEO appointment (p-value)	0.03			

Notes: *** (**) (*) denotes significance at the 1%, 5% and 10% level, respectively. All columns include firms fixed effects, country by year dummies and noise controls. We include all available years with information on sales, employment and capital for all firms with observable data at least one year after the CEO was appointed. Noise controls are a full set of dummies to denote the week in the year in which the data was collected, a reliability score assigned by the interviewer at the end of the survey week and a dummy taking value one if the data was collected through the PA of the CEO, rather than the CEO himself. Country by year dummies are included in all columns. Industry controls are 2 digit SIC dummies. All columns weighted by the week representativeness score assigned by the CEO at the end of the interview week. Errors clustered by firm and before/after CEO appointment period.

4.3 Robustness Checks

4.3.1 Managers or Management?

So far we have interpreted the CEO index primarily in terms of manager-specific behavior. However, what CEOs do with their time may also reflect broader differences in management processes across firms. For example, the propensity to engage in cross-functional coordination activities (vs. purely operational tasks) captured by higher values of the CEO index may be facilitated by the presence of systematic monitoring systems. To investigate this issue, we matched the CEO behavior index with detailed information on the type of management practices adopted in the firm. The management data was collected using the basic approach of the World Management Survey (Bloom et al., 2016). The survey methodology is based on semi-structured double blind interviews with plant level managers, run independently from the CEO time use survey.[31] To our knowledge, this is the first time that data on middle level management practices and information on CEO behavior is systematically analyzed.[32]

We start by looking at the correlation between the CEO behavior index and the management practices data in a simple specification including country and industry (SIC 1 level, given the smaller sample for which we are able to conduct this analysis) dummies, controls for log firm and plant employment (since the management data is collected at the plant level) and interview noise controls, using the weighting scheme described in previous specifications.[33] Table 6 Column (1) shows that higher values of the CEO behavior index are significantly correlated with a higher management score - a one standard deviation in management is associated with 0.059 increase in the CEO behavior index, or 18% of a standard deviation. Columns (2) and (3) show that this result is driven primarily by the sections of the management score measuring processes relative to operations, monitoring and targets, rather than people management practices (e.g. use of financial and non financial rewards in managing employees).

[31] We collected the majority of the data in the Summer of 2013. A small share of the management data (16 observations out of a total of 191) was collected between 2006 and 2012 in the context of the larger WMS survey waves. We include this data in the analysis only if the CEO was in office at the time in which it was collected, and include wave dummies in all specifications.

[32] Bloom et al. (2016) analyze the correlation between management practices and employees' wage fixed effects and find evidence of sorting of employees with higher fixed effects in better managed firms. The analysis also includes a subsample of top managers, but due to data confidentiality it excludes from the sample highest paid individuals, who are likely to be CEOs.

[33] Given the limited number of firms in the sample we cannot include a full set of week dummies in the vector of noise controls as in previous specifications. We also include two measures of interview noise drawn from the management interviews, namely a variable denoting the duration of the management interview and the overall reliability of the interview as assessed by the interviewer.

We then turn to analyzing both the CEO behavior index and the management variables in the context of the production function of Equation (1). Column (4) shows that the CEO behavior index is positive and statistically significant even in the smaller sample of 146 firms with both management and CEO data (coefficient 0.474, standard error 0.213). Column (5) shows that the management index is also correlated with labor productivity within the same sample (coefficient 0.18, standard error 0.073). Column (6) shows that the two variables retain a similar magnitude and significance level even when both included in the production function regression. The magnitude of the coefficients is also similar: a standard deviation change in the CEO behavior index is associated with an increase of 0.16 log points in sales, versus the 0.17 change implied by a standard deviation change in the management score.

To summarize, even if management and the CEO behavior index are positively correlated among each other, they appear to be independently correlated with performance. The latter finding suggests that the positive relationship between the CEO index and firm performance is not entirely a reflection of the management practices adopted by the firm when the CEO is in office.

4.4 Pure vertical differentiation vs. horizontal differentiation with misallocation.

The analysis above allows us to reject the null of pure horizontal differentiation (Proposition 1), namely that all firms get CEOs whose behavior is optimal for the firm. However, within the alternative hypothesis (Proposition 2), there are two nested cases. The simpler one is pure vertical differentiation (Corollary 1): namely no firm wants low-index CEOs, but some low-index CEOs manage to get hired and, when they do, they invariably lower productivity. The more complex case included in Proposition 2 is when elements of vertical and horizontal differentiation coexist. Some firms do prefer low-index CEOs, they get them, and they achieve their productivity potential. The remaining firms prefer high-index CEOs and they do not necessarily get their ideal type.

As the theoretical framework makes clear, there are two alternative hypotheses. The first is horizontal differentiation with mismatches, namely some firms demand low-index CEOs, but some of these are assigned in equilibrium to firms that demand high-index CEOs. In other words, low-index CEOs exist because some firms demand them but their supply outstrips demand and some of them end up in firms that would be better off with high-index CEOs. The second is pure vertical differentiation, namely *all* firms need high-index CEOs so that all low-index CEOs are incorrectly assigned. These two alternative hypothesis are

Table 6: Management Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	CEO behavior index			Log(sales)		
CEO behavior index				0.474**		0.417**
				(0.213)		(0.206)
Management (z-score)	0.059**				0.180**	0.163**
	(0.029)				(0.073)	(0.069)
Operations, Monitoring, Targets (z-score)		0.062**				
		(0.029)				
People (zscore)			0.044			
			(0.032)			
log(employment)	0.101***	0.103***	0.101***	0.640***	0.640***	0.623***
	(0.030)	(0.030)	(0.031)	(0.080)	(0.089)	(0.084)
log(capital)				0.242***	0.235**	0.237***
				(0.088)	(0.090)	(0.087)
log(materials)				0.268**	0.302**	0.258**
				(0.116)	(0.129)	(0.128)
Adjusted R-squared	0.144	0.145	0.133	0.839	0.842	0.850
Number of firms	191	191	191	146	146	146

Note: *** (**) (*) denotes significance at the 1%, 5% and 10% level, respectively. All columns include industry dummies and a restricted set of noise controls. Columns (1) to (3) include country dummies. Columns (4) to (6) include country by year dummies. Management is the standardized value of the Bloom and Van Reenen (2007) management score, Operations, Monitoring and Targets" and "People" are subcomponents of the main management score. Noise controls are a reliability score assigned by the interviewer at the end of the survey week and a dummy taking value one if the data was collected through the PA of the CEO, rather than the CEO himself, as well a variable capturing the reliability of the management score (as assessed by the interviewer) and the duration of the management interview. In columns (4) to (6) we include at most 5 years of data for each firm and build a simple average across output and all inputs over this period. Industry controls are 1 digit SIC dummies. All columns weighted by the week representativeness score assigned by the CEO at the end of the interview week. Errors clustered at the 2 digit SIC level.

observationally equivalent in reduced form as they both yield a positive correlation between the behavior index and firm performance. To assess which of the two fits our data best we estimate the share of mismatched CEOs- firms pairs structurally.

The structural model

The main data input of the model is the firms' conditional productivity, that is the residuals of a regression of productivity on firm characteristics. We denote this residual $\widehat{\varepsilon}_{if}$. In line with the theory model, we adopt the statistical model $\widehat{\varepsilon}_{if} = \lambda_f + \mathbb{1}(\tau_f = x_i)\Delta + v_{if}$, where λ_f is a "baseline" productivity; $\tau_f \in \{0, 1\}$ is the firm's type; $x_i \in \{0, 1\}$ is the CEO's behavior; and Δ is the productivity difference between firms with the right CEO and firms with the wrong CEO. Throughout this section we assume the conditions of Proposition 2 hold. Since our CEO behavior index is continuous, while in the model x_i is binary, we discretize the former to obtain the empirical proxy $\widehat{x}_i = \mathbb{1}\left(\widehat{\theta}_i >= 0.5\right)$. While we treat $\widehat{x}_i$ as observed data, τ_f is a random variable. Since all CEOs with $\widehat{x}_i = 1$ are correctly assigned, whenever we observe $\widehat{x}_i = 1$ we also must have $\tau_f = 1$. In contrast, only a share q of CEOs with $\widehat{x}_i = 0$ are correctly assigned. In the case of pure vertical differentiation $q = 0$, that is all CEOs with $\widehat{x}_i = 0$ are misassigned because there is no demand for their behavior. Vice versa, in the case of pure horizontal differentiation $q = 1$ that is all CEOs with $\widehat{x}_i = 0$ are assigned to firms that demand their behavior. In general, when we observe $\widehat{x}_i = 0$, the CEO is correctly assigned and $\tau_f = 0$ with probability $0 \leq q \leq 1$; otherwise, with probability $1 - q$ the CEO is misassigned and $\tau_f = 1$.

As for the baseline productivity, we model $\lambda_f = x_{c(f)}$ where $c(f)$ denotes the country in which firm f operates. Finally, we treat v_{if} as a mean-zero normal random variable whose variance is both country and assignment specific: $\sigma^2_{1,c(f)}$ ($\sigma^2_{0,c(f)}$) is the standard deviation of residuals in an efficient (inefficient) CEO-firm pair.

Given these observations, the likelihood function can be written as

$$\prod_{f\in\Theta(0)} \left\{ \begin{array}{c} \frac{q}{\sqrt{2\pi}\sigma_{1,c(f)}} \exp\left[-\frac{1}{2\sigma^2_{1,c(f)}} \left(\widehat{\varepsilon}_{if} - x_{c(f)} - \Delta\right)^2\right] + \\ \frac{1-q}{\sqrt{2\pi}\sigma_{0,c(f)}} \exp\left[-\frac{1}{2\sigma^2_{0,c(f)}} \left(\widehat{\varepsilon}_{if} - x_{c(f)}\right)^2\right] \end{array} \right\} \times$$

$$\prod_{f\in\Theta(1)} \frac{1}{\sqrt{2\pi}\sigma_{1,c(f)}} \exp\left[-\frac{1}{2\sigma^2_{1,c(f)}} \left(\widehat{\varepsilon}_{if} - x_{c(f)} - \Delta\right)^2\right] \tag{3}$$

where $\Theta(0)$ and $\Theta(1)$ are the sets of firms managed by CEOs with behaviors 0 and 1, respectively. Behavior-1 CEOs are always efficiently assigned to type-1 firms and their residuals are drawn from a normal distribution with mean $x_{c(f)} + \Delta$; in contrast, firms managed by behavior-0 CEOs have their residuals drawn from a mixture of two normals, one with mean $x_{c(f)} + \Delta$ if the assignment is efficient and another with mean $x_{c(f)}$ if the assignment is inefficient. The mixing probability is simply q, the probability that CEOs with behavior 0 are assigned to type-0 firms. We use the EM algorithm to maximize (3).[34]

Results

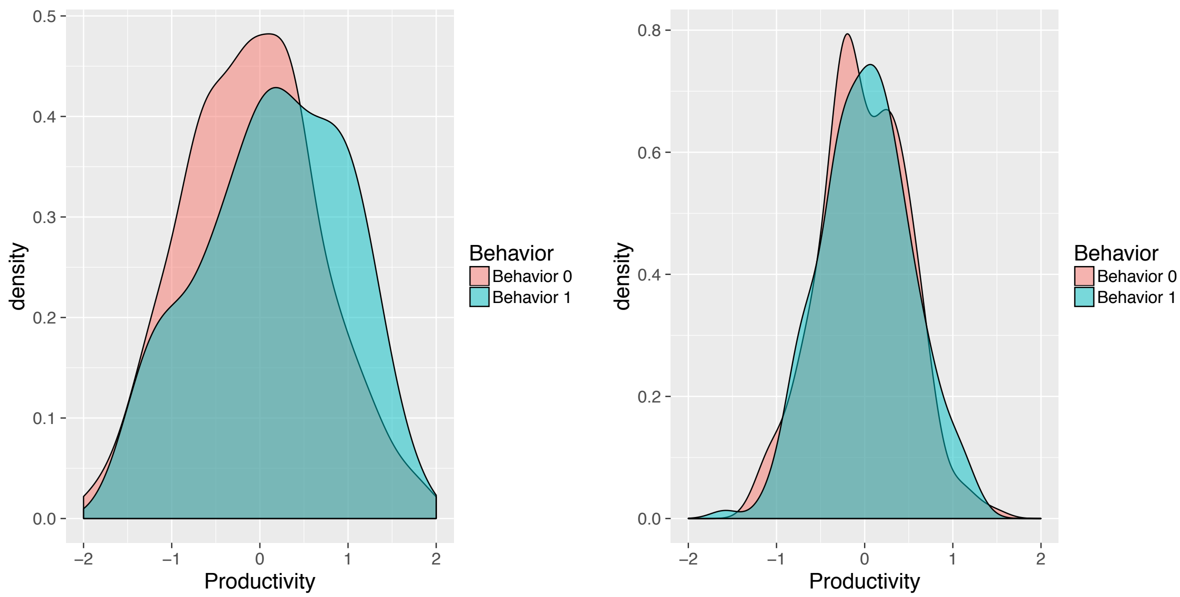
The estimated q is equal to 0.737. To test whether the data are consistent with pure vertical differentiation, we impose the restriction $q = 0$ and maximize (3) with respect to the remaining parameters. A simple likelihood ratio test rejects the null hypothesis of vertical differentiation with p-value is 0.00859. For completeness, we also fit a model with pure horizontal differentiation in which $q = 1$, which we also reject in line with the reduced form results. Overall, then, a model with heterogenous firms and assignment frictions fits the data significantly better than one without firm heterogeneity (pure vertical differentiation) or one without such frictions (pure horizontal differentiation).

5 Quantifying productivity losses due to misallocation.

The structural model also allows us to quantify the productivity losses due to the misallocation of CEOs to firms. We do so both in the sample as a whole and separately for high- and middle-income regions. The latter split is informed by the recent literature on the misallocation of factors of production as a key driver of economic development and serves as a "sanity" check because if misallocation is driven by frictions, we should observe more misallocation where frictions are more severe, that is in the poorest countries in our sample (India and Brazil)

To quantify the share of mismatches we first derive ϕ , i.e. the share of type-1 firms, from

[34] As is well known, direct maximization of (3) is difficult since there is no way to remove the sum over the two normals distributions involving q from inside the logarithms in the log-likelihood function. The standard version of the EM algorithm iteratively computes the probability that observations belong to each latent state (in our case, an efficient or inefficient assignment) in the E-step, then maximizes the expected value of (3) with respect to these probabilities in the M-step. This procedure is guaranteed to increase the value of likelihood at each iteration, and so converges to a local maximum. The only difference here is that the structure of the economic model implies there is no uncertainty with respect to the assignment of CEOs with observed behavior 1, so we only apply the E-step to the set of CEOs in Θ_0. Further details of our approach are in the appendix.



(a) Low/Middle-Income Countries

(b) High-Income Countries

Figure 6: Productivity Residuals Densities by Regional Income Level
Notes: This figure plots the estimated kernel densities of the productivity residuals $\widehat{\varepsilon}_{if}$ (demeaned at the country level) by regional income and by observed CEO behavior. The low/middle income countries are Brazil and India, while the high income countries are France, Germany, the United Kingdom, and the United States.

the market clearing condition. Overall we observe a share $\widehat{s} = 0.339$ of CEOs who adopt behavior 1. We must then have $\phi = \widehat{s} + (1 - \widehat{s})(1 - q)$. The right-hand side of this expression is the total share of CEOs assigned to type-1 firms: all behavior-1 CEOs and a portion $1 - q$ of behavior-0 CEOs. Plugging in for $\widehat{s}$ and q, we obtain $\phi = 0.513$ so that slightly over half of firms have type 1. This in turn implies that a share $\phi - \widehat{s} = 0.174$ of firms are misassigned in our data, leading to an overall productivity loss of 0.096 ($= 0.174 * \Delta$) log points.

To explore differences along the development path, figure 6 the densities of conditional productivity by income level. The figure shows a marked difference between the two sets of countries. In the poorer countries the behavior 1 density first-order stochastically dominates the behavior 0 density while in the richer countries (France, Germany, UK, and US) the densities are much more similar. This indicates that observed CEO behavior in high-income countries is much more likely to correspond to the firm's needs than in low/middle-income countries, i.e. that assignment frictions are substantially lower.

To quantify these effects, we allow the q parameter in the likelihood function (3) to vary according to whether the firm operates in a low/middle- or high-income country. The

	Estimated Parameters		Derived Parameters		
	Δ	q	$\widehat{s}$	ϕ	share firms misassigned
low/middle income	0.56	0.434	0.213	0.658	0.445
high income	0.56	0.803	0.481	0.583	0.102

Table 7: Estimation Results by Region

estimation results are in table 7. Confirming the notable differences in figure 6, behavior-0 CEOs in low/middle income countries are efficiently assigned with probability 0.434, while the corresponding probability for behavior-0 CEOs in high income countries is 0.803. The derived parameters in the table are obtained using the same steps as described above. We observe a much larger share of behavior-1 CEOs in high income countries (0.481 vs. 0.213). One explanation is that firms in high-income countries have higher demand for coordinating rather than micro-managing behavior. However, the ϕ parameters we extract are very similar in both regions (if anything there is somewhat higher demand for coordinating behavior in poorer countries).[35] Instead, the main difference between regions is that type-1 firms in low/middle-income countries are unable to locate and hire behavior-1 CEOs. It is important to reiterate that this is not necessarily due to scarcity of type-1 CEOs in the population per se. Rather, barriers to the allocation of talent might prevent the right individuals from entering the CEO job market. Regardless of the deeper cause, the share of inefficiently matched type-1 firms in these countries is 0.445, compared to 0.102 in high-income countries. While there is still a sizable number of inefficient assignments in richer countries, the share in poorer countries is over four times as large.

To conclude, we use our estimates to quantify how much productivity in India and Brazil would increase if the assignment process were as efficient as in the richer countries in the sample. This implies building a counterfactual where q increases from 0.434 to 0.803, which requires the share of high index CEOs to increase from 0.213 to .574 to maintain market clearing, and which yields a drop in the share of mismatches from 0.445 to 0.084. Given that the productivity difference between a good and a bad match is 0.56, productivity would increase by 0.202 log points.

[35] Also of note is that the ϕ estimates suggest that horizontal differentiation exists in both regions since firms demand substantial shares of both behaviors. We have repeated the same exercise as in table **??** but for each region separately. While the power of the chi-squared tests is lower due to reduced sample size, we are able to reject pure vertical and horizontal differentiation at a 10% significance level in both regions.

6 Conclusions

This paper combines a new survey methodology with a machine learning algorithm to measure the behavior of CEOs in large samples. We show that CEOs differ in their behavior along a number of dimensions. Guided by a simple firm-CEO assignment model we show that while there is no "best practice", that is a behavior that is optimal for all the firms, there is evidence of significant matching frictions in the assignment of CEOs to firms. In our sample of manufacturing firms across six countries we estimate that 17% of firm-CEO pairs are mismatched and that mismatches are found in all regions but are more frequent in emerging economies. The consequences for productivity are large: differences in match quality explain 12% of the productivity gap between firms in high and middle-low income countries in our sample.

While this paper has intentionally taken an agnostic approach to leadership, an obvious next step would be to explore in more detail the precise mechanisms through which different administrative behaviors affect firm performance, and why different firms need different behaviors. The CEO behavior that according to our CEO-firm assignment model and our data is scarcer in the population of actual CEOs (and hence produces a better average performance) features a longer planning horizon, larger multi-functional meetings, a focus on higher-level executives and non-production functions. One tentative interpretation is that a CEO that displays this pattern of behavior coordinates at a high level, he delegates operational tasks to other executives and spends his time ensuring good communication in the top management team. Within the same interpretation, the other pattern of behavior captures coordination at a lower level, CEOs who display this behavior act as *micromanager*s, who tend to intervene directly in operational aspects, prefer one-on-one meetings with a variety of internal and external constituents, and put less emphasis on long term planning.

To the best of our knowledge, this dichotomy has not been directly addressed by any of the existing literature on leadership - within and outside economics - although the general idea of leader types is present in recent papers in the economic leadership literature.[36] Future

[36] Hermalin (1998) and Hermalin (2007) proposes a rational theory of leadership, whereby the leader possesses private non-verifiable information on the productivity of the venture that she leads. In the dynamic version of the model, the leader can develop a reputation for honestly announcing the true state of the world. In practice, one way of strengthening this reputation is to have formal gatherings where the leader is held accountable for her past announcements. Van den Steen (2010) highlights the importance of shared beliefs in organizations. Shared beliefs lead to more delegation, less monitoring, higher utility, higher execution effort, faster coordination, less influence activities, and more communication. Bolton et al. (2013) propose a model of resoluteness. A resolute leader has a strong, stable vision that makes her credible among her followers. This helps align the followers' incentives and generates higher effort and performance. Finally,

work could utilize information about CEO behavior to inform alternative leadership models, and in particular explore in detail the underlying CEO-firm matching function, which is not dealt with explicitly in the current paper. Furthermore, a possible next step of this research would be to extend the data collection to the diaries of multiple managerial figures beyond the CEO. This approach would allow us to further explore the importance of managerial interactions and team behavior (Hambrick and Mason, 1984), which are now largely absent from our analysis. We leave these topics for further research.

Dessein and Santos (2016) explore the interaction between CEO characteristics, CEO attention allocation, and firm behavior: small differences in managerial expertise may be amplified by optimal attention allocation and result in dramatically different firm behavior.

References

Bandiera, O., Guiso, L., Prat, A., and Sadun, R. (2012). What Do CEOs Do? CEP Discussion Papers dp1145, Centre for Economic Performance, LSE.

Bandiera, O., Guiso, L., Prat, A., and Sadun, R. (2015). Matching Firms, Managers, and Incentives. *Journal of Labor Economics*, 33(3):623 – 681.

Bandiera, O., Prat, A., and Sadun, R. (2013). Managing the Family Firm: Evidence from CEOs at Work. Harvard Business School Working Papers 14-044, Harvard Business School.

Bennedsen, M., Nielsen, K. M., Perez-Gonzalez, F., and Wolfenzon, D. (2007). Inside the Family Firm: The Role of Families in Succession Decisions and Performance. *The Quarterly Journal of Economics*, 122(2):647–691.

Bertrand, M. and Schoar, A. (2003). Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics*, 118(4):1169–1208.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bloom, N., Sadun, R., and Van Reenen, J. (2016). Management as a Technology? mimeograph.

Bloom, N. and Van Reenen, J. (2007). Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

Bolton, P., Brunnermeier, M. K., and Veldkamp, L. (2013). Leadership, Coordination, and Corporate Culture. *The Review of Economic Studies*, 80(2):512–537.

Chang, J., Gerrish, S., Boyd-Graber, J. L., and Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems*.

Coase, R. (1937). The nature of the firm. *Economica*, 4, New Series(16):386–405.

Dessein, W. and Santos, T. (2016). Managerial Style and Attention. mimeograph.

Gabaix, X. and Landier, A. (2008). Why has CEO Pay Increased So Much? *The Quarterly Journal of Economics*, 123(1):49–100.

Griffiths, T. L. and Steyvers, M. (2004). Finding Scientific Topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1):5228–5235.

Halac, M. and Prat, A. (2014). Managerial Attention and Worker Engagement. CEPR Discussion Papers 10035, C.E.P.R. Discussion Papers.

Hambrick, D. C. and Mason, P. A. (1984). Upper Echelons: The Organization as a Reflection of Its Top Managers. *The Academy of Management Review*, 9(2):193–206.

Hansen, S., McMahon, M., and Prat, A. (2014). Transparency and Deliberation within the FOMC: a Computational Linguistics Approach. Discussion Paper 9994, CEPR.

Hermalin, B. E. (1998). Toward an Economic Theory of Leadership: Leading by Example. *American Economic Review*, 88(5):1188–1206.

Hermalin, B. E. (2007). Leading for the Long Term. *Journal of Economic Behavior & Organization*, 62(1):1–19.

Kaplan, S. N., Klebanov, M. M., and Sorensen, M. (2012). Which CEO Characteristics and Abilities Matter? *The Journal of Finance*, 67(3):973–1007.

Kaplan, S. N. and Sorensen, M. (2016). Are CEOs Different? Characteristics of Top Managers. mimeograph.

Kotter, J. P. (1999). *John Kotter on What Leaders Really Do.* Harvard Business School Press, Boston.

Luthans, F. (1988). Successful vs. Effective Real Managers. *Academy of Management Executive*, 2(2):127–132.

Malmendier, U. and Tate, G. (2005). CEO Overconfidence and Corporate Investment. *Journal of Finance*, 60(6):2661–2700.

Malmendier, U. and Tate, G. (2009). Superstar CEOs. *The Quarterly Journal of Economics*, 124(4):1593–1638.

Mintzberg, H. (1973). *The Nature of Managerial Work.* Harper & Row., New York.

Mullins, W. and Schoar, A. (2013). How do CEOs see their Role? Management Philosophy and Styles in Family and Non-Family Firms. NBER Working Papers 19395, National Bureau of Economic Research, Inc.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2):326–65.

Tervio, M. (2008). The Difference That CEOs Make: An Assignment Model Approach. *American Economic Review*, 98(3):642–68.

Van den Steen, E. (2010). On the origin of shared beliefs (and corporate culture). *RAND Journal of Economics*, 41(4):617–648.

A Data Appendix

A.1 The Time Use Survey

The time use survey took place in two stages: in the Spring of 2011 a team of 15 analysts based in Mumbai and led by one of our project managers collected data on India, while the rest of the countries were covered in a second survey wave in the Spring of 2013 by a team of 40 enumerators based at the London School of Economics. To ensure comparability, we adopted the same protocol and retained the same project manager across both waves. The enumerators where typically graduate students (often MBAs) recruited specifically for this project. All enumerators were subject to a common intensive training on the survey methodology for three days at the beginning of the project, plus weekly team progress reviews and one to one conversations with their supervisors to discuss possible uncertainties with respect to the classification of the time use data. Each interview was checked off at the end of the week by one supervisor, who would make sure that the data was complete in every field, and that the enumerator had codified all the activities according to the survey protocol. Each enumerator ran on average 30 interviews.

Each enumerator was allocated a random list of about 120 companies, and was in charge of calling up the numbers of his or her list to convince the CEO to participate in the survey, and to collect the time use data in the week allocated to the CEO. One project manager, five full time supervisors and one additional manager working on a part time basis led the survey team.

We actively monitored and coached the enumerators throughout the project, which intensified their persistence in chasing the CEOs and getting them to participate. We also offered the CEOs a personalized analysis of their use of time (which was sent to them in January 2012 to the Indian CEOs and in June 2014 to the rest of the countries) to give them the ability to monitor their time allocation, and compare it with peers in the industry.

A.2 Sampling Frame

The sampling frame was drawn from ORBIS, an extensive commercial data set that contains company accounts for several millions of companies around the world. Our sampling criteria were as follows. First, we restricted the sample to manufacturing and additionally kept firms that were classified as "active" in the year prior to the survey (2010 in India and 2012 for the other countries) and with available recent accounting data.[37]These conditions restricted our sample to 11,500 firms. Second, we further restricted the sample to companies for which we could find CEOs contact details. To gather contact information we hired a team of research assistants based in Mumbai, London and Boston who verified the CEOs names and found their phone numbers and emails. This restricted the sample to 7,744 firms.

[37]For the Indian sample, we also restricted the sample to firms headquartered in the fifteen main Indian states. This excluded firms located in Assam, Bihar, Chandigarh, Chhattisgarh, Dadra, Daman and Diu, Goa, Himachal Pradesh, Jammu and Kashmir, Jharkhand, Orissa and Uttarakhand, each of which accounts for less than 3% of Indian GDP.

Of these, 907 later resulted not to be eligible for the interviews upon the first telephonic contact (the reasons for non eligibility included recent bankruptcy or the company not being in manufacturing), and 310 were never contacted because the project ended before this was possible. The final number of eligible companies was thus 6,527, with median yearly sales of $53,000,000. Of these, we were able to secure an interview with 1,131 CEOs, although 17 CEOs dropped out before the end of the data collection week for personal reasons and were thus removed from the sample before the analysis was conducted.

The selection analysis in Table A1 shows that firms in the final sample have on average slightly lower log sales relative to the sampling frame (coefficient 0.071, standard error 0.011). However, we do not find any significant selection effect on performance variables, such as labor productivity (sales over employees) and return on capital employed (ROCE).

B The Inference Algorithm and MCMC Estimation

The Dirichlet distribution of dimensionality M is defined on the $M-1$-simplex, and provides a flexible means of modeling the probability of the weights for multinomial or categorical distributions. (The Dirichlet with $M = 2$ corresponds to the beta distribution). Symmetric Dirichlet distributions are parameterized by a scalar α. When $\alpha = 1$, the Dirichlet places uniform probability on all elements of simplex; when $\alpha < 1$ it places more weight on the corners of the simplex and so generates multinomial weights that tend to have a few large values and many small values; and when $\alpha > 1$ it places more weight on the center of the simplex and so generates multinomial weights that are similar in magnitude. Hereafter, we let α denote the parameter associated with the symmetric Dirichlet prior on the behavioral indices, and η the parameter associated with the prior on the behaviors. For the hyper-parameters, we set $\alpha = 1$, which corresponds to a uniform prior on each CEO's behavioral index. We also set $\eta = 0.1$. As noted above, this means the prior on pure behaviors places more weight on probability vectors that have their mass concentrated on a limited number of elements of X. In other words, we set the prior so that behaviors feature some combinations prominently, but put little weight on many others. This allows us to more sharply separate behaviors than would be the case for larger values of η.

As is typical with Bayesian models for high-dimensional data, exact posterior inference for LDA is intractable. One must therefore use approximate posterior inference algorithms, and we follow the Markov Chain Monte Carlo (MCMC) approach of Griffiths and Steyvers (2004) to sample the behaviors associated with each unit of time.

The general idea of MCMC estimation is to randomly seed the model with initial values for the behaviors associated to time units $z_{i,t}$, perform initial sampling iterations while the Markov Chain "burns in" to its stationary distribution, and then draw samples every nth iteration thereafter. The gap between sample draws is called a thinning interval, and is introduced to reduce autocorrelation between samples. The samples are then averaged to form estimates as in Monte Carlo simulations.

The specific procedure we adopt is:[38]

[38]We run five chains beginning from five different seeds, and select the one for analysis that has the best

1. Randomly allocate to each time block a behavior drawn uniformly from β^k with $k = 0, 1$.

2. For each time block in sequence, draw a new behavior using multinomial sampling. The probability that block t for CEO i is assigned to behavior k is increasing in:

   (a) The number of other blocks for CEO i that are currently assigned to k.

   (b) The number of other occurrences of the feature combination $y_{i,t}$ in the entire dataset that is currently assigned to k.

3. Repeat step 2 5,000 times as a burn in phase.

4. Repeat step 2 5,000 more times, and store every 50th sample.

Steps 2a and 2b mean that feature combinations that regularly co-occur in CEOs' time use will be grouped together to form behaviors. Also, step 2a means that feature combinations within individual CEOs will tend to be concentrated rather than spread across behaviors.

Many combinations are rare: there are 183 combinations that appear in just one time block, and 430 that appear in two. Since inference in LDA relies on co-occurrence, the assignment of such rare combinations to behaviors is noisy. For this reason, we drop any combination that is not present in at least 30 CEOs' time use. This leaves 654 combinations and 98,347 time blocks in the baseline analysis, which represents around 77% of the 127,660 interactive activities. Table A.1 shows average CEO time shares across features on this subsample, which are very similar to those of the whole sample reported in table.[39]

For each draw in step 4, the estimate $\widehat{\theta}_i^k$ is proportional to the total number of time units of CEO i allocated to behavior k plus the prior α, and the estimate $\widehat{\beta}_k^f$ is proportional to the total number of times x_f is allocated to behavior k plus the prior η. We then average these estimates across all draws, to form the final objects we analyze in the paper.

To make the inference procedure more concrete, consider a simplified dataset with three CEOs and an activity feature set $X = \{\text{unplanned,planned}\} \times \{\text{size1}, \text{size2+}\}$. Table A.3 tabulates the number of time blocks of each CEO according to their value of x_f and their allocation across two behaviors—which we denote B0 and B1—at different points in a Markov chain. The row sums within each value of $x_f \in X$ represent the total number of time blocks of a CEO associated to x_f. CEO A's time is dominated by planned activities with two or more people (162 out of 168 time blocks have x_f = size2+planned); CEO B's time is dominated by unplanned activities; while CEO C has a broader distribution of time use across feature combinations.

Table A.3 represents the random seed from which sampling begins. Since behavior assignments are drawn uniformly, each CEO's time is split roughly evenly between behaviors. The last column shows the behavioral indices derived from these assignments, which is around

goodness-of-fit across the draws we take after burn in.

[39]For robustness, we have also kept combinations present in 15 and, alternatively, 45 CEOs' time use, and find very similar results.

Table A.1: Raw average time shares for all CEOs on Estimation Subsample

(a) Distribution of time within features

Type		**Duration**		**Planned**		**Participants**	
value	share	value	share	value	share	value	share
meeting	0.803	1hr+	0.657	planned	0.764	size2+	0.553
site_visit	0.06	1hr	0.188	unplanned	0.236	size1	0.427
phone_call	0.054	30m	0.139			missing	0.019
business_meal	0.049	15m	0.017				
public_event	0.015						
conference_call	0.013						
workrelated_leisure	0.005						
video_conference	0.001						

(b) Distribution of time across functions

Inside Functions		**Outside Functions**	
function	share	function	share
production	0.35	clients	0.103
mkting	0.206	suppliers	0.064
finance	0.147	others	0.05
groupcom	0.073	associations	0.031
hr	0.063	consultants	0.026
bunits	0.042	govoff	0.016
board	0.031	banks	0.013
other	0.029	compts	0.012
admin	0.029	pemployee	0.01
cao	0.023	lawyers	0.008
coo	0.017	investors	0.005
strategy	0.011		
legal	0.008		

Notes: The top table shows the amount of time the average CEO spends on different options within features for the 98,347 15-minute units of time in the baseline estimation exercise excluding rare combinations. The bottom table shows the amount of time the average CEO spends with different functions on the same subsample.

Table A.2: Estimated average time shares for all CEOs on Estimation Subsample

(a) Distribution of time within features

Type		**Duration**		**Planned**		**Participants**	
value	share	value	share	value	share	value	share
meeting	0.801	1hr+	0.687	planned	0.782	size2+	0.573
site_visit	0.062	1hr	0.176	unplanned	0.218	size1	0.411
business_meal	0.053	30m	0.123			missing	0.017
phone_call	0.047	15m	0.014				
public_event	0.017						
conference_call	0.012						
workrelated_leisure	0.006						
video_conference	0.001						

(b) Distribution of time across functions

Inside Functions		**Outside Functions**	
function	share	function	share
production	0.355	clients	0.104
mkting	0.208	suppliers	0.068
finance	0.144	others	0.05
groupcom	0.081	associations	0.033
other	0.077	consultants	0.026
hr	0.062	govoff	0.015
bunits	0.041	compts	0.014
board	0.032	banks	0.013
admin	0.029	pemployee	0.01
cao	0.022	lawyers	0.008
coo	0.015	investors	0.006
strategy	0.01		
legal	0.008		

Notes: The top table shows the estimated amount of time the average CEO spends on different options within features for the baseline estimation exercise. The bottom table shows the estimated amount of time the average CEO spends with different functions on the same subsample. These estimated shares are derived from the marginal distributions computed from the estimated behaviors, and the estimated CEO behavioral indices.

0.5 for all CEOs. The last row shows the estimated probability that each x_f appears in each behavior, which begins around the empirical frequency of x_f in the overall sample.

As sampling proceeds from the random seed, time units are re-allocated between behaviors. size1unplanned and size2+unplanned activities begin to be pulled into B0, while size1planned and size2+planned activities are pulled into B1. As this happens, A's behavioral index moves towards one, C's moves towards zero, and B's remains around 0.5. This shows the importance of allowing CEOs to mix behaviors, as forcing B into one of the two behaviors would not capture the full heterogeneity of his or her time use.

In such a small dataset, the chain converges quickly and by the fifth iteration stabilizes. The only time units whose assignments vary substantially in further sampling are the two that CEO A spends in size2+unplanned activities. This combination is both strongly associated with B0—which favors sampling its value to 0—and present in a CEO's time use that is strongly associated to B1—which favors sampling its value to 1. Averaging over numerous draws accounts for this uncertainty.

C Proof of Proposition 2

We verify that the situation described in the proposition corresponds to a Bayesian equilibrium. To simplify notation re-normalize all variables so that $\Delta = 1$.

First note, that if $B > 1$, all CEOs will choose the behavior that is optimal for the firm that hires them. This means that CEO behavior only depends on firm type. Therefore, in what follows we assume that governance is sufficiently poor, so $B < c$.

In that case, when a CEO is hired, her utility is $\bar{w} + B$ if she works for a firm of the same type and $\bar{w}$ if she works for a firm of a different type. To simplify notation, further normalize $\bar{w} + B = 1$. Hence the utility of a correctly matched CEO is one and the utility of a mismatched CEO is

$$b \equiv \frac{\bar{w}}{\bar{w} + B}.$$

Note that b is a measure of the quality of governance, with $b = 1$, being the worst level of governance.

A type-0 firm faces an abundant supply of type-0 CEOs. As all the applications it receives come from type-0 CEOs, the firm will simply hire the first applicant. A type-1 firm instead may receive applications from both CEO types. If k is sufficiently low, the optimal policy consists in waiting for the first candidate with $s = 1$ and hire him.

We now consider CEOs. Suppose that all type-1 CEOs apply to type-1 firms and type-0 CEOs apply to type-1 firms with probability z and to type-0 firms with probability $1 - z$.

If a type-0 CEO applies to a type-0 firm, he will get a job if and only if his application is downloaded. The mass of type-0 firms is $1 - \phi$. The mass of type-0 CEOs applying to type-0 firms is $(1 - \gamma)(1 - z)m$. The probability the CEO is hired is

$$P_0 = \frac{1 - \phi}{(1 - \gamma)(1 - z)m}.$$

Table A.3: Example of MCMC estimation of allocation of time blocks to behaviors

(a) Random Seed

	size1unplanned		size1planned		size2+unplanned		size2+planned		
CEO	B0	B1	B0	B1	B0	B1	B0	B1	$\widehat{\theta}_i$
A	0	0	1	3	0	2	82	80	0.506
B	9	4	1	0	5	4	12	19	0.5
C	35	43	0	0	38	30	0	0	0.5
	0.24	0.254	0.011	0.017	0.235	0.195	0.513	0.535	
	$\widehat{\beta}_0^1$	$\widehat{\beta}_1^1$	$\widehat{\beta}_0^2$	$\widehat{\beta}_1^2$	$\widehat{\beta}_0^3$	$\widehat{\beta}_1^3$	$\widehat{\beta}_0^4$	$\widehat{\beta}_1^4$	

(b) Iteration 2

	size1unplanned		size1planned		size2+unplanned		size2+planned		
CEO	B0	B1	B0	B1	B0	B1	B0	B1	$\widehat{\theta}_i$
A	0	0	4	0	2	0	35	127	0.753
B	10	3	1	0	5	4	4	27	0.625
C	73	5	0	0	63	5	0	0	0.074
	0.421	0.047	0.026	0.001	0.355	0.053	0.198	0.899	
	$\widehat{\beta}_0^1$	$\widehat{\beta}_1^1$	$\widehat{\beta}_0^2$	$\widehat{\beta}_1^2$	$\widehat{\beta}_0^3$	$\widehat{\beta}_1^3$	$\widehat{\beta}_0^4$	$\widehat{\beta}_1^4$	

(c) Iteration 5

	size1unplanned		size1planned		size2+unplanned		size2+planned		
CEO	B0	B1	B0	B1	B0	B1	B0	B1	$\widehat{\theta}_i$
A	0	0	0	4	2	0	0	162	0.982
B	13	0	0	1	9	0	0	31	0.589
C	78	0	0	0	68	0	0	0	0.007
	0.535	0.001	0.001	0.026	0.464	0.001	0.001	0.973	
	$\widehat{\beta}_0^1$	$\widehat{\beta}_1^1$	$\widehat{\beta}_0^2$	$\widehat{\beta}_1^2$	$\widehat{\beta}_0^3$	$\widehat{\beta}_1^3$	$\widehat{\beta}_0^4$	$\widehat{\beta}_1^4$	

Notes: This table shows the allocation of three CEO's time use to behaviors at different points in an example Markov chain. The algorithm samples each unit of time into one of two behaviors, from which we derive estimates of the behavioral index $\widehat{\theta}_i$ and behaviors $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\boldsymbol{\beta}}_1$. In this simple example, the chain converges within a few iterations.

If instead a type-0 CEO applies to a type-1 firm, he will get a job if and only if his application is considered and the firm does not detect deception. Computing the first probability requires an additional step, because some firms consider more than one application before they find an application which passes the screening process.

The probability that a type-1 firm application is accepted if it is considered is:

$$H = \frac{(1-\gamma)\, z\, (1-\rho) + \gamma}{(1-\gamma)\, z + \gamma}.$$

The mass of applications that are downloaded by type-1 firms is therefore:

$$\phi \left(1 + (1-H) + (1-H)^2 + ...\right) = \phi \frac{1}{H}.$$

Given that the mass of applicants to type-1 firms is $m\left((1-\gamma)\, z + \gamma\right)$, the probability that an application is considered is

$$\frac{\phi}{m\left(\gamma + (1-\gamma)\, z\right) H} = \frac{\phi}{m\left((1-\gamma)\, z\, (1-\rho) + \gamma\right)}$$

The probability that a type-0 applicant passes the screening process is $1-\rho$. Thus, the probability that a type-0 applicant is hired by a type-1 firm is

$$P_1 = \frac{(1-\rho)\, \phi}{m\left((1-\gamma)\, z\, (1-\rho) + \gamma\right)}.$$

In the equilibrium under consideration a type-0 CEO must be indifferent between applying to the two types of firms. As the benefit of being hired by a same-type firm is one, while the benefit of being hired by a type-1 firm is b, the indifference condition is $P_0 = bP_1$, which yields:

$$\frac{1-\phi}{(1-\gamma)\,(1-z)} = \frac{(1-\rho)\, \phi b}{\left((1-\gamma)\, z\, (1-\rho) + \gamma\right)},$$

yielding

$$z = \frac{(1-\gamma)\,(1-\rho)\, \phi b - (1-\phi)\, \gamma}{(1-\phi+\phi b)\,(1-\gamma)\,(1-\rho)}.$$

The solution of z will be positive – meaning that some 0-types will apply to 1-firms – if

$$\rho < 1 - \frac{(1-\phi)\, \gamma}{(1-\gamma)\, \phi b},$$

which is satisfied as long as ρ is not too high, b is not too low, and γ is sufficiently smaller than ϕ. For instance, the combination of $\rho = 0$, $b = 1$, and $\phi > \gamma$ would work.

Type-1 CEOs always produce 1, while the average productivity of a type-0 CEO is equal to the probability that he is matched with a type 0 firm, which is

$$\frac{1-z}{1-z+z\left(1-\rho\right)}.$$

By replacing z, we find the average productivity of a type-0 CEO:

$$\frac{\left(1-\phi\right)\left(\left(1-\gamma\right)\left(1-\rho\right)+\gamma\right)}{\left(1-\phi\right)\left(1-\gamma\right)\left(1-\rho\right)+\left(1-\phi\right)\gamma+\left(\left(1-\gamma\right)\left(1-\rho\right)\phi b-\left(1-\phi\right)\gamma\right)\left(1-\rho\right)},$$

which is smaller than one whenever $\rho < 1$.

Finally, note that the difference between the profit (including CEO compensation) of a correctly matched firm and an incorrectly matched one is $1-B$.

D Dynamic Implications of the assignment Model

We now explore the dynamic implications of our CEO-firm assignment model. Suppose that we know the behavior of the current CEO, but not the type of the firm and the behavior of the previous CEO. What can we say about the evolution of firm performance over time?

Let us assume that the conditions for Proposition 2 are satisfied. There are two types of CEOs ($i \in \{0,1\}$) and two types of firms ($j \in \{0,1\}$). We assume that the abundant CEO type is $i=0$. Using a reduced form expression from the previous section, assume that the performance of a firm is $y_j + x_{ij}$, where $x_{ij} = 1$ if the firm type and the CEO type match ($i = j$) and $x_{ij} = 0$ if there is a mismatch $(i \neq j)x_{ij}$, and the term y_j indicate that the two firm types may have different baseline productivities.

Let us consider a firm whose CEO is replaced at time 0. Let x_{ij}^{old} and x_{ij}^{new} denote the match quality of the previous CEO and the current CEO, respectively. The performance of the firm at time $t < 0$ was determined uniquely by the performance of the old CEO (thus assuming that he had been in the job sufficiently long). The performance at $t \geq 0$ is given by

$$Y_t = y_j + \left(1-\alpha_t\right) x_{ij}^{\text{old}} + \alpha_t x_{ij}^{\text{new}},$$

where α_t is increasing and s-shaped in t. Namely,$\alpha_0 = 0$,$\alpha_t' > 0$, $\lim_{t\to 0+}\alpha_t' = 0$, $\lim_{t\to\infty}\alpha_t = 1$, and $\alpha"_t > 0$ if t is low and $\alpha"_t < 0$ if t is high. As time passes, the company's performance is determined more and more by the type of the new CEO as his tenure increases. The s-shaped assumption captures the idea that the effect of a new CEO is limited in the beginning, it increases with time, but then it reaches a stable plateau.

Consider a large sample of firms. Suppose we observe the type of the current CEO, but we do not observe the type of the previous CEO, nor the type of the firm. What can we say about them?

If the current CEO belongs to the scarce type, we know for sure that the firm has type-1. The previous CEO was the scarce type too with probability π and the abundant-type with probability $1-\pi$.[40]

Focus on performance growth, taking $t=0$ as the baseline year: $\Delta Y_t = Y_t - Y_0$. If the current CEO belongs to the scarce type, we have

$$\Delta Y_t\left(i^{\text{new}}=1\right)=\begin{cases}0 & \text{if } t<0\\ \left(\left(1-\alpha_t\right)E\left[x_{ij}^{\text{old}}|x_{ij}^{\text{new}}=1\right]+\alpha_t\right)-E\left[x_{ij}^{\text{old}}|x_{ij}^{\text{new}}=1\right] & \text{if } t>0\end{cases}$$

but note that $E\left[x_{ij}^{\text{old}}|x_{ij}^{\text{new}}=1\right]=\pi<1$. Therefore,

$$\Delta Y_t\left(i^{\text{new}}=1\right)=\begin{cases}0 & \text{if } t<0\\ \alpha_t\left(1-\pi\right) & \text{if } t>0\end{cases},$$

which implies that average performance is flat before the CEO replacement and follows $\alpha_t\left(1-\pi\right)$ thereafter.

If instead we consider a sample of firms run by abundant-type CEOs, a specular argument applies: we would observe that the average performance decreases after the current CEO is hired and follows a similarly s-shaped curve. Therefore we have:

Proposition 3 *The average performance of a sample of firms who are currently run by scarce-type (abundant-type) CEOs was flat before the new CEOs were hired and it becomes increasing (decreasing) and s-shaped thereafter.*

Figure A.1depicts the average performance of a set of firms run by scarce-type CEOs, $\Delta Y_t\left(i^{\text{new}}=1\right)$, under the assumption that α_t is a sigmoid function ($\alpha_t = t/\sqrt{1+t^2}$) and $\pi=\frac{1}{2}$. The average effect of having a scarce-type CEO is positive, gradual, and s-shaped. This result implies that if we observe a set of firms run by scarce type CEOs who were all hired at the same date, we should predict that the average performance of those firms is constant before the CEOs are hired, almost constant right after they are hired, and increasing and s-shaped afterwards.

E LDA and Clustering

E.0.1 LDA vs. other dimensionality-reduction techniques

We have performed PCA and k-means analysis on the key marginals that emerge from LDA. For each CEO, we count the number of engagements that: (1) last longer than one hour; (2)

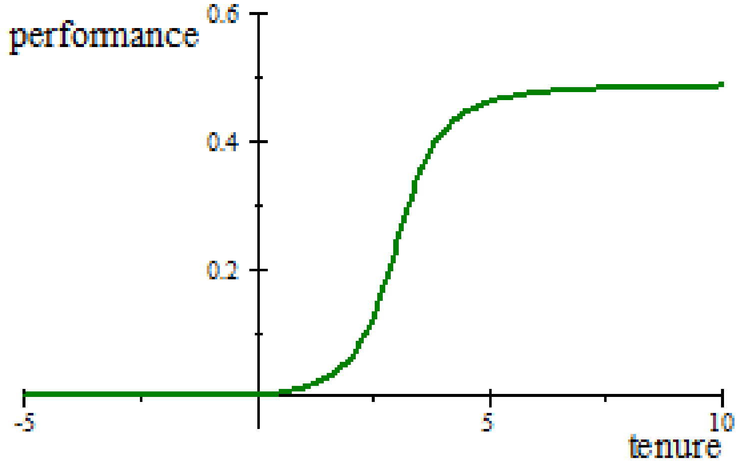
[40]This probability is given in equilibrium by

$$\pi=\frac{\gamma}{\gamma+\left(1-\gamma\right)z},$$

where

$$z=\frac{\left(1-\gamma\right)\left(1-\rho\right)\phi-\gamma\left(1-\phi\right)}{\left(1-\gamma\right)\left(1-\rho\right)}.$$

Figure A.1: Average performance of firms run by type-1 CEOs

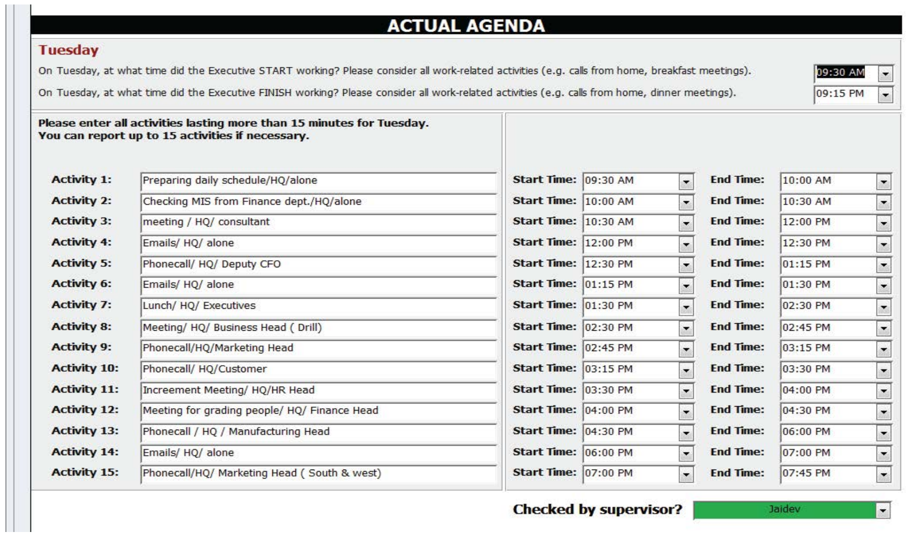


are planned; (3) involve two or more people; (4) involve outsiders alone; (5) involve high-level inside functions; and (6) involve more than one function. The first principal component in PCA analysis explains 36% of the variance in this feature space and places a positive weight on all dimensions except (4). Meanwhile, k-means clustering produces one centroid with higher values on all dimensions except (4) (and, ipso facto, a second centroid with a higher value for (4) and lower values for all others). Hence the patterns identified using simpler methods validate the key differences from LDA with two pure behaviors. Note that LDA is still a necessary first step in this analysis because it allows us to identify the important marginals along which CEOs vary.[41] In the appendix we show the relationship between these alternative ways of classifying CEOs and performance.

F Appendix Figures and Tables

[41]We have experimented with PCA and k-means on the 654-dimensional feature space over which we estimate the LDA model, but the results are much harder to analyze as described above.

Figure A.2: Survey Instrument



ACTUAL AGENDA

Tuesday

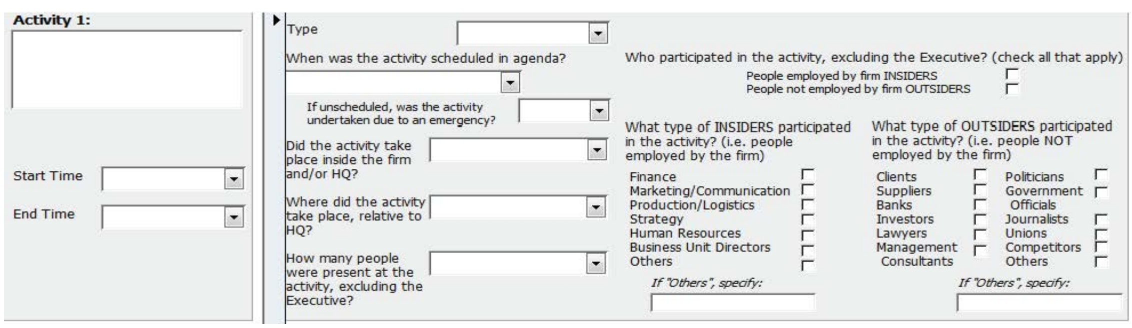
On Tuesday, at what time did the Executive START working? Please consider all work-related activities (e.g. calls from home, breakfast meetings). 09:30 AM

On Tuesday, at what time did the Executive FINISH working? Please consider all work-related activities (e.g. calls from home, dinner meetings). 09:15 PM

Please enter all activities lasting more than 15 minutes for Tuesday.
You can report up to 15 activities if necessary.

Activity 1:	Preparing daily schedule/HQ/alone	**Start Time:**	09:30 AM	**End Time:**	10:00 AM
Activity 2:	Checking MIS from Finance dept./HQ/alone	**Start Time:**	10:00 AM	**End Time:**	10:30 AM
Activity 3:	meeting / HQ/ consultant	**Start Time:**	10:30 AM	**End Time:**	12:00 PM
Activity 4:	Emails/ HQ/ alone	**Start Time:**	12:00 PM	**End Time:**	12:30 PM
Activity 5:	Phonecall/ HQ/ Deputy CFO	**Start Time:**	12:30 PM	**End Time:**	01:15 PM
Activity 6:	Emails/ HQ/ alone	**Start Time:**	01:15 PM	**End Time:**	01:30 PM
Activity 7:	Lunch/ HQ/ Executives	**Start Time:**	01:30 PM	**End Time:**	02:30 PM
Activity 8:	Meeting/ HQ/ Business Head (Drill)	**Start Time:**	02:30 PM	**End Time:**	02:45 PM
Activity 9:	Phonecall/HQ/Marketing Head	**Start Time:**	02:45 PM	**End Time:**	03:15 PM
Activity 10:	Phonecall/ HQ/Customer	**Start Time:**	03:15 PM	**End Time:**	03:30 PM
Activity 11:	Increement Meeting/ HQ/HR Head	**Start Time:**	03:30 PM	**End Time:**	04:00 PM
Activity 12:	Meeting for grading people/ HQ/ Finance Head	**Start Time:**	04:00 PM	**End Time:**	04:30 PM
Activity 13:	Phonecall / HQ / Manufacturing Head	**Start Time:**	04:30 PM	**End Time:**	06:00 PM
Activity 14:	Emails/ HQ/ alone	**Start Time:**	06:00 PM	**End Time:**	07:00 PM
Activity 15:	Phonecall/HQ/ Marketing Head (South & west)	**Start Time:**	07:00 PM	**End Time:**	07:45 PM

Checked by supervisor? Jaidev



Activity 1:

Start Time

End Time

Type

When was the activity scheduled in agenda?

If unscheduled, was the activity undertaken due to an emergency?

Did the activity take place inside the firm and/or HQ?

Where did the activity take place, relative to HQ?

How many people were present at the activity, excluding the Executive?

Who participated in the activity, excluding the Executive? (check all that apply)

People employed by firm INSIDERS

People not employed by firm OUTSIDERS

What type of INSIDERS participated in the activity? (i.e. people employed by the firm)

Finance
Marketing/Communication
Production/Logistics
Strategy
Human Resources
Business Unit Directors
Others

If "Others", specify:

What type of OUTSIDERS participated in the activity? (i.e. people NOT employed by the firm)

Clients
Suppliers
Banks
Investors
Lawyers
Management Consultants
Politicians
Government Officials
Journalists
Unions
Competitors
Others

If "Others", specify:

Table A.4: CEO behavior index and firm characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			CEO behavior index			
log(employment)	0.055***	0.053***	0.055***	0.056***	0.052***	0.053***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
COO Dummy		0.066***	0.062***	0.062***	0.064***	0.056**
		(0.021)	(0.021)	(0.021)	(0.021)	(0.025)
task abstraction (industry)			0.032**	0.035**	0.030**	
			(0.012)	(0.013)	(0.013)	
capital intensity (industry)				-0.006	-0.009	
				(0.018)	(0.018)	
homogeneous product (industry)				-0.031	-0.027	
				(0.030)	(0.031)	
log(CEO tenure)					-0.019*	-0.015
					(0.010)	(0.010)
CEO has an MBA					0.048**	0.062**
					(0.023)	(0.026)
Family CEO					-0.032	-0.035
					(0.021)	(0.024)
Adjusted R-squared	0.242	0.248	0.252	0.251	0.261	0.265
Observations	1114	1114	1114	1114	1114	1114
Industry FE	n	n	n	n	n	y

Notes: *** (**) (*) denotes significance at the 1%, 5% and 10% level, respectively. All columns include country fixed effects and noise controls. The COO dummy takes value one if the firm employs a COO, "task abstraction" is an industry metric drawn from Autor et al (2003), with higher values denoting a higher intensity of abstract tasks in production. "Capital intensity" denotes the average industry level value of capital over labor, built from the NBER manufacturing database (aggregated between 2000 and 2010). "Homogeneous product" is an industry dummy drawn from Rausch (1999). "Log CEO tenure" is the log of 1+number of years CEO is in office, "CEO has an MBA"is a dummy taking value one is the CEO has attained an MBA degree or equivalent postgraduate qualification. "Family CEO" denotes CEOs who are affiliated with the owning family. Noise controls are a full set of dummies to denote the week in the year in which the data was collected, a reliability score assigned by the interviewer at the end of the survey week and a dummy taking value one if the data was collected through the PA of the CEO, rather than the CEO himself.. Industry controls are 2 digit SIC dummies. All columns weighted by the week representativeness score assigned by the CEO at the end of the interview week. Errors clustered at the 2 digit SIC level.